



Marina Sequeiros Dias

**O uso de Máquina de Suporte Vetorial para
Regressão (SVR) na Estimação da Estrutura a
Termo da Taxa de Juros do Brasil**

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-graduação em Matemática do Departamento de Matemática da PUC-Rio

Orientador : Prof. Hélio Côrtes Vieira Lopes

Co-Orientador: Prof. Luciano Vereda Oliveira

Rio de Janeiro
Fevereiro de 2007

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.



Marina Sequeiros Dias

**O uso de Máquina de Suporte Vetorial para
Regressão (SVR) na Estimação da Estrutura a
Termo da Taxa de Juros do Brasil**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-graduação em Matemática do Departamento de Matemática do Centro Técnico Científico da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Hélio Côrtes Vieira Lopes

Orientador

Departamento de Matemática — PUC-Rio

Prof. Luciano Vereda Oliveira

Co-Orientador

IAPUC — PUC-Rio

Prof. Cristiano Augusto C. Fernandes

Departamento de Engenharia Elétrica – PUC-Rio

Prof. Marco Antônio Grivet Mattoso Maia

Centro de Estudos em Telecomunicações – PUC-Rio

Prof. José Eugênio Leal

Coordenador Setorial do Centro Técnico Científico — PUC-Rio

Rio de Janeiro, 27 de Fevereiro de 2007

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

Marina Sequeiros Dias

Graduou-se em Licenciatura em Matemática na UERJ (Universidade do Estado do Rio de Janeiro) em 2004.

Ficha Catalográfica

Dias, Marina Sequeiros

O uso de Máquina de Suporte Vetorial para Regressão (SVR) na Estimação da Estrutura a Termo da Taxa de Juros do Brasil / Marina Sequeiros Dias; orientador: Hélio Côrtes Vieira Lopes; co-orientador: Luciano Vereda Oliveira. — Rio de Janeiro : PUC-Rio, Departamento de Matemática, 2007.

v., 77 f: il. ; 29,7 cm

1. Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Matemática.

Inclui referências bibliográficas.

1. Matemática – Tese. 2. Séries Temporais. 3. Modelos de previsão. 4. Taxa de Juros. 5. Máquina de Suporte Vetorial para Regressão. 6. Aprendizagem estatística. I. Lopes, Hélio Côrtes Vieira. II. Oliveira, Luciano Vereda. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Matemática. IV. Título.

CDD: 510

Agradecimentos

À Deus, em primeiro lugar.

Aos professores Hélio Côrtes Vieira Lopes e Luciano Vereda Oliveira pela orientação.

Aos meus pais, Benjamin e Maria das Dores, por colaborarem e incentivarem os meus estudos.

Ao meu namorado Eduardo pelo seu companherismo e apoio aos meus estudos e projetos.

Ao meu irmão Adriano por acompanhar e ajudar nos meus estudos.

À Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio) pela oportunidade de cursar o Mestrado e pela bolsa de isenção.

Ao Iapuc pelo apoio.

A CAPES e a FAPERJ pela concessão da bolsa de mestrado.

Aos meus amigos, por estarem sempre participando de todos os momentos de minha vida, em especial, aos amigos que fiz na PUC.

Às secretárias Creuza Nascimento e Katia Beatriz Aguiar pelo carinho e aos Auxiliares Administrativos do Departamento de Matemática da PUC-Rio pela colaboração de sempre.

Resumo

Dias, Marina Sequeiros; Lopes, Hélio Côrtes Vieira; Oliveira, Luciano Vereda. **O uso de Máquina de Suporte Vetorial para Regressão (SVR) na Estimação da Estrutura a Termo da Taxa de Juros do Brasil**. Rio de Janeiro, 2007. 77p. Dissertação de Mestrado — Departamento de Matemática, Pontifícia Universidade Católica do Rio de Janeiro.

Nessa dissertação um novo método para previsão da Estrutura a Termo da Taxa de Juros Brasileira - ETTJ brasileira - conhecido como Máquina de Suporte Vetorial para Regressão é investigado, comparando-o com os métodos tradicionais, tais como modelos VAR (Vetor Auto-regressivo) e ECM (Modelos de Correção de Erros). Utiliza-se além dos retornos de títulos de renda fixa, algumas variáveis macro-econômicas, que conforme sugerido no artigo de Evans e Marshall (1998) e verificado para economia brasileira no artigo de Fukuda, Vereda e Lopes (2006) melhoram a previsão dos retornos de títulos de renda fixa no longo prazo. O experimento mostra uma melhora considerável do SVR sobre os modelos tradicionais mencionados no longo prazo, atuando ainda como ótimo indicador da direção das taxas em praticamente todos os horizontes de previsão. Para tal avaliação, foram utilizados os critérios de raiz do erro quadrado médio, erro absoluto médio, simetria direcional e simetria direcional ponderada, correta tendência para cima e correta tendência para baixo além do teste U de Theil, que faz uso da raiz do erro quadrado médio para verificar se ocorre uma melhora significativa de um modelo sobre outro. Uma vez que não existe uma maneira estruturada para escolha dos parâmetros livres do SVR, a escolha dos mesmos foi feita através de uma função do software R, que faz uma pesquisa em um domínio retangular fornecido pelo usuário. A análise dos resultados mostra que SVR é uma técnica promissora para previsão dos retornos de títulos de renda fixa, sugerindo-se ainda melhorar as escolhas dos parâmetros livres do SVR uma vez que os mesmos são meios poderosos de regularização e adaptação do ruído aos dados.

Palavras-chave

Séries Temporais. Modelos de previsão. Taxa de Juros. Máquina de Suporte Vetorial para Regressão. Aprendizagem estatística.

Abstract

Dias, Marina Sequeiros; Lopes, Hélio Côrtes Vieira; Oliveira, Luciano Vereda. **the use of Support Vector Regression (SVR) in Estimating the Brazilian Term Structure of Interest Rates.** Rio de Janeiro, 2007. 77p. MsC Thesis — Department of Mathematics, Pontifícia Universidade Católica do Rio de Janeiro.

In this dissertation a new method for the prediction of the Brazilian Term Structure of Interest Rates - Brazilian ETTJ - known as Support Vector Regression is investigated. This is compared with the traditional methods used in this set up, such as VAR models (Vector Autoregressive) and ECM (Error Correction Models). Besides the interest rates, some macroeconomic variables are also used, as it was suggested in a work from Evans and Marshall(1998) and verified for brazilian economy in a work from Fukuda, Vereda and Lopes (2006), the inclusion of macroeconomic variables can improve the prediction of the interest rates in long term forecasts. The experiment show some improvements in using SVR in the long term in relation to the traditional methods mentioned, acting like a really good predictor of the direction of the interest rates along the short and long term forecasts. To make these assertions, we make use of some tests like the root mean squared error, mean absolute error, directional symmetry and weighted directional symmetry, Correct Up trend and Corret Down trend besides Theil U test, which uses the root mean squared error to verify if there is some significant improvement between two models. As there is not a structured way to choose the free parameters of SVR, a function in the R software was used in order to make a grid search over a supplied parameter ranges. The analysis of the results demonstrate that SVR is a promising technique to prediction of interest rates, suggestions are also made in order to get better the choices of the free SVR parameters once they are powerful means of regularization and adaptation to the noise in the data.

Keywords

Time series. Prediction models. Interest rates. Support Vector Regression. Statistical Learning.

Sumário

1	Introdução	11
1.1	Motivação	13
1.2	Objetivos	14
1.3	Estrutura da dissertação	15
2	Um pouco da teoria de otimização	16
2.1	Teoremas da teoria de otimização	17
3	Máquinas de Suporte Vetorial para regressão	20
3.1	Introdução à teoria de aprendizagem supervisionada	20
3.2	O Desenvolvimento da teoria SVM	22
3.3	Teoria de SVMs na aproximação por regressão	24
3.4	Como obter b	33
3.5	Os Núcleos	33
3.6	Arquitetura de um SVR	38
3.7	O SVR no <i>software</i> R	39
4	Descrição dos dados	40
5	Modelos de previsão	42
5.1	Previsão de séries temporais	42
5.2	Modelo de previsão usando SVR	42
5.3	Modelo de previsão usando Vetor Auto-regressivo (VAR) e Modelo de Correção de Erros (ECM)	45
5.4	Critérios de desempenho	45
6	Resultados experimentais	48
7	Conclusão e trabalhos futuros	62
	Referências Bibliográficas	75

Lista de figuras

1.1	Modelo de aprendizagem supervisionada a partir de exemplos (Vapnik, 2000(30)).	12
3.1	A máquina de vetores de suporte (Vapnik, 1998 (29)).	23
3.2	Hiperplano linear de separação ótima para dados linearmente separáveis (esquerda) e dados não separáveis (direita). Os vetores suporte estão circulados (Burges (1998) (3)).	23
3.3	Função de custo ε -insensível linear (Schölkopf, Smola, 2001(18)).	27
3.4	Hiperplano separador linear com a função de custo ε -insensível (linear)(Schölkopf, Smola, 2001(18)).	28
3.5	A função de custo ε -insensível quadrática (Cristianini, Shawe-Taylor, 2000 (7)).	29
3.6	O mapeamento dos dados de entrada do espaço de dimensão R^n para um espaço característico com alta dimensionalidade R^N , onde uma regressão linear é efetuada (Smola, 1996 (22)).	34
3.7	Arquitetura de máquina de vetor suporte (Schölkopf, Smola, 2001 (18)).	38
7.1	Valores previstos nos dados de teste de $\log_s 30$ (1 mês).	63
7.2	Valores previstos nos dados de teste de $\log_s 60$ (1 mês).	63
7.3	Valores previstos nos dados de teste de $\log_s 90$ (1 mês).	63
7.4	Valores previstos nos dados de teste de $\log_s 120$ (1 mês).	64
7.5	Valores previstos nos dados de teste de $\log_s 180$ (1 mês).	64
7.6	Valores previstos nos dados de teste de $\log_s 360$ (1 mês).	64
7.7	Valores previstos nos dados de teste de $\log_s 720$ (1 mês).	65
7.8	Valores previstos nos dados de teste de $\log_s 30$ (3 meses).	65
7.9	Valores previstos nos dados de teste de $\log_s 60$ (3 meses).	65
7.10	Valores previstos nos dados de teste de $\log_s 90$ (3 meses).	66
7.11	Valores previstos nos dados de teste de $\log_s 120$ (3 meses).	66
7.12	Valores previstos nos dados de teste de $\log_s 180$ (3 meses).	66
7.13	Valores previstos nos dados de teste de $\log_s 360$ (3 meses).	67
7.14	Valores previstos nos dados de teste de $\log_s 720$ (3 meses).	67
7.15	Valores previstos nos dados de teste de $\log_s 30$ (6 meses).	67
7.16	Valores previstos nos dados de teste de $\log_s 60$ (6 meses).	68
7.17	Valores previstos nos dados de teste de $\log_s 90$ (6 meses).	68
7.18	Valores previstos nos dados de teste de $\log_s 120$ (6 meses).	69
7.19	Valores previstos nos dados de teste de $\log_s 180$ (6 meses).	69
7.20	Valores previstos nos dados de teste de $\log_s 360$ (6 meses).	69
7.21	Valores previstos nos dados de teste de $\log_s 720$ (6 meses).	70
7.22	Valores previstos nos dados de teste de $\log_s 30$ (9 meses).	70
7.23	Valores previstos nos dados de teste de $\log_s 60$ (9 meses).	70
7.24	Valores previstos nos dados de teste de $\log_s 90$ (9 meses).	71
7.25	Valores previstos nos dados de teste de $\log_s 120$ (9 meses).	71
7.26	Valores previstos nos dados de teste de $\log_s 180$ (9 meses).	71
7.27	Valores previstos nos dados de teste de $\log_s 360$ (9 meses).	72

7.28	Valores previstos nos dados de teste de $\log_s 720$ (9 meses).	72
7.29	Valores previstos nos dados de teste de $\log_s 30$ (12 meses).	72
7.30	Valores previstos nos dados de teste de $\log_s 60$ (12 meses).	73
7.31	Valores previstos nos dados de teste de $\log_s 90$ (12 meses).	73
7.32	Valores previstos nos dados de teste de $\log_s 120$ (12 meses).	73
7.33	Valores previstos nos dados de teste de $\log_s 180$ (12 meses).	74
7.34	Valores previstos nos dados de teste de $\log_s 360$ (12 meses).	74
7.35	Valores previstos nos dados de teste de $\log_s 720$ (12 meses).	74

Lista de tabelas

3.1	Funções de custo comumente usadas.	29
5.1	Valores escolhidos para cada parâmetro em cada um dos modelos.	44
5.2	Métricas utilizadas (a_i e p_i representam os valores atuais e previstos, respectivamente).	46
6.1	Resultados para previsões 1 passo à frente.	53
6.2	Teste de Theil para previsões 1 passo à frente.	53
6.3	Resultados para previsões 3 passos à frente.	55
6.4	Teste de Theil para previsões 3 passos à frente.	55
6.5	Resultados para previsões 6 passos à frente.	57
6.6	Teste de Theil para previsões 6 passos à frente.	57
6.7	Resultados para previsões 9 passos à frente.	59
6.8	Teste de Theil para previsões 9 passos à frente.	59
6.9	Resultados para previsões 12 passos à frente.	61
6.10	Teste de Theil para previsões 12 passos à frente.	61

1

Introdução

No Brasil, o grande aumento na adesão a planos de previdência privada nos últimos tempos implicou em maiores preocupações de bancos e seguradoras em fazer investimentos visando garantir o pagamento de benefícios no longo prazo. No entanto, o cenário da economia que hoje em dia é muito favorável pode não o ser daqui a alguns anos. Uma boa previsão do comportamento das taxas de juros é uma tarefa fundamental. Neste contexto, conseqüentemente, especificar e estimar modelos que forneçam previsões confiáveis e que acompanhem a curva de retornos com eficiência se torna cada vez mais importantes. Nesse sentido, modelos tradicionalmente propostos incluem Vetor Auto-regressivo (VAR) e Modelos de Correção de Erros (ECM). Nessa dissertação, procura-se desenvolver um novo modelo baseado na metodologia de aprendizagem estatística supervisionada.

A teoria estatística de aprendizagem busca extrair tendências e padrões importantes dos dados e entender o que os mesmos querem dizer. Conforme colocado por Vapnik (2000)(30), pode-se descrever o problema geral de aprendizagem por exemplos através de três componentes, conforme ilustrado na figura 1.1:

- Um gerador (G) de vetores aleatórios $\mathbf{x} \in \mathbb{R}^n$, obtidos independentemente a partir de uma distribuição de probabilidade fixa e desconhecida $F(\mathbf{x})$.
- Um supervisor (S) que retorna um valor de saída y para cada vetor de entrada \mathbf{x} , de acordo com uma função de distribuição condicional $F(y | \mathbf{x})$, também fixa e desconhecida.
- Uma máquina de aprendizagem (LM) capaz de implementar um conjunto de funções $f(\mathbf{x}, \alpha)$, $\alpha \in A$, onde A é um conjunto de parâmetros.

Dado um conjunto de funções $f(\mathbf{x}, \alpha)$, $\alpha \in A$, o problema de aprendizagem é escolher aquela que melhor se aproxima da resposta do supervisor. A seleção da função desejada é baseada em um conjunto de treinamento composto de l observações $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_l, y_l)$, independentes e identicamente distribuídas (iid) obtidas de acordo com $F(\mathbf{x}, y) = F(\mathbf{x})F(y | \mathbf{x})$.

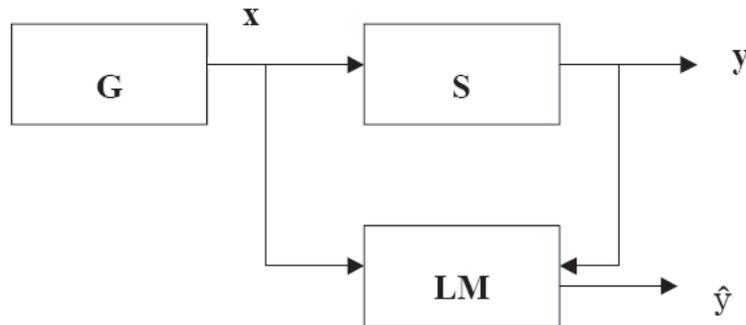


Figura 1.1: Modelo de aprendizagem supervisionada a partir de exemplos (Vapnik, 2000(30)).

Por sua vez, a aprendizagem supervisionada procura aprender uma função que forneça a saída desejada a partir de um conjunto de entradas usando exemplos (pares de entrada e saída desejada). Quando existe uma função dos dados de entrada para os dados de saída tal função é chamada de função alvo e a estimativa de tal função é escolhida dentre um conjunto de funções candidatas, conhecido como um conjunto de hipóteses ou espaço de hipóteses. Em aprendizagem supervisionada, também se deve ter boas propriedades de generalização para dados não utilizados no treinamento, chamados dados de teste.

Um método de aprendizagem supervisionada para classificação originalmente desenvolvido por Vapnik e colaboradores, conhecido como Máquina de Suporte Vetorial (SVM) utiliza um espaço de hipóteses de funções lineares em um espaço característico de alta dimensionalidade e tem sido aplicado com sucesso em tarefas, como por exemplo, OCR (reconhecimento de caráter ótico) (30, 20), classificação de tipos de rocha em geologia e geofísica (12, 13, 31) e, recentemente, em regressão (19, 10) e até alcançando êxito na previsão de séries temporais (25, 11). Por essa razão, estudamos Máquinas de Suporte Vetorial com ênfase na sua extensão para Regressão (SVR) visando sua aplicação em previsão de séries temporais.

Investiga-se ainda a utilização do SVR na previsão de retornos de títulos de renda fixa com diversas maturidades. Nessa formulação, além dos retornos, a informação contida em algumas variáveis macro-econômicas é explorada a fim de melhor prever o desenvolvimento da estrutura a termo. A idéia de utilizar variáveis macro-econômicas e os retornos foi desenvolvida para a economia brasileira (9) e verificou-se que, a informação contida nessas variáveis pode

melhorar o desempenho da previsão no longo prazo. Assim, formula-se o modelo SVR e os modelos tradicionalmente propostos (VAR e ECM), incluindo as observações passadas dos retornos e de algumas variáveis macroeconômicas já que o objetivo é obter um bom desempenho preditivo no longo prazo, que é de fundamental importância para fins atuariais.

1.1 Motivação

Empresas de fundos de pensão têm como objetivo conseguir o máximo de lucro dos prêmios de seguro de pensão, evitando níveis de risco significantes, a fim de garantir que seus compromissos com o pagamento dos benefícios futuros sejam cumpridos. Ademais, dependendo da situação da instituição e das preferências de seus participantes, podem ocorrer diferentes estratégias de investimentos. Portanto, modelos que estudem o comportamento dos ativos e passivos ao longo do tempo (ALM) e aplicados a fundos de pensão estão diretamente associados aos riscos de longo prazo, os quais estão relacionados aos riscos de mercado, como o risco associado à variação da taxa de juros.

A Estrutura a Termo da Taxa de Juros (ETJT) é um indicador de grande importância na economia uma vez que, conceitualmente, representa a relação, em determinado instante, entre o prazo para o vencimento e a taxa de retorno dos títulos oriundos de uma mesma classe de risco, desempenhando um papel central na determinação de ativos (ou passivos) que possam ser representados por um fluxo de caixa.

Modelos econométricos realistas são naturalmente necessários, os quais devem levar em conta a evolução das variáveis macro-econômicas principais para prever as trajetórias dos retornos de investimentos e fluxo de responsabilidades. Na verdade, uma tentativa de resolver problemas de casamento de ativos e passivos (ALM) deve levar em consideração estimativas da estrutura a termo da taxa de juros, motivo pelo qual se faz obrigatório a existência de modelos que sejam confiáveis e bons preditores. Os modelos de Vetores Auto-regressivos (VAR) e Modelos de Correção de Erros (ECM) são as ferramentas tipicamente usadas para tais objetivos.

Esses modelos são freqüentemente utilizados para descrever o comportamento de um pequeno conjunto representativo de ativos composto, por exemplo, por títulos de renda fixa de curto e longo prazos, ações, móveis e indicadores de crescimento de salários e preços (cf. Boender, van Aalst e Heemskerk em (1) e o trabalho de Dert (8)). Assim, os modelos fornecem uma primeira avaliação dos retornos de investimentos desde que a maior parte da carteira típica de um fundo de previdência possa ser classificada nas categorias citadas.

Além disso, os modelos também fornecem uma primeira avaliação da evolução das contribuições e dos passivos, que estão relacionados ao crescimento dos salários e dos preços.

Esses modelos, no entanto, incorrem em algumas desvantagens relevantes:

- Não conseguem produzir uma descrição confiável da trajetória futura do valor de mercado das carteiras devido ao número limitado de categorias de ativos, dado que, na prática, essas carteiras não se limitam a ativos que acompanham os índices que aparecem nos modelos mais convencionais;
- Não fornecem os fatores de desconto necessários para calcular o preço de mercado dos diversos ativos de renda fixa disponíveis, sejam eles " *pure discount bonds*" ou " *coupon bearing bonds*"; e
- Não conseguem tratar adequadamente a questão do risco de reinvestimento, que necessariamente requer previsões para a rentabilidade de títulos de diversas maturidades.

Exercícios de previsão são geralmente realizados baseados nas hipóteses de que as variáveis relevantes (ou a combinação linear delas) são estacionárias e/ou combinadas por relações lineares. Estas restrições, entretanto, não são consistentes com as características de muitas séries temporais financeiras e econômicas, as quais não possuem médias e variâncias constantes. Um bom exemplo seriam as taxas de juros, nosso principal foco, as quais exibem períodos sucessivos de muita volatilidade e, pelo menos em princípio poderiam estar relacionadas umas com as outras e com outras variáveis macro-econômicas de diferentes maneiras dependendo do nível de volatilidade. Devido a essas características, é proposto aplicar um método de aprendizagem supervisionada chamado Máquina de Suporte Vetorial para Regressão (SVR) como uma técnica de modelagem, visando obter uma estimação mais precisa da curva de retornos, que é uma condição necessária para recuperação dos fatores de desconto e análise do risco de reinvestimento, desde que tal técnica facilmente acomoda relações não-lineares entre as variáveis relevantes.

1.2

Objetivos

Máquinas de Suporte Vetorial para Regressão são uma nova metodologia de aprendizagem supervisionada não-paramétrica para estimação de regressão. Dentre as características principais da técnica, destacam-se o bom desempenho de generalização e o fato de que treinar SVR é equivalente a resolver um problema de programação quadrática cuja solução é sempre única e global, não existindo portanto problemas de existência de mínimos locais. Por fim,

consegue-se ainda uma representação esparsa da solução, pois a solução do SVR depende apenas de um subconjunto dos dados de entrada referidos como vetores de suporte.

Por todos esses motivos, examina-se a viabilidade de sua aplicação na previsão de retornos de investimentos em títulos de renda fixa de várias maturidades comparando com os modelos VAR e ECM, utilizando diversas métricas.

1.3

Estrutura da dissertação

No capítulo seguinte, algumas definições e teoremas da teoria de otimização que serão importantes para melhor entendimento da teoria SVR serão apresentados.

No capítulo 3 uma introdução a teoria de aprendizagem supervisionada é feita, bem como as vantagens e desvantagens de tal metodologia. A teoria de Máquina de Suporte Vetorial é estudada e uma análise em detalhes de sua extensão para regressão é realizada. Defini-se o princípio da minimização do risco empírico e sua importância para a solução do problema de otimização envolvido, o qual é descrito em detalhes, explicando-se ainda as vantagens em utilizar a representação dual do problema e sua solução. Algumas funções de custo possíveis são apresentadas e a definição da função de custo implementada, a saber, a função de custo ε -insensível é estudada. Considera-se a necessidade de introduzir as funções núcleo para incorporar as relações não-lineares dos dados. As propriedades das funções núcleo são analisadas bem como sua principal caracterização pelo Teorema de Mercer. Estuda-se ainda o uso do *software* R, que foi utilizado para a implementação do SVR.

No capítulo 4, é feita uma descrição dos dados utilizados e no capítulo 5, os modelos SVR com as medidas de erro utilizadas são apresentados, bem como a divisão dos dados de acordo com a teoria de aprendizagem.

No capítulo 6, os resultados e tabelas com as métricas usadas para comparar os métodos estudados são mostrados. Por fim, no capítulo 7, conclui-se o trabalho, alguns trabalhos futuros são discutidos e os gráficos comparativos entre os métodos tradicionais e os modelos SVR são exibidos.

2

Um pouco da teoria de otimização

Algumas definições importantes da teoria de otimização são introduzidas. Para maiores detalhes, veja Vapnik, 1998 (29) e Cristianini e Shawe-Taylor, 2000(7).

Definição 2.1 (Problema de otimização primal) *Considere as funções $f; g_i, i = 1, \dots, m$; e $h_j, j = 1, \dots, n$, definida em um domínio $\Psi \subset \mathbb{R}^n$,*

$$\begin{aligned} & \text{minimize} && f(x), && x \in \Psi \\ & \text{sujeito a} && g_i(x) \leq 0, && i = 1, \dots, m \\ & && h_j(x) = 0, && j = 1, \dots, n \end{aligned}$$

onde $f(x)$ é a função objetivo e as relações restantes são chamadas, respectivamente, restrições de desigualdade e de igualdade.

A região do domínio onde a função objetivo está definida e onde todas as restrições são satisfeitas é denominada região viável e é denotada por:

$$R = \{x \in \Psi : g(x) \leq 0 \text{ e } h(x) = 0\}$$

Uma solução para o problema de otimização é um ponto $x^* \in \mathbb{R}$ tal que não existe outro ponto $x \in \mathbb{R}$ para o qual $f(x) < f(x^*)$. Tal ponto é chamado mínimo global. Um mínimo local é um ponto x^* para o qual $f(x^*) \leq f(x)$ para todo x com $\|x - x^*\| < \varepsilon$.

Um problema de otimização no qual a função objetivo, as restrições de igualdade e desigualdade são todas funções lineares é chamado um problema de programação linear. Caso a função objetivo seja quadrática e as restrições lineares, será denominado de problema de programação quadrática.

Definição 2.2 *Uma restrição de desigualdade $g_i(x) \leq 0$ é dita ser ativa se a solução ótima x^* , satisfaz $g(x^*) = 0$, caso contrário é dita ser uma restrição inativa. As restrições de igualdade serão sempre ativas.*

Definição 2.3 (conjunto convexo) *Um conjunto X é chamado convexo se para todo $x, x' \in X$ e para todo $\lambda \in [0, 1]$, temos: $\lambda x + (1 - \lambda)x' \in X$.*

Definição 2.4 (função convexa) Uma função f definida em um conjunto X , o qual não precisa ser convexo, é chamada convexa se e somente se para todo $x, x' \in X$ e para todo $\lambda \in [0, 1]$ tal que $\lambda x + (1 - \lambda)x' \in X$, temos $f(\lambda x + (1 - \lambda)x') \leq \lambda f(x) + (1 - \lambda)f(x')$ (*).

Uma função f é chamada estritamente convexa se para $x \neq x'$ e $\lambda \in (0, 1)$, (*) é a desigualdade estrita.

Se uma função f é convexa, qualquer mínimo local x^* de um problema de otimização irrestrito com função objetivo f é também um mínimo global, pois para qualquer $y \neq x^*$ por definição de mínimo local existe θ suficientemente próximo de 1 tal que

$$f(x^*) \leq f(\theta x^* + (1 - \theta)y) \leq \theta f(x^*) + (1 - \theta)f(y). \text{ Logo, } f(x^*) \leq f(y).$$

Definição 2.5 Um problema de otimização no qual o conjunto Ψ , a função objetivo e todas as restrições são convexas é dito ser convexo.

Conforme será visto, treinar uma máquina SVR vai significar resolver um problema de otimização quadrática convexa.

2.1

Teoremas da teoria de otimização

Os principais resultados que serão utilizados para desenvolver soluções eficientes para a tarefa de otimizar SVRs são descritos. Maiores detalhes e demonstrações dos teoremas apresentados podem ser encontradas em (5).

O primeiro método analítico para resolver problemas de otimização foi desenvolvido por Fermat em 1629, o qual descreveu um método para obter máximos e mínimos de uma função sem restrições.

Teorema 2.6 (Fermat) Considere uma função $f : \mathbb{R}^n \rightarrow \mathbb{R}$ de classe C^1 . Uma condição necessária para que $x^* = (x_1^*, \dots, x_n^*)$ seja um mínimo para f é que $\frac{\partial f(x^*)}{\partial x_i} = 0, i = 1, \dots, n$. Esta condição juntamente com a convexidade de f constitui uma condição suficiente.

Em 1788, Lagrange sugeriu um método para resolver problemas de otimização com restrições de igualdade. Inicialmente, precisa-se definir uma função que incorpore tanto a função objetivo quanto às restrições, conhecida como Lagrangiana.

Definição 2.7 Dado um problema de otimização com função objetivo $f(x)$ e restrições de igualdade $h_i(x) = 0, i = 1, \dots, m$, definimos a função Lagrangiana como:

$L(x, \beta) = f(x) + \sum_{i=1}^m \beta_i h_i(x)$ onde os coeficientes β_i são chamados os multiplicadores de Lagrange.

Teorema 2.8 (Lagrange) Uma condição necessária para que um ponto $x^* = (x_1^*, \dots, x_n^*)$ seja um mínimo para $f : \mathbb{R}^n \rightarrow \mathbb{R}$ sujeito a $h_i(x) = 0$, com f, h_i de classe C^1 é que as seguintes condições, chamadas condições de estacionariedade, sejam satisfeitas:

$$\begin{aligned} \frac{\partial L(x^*, \beta^*)}{\partial x_i} &= 0 \\ \frac{\partial L(x^*, \beta^*)}{\partial \beta_i} &= 0, \end{aligned}$$

para algum valor de β^* . As condições anteriores serão também suficientes caso $L(x, \beta^*)$ seja uma função convexa de x .

Observação: Uma vez que as restrições são iguais a zero, o valor do Lagrangiano no ponto ótimo será igual ao valor da função objetivo, ou seja,

$$L(x^*, \beta^*) = f(x^*)$$

Considera-se o problema de otimização com restrições de igualdade e desigualdade e define-se o Lagrangiano generalizado.

Definição 2.9 Dado um problema de otimização com domínio $\Psi \subset \mathbb{R}^n$,

$$\begin{aligned} &\text{minimize } f(x), && x \in \Psi \\ &\text{sujeito a } g_i(x) \leq 0, && i = 1, \dots, m \\ & && h_j(x) = 0, \quad j = 1, \dots, n \end{aligned}$$

Definimos a função de Lagrange generalizada como:

$$\begin{aligned} L(x, \alpha, \beta) &= f(x) + \sum_{i=1}^m \alpha_i g_i(x) + \sum_{j=1}^n \beta_j h_j(x) \\ &= f(x) + \alpha'g(x) + \beta'h(x). \end{aligned}$$

Assim, o problema dual de Lagrange pode ser definido.

Definição 2.10 O problema dual de Lagrange do problema primal da definição

2.1 é o seguinte problema:

$$\begin{aligned} &\text{maximize } \theta(\alpha, \beta), \\ &\text{sujeito a } \alpha \geq 0, \\ &\text{onde } \theta(\alpha, \beta) = \inf_{x \in \Psi} L(x, \alpha, \beta). \end{aligned}$$

Teorema 2.11 (Dualidade Fraca) *Considere que $x \in \Psi$ seja uma solução viável do problema primal da definição 2.9 e (α, β) uma solução viável do problema dual da definição 2.10. Então $f(x) \geq \theta(\alpha, \beta)$.*

Corolário 2.12 *Seja x^* e (α^*, β^*) são soluções viáveis para os problemas primal e dual, respectivamente, e suponha que $f(x^*) = \theta(\alpha^*, \beta^*)$ onde $\alpha^* \geq 0$ e $g(x^*) \leq 0, h(x^*) = 0$, então x^* e (α^*, β^*) são soluções ótimas para o problema primal e dual, respectivamente.*

Teorema 2.13 (Dualidade Forte) *Se o problema primal P tem uma solução ótima x^* então o dual D tem uma solução ótima (α^*, β^*) tal que $f(x^*) = \theta(\alpha^*, \beta^*)$*

Em 1951, Kuhn e Tucker estenderam o método de multiplicadores de Lagrange para permitir restrições de desigualdades.

Teorema 2.14 (Kuhn-Tucker) *Considere o problema de otimização com domínio convexo $\Psi \subset R^n$,*

$$\begin{aligned} & \text{minimize} && f(x), && x \in \Psi \\ & \text{sujeito a} && g_i(x) \leq 0, && i = 1, \dots, m \\ & && h_j(x) = 0, && j = 1, \dots, n \end{aligned}$$

Com f de classe C^1 , convexa, g_i, h_j funções afins, as condições necessárias e suficientes para que x^ seja um ótimo é a existência de α^*, β^* tal que:*

$$\begin{aligned} \frac{\partial L(x^*, \alpha^*, \beta^*)}{\partial x} &= 0, \\ \frac{\partial L(x^*, \alpha^*, \beta^*)}{\partial \beta} &= 0, \\ \alpha_i g_i(x) &= 0, && i = 1, \dots, m, \\ g_i(x) &\leq 0, && i = 1, \dots, m, \\ \alpha_i^* &\geq 0, && i = 1, \dots, m, \end{aligned}$$

A terceira condição é conhecida como condição complementar de Karush-Kuhn-Tucker (KKT).

3

Máquinas de Suporte Vetorial para regressão

3.1

Introdução à teoria de aprendizagem supervisionada

Na aprendizagem supervisionada, o objetivo é prever o valor de uma função para qualquer entrada válida utilizando um certo número de exemplos, isto é, dados de treinamento. Tais dados exemplificam as relações entre a entrada e saída. A função de saída pode ser um valor em um espaço discreto, utilizado para classificação ou pode ser um valor em um espaço contínuo, chamado regressão. A solução é escolhida dentre um conjunto de funções candidatas que mapeiam o espaço de entrada no domínio de saída. Usualmente um conjunto particular de funções candidatas é escolhido e denominado *hipótese* antes de se aprender a função correta. O algoritmo que tem com entrada os dados de treinamento e seleciona uma hipótese do espaço de *hipóteses* é referido como algoritmo de aprendizagem.

Para resolver um problema de aprendizagem supervisionada, deve-se seguir os passos:

1. Obter dados de entrada que sejam representativos do que se quer como resposta.
2. Determinar a representação dos dados de entrada, uma vez que, a precisão da função aprendida depende de como os objetos de entrada estão representados. Geralmente, o dado de entrada é transformado em um vetor num espaço vetorial com produto interno que chamamos de *espaço característico*.
3. Determinar a estrutura da função a ser aprendida e o correspondente algoritmo de aprendizagem. Nesta dissertação, utilizamos SVR.
4. O modelo. Utilizar um algoritmo de aprendizagem no conjunto de treinamento. Os parâmetros do algoritmo podem ser ajustados otimizando o desempenho num subconjunto do conjunto de treinamento, chamado conjunto de validação, ou via validação cruzada. Após a escolha dos

parâmetros e aprendizagem, o desempenho do algoritmo pode ser avaliado em um conjunto de teste, o qual é um conjunto separado do conjunto de treino.

3.1.1

Vantagens e desvantagens da metodologia

Um ponto marcante da teoria de aprendizagem supervisionada é o grande número de problemas práticos que podem ser resolvidos. Ademais, nesta metodologia basta coletar alguns pares de dados de entrada/saída e usar o algoritmo para aprendizagem da função que relaciona a entrada com a saída.

No entanto, muitas dificuldades merecem uma análise e estudos cuidadosos. Conforme destacado por Cristianini e Shawe-Taylor (7), uma das dificuldades está na escolha da classe de funções das quais o mapeamento dos dados de entrada e saída deve ser procurado, já que a classe deve ser rica o suficiente a fim de que o mapeamento ou uma aproximação seja encontrada, porém a classe não deve ser muito grande para que a complexidade do aprendizado dos exemplos não se torne proibitiva, particularmente quando se leva em consideração o número de exemplos necessários para fazer inferências estatisticamente confiáveis em um ampla classe de funções.

Na prática, esses problemas se traduzem em dificuldades de aprendizagem específicas. Uma delas se refere a ineficiência do algoritmo no caso de existência de mínimos locais. Em segundo lugar, a complexidade da descrição da função de saída, que pode freqüentemente ficar muito grande e sem utilidade na prática. O terceiro problema ocorre se apenas um número limitado de exemplos de treinamento está disponível. Então enriquecer a classe de hipóteses levará a problemas de excesso de ajuste (" *overfitting* ") e, portanto, a uma generalização ruim. O quarto problema decorre do fato de que o algoritmo de aprendizado freqüentemente é controlado por um grande número de parâmetros cuja escolha é baseada em heurísticas, tornando o sistema difícil e não confiável para usar.

Apesar das dificuldades apresentadas, a metodologia de aprendizagem tem obtido muito sucesso em problemas de interesse prático. A Máquina de Suporte Vetorial é uma ferramenta proeminente nessa área de estudos no qual os quatro problemas apresentados de eficiência no treinamento, eficiência nos dados de teste, excesso de ajuste (" *overfitting* ") e ajuste de parâmetros do algoritmo são evitados. Deve-se ressaltar que o problema de excesso de ajuste é parcialmente resolvido pelo SVR.

3.2

O Desenvolvimento da teoria SVM

No início da década de 60, o algoritmo de máquinas de suporte vetorial foi desenvolvido para construir hiperplanos separadores para problemas de reconhecimento de padrões (Vapnik e Lerner, 1963 (28), Vapnik e Chervonenkis, 1964 (27)). Embora suas características já tivessem sido apresentadas e utilizadas desde a década de 60, a teoria SVM foi introduzida pela primeira vez em um artigo na Annual Conference on Computational Learning Theory (COLT) em 1992 (2).

Na década de 90, o método foi generalizado para construir funções separadoras não lineares (porém lineares em um espaço característico) (Boser et al. 1992 (2), Cortes (6) e Vapnik 1995 (30)). Em 1995, o método foi generalizado para estimar funções de valor real (regressão) (Vapnik, 1995)(30).

Os dois livros recentes escritos por Vapnik(30, 29) fornecem um conhecimento teórico muito extenso no desenvolvimento do conceito de Máquinas de Suporte Vetorial. Podemos ainda destacar dois tutoriais. Um tutorial sobre classificadores baseados em suporte vetorial foi publicado por Burges (1998) (3) e outro sobre Máquinas de Suporte Vetorial para Regressão foi publicado por Smola e Schölkopf em 2003 (23).

A Máquina de Suporte Vetorial é um sistema de aprendizado treinado com um algoritmo de otimização baseado na teoria estatística de aprendizagem, que implementa a seguinte idéia (veja figura 3.1): vetores do espaço de entrada são mapeados não linearmente para um espaço característico de alta dimensionalidade, através de um mapeamento escolhido a priori e, nesse espaço uma superfície de decisão linear é construída, constituindo um hiperplano de separação ótima de exemplos, como por exemplo, a separação binária entre exemplos que possuem rótulos positivos e negativos, tal que a margem de separação seja máxima.

O objetivo da classificação de vetores de suporte está em conceber uma maneira computacionalmente eficiente, ou seja, capaz de lidar com diferentes tamanhos de amostras e capaz de aprender “bons” hiperplanos com separação ótima em um espaço característico de grande dimensionalidade, onde um bom hiperplano de separação se refere àquele que tem grande habilidade de generalização. A teoria de generalização nos fornece um guia sobre como controlar a capacidade e prevenir problemas de excesso de ajustes (“*overfitting*”) através do controle das medidas da margem do hiperplano. Já a teoria de otimização fornece as técnicas matemáticas necessárias para encontrar tais hiperplanos. Para maiores detalhes, consulte Cristianini e Shawe-Taylor (7).

O modelo de SVM mais simples é o classificador binário de margem

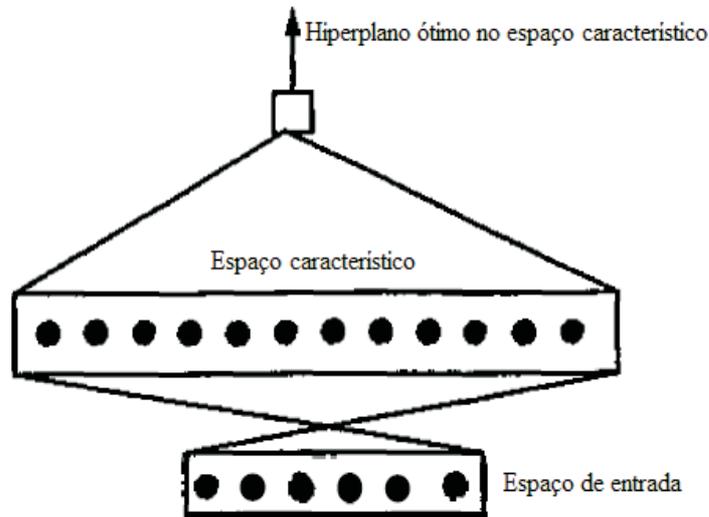


Figura 3.1: A máquina de vetores de suporte (Vapnik, 1998 (29)).

máxima, o qual funciona apenas para dados linearmente separáveis no espaço característico e por isso não se aplica em muitos problemas do mundo real. No entanto, esse classificador é o ponto de partida para o entendimento de SVMs mais complicados, podendo ser estendido para lidar com dados não separáveis, permitindo que alguns erros ocorram, mas com uma certa penalidade (veja figura 3.2).

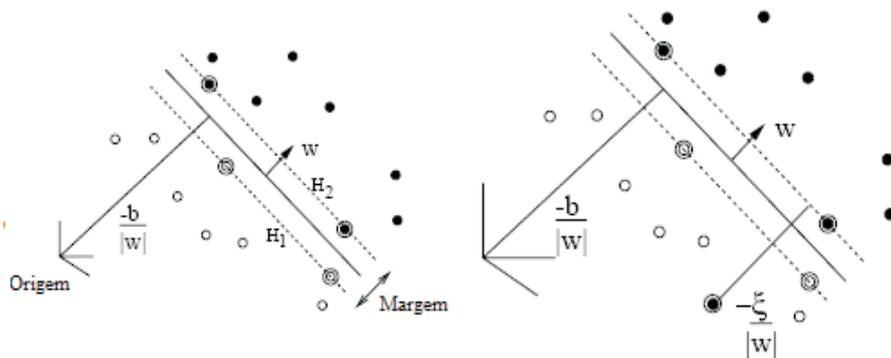


Figura 3.2: Hiperplano linear de separação ótima para dados linearmente separáveis (esquerda) e dados não separáveis (direita). Os vetores suporte estão circulados (Burges (1998) (3)).

O estudo de SVMs constitui uma área de pesquisa ativa. Sua utilização em regressão e em séries temporais é muito promissora, como podemos destacar a dissertação de doutorado de Martin (2005) (14) e o artigo de Silva e Ferreira (2006) (21) na área de energia elétrica.

3.3

Teoria de SVMs na aproximação por regressão

Considere um conjunto de l amostras $\{x_i\}$ onde $x_i \in \mathcal{X}$ (por exemplo, $\mathcal{X} = \mathbb{R}^n$) são os dados de entrada e $y_i \in \mathbb{R}$, os valores alvos para $i = 1, 2, 3, \dots, l$. Nesta aplicação, por exemplo, o espaço de entrada é composto por taxas *swap* de diferentes maturidades juntamente com algumas variáveis macroeconômicas em meses subsequentes e os valores de saída correspondem a uma taxa *swap* com vencimento específico, em meses subsequentes. Assume-se que os dados são iid (independentes e identicamente distribuídos) e que existe alguma distribuição de probabilidade desconhecida $P(x, y)$ de onde os dados são obtidos.

O objetivo é aprender o mapeamento $x_i \rightarrow y_i$, ou melhor, encontrar uma função a qual aproxime o mapeamento dos dados de entrada (os x_i 's) para os dados de saída (os y_i 's), os quais são números reais. Essa aprendizagem é feita com base em dados de treinamento. Em outras palavras, deseja-se encontrar uma função f que relacione a entrada com a saída ($x \rightarrow f(x)$). Também será determinado um critério para verificar a qualidade das estimativas vindas dos dados. A diferença entre o valor dado por $f(x_i)$ e o valor de y_i do arranjo dos dados de treinamento é chamada resíduo da saída e é uma indicação da precisão do ajuste no ponto x_i . Deve-se decidir como medir a importância desta precisão, pois resíduos pequenos podem ser aceitáveis mas é desejável que resíduos grandes sejam evitados.

Para considerar dados de treinamento que apresentem ruídos é necessário estabelecer funções de custo (também chamadas de funções de perda) que penalizem tais singularidades. Será determinada uma função de custo a qual determinará como penalizar os erros de estimação. A função de custo determina a importância da precisão e na maioria dos casos será do tipo $c(x, y, f(x)) = c(f(x) - y)$ e pode ser considerada como o custo decorrido de uma falha para prever precisamente uma dada variável.

O objetivo é encontrar a função f que minimiza o valor esperado da perda, fornecido pelo risco funcional:

$$R[f] = \int c(x, y, f(x)) dP(x, y), \quad (3-1)$$

onde $c(x, y, f(x))$ denota uma função de custo.

Uma vez que a distribuição $P(x, y)$ é desconhecida, a utilização do risco funcional é inútil. Dessa maneira, pode-se apenas usar os dados de treinamento (x_i, y_i) para estimar uma função f que de alguma maneira esteja “próxima” daquela que minimiza $R[f]$.

3.3.1

Princípio indutivo da minimização do risco empírico

Uma aproximação possível para o risco funcional consiste em substituir a integral por uma estimativa empírica, a fim de obter o chamado risco funcional empírico,

$$R_{emp}[f] := \frac{1}{l} \sum_{i=1}^l c(x_i, y_i, f(x_i)). \quad (3-2)$$

Observe que a distribuição de probabilidade não aparece nesta formulação. Este risco é definido apenas como uma medida do erro médio no conjunto de treinamento (para um número de observações finitas e fixas).

Uma tentativa seria encontrar uma função f que minimizasse o risco empírico (princípio da minimização do risco empírico), ou seja,

$$f_0 := \arg \min_{f \in F} R_{emp}[f] \text{ para uma classe de funções } F.$$

É possível fornecer condições a máquina de aprendizagem que garantam que assintoticamente (quando $l \rightarrow \infty$), o risco empírico irá convergir para o risco esperado. No entanto, um dos problemas que decorrem da tentativa de usar uma função f que minimize o risco empírico de uma classe de funções F ocorre quando esta classe de funções é muito rica, o que acontece, por exemplo, quando se tem poucos dados em espaços de grande dimensionalidade, podendo ocorrer grandes desvios, excesso de ajuste (" *overfitting* ") e propriedades de generalização ruins, ou seja, pode-se obter uma função $f \in F$ que faz uma ótima predição dos valores y_i nos dados de treino e, no entanto, não será possível garantir um bom desempenho nos dados de teste.

Uma tentativa de resolver o problema de minimizar o risco empírico seria restringir a classe de soluções admissíveis, ou seja, reduzir a classe F para um conjunto compacto. Essa técnica foi introduzida por Tikhonov e Arsenin em 1963 (26) para resolver problemas inversos e desde então tem sido aplicada a máquinas de aprendizagem com sucesso.

No entanto, ao invés de especificar o conjunto compacto F , um termo de estabilização ou regularização $\Omega[f]$ é adicionado à função objetivo original para limitar a complexidade da classe de funções F , constituindo a classe de risco funcional regularizada:

$$R_{reg}[f] := R_{emp}[f] + \lambda \Omega[f]. \quad (3-3)$$

onde $\lambda > 0$ é a chamada constante de regularização e especifica a troca entre a minimização do risco empírico e a simplicidade garantida por pequenos valores de $\Omega[f]$.

Usualmente $\Omega[f]$ é escolhido para ser convexo uma vez que o $R_{emp}[f]$ também é convexo e assim obtém-se um único mínimo global.

O princípio de minimização do risco empírico é um dos principais assuntos dos livros de Vapnik (30, 29) e maiores detalhes também podem ser obtidos em (18), onde é feita uma discussão das relações envolvendo regularização, Espaço de Hilbert de Núcleo Reproduzível (Reproducing Kernel Hilbert Spaces - RKHS), espaços característicos e operadores de regularização e suas relações com núcleos para vetores de suporte.

3.3.2

O problema de otimização

Inicialmente, buscam-se funções lineares f da seguinte maneira:

$f(x) = \langle w, x \rangle + b$, com $w \in \mathcal{X}$, $b \in \mathbb{R}$, onde $\langle \cdot, \cdot \rangle$ denota o produto interno em \mathcal{X} .

Com relação a escolha do termo de regularização, considera-se $\Omega[f] := \frac{1}{2}\|w\|^2$ uma escolha comum para classificação e regressão. Em regressão, a interpretação geométrica para minimizar $\frac{1}{2}\|w\|^2$ é obter uma função linear aderente com qualidades de aproximação suficientes. Logo,

$$R_{reg}[f] := R_{emp}[f] + \frac{\lambda}{2}\|w\|^2 = \frac{1}{l} \sum_{i=1}^l c(x_i, y_i, f(x_i)) + \frac{\lambda}{2}\|w\|^2 \quad (3-4)$$

onde $\lambda > 0$ tem que ser escolhido baseado nos dados disponíveis $\lambda(l, (x_i, y_i))$, $i = 1, \dots, l$.

Minimizar (3-4) captura a principal idéia da teoria de aprendizagem, ou seja, com o objetivo de se conseguir um risco pequeno, é necessário controlar tanto o erro de treinamento quanto à complexidade do modelo através da explicação dos dados com um modelo simples.

A função de custo deve ser escolhida de forma a melhor representar as incertezas (ruídos) do problema de otimização proposto. Para um grande número de funções de custo, a equação (3-4) pode ser minimizada resolvendo-se um problema de otimização quadrática, o qual possui solução única.

Primeiramente, estuda-se uma regressão com vetores de suporte com o intuito de aprender uma função f que tenha no máximo um desvio ε em x_i em relação aos valores alvos y_i para todos os dados de treinamento, ou seja, a função f poderá apresentar erros no intervalo $[y_i - \varepsilon, y_i + \varepsilon]$, $i = 1, \dots, l$ mas valores de erros superiores a ε não serão aceitos. Ao mesmo tempo, a função a ser aprendida deve apresentar a máxima aderência possível.

A aderência de f pode ser conseguida pela determinação de um fator w que seja o menor possível, o que pode corresponder à minimização da norma euclidiana de w , isto é, $\|w\|^2 = \langle w, w \rangle$. Tal função deve ainda satisfazer as restrições de erro $|f(x_i) - y_i| \leq \varepsilon$ para todo $i = 1, \dots, l$ onde $S = \{(x_1, y_1), \dots, (x_l, y_l)\}$ compõem os dados de treinamento. Assim, obtém-se o seguinte problema de otimização convexa:

$$\begin{aligned} & \text{minimize}_{w,b} \quad \frac{1}{2} \|w\|^2 \\ & \text{sujeito a} \quad \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon, & i = 1, \dots, l \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon, & i = 1, \dots, l \end{cases} \end{aligned} \quad (3-5)$$

O problema (3-5) supõe a existência de uma função f que aproxima todos os pares (x_i, y_i) com precisão ε especificada a priori, em outras palavras, que o problema de otimização é viável. Mas, algumas vezes esse pode não ser o caso ou deseja-se que alguns erros ocorram. Por essa razão, estuda-se um tipo de função denominada função de custo ε -insensível, a qual introduz variáveis de folga não-negativas ξ_i, ξ_i^* , $i = 1, \dots, l$ que consideram os pontos situados fora da margem $|f(x_i) - y_i| \leq \varepsilon$ com uma certa penalidade, separando os dados de treinamento com um número mínimo de erros além de manter restrições que de outra forma seriam inviáveis no problema de otimização anterior. Essa função de custo foi desenvolvida por Cortes e Vapnik, 1995 [6]. A função de custo é descrita a seguir (Veja figura 3.3):

$$|\xi|_\varepsilon := \begin{cases} 0 & \text{se } |\xi| \leq \varepsilon \\ |\xi| - \varepsilon & \text{caso contrário} \end{cases} \quad (3-6)$$

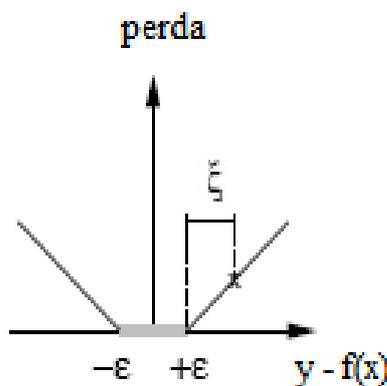


Figura 3.3: Função de custo ε -insensível linear (Schölkopf, Smola, 2001(18)).

Consegue-se a seguinte formulação do problema:

$$\begin{aligned} & \text{minimize}_{w,b} \quad \frac{1}{2} \| w \|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \\ & \text{sujeito a} \quad \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i, & i = 1, \dots, l, \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^*, & i = 1, \dots, l, \\ \xi_i, \xi_i^* \geq 0, & i = 1, \dots, l. \end{cases} \end{aligned} \quad (3-7)$$

As duas variáveis de folga ξ_i e ξ_i^* são introduzidas, uma por estar mais de ε abaixo do valor alvo e a outra por exceder o valor alvo por mais de ε , respectivamente. O vetor w e o escalar b são a solução do problema de otimização apresentado. A constante $C > 0$ é escolhida pelo usuário e determina o compromisso entre a aderência de f e o montante de desvios maiores do que ε tolerados. Deve-se atentar para o fato de que cada escolha de C corresponde à escolha de um valor para $\| w \|^2$ e então minimizar ξ_i, ξ_i^* para aquela escolha de w .

Consegue-se um problema cuja função objetivo é quadrática e possui restrições que são desigualdades lineares, logo, um problema de otimização quadrática. Além disso, ele é convexo e, portanto, terá solução única.

Na regressão com vetores suporte uma precisão epsilon é especificada a priori, gerando um região tubular com raio ε em torno dos dados. O compromisso entre a complexidade do modelo e os pontos situados fora da região tubular (com variáveis de folga ξ_i, ξ_i^* positivas) é determinado pela minimização de (3-7). A figura 3.4 descreve a situação.

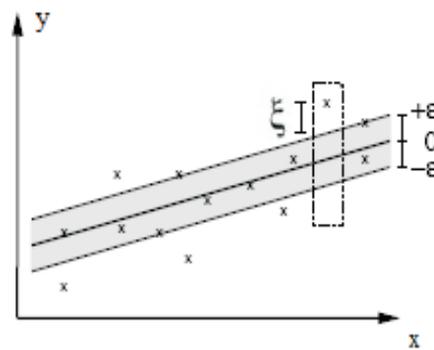


Figura 3.4: Hiperplano separador linear com a função de custo ε -insensível (linear)(Schölkopf, Smola, 2001(18)).

No caso anterior os desvios são penalizados de forma linear. Analogamente, pode-se considerar uma função de custo ε -insensível quadrática conforme mostra a figura 3.5 .

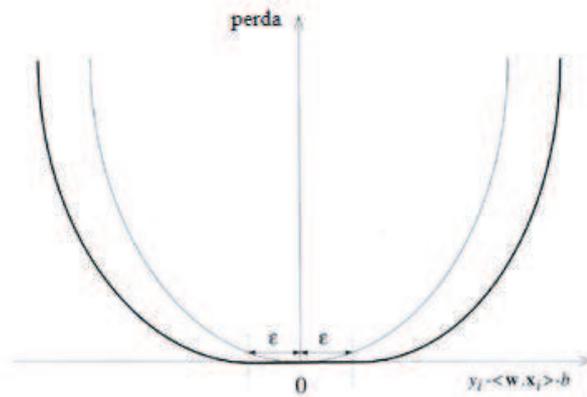


Figura 3.5: A função de custo ε -insensível quadrática (Cristianini, Shawe-Taylor, 2000 (7)).

O problema primal baseado na função de custo ε -insensível quadrática pode ser definido da seguinte maneira:

$$\begin{aligned}
 & \text{minimize}_{w,b} \quad \|w\|^2 + C \sum_{i=1}^l (\xi_i^2 + \xi_i^{*2}), \\
 & \text{sujeito a} \quad \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i, & i = 1, \dots, l, \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^*, & i = 1, \dots, l, \\ \xi_i, \xi_i^* \geq 0, & i = 1, \dots, l. \end{cases} \quad (3-8)
 \end{aligned}$$

Cada escolha da função de custo irá resultar em uma estratégia diferente para realizar a regressão. Tais funções devem ser convexas a fim de garantir a existência e unicidade (para convexidade estrita) da solução do problema de otimização. Na tabela seguinte, serão apresentadas algumas funções de custo comumente usadas.

	Função de custo
ε -insensível	$c(\xi) = \xi _\varepsilon$
Laplaciano	$c(\xi) = \xi $
Gaussiano	$c(\xi) = \frac{1}{2}\xi^2$
Perda Robusta de Huber	$c(\xi) = \begin{cases} \frac{1}{2\sigma}(\xi)^2 & \text{se } \xi \leq \sigma \\ \xi - \frac{\sigma}{2} & \text{caso contrário} \end{cases}$
Polinomial	$c(\xi) = \frac{1}{p} \xi ^p$
Polinomial por partes	$c(\xi) = \begin{cases} \frac{1}{p\sigma^{p-1}}(\xi)^p & \text{se } \xi \leq \sigma \\ \xi - \sigma \frac{p-1}{p} & \text{caso contrário} \end{cases}$

Tabela 3.1: Funções de custo comumente usadas.

No modelo SVR, faz-se uso da função de custo ε -insensível linear, embora outras funções possam ser usadas e o desempenho com cada uma dessas funções comparada, como pode ser verificado em (15) onde são usadas as funções de custo ε -insensível e a robusta de Hubber em dois conjuntos de dados diferentes.

3.3.3 Conversão para o Dual

Sabe-se que trabalhar com restrições de desigualdades diretamente é difícil. Assim, será obtida uma descrição dual alternativa que geralmente é mais fácil de resolver que o primal, através de uma reformulação Lagrangiana do problema com a introdução de multiplicadores de Lagrange ou variáveis duais para cada uma das restrições em (3-7). Lembre-se de que para restrições da forma $g_i \geq 0$, as equações são multiplicadas por multiplicadores de Lagrange positivos e subtraídas da função objetivo para formar o Lagrangiano. Para restrições de igualdade, os multiplicadores de Lagrange são irrestritos. O Lagrangiano é dado a seguir:

$$L := \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) - \sum_{i=1}^l (\eta_i \xi_i + \eta_i^* \xi_i^*) - \sum_{i=1}^l \alpha_i (\varepsilon + \xi_i - y_i + \langle w, x_i \rangle + b) - \sum_{i=1}^l \alpha_i^* (\varepsilon + \xi_i^* + y_i - \langle w, x_i \rangle - b) \quad (3-9)$$

onde $\eta_i, \eta_i^*, \alpha_i, \alpha_i^*$ são os multiplicadores de Lagrange. Portanto, as variáveis duais devem satisfazer as restrições de positividade, isto é, $\eta_i, \eta_i^*, \alpha_i, \alpha_i^* \geq 0$.

O problema de otimização pode ser visto como sendo a minimização da função de Lagrange com relação às variáveis primais ou maximização em relação aos multiplicadores de Lagrange.

A função de Lagrange L será minimizada em relação às variáveis primais w, b, ξ_i, ξ_i^* calculando as derivadas parciais de L com respeito a cada uma dessas variáveis primais e igualando-as a zero, obtendo assim:

$$\partial_b L = \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \quad (3-10)$$

$$\partial_w L = w - \sum_{i=1}^l (\alpha_i - \alpha_i^*) x_i = 0 \quad (3-11)$$

$$\partial_{\xi_i} L = C - \alpha_i - \eta_i = 0 \quad (3-12)$$

$$\partial_{\xi_i^*} L = C - \alpha_i^* - \eta_i^* = 0 \quad (3-13)$$

Substituindo 3-10, 3-11, 3-12 e 3-13 em 3-9 resolve-se o problema de otimização dual 3-14. A condição 3-10 implica na eliminação da variável primal b . Através das condições 3-12 e 3-13 temos $\eta_i = C - \alpha_i$ e $\eta_i^* = C - \alpha_i^*$ e, a eliminação das variáveis duais η_i e η_i^* .

$$\begin{aligned} & \text{maximize} && \begin{cases} -\frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \langle x_i, x_j \rangle \\ -\varepsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) + \sum_{i=1}^l y_i (\alpha_i - \alpha_i^*) \end{cases} && (3-14) \\ & \text{sujeito a} && \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \text{ e } \alpha_i, \alpha_i^* \in [0, C] \end{aligned}$$

A condição (3-11) mostra que:

$$w = \sum_{i=1}^l (\alpha_i - \alpha_i^*) x_i, \text{ portanto, } f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \langle x_i, x \rangle + b. \quad (3-15)$$

A observação da equação 3-15 permite destacar alguns pontos que serão importantes na extensão para o caso não-linear:

- w é descrito como a combinação linear de pontos de treinamento x_i onde α_i, α_i^* são soluções do problema dual. Essas variáveis duais possuem uma interpretação intuitiva como forças que agem na estimativa $f(x_i)$ a fim de que a mesma se aproxime das medições alvo desejadas y_i .
- o algoritmo completo pode ser descrito em termos de produto interno entre os dados. Ao se obter $f(x)$ não será preciso obter w explicitamente.

3.3.4

Condições complementares de Karush-Kuhn-Tucker (KKT)

Pode-se obter uma relação importante entre as soluções primal e dual através das condições complementares de KKT (veja 2.14).

As condições de KKT para o problema de otimização apresentado são:

$$\alpha_i(\varepsilon + \xi_i - y_i + \langle w, x_i \rangle + b) = 0, \quad i = 1, \dots, l, \quad (3-16)$$

$$\alpha_i^*(\varepsilon + \xi_i^* + y_i - \langle w, x_i \rangle - b) = 0, \quad i = 1, \dots, l, \quad (3-17)$$

$$(C - \alpha_i)\xi_i = 0, \quad i = 1, \dots, l, \quad (3-18)$$

$$(C - \alpha_i^*)\xi_i^* = 0, \quad i = 1, \dots, l, \quad (3-19)$$

$$\xi_i \xi_i^* = 0, \quad i = 1, \dots, l, \quad (3-20)$$

$$\alpha_i \alpha_i^* = 0, \quad i = 1, \dots, l \quad (3-21)$$

Pode-se extrair algumas conclusões úteis:

- Da condição (3-21), nunca existirá um conjunto de variáveis duais α_i e α_i^* que sejam ambas não-nulas, pois isto implicaria em uma movimentação não nula da função de otimização em ambas as direções.
- Apenas exemplos (x_i, y_i) com o correspondente $\alpha_i = C$, $\alpha_i^* = C$ estão fora do tubo ε -insensível por causa das condições (3-18) e (3-19).
- Das condições (3-16) e (3-17), as variáveis duais serão não-nulas somente quando $|f(x_i) - y_i| \geq \varepsilon$, ou seja, apenas quando as amostras estiverem fora do tubo ε -insensível. Em outras palavras, para $|f(x_i) - y_i| < \varepsilon$, o segundo fator é não-nulo e, portanto α_i, α_i^* se anulam para satisfazer as condições complementares de KKT. Logo, consegue-se uma representação esparsa de w em termos de x_i já que não será preciso todos os x_i a fim de descrever w , só serão considerados os pontos para os quais $\alpha_i, \alpha_i^* > 0$. Tais pontos são chamados de **vetores suporte**.

Observação: Quanto menos ruído os dados tiverem, mais esparsa será a representação do problema, uma vez que mais dados estarão na região $|f(x_i) - y_i| < \varepsilon$. Assim, a representação esparsa dos dados se deve ao uso da função de custo especial ε -insensível, a qual também introduz um viés sistemático desde que, se os valores de ε são muito grandes, há uma tendência de falta de ajuste (" *underfitting* "), por exemplo, no caso extremo de ε muito grande a regressão resultante será uma constante. Entretanto, outra função de custo poderia ter sido usada sacrificando a dispersão dos dados, como por exemplo, a função de custo robusta de Hubber que tem a vantagem de não introduzir um viés adicional ao contrário da ε -insensível .

3.4

Como obter b

Nota-se que w é explicitamente determinado pelo procedimento de treinamento, ao contrário de b . No entanto, usando as condições complementares de KKT pode-se determinar b . Primeiramente, observa-se das condições (3-18) e (3-19) que $\xi_i = \xi_i^* = 0$ se $\alpha_i, \alpha_i^* < C$. Logo, pode-se simplesmente tomar todos os pontos de treinamento para os quais $0 < \alpha_i, \alpha_i^* < C$ e usar as condições (3-16) e (3-17) para obter b da seguinte maneira:

$$b = y_i - \langle w, x_i \rangle - \varepsilon \quad \text{para } 0 < \alpha_i < C$$

$$b = y_i - \langle w, x_i \rangle + \varepsilon \quad \text{para } 0 < \alpha_i^* < C$$

Um único x_i seria suficiente, no entanto para propósitos de estabilidade, é recomendável tomar uma média sobre todos os pontos x_i . Para isso, considera-se $\delta_i = \varepsilon * \text{signal}(\alpha_i - \alpha_i^*)$ onde ε é o erro de predição dado por $f(x_i) - y_i$ e b é calculado como sendo a média em todos os $i = 1, \dots, l$ de $\{\delta_i + y_i - \langle w, x_i \rangle\}$.

3.5

Os Núcleos

Até o momento foram considerados modelos lineares no espaço de entrada. Representar f como um modelo linear pode ser uma aproximação mais conveniente e até necessária em alguns casos, visto que o modelo linear é fácil de interpretar além de ser a aproximação de 1^a ordem de Taylor para f e ainda, nos casos em que se tem poucos dados, o modelo linear pode ser uma alternativa para ajustar os dados sem excesso de ajuste ("overfitting"). No entanto, aplicações complexas do mundo real requerem um espaço de hipóteses mais expressivo do que funções lineares. Muitas vezes, ao invés de uma combinação linear simples dos atributos dados, características mais abstratas dos mesmos precisam ser exploradas e as representações por núcleos oferecem uma solução alternativa.

Com o objetivo de aprender relações não-lineares com uma máquina linear, precisa-se selecionar um conjunto de características não-lineares e reescrever os dados nesta nova representação, o que é equivalente a aplicar um mapeamento não-linear fixo dos dados para um espaço característico de alta dimensionalidade onde uma máquina linear possa ser usada. Assim, pode-se construir máquinas não-lineares em dois passos: um mapeamento não-linear fixo transforma os dados em um espaço característico com alta dimensionalidade - o uso de máquinas lineares na representação dual torna possível realizar este passo implicitamente. Em seguida, uma regressão linear neste novo espaço é realizada.

A idéia básica em SVR consiste em mapear os dados de entrada $x \in \mathbb{R}^n$

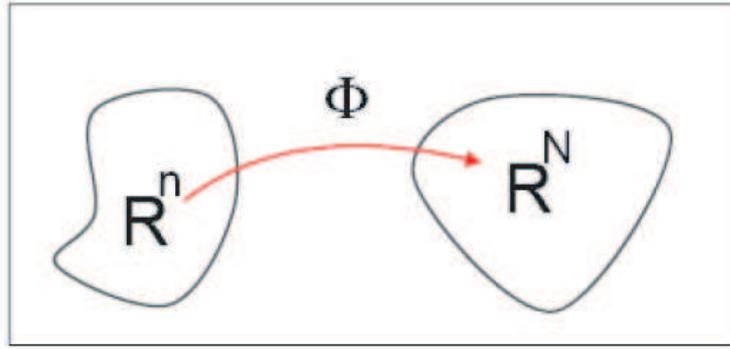


Figura 3.6: O mapeamento dos dados de entrada do espaço de dimensão R^n para um espaço característico com alta dimensionalidade R^N , onde uma regressão linear é efetuada (Smola, 1996 (22)).

em um espaço característico com alta dimensionalidade \mathcal{F} via um mapeamento não-linear $\phi : \mathbb{R}^n \rightarrow \mathcal{F}$ aumentando o poder de generalização da máquina de aprendizagem e, aplicar uma regressão linear neste novo espaço. (Veja figura 3.6, onde $\mathcal{F} = \mathbb{R}^N$).

Definição 3.1 (Função núcleo) *Uma função núcleo é uma função K tal que para todo $x, y \in \mathcal{X}$ tem-se $K(x, y) = \langle \phi(x), \phi(y) \rangle$ onde ϕ representa o mapeamento de \mathcal{X} para um (produto interno) espaço característico \mathcal{F} .*

Ao substituir $\langle x_i, x_j \rangle$ por $K(x_i, x_j)$ no algoritmo de vetores de suporte 3-14, todas as considerações anteriores serão válidas desde que ainda será calculada uma regressão linear, porém em um espaço de dimensão infinita e o tempo computacional necessário será praticamente o mesmo. O problema dual de otimização será dado da maneira a seguir:

$$\begin{aligned} & \text{maximize} \quad \begin{cases} -\frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(x_i, x_j) \\ -\varepsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) + \sum_{i=1}^l y_i(\alpha_i - \alpha_i^*) \end{cases} & (3-22) \\ & \text{sujeito a} \quad \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \quad e \quad \alpha_i, \alpha_i^* \in [0, C] \end{aligned}$$

e a função estimada será obtida do seguinte modo:

$$w = \sum_{i=1}^l (\alpha_i - \alpha_i^*)\phi(x_i) \quad e \quad f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*)K(x_i, x) + b. \quad (3-23)$$

Ao introduzir a idéia de função núcleo no problema, w não será mais descrito explicitamente.

Um dos pontos marcantes da teoria do SVR é a não necessidade de conhecer a função ϕ explicitamente, pois através da representação dual do problema de otimização, a função decisão pode ser expressa usando o produto interno entre os pontos de treinamento e os pontos de teste $f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \langle \phi(x_i), \phi(x) \rangle + b$. Ademais, existe uma maneira de calcular diretamente o produto interno $\langle \phi(x_i), \phi(x) \rangle$ no espaço característico como uma função dos valores de entrada. Este método de cálculo direto é chamado de função núcleo, cuja definição foi dada anteriormente.

A vantagem de escrever o problema de otimização na forma dual ocorre, pois, nessa reformulação o número de parâmetros independe do número de atributos e sim do tamanho da amostra. A partir da escolha de uma função núcleo, pode-se realizar um mapeamento não-linear implicitamente para um espaço característico de grande dimensionalidade sem aumentar o número de parâmetros ajustáveis, pois a função núcleo calcula o produto interno de dois vetores característicos correspondentes a duas entradas. Logo, defini-se a função núcleo diretamente e, implicitamente, o espaço característico.

3.5.1 Propriedades das funções núcleo

Pela definição de função núcleo consegue-se uma função que retorna o produto interno entre as imagens de dois valores de entrada em um espaço característico. Assim, as propriedades que podem ser atribuídas às funções núcleo são aquelas que satisfazem as do produto interno.

Funções núcleo podem ser consideradas como produtos internos generalizados. Pode-se mostrar que qualquer produto interno é uma função núcleo, mas, a propriedade de linearidade do produto interno não ocorre em funções núcleo. No entanto, algumas propriedades de produto interno que possuem uma generalização para funções núcleo serão verificadas.

– Simetria:

$$K(x, y) = \langle \phi(x), \phi(y) \rangle = \langle \phi(y), \phi(x) \rangle = K(y, x)$$

– Desigualdade de Cauchy-Schwarz:

$$\begin{aligned} K(x, y)^2 &= \langle \phi(x), \phi(y) \rangle^2 \leq \| \phi(x) \|^2 \| \phi(y) \|^2 \\ &= \langle \phi(x), \phi(x) \rangle \langle \phi(y), \phi(y) \rangle = K(x, x) K(y, y) \end{aligned}$$

O núcleo $K(x, y)$ pode ainda ter uma interpretação de medição de distância no espaço de entrada entre os exemplos x e y conforme estudos de Schölkopf (17).

$$d^2(x, y) = (\phi(x) - \phi(y))^2 = K(x, x) - 2K(x, y) + K(y, y).$$

É preciso definir quais funções $K(x, y)$ corresponderão a um produto interno em algum espaço característico \mathcal{F} . O teorema de Mercer caracteriza essas funções.

Teorema 3.2 (Mercer) *Existe um mapeamento ϕ e uma expansão da forma:*

$$K(x, y) = \langle \phi(x), \phi(y) \rangle = \sum_i \phi(x)_i \phi(y)_i \quad (3-24)$$

se e somente se, $\forall g(x)$ tal que

$$\int g(x)^2 dx \quad \text{é finito} \quad (3-25)$$

$$\text{implica} \int K(x, y)g(x)g(y)dxdy \geq 0 \quad (3-26)$$

Em muitos casos não deve ser fácil satisfazer a condição de Mercer. No entanto, a condição (3-26) deve valer para toda $g(x)$ com norma L_2 , ou seja, que satisfaça a equação (3-25).

A condição de Mercer informa se um determinado núcleo constitui ou não um produto interno em algum espaço característico \mathcal{F} , embora não mostre como construir a função Φ e até mesmo que espaço característico teremos.

Alguns núcleos usados em SVR para análise de séries temporais são estudados por Stefan Rüping (24). Exemplos de núcleos:

- Função Núcleo linear:

$$K(x, y) = \langle x, y \rangle.$$

Esta é a função núcleo mais simples. A função decisão toma a forma $f(x) = \langle w, x \rangle + b$. O núcleo linear utilizado para prever séries temporais, isto é, $x_t = f(x_{t-1}, x_{t-2}, x_{t-3}, \dots, x_{t-k})$ tem como modelo resultante um modelo auto-regressivo de ordem k (AR(k)).

- Função Núcleo polinomial:

$$K(x, y) = \langle x, y \rangle^d.$$

$$K(x, y) = (\langle x, y \rangle + 1)^d.$$

- Função Núcleo RBF: (Função Base Radial): São funções núcleo que podem ser escritas na forma $K(x, y) = f(d(x, y))$, onde d é uma métrica em \mathcal{X} e f é uma função em \mathbb{R}^+ . Usualmente a métrica vem do produto interno $d(x, y) = \|x - y\| = \sqrt{\langle x - y, x - y \rangle}$. Como exemplo de função base radial, podemos definir o núcleo função base radial Gaussiano:

$$K(x, y) = \exp(-\gamma \|x - y\|^2).$$

Neste caso, a similaridade entre dois exemplos é julgada pela distância euclidiana.

- Função Núcleo de Fourier :

$$K_F(x, y) = \frac{1 - q^2}{2(1 - 2q \cos(x - y)) + q^2}$$

Uma transformação que pode ser feita em dados de séries temporais é a transformada de Fourier. Tal representação é útil se a informação das séries temporais não recaem em valores individuais em cada ponto de tempo, porém na frequência de alguns eventos. O produto interno da expansão de Fourier de duas séries temporais pode ser calculado diretamente pela expansão de Fourier regularizada conforme descrito por Vapnik em (29) com o resultado do cálculo mostrado acima.

Pode-se ainda criar núcleos mais complexos a partir de outros mais simples. As demonstrações serão omitidas podendo ser encontradas em (7).

Proposição 3.3 *Sejam K_1 e K_2 funções núcleo em $X \times X$, $X \subseteq \mathbb{R}^n$, $a \in \mathbb{R}^+$ e f é uma função de valor real em X , $\Phi : X \rightarrow \mathbb{R}^m$, com o núcleo K_3 definido em $\mathbb{R}^m \times \mathbb{R}^m$ e \mathbf{B} uma matriz $n \times n$ simétrica (semi)definida positiva. Então as seguintes funções são funções núcleo:*

1. $K(x, z) = K_1(x, z) + K_2(x, z)$,
2. $K(x, z) = aK_1(x, z)$,
3. $K(x, z) = K_1(x, z)K_2(x, z)$,
4. $K(x, z) = f(x)f(z)$,
5. $K(x, z) = K_3(\phi(x), \phi(z))$,
6. $K(x, z) = x'\mathbf{B}z$.

Corolário 3.4 Dado K_1 uma função núcleo em $X \times X$, $x, z \in X$ e $p(x)$ um polinômio com coeficientes positivos. Então as seguintes funções também são funções núcleo:

1. $K(x, z) = p(K_1(x, z))$,
2. $K(x, z) = \exp(K(x, z))$,
3. $K(x, z) = \exp(-\gamma \|x - z\|^2)$.

Para maiores informações sobre núcleos, consulte Schölkopf e Smola (18).

3.6 Arquitetura de um SVR

A figura 3.7 descreve os passos do SVR. A entrada x (para a qual uma predição deve ser feita) e os vetores suporte $x_i, i = 1, \dots, n$ são mapeados não linearmente pela função Φ para um espaço característico, onde os produtos internos são calculados. Isto corresponde ao cálculo de funções núcleo $K(x, x_i)$. Finalmente, os resultados são combinados através dos pesos $v_i = (\alpha_i^* - \alpha_i)$ e somados com o termo constante b para fornecer a previsão de x .

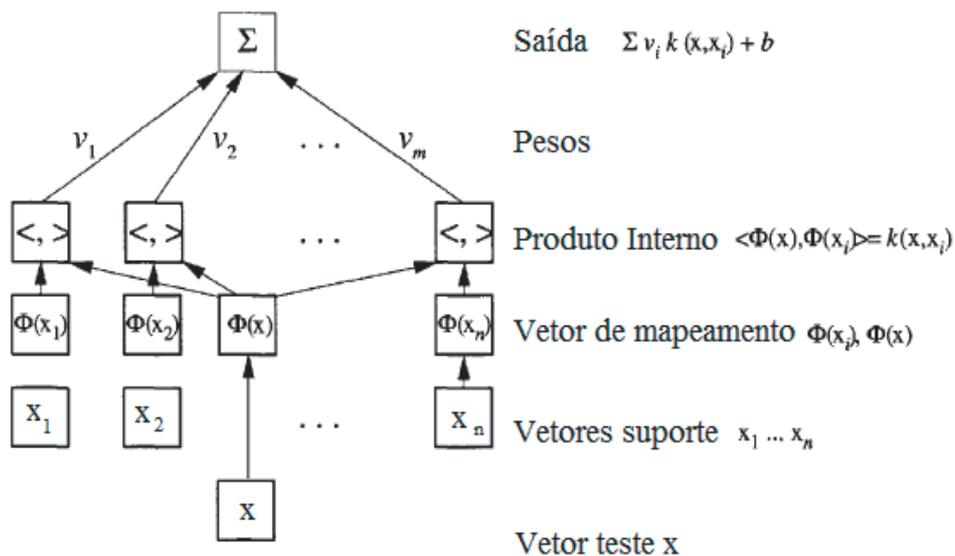


Figura 3.7: Arquitetura de máquina de vetor suporte (Schölkopf, Smola, 2001 (18)).

3.7

O SVR no software R

O R (16) é um software livre usado para análise estatística de dados, oferecendo grande variedade de técnicas estatísticas (modelos lineares e não-lineares, modelos de séries temporais, testes estatísticos clássicos, entre outros) além de produzir gráficos de qualidade. O programa está disponível em <http://www.r-project.org>, onde também encontram-se textos e tutoriais, além do código fonte para compilação e os executáveis já compilados para diferentes sistemas operacionais.

Utiliza-se a implementação LIBSVM de Chang e Lin (4) empregando o software do R-project (16), após a instalação do pacote e1071 para estimar modelos SVR. Este pacote pode ser usado tanto para classificação quanto para regressão. A matriz de dados \mathbf{X} e um vetor de respostas \mathbf{y} são definidos e, dependendo do tipo de vetor \mathbf{y} definido, o tipo padrão a ser empregado é uma C -classificação ou uma ε -regressão. Também é necessário definir a função núcleo a ser utilizada. Neste estudo, utiliza-se a função núcleo linear e a função núcleo Gaussiano (default).

Deve-se ainda definir os parâmetros livres do núcleo (para o núcleo Gaussiano, o valor default: $\gamma=1/(\text{dimensão dos dados})$), o valor do ε da função de custo ε -insensível linear (default: 0.1) e o valor da constante C (default: 1).

Nesse pacote, pode-se ainda utilizar a função *tune* que ajusta os hiperparâmetros do SVR fazendo uma pesquisa em um domínio retangular com os valores da busca fornecidos pelo usuário. Essa função usa como avaliação da performance para regressão, o erro quadrado médio.

4

Descrição dos dados

Nesta seção, os dados que são utilizados para estimar os modelos VAR, ECM e SVR são apresentados. Devido às limitações dos mercados financeiros brasileiros, não existem bancos de dados análogos aos arquivos "Fama CRSP zero coupon" e aos arquivos "Fama CRSP Treasury Bill", os quais são os recursos mais comuns para o retorno de "zero coupon bonds" nos Estados Unidos. Tal fato ocorre porque pelo menos até muito recentemente, "pure discount bonds" de vencimentos maiores do que 12 meses tinham muito pouca liquidez no mercado brasileiro, onde investidores tinham forte preferência por títulos indexados ao dólar norte-americano, a índices de inflação e à taxa Selic. O melhor que se pode fazer nessa situação é usar cotas para contratos de taxa de juros de *swap*, no qual uma taxa de juros fixa e um fluxo de taxas *overnight* para um dado intervalo de tempo são trocadas (em outras palavras, é a forma usual "fixo contra flutuante" para contratos *swap*).¹

Interessados em estudar o comportamento da curva de juros brasileira acabam recorrendo às cotações dos contratos de *swap* DI-pré negociados na Bolsa de Mercadorias e Futuros (BM&F).

Um contrato de *swap* DI-pré consiste em um acordo onde as partes trocam a remuneração incidente sobre as suas posições ativas em um dado período. Ao assumir a posição comprada, o investidor se obriga a remunerar aquele que assumiu a posição vendida com a variação acumulada do CDI (Certificado de Depósito Interbancário) de um dia ao longo do período de vigência do contrato. Dessa maneira, se a variação acumulada do CDI de um dia superar o limite acordado entre as partes (que nada mais é do que a cotação do *swap* DI-pré no momento em que o contrato é celebrado), então a parte comprada paga a diferença para a parte vendida. O contrário ocorre caso a variação acumulada do CDI de um dia seja menor do que o limite citado.

Já o CDI é um título emitido por instituições financeiras e que é transacionado exclusivamente entre elas no mercado interbancário. As instituições que necessitam de recursos para complementar as suas reservas bancárias

¹ver <http://www.bmf.com.br/portal/pages/contratos2/pdf/SwapContracts.pdf> para maiores detalhes sobre contratos de *swap* brasileiros.

emitem esses títulos, que são adquiridos por outras instituições com reservas em excesso. Quando o título possui maturidade de apenas um dia, a sua taxa de emissão é chamada "overnight", ou simplesmente "over".

A taxa do CDI de um dia é bastante influenciada pela taxa Selic, que tem um valor de referência fixado pelo Banco Central do Brasil (BCB) a cada reunião do Comitê de Política Monetária (COPOM), órgão do BCB onde são tomadas as decisões de política monetária. Ela é apurada no Sistema Especial de Liquidação e Custódia (Selic, daí seu nome) e corresponde à taxa média das operações de financiamento por um dia baseadas em títulos públicos federais e que são registradas no referido sistema ou em câmaras de compensação e liquidação de ativos. Essas transações ocorrem sob a forma de operações compromissadas, ou seja, operações onde a parte vendedora vende títulos públicos federais para a parte compradora assumindo o compromisso de recomprar os referidos títulos no dia útil seguinte (obviamente a parte compradora também assume o compromisso de vender os títulos à parte vendedora nesse mesmo momento). Estão aptas a realizar operações compromissadas com prazo de um dia útil, instituições financeiras tais como bancos e sociedades corretoras de títulos e valores mobiliários.

A base de dados estudada compreende uma amostra de taxas continuamente compostas de contratos *swap* DI-pré os quais começam em junho de 1999 e terminam em junho de 2006². Séries diárias são convertidas em uma frequência mensal pelo cálculo das taxas médias para cada mês e depois são calculados os logaritmos naturais das taxas brutas mensais. Considera-se sete diferentes intervalos de tempo entre a data de transação e a data de vencimento, os quais são um, dois, três, quatro, seis, doze e vinte e quatro meses. Estas taxas servem como substitutas para os retornos de "zero coupon bonds" que são comumente usados na literatura.

As variáveis macroeconômicas são inflação (conforme medida pelo índice IGP-DI (Índice Geral de Preços - Disponibilidade Interna), o hiato do produto (medido pelo desvio de um índice de produção industrial de sua tendência) e a taxa Selic (a qual é equivalente às taxas dos fundos FED norte americana). As três séries são calculadas como o logaritmo natural das taxas brutas e são medidas em uma frequência mensal.

² As observações do período de regime de taxa cambial controlada que durou até janeiro de 1999 foram excluídas, quando esta foi substituída pelo regime de taxa cambial flutuante que permanece até hoje. Também foram excluídas as observações entre janeiro de 1999 e maio de 1999, que foi um período de transição e a amostra possuía muitos ruído.

5

Modelos de previsão

5.1

Previsão de séries temporais

Um método comum para escrever o problema de previsão de séries temporais em uma estrutura de estimação por regressão é considerar as séries temporais como sistemas dinâmicos. Seja uma série temporal $x(t)$, se $N \in \mathbb{N}$ e $\tau > 0$ são escolhidos adequadamente, então podemos prever $x(t)$ a partir de $(x(t - \tau), \dots, x(t - N_\tau)) \in \mathbb{R}^N$. Portanto, pode-se considerar um problema de regressão onde o conjunto de treinamento consiste de entradas $(x(t - \tau), \dots, x(t - N_\tau))$ e saídas $x(t)$ para diferentes valores de t . Porém, conforme destacado por Schölkopf e Smola (2001), várias características de previsão de séries temporais tornam o problema difícil para esta aproximação ingênua para regressão. Dentre elas:

- Séries temporais são freqüentemente não estacionárias, ou seja, a distribuição de séries temporais muda com o tempo e como conseqüência, exemplos de treinamento que são gerados como descrito acima se tornam menos úteis se eles são tomados de um passado distante.
- Os diferentes exemplos de treinamento não são iid, que é uma das hipóteses assumidas por modelos de aprendizagem estatística, como o SVR.

Em contrapartida, excelentes resultados têm sido obtidos usando SVR em problemas de previsão de séries temporais (25, 11) e tais resultados não podem ser desconsiderados.

5.2

Modelo de previsão usando SVR

Foram utilizadas uma e três defasagens para cada um dos modelos propostos e cada modelo foi identificado com o número de defasagens colocado após o nome dos mesmos. Assim, SVR1 corresponde a Máquina de Suporte Vetorial para Regressão com uma defasagem enquanto que SVR3 corresponde a três defasagens.

O primeiro passo para utilizar o SVR é definir qual núcleo será utilizado. Nesse estudo, opta-se por usar uma função núcleo Gaussiano, uma vez que funções núcleo Gaussianos, ou seja, $K(x, y) = \exp(-\gamma \|x - y\|^2)$, são úteis quando não se tem nenhum conhecimento dos dados disponíveis [11]. Também estima-se SVRL1 e SVRL3 que correspondem a uma regressão usando o núcleo linear, ou seja, $K(x, y) = \langle x, y \rangle$. O objetivo dessa estimação é verificar a contribuição da não-linearidade para o modelo. Os números 1 e 3 depois do nome do método têm o mesmo significado anterior. Uma das vantagens do SVR linear é que não existe parâmetro para ser ajustado além da constante C e do parâmetro ε . Para o SVR com núcleo Gaussiano existe um parâmetro adicional a ser ajustado, o parâmetro do núcleo γ .

No modelo SVR, cada entrada corresponde a um vetor com os valores de cada uma das sete taxas *swap* e cada uma das três variáveis macro-econômicas no instante (t-1), para o caso de uma defasagem. No caso de três defasagens, além dos valores em (t-1), observa-se ainda os valores de cada uma das variáveis em (t-2) e (t-3), ou seja, existem 10 e 30 valores de entrada para uma e três defasagens, respectivamente. Já os valores alvos desejados corresponderão a um único valor da taxa de *swap* a qual se quer determinar no instante t.

Veja alguns exemplos.

1. Caso o objetivo seja prever a taxa de *swap* 30 um passo à frente usando uma defasagem, os dados de entrada devem estar representados da seguinte maneira:

$x_{t-1} = (S30_{(t-1)}, S60_{(t-1)}, S90_{(t-1)}, S120_{(t-1)}, S180_{(t-1)}, S360_{(t-1)}, S720_{(t-1)}, Selic_{(t-1)}, hiato_{(t-1)}, igpdi_{(t-1)})$. Com esses dados o valor estimado $f(x_{t-1}) = S30_t^*$ é obtido, o qual poderá ser comparado com o valor alvo desejado $y_t = S30_t$.

2. E se o objetivo for usar três defasagens?

A única diferença estaria no vetor de entrada: $x_{t-1} = (S30_{(t-1)}, S30_{(t-2)}, S30_{(t-3)}, S60_{(t-1)}, S60_{(t-2)}, S60_{(t-3)}, S90_{(t-1)}, S90_{(t-2)}, S90_{(t-3)}, S120_{(t-1)}, S120_{(t-2)}, S120_{(t-3)}, S180_{(t-1)}, S180_{(t-2)}, S180_{(t-3)}, S360_{(t-1)}, S360_{(t-2)}, S360_{(t-3)}, S720_{(t-1)}, S720_{(t-2)}, S720_{(t-3)}, Selic_{(t-1)}, Selic_{(t-2)}, Selic_{(t-3)}, hiato_{(t-1)}, hiato_{(t-2)}, hiato_{(t-3)}, igpdi_{(t-1)}, igpdi_{(t-2)}, igpdi_{(t-3)})$.

3. E se o interesse está em fazer previsões três passos à frente usando uma defasagem?

Seria preciso calcular as estimativas de cada uma das dez variáveis para o instante t, como no exemplo 1. Em seguida, essas estimativas seriam

usadas como elementos do vetor de entrada para o cálculo da previsão de cada uma das variáveis no instante seguinte ($t+1$), logo uma previsão dois passos à frente seria obtida. Por último, as estimativas das variáveis obtidas no último passo seriam utilizadas como elementos do vetor de entrada para se obter a estimativa procurada, ou seja, a previsão três passos à frente da taxa *swap* desejada. É importante notar que embora o vetor de entrada seja o mesmo para cada etapa de previsão, o modelo SVR mudará, posto que cada previsão de uma taxa *swap* diferente terá valores diferentes para os parâmetros ε, γ e C .

Para se fazer previsões com seis, nove e doze passos à frente, o procedimento é análogo ao anterior, e para usar três defasagens, basta modificar o vetor de entrada.

A escolha dos parâmetros C, ε e γ (para a função núcleo Gaussiano) é feita baseada num conjunto de validação através de uma pesquisa em um domínio retangular utilizando um conjunto de valores fornecidos pelo usuário. Tal pesquisa é feita utilizando a função *tune* do pacote *e1071* do *software* R. Para cada uma das dez variáveis do modelo, para cada um dos modelos com uma e três defasagens e para ambas funções núcleo é feita essa busca dos parâmetros.

Os valores usados para busca dos parâmetros são exibidos a seguir.

γ (função núcleo gaussiano): (0.0001, 0.0003, 0.0005, 0.0007, 0.0009, 0.001, 0.003, 0.005, 0.007, 0.009, 0.01, 0.03, 0.05, 0.07, 0.09, 0.1, 0.3, 0.5).

ε : (0.0001, 0.0003, 0.0005, 0.0007, 0.0009, 0.001, 0.003, 0.005, 0.007, 0.009, 0.01, 0.03, 0.05, 0.07, 0.09, 0.1, 0.3, 0.5).

C : 1-200 (de 5 em 5).

A tabela 5.1 exhibe os parâmetros escolhidos.

PARÂMETROS SVR1										
	<i>log_s30</i>	<i>log_s60</i>	<i>log_s90</i>	<i>log_s120</i>	<i>log_s180</i>	<i>log_s360</i>	<i>log_s720</i>	<i>log_selic</i>	<i>log_hiato</i>	<i>log_igpdi</i>
γ	0.01	0.007	0.0009	0.003	0.005	0.003	0.005	0.005	0.01	0.0001
C	26	191	161	146	21	41	21	101	196	191
ε	0.1	0.0001	0.09	0.07	0.1	0.3	0.3	0.03	0.3	0.3
PARÂMETROS SVR3										
	<i>log_s30</i>	<i>log_s60</i>	<i>log_s90</i>	<i>log_s120</i>	<i>log_s180</i>	<i>log_s360</i>	<i>log_s720</i>	<i>log_selic</i>	<i>log_hiato</i>	<i>log_igpdi</i>
γ	0.0005	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.005	0.1	0.1
C	136	196	196	176	196	96	156	16	6	11
ε	0.0001	0.3	0.0001	0.0001	0.01	0.007	0.009	0.07	0.3	0.07
PARÂMETROS SVRL1										
	<i>log_s30</i>	<i>log_s60</i>	<i>log_s90</i>	<i>log_s120</i>	<i>log_s180</i>	<i>log_s360</i>	<i>log_s720</i>	<i>log_selic</i>	<i>log_hiato</i>	<i>log_igpdi</i>
C	1	6	9	11	76	41	16	146	1	1
ε	0.1	0.1	0.0007	0.01	0.03	0.1	0.1	0.0003	0.05	0.1
PARÂMETROS SVRL3										
	<i>log_s30</i>	<i>log_s60</i>	<i>log_s90</i>	<i>log_s120</i>	<i>log_s180</i>	<i>log_s360</i>	<i>log_s720</i>	<i>log_selic</i>	<i>log_hiato</i>	<i>log_igpdi</i>
C	1	11	1	56	91	11	1	1	1	36
ε	0.05	0.03	0.0001	0.05	0.0007	0.007	0.03	0.0003	0.1	0.005

Tabela 5.1: Valores escolhidos para cada parâmetro em cada um dos modelos.

5.2.1

Metodologia de aprendizagem

Todos os 7 retornos de contratos *swap* DI-pré e todas as variáveis macro-econômicas são divididas em três partes de acordo com a seqüência no tempo. A primeira parte (dados de junho de 1999 a janeiro de 2004) é usada para treinamento, a segunda parte para validação é usada para selecionar os parâmetros ótimos para SVRs (dados de fevereiro de 2004 a janeiro de 2005) e a última parte é usada para testes (dados de fevereiro de 2005 a junho de 2006). Cada uma das partes possui respectivamente 65%, 15% e 20% dos dados.

5.3

Modelo de previsão usando Vetor Auto-regressivo (VAR) e Modelo de Correção de Erros (ECM)

Emprega-se o *software* EVIEWS 4, um programa de análise estatística e econométrica, que possui ferramentas para realizar previsões, a fim de modelarmos dois modelos de vetores auto-regressivos com uma e três defasagens.

A partir da observação dos gráficos de cada série e da realização de testes da raiz unitária de *Dickey-Fuller Augmented* foi verificado que as séries possuíam raiz unitária. Logo, também serão empregados dois modelos de correção de erros com uma e três defasagens. Testes de cointegração de *Johansen* indicaram a existência de 6 relações cointegrantes para uma defasagem e quatro relações cointegrantes para três defasagens.

Ambos os modelos, VAR e ECM, foram estimados e previsões foram realizadas a fim de serem comparados com os modelos de aprendizagem SVR.

Mais informações do *software* podem ser encontradas em <http://www.eviews.com>.

5.4

Critérios de desempenho

A performance de predição baseia-se nas métricas estatísticas seguintes:

- RMSE (Raiz do erro quadrado médio), MAE (erro absoluto médio): medem os desvios entre os valores previstos e atuais. Quanto menor esses valores, mais próximos estão os valores das séries temporais previstas dos valores atuais.
- DS (simetria direcional): fornece uma indicação de que a direção prevista está correta, dado na forma percentual e um valor grande sugere um preditor melhor.

$RMSE = \sqrt{\sum_{i=1}^n \frac{1}{n} (a_i - p_i)^2}$
$MAE = \frac{1}{n} * \sum_{i=1}^n a_i - p_i $
$DS = \frac{100}{n} * \sum_{i=1}^n d_i$ $d_i = \begin{cases} 1, & se(a_i - a_{i-1})(p_i - p_{i-1}) \geq 0 \\ 0, & \text{caso contrário} \end{cases}$
$WDS = \frac{\sum_{i=1}^n d'_i a_i - p_i }{\sum_{i=1}^n d_i a_i - p_i }$ $d'_i = \begin{cases} 0, & se(a_i - a_{i-1})(p_i - p_{i-1}) \geq 0 \\ 1, & \text{caso contrário} \end{cases}$ $d_i = \begin{cases} 1, & se(a_i - a_{i-1})(p_i - p_{i-1}) \geq 0 \\ 0, & \text{caso contrário} \end{cases}$
$CP = \frac{100}{n} \sum_{i=1}^n d_i$ $d_i = \begin{cases} 1, & se(p_i - p_{i-1}) > 0, (a_i - a_{i-1})(p_i - p_{i-1}) \geq 0 \\ 0, & \text{caso contrário} \end{cases}$
$CD = \frac{100}{n} \sum_{i=1}^n d_i$ $d_i = \begin{cases} 1, & se(p_i - p_{i-1}) < 0, (a_i - a_{i-1})(p_i - p_{i-1}) \geq 0 \\ 0, & \text{caso contrário} \end{cases}$
$U = \frac{RMSE_i}{RMSE_j}$

Tabela 5.2: Métricas utilizadas (a_i e p_i representam os valores atuais e previstos, respectivamente).

- WDS (simetria direcional ponderada): mede a magnitude do erro e da direção de previsão, penalizando os erros relacionados à direção prevista incorretamente e recompensa aqueles associados com as direções previstas corretamente. Quanto menor o valor de WDS, melhor a performance de previsão tanto em magnitude quanto na direção.
- CP e CD (tendência para cima ou para baixo correta, respectivamente): medem a correta previsão de tendência para cima e para baixo, respectivamente, e em termos percentuais.
- Estatística de Theil-U : mede o desempenho de previsão relativa entre dois modelos, utilizando o RMSE como métrica.

As definições desses critérios estão ilustradas na tabela 5.2.

Em relação a estatística de Theil-U, $RMSE_{i(j)}$ representa a raiz do erro quadrado médio do método $i(j)$; $i = \text{SVR1}$ ou SVR3 e $j = \text{SVR1}$, SVR3 , VAR1 , VAR3 , ECM1 , ECM3 , SVRL1 , SVRL3 ($i \neq j$). Se $U < 1$ indica que a primeira previsão tem um RMSE menor do que o da segunda.

Os critérios DS, WDS, CP e CD fornecem uma boa medida da consistência na previsão da direção das taxas.

6

Resultados experimentais

Este capítulo tem por objetivo estudar os resultados. O procedimento para examinar previsões fora da amostra é muito convencional. Estima-se cada um dos modelos para cada uma das taxas de *swap*, usando observações de junho de 1999 até e inclusive janeiro de 2005. Em seguida, calcula-se a previsão k meses a frente ($k=1, 3, 6, 9, 12$ meses) das sete taxas *swap*. O procedimento é repetido seis vezes adicionando uma observação a amostra até que a mesma compreenda dados de junho de 1999 até junho de 2005. As observações que não são usadas durante o processo de estimação são utilizadas para calcular os erros de previsão. Os critérios usados para comparação dos modelos foram anteriormente descritos.

Na tabela 6.1 são exibidos os resultados para as sete taxas de *swap* para previsões um passo à frente e verifica-se que o método ECM1 apresenta os menores erros de previsão para as taxas de *swap* 30, *swap* 60, *swap* 90 e *swap* 120 enquanto que o método SVRL1 apresenta menores erros de previsão para as *swaps* 180 e 360. Para *swap* 720, o SVRL1 e SVRL3 apresentam erros ligeiramente superiores ao VAR1, mas comparáveis a este último. Destaca-se o desempenho da previsão do ECM1 para *swaps* 90 e 120, pois apresentam não apenas os menores erros de previsão como também os melhores resultados para DS e WDS. Nestes dois últimos testes, para as *swaps* 90 e 120, também o VAR1 se destaca.

Com relação à consistência das previsões, os testes DS e WDS são bons indicadores da direção corretamente prevista das taxas. Embora o SVR1 e SVR3 não apresentem os menores erros médios quadrados de previsão, observamos que SVR3 apresenta os melhores resultados para DS e WDS para as *swaps* 30, 60 (onde o SVR1 e SVRL3 também têm ótimo desempenho), 360 e 720 e, resultados comparáveis para DS e WDS para taxa *swap* 180 onde SVRL1 apresenta menor WDS.

A tabela 6.2 mostra o coeficiente de Theil-U, que se baseia no RMSE e assim, pode mais uma vez ser verificado o desempenho ruim do SVR1 e do SVR3 para um mês de previsão. Basta observar o desempenho deles em relação ao ECM1. Os métodos SVR1 e SVR3 só superam alguns métodos

e apenas para as *swaps* 120, 180, 360 e 720, como pode ser constatado pelo desempenho de ambos em relação ao ECM3 para *swap* 360. Pode-se mencionar o bom desempenho do SVR1 para *swap* 360 em relação aos outros métodos, com exceção do SVRL1 que apresenta uma melhora de 15% na previsão quando comparado ao SVR1. Isso poderia ser explicado utilizando um simples feito que se conhece da Matemática: localmente as funções diferenciáveis podem ser bem aproximadas por uma função linear.

A tabela 6.3 mostra os resultados para previsões três passos à frente. Os métodos ECM1 e VAR3 apresentam resultados comparáveis para os menores erros de previsão para as *swaps* 30, 60, 90, 120 e 180. Os menores erros para as taxas 360 e 720 são dados por SVR1 e VAR1, respectivamente.

O método SVRL1 apresenta menor WDS e maior DS para *swaps* 360 e 720 e os melhores resultados para DS e WDS para *swaps* 120 e 180 são dados pelo VAR1. Para *swaps* 30, 60, 90, os melhores resultados desses testes são dados, respectivamente, por SVRL3, SVR3 e SVR1.

A tabela 6.4 mostra a estatística Theil-U, os resultados obtidos são análogos aos casos de previsões um passo à frente, ou seja, o SVR1 e SVR3 só superam alguns métodos. Como por exemplo, pode-se observar a razão entre o SVR1 e o ECM3 para *swap* 360 com uma melhora de 83% e a razão SVR3/ECM3 para *swap* 360 com melhora de 69%. Também pode-se ressaltar o desempenho do ECM1 sobre SVR1 e SVR3, como exemplo, basta verificar a razão entre SVR1 e ECM1 para *swap* 30 com melhora de 67% na previsão, bem como a razão entre SVR3 e ECM1 para a mesma *swap* e melhora de 72% usando o modelo ECM1.

A tabela 6.5 mostra os testes para seis passos à frente. Para *swaps* 30, 60, 90 e 720, o método ECM1 ainda apresenta os menores erros RMSE e MAE, no entanto para DS e WDS os melhores resultados são dados por SVR3 para *swap* 30, já para as *swaps* 60 e 90 o SVR3 apresenta maior DS embora ECM1 ainda se destaque no WDS. Já para *swap* 720, os melhores resultados para DS e WDS são dados pelo VAR3. Deve-se ressaltar o melhor desempenho do SVR1 para *swaps* 120 e 180 em todos os testes, bem como o desempenho do SVR3 na previsão da *swap* 360 para todos os testes, com exceção do WDS em que o SVR1 se destaca.

Não se pode deixar de analisar um resultado importante: para horizonte de seis meses e períodos mais longos, que podem ser avaliados nas tabelas 6.7 e 6.9, o SVRL1 e SVRL3 apresentam os maiores erros de previsão, principalmente para SVRL3 no horizonte de 12 meses de previsão. Por essa razão, foram obtidos valores muito grandes nos desvios dos valores atuais dos previstos, tornando os resultados obtidos para WDS inconsistentes e, portanto

os mesmos devem ser desconsiderados. Por esse motivo, observar um único resultado de teste pode ser inconclusivo e levar a erros nas escolhas. Assim, é importante uma análise conjunta dos resultados obtidos.

Na tabela 6.6, conforme visto na tabela 6.5, observa-se que o SVR1 se destaca em relação a todos os outros métodos na previsão das taxas de *swaps* 120 e 180, apresentando uma melhora entre 19% a 76%, desconsiderando-se as melhoras em relação ao SVRL1 e SVRL3. Já o SVR3 se destaca na previsão das taxas *swap* 360 e 720 em relação aos outros métodos e apresenta uma melhora na previsão de 5% a 91%, com exceção do ECM1 que apresenta uma melhora na previsão da taxa de *swap* 720 em relação ao SVR3 de 33%.

Na tabela 6.7 estão os resultados para previsões nove passos à frente. Neste caso, os melhores resultados para todos os testes para a taxa *swap* 30 é dado pelo SVR1, já para as *swaps* 60, 90, 120, 180 e 360 o SVR3 se destaca. A única exceção é a taxa de *swap* 720, cujos menores valores para RMSE e MAE são dados pelo ECM1, embora SVR1 e SVR3 apresentem os melhores resultados para DS e WDS.

A tabela 6.8 ressalta a superioridade do SVR1 e SVR3 com poucas exceções em relação aos outros métodos, apresentando uma melhora de 4% a 81%. O SVR1 é melhor do que os outros métodos em relação à taxa de *swap* 30 com melhora de 4 a 68% na previsão, já em relação às taxas de *swap* 60, 90, 120, 180 e 360 o SVR3 tem uma performance superior variando entre 15% a 81% e para *swap* 720, o VAR3 e ECM1 superam o SVR3 em 22% e 36%, respectivamente. Mais uma vez os resultados em relação a SVRL1 e SVRL3 são desconsiderados.

A tabela 6.9, exibe os resultados para previsões doze passos à frente. O SVR3 apresenta os menores erros RMSE e MAE para as taxas de *swap* 30, 60 e 90. Já para DS e WDS, os melhores resultados para as respectivas taxas são dados por VAR1, SVR3 e SVR1. Destaca-se os menores erros de previsão obtidos pelo VAR3 para as taxas de *swap* 120, 180, 360 e 720, enquanto que os melhores valores para DS e WDS são dados por SVR3 (taxas de *swaps* 120 e 180), SVR1 (taxa de *swap* 360) e VAR3 (taxa de *swap* 720).

Na tabela 6.10, observa-se que o SVR3 se destaca em relação a todos os outros métodos para taxas de *swap* 30, 60, e 90, apresentando uma melhora entre 4% e 75%. Para taxa de *swap* 120 e 180, o SVR3 é superado pelo VAR3 que apresenta uma melhoria de 24% (taxa de *swap* 120) e 42% (taxa de *swap* 180) e para as taxas de *swap* 360 e 720 é superado pelo VAR3 (60% e 75%) e ECM1 (20% e 59%).

Também gráficos dos valores previstos nos dados de teste para cada método e cada uma das sete taxas de *swap*, considerando todos horizontes

de previsão são apresentados. Os métodos SVRL1 e SVRL3 só foram incluídos para previsões um e três passos à frente, devido a sua performance ruim na previsão para os horizontes de previsão subsequentes, conforme evidenciado nos resultados dos erros de previsão nas tabelas.

Resumindo a análise dos resultados: o método ECM1 apresenta-se como bom preditor para taxas curtas e intermediárias (taxas de *swap* 30, 60, 90, 120 e 180) no curto e médio prazo (um e três passos à frente na previsão) e também mostra bom desempenho para taxa de *swap* 720 para horizontes de previsão de seis e nove meses. Já o VAR3 se destacou na previsão das taxas 180, 360 e 720 em doze passos à frente. Por sua vez, o método SVR apresenta um bom desempenho para taxas intermediárias e longas (taxas de *swap* 120, 180, 360 e 720) na previsão de curto e médio prazo (um, três e seis meses) e ótimo desempenho no longo prazo (nove e doze meses) em relação a quase todas as taxas, além de atuar como ótimo indicador da direção das taxas em todos os horizontes de previsão. Também merece destaque o melhor desempenho do SVR3 sobre o SVR1 principalmente no longo prazo, ou seja, incorporar mais informação aumenta o poder preditivo. Vale lembrar o fato que treinar uma SVR depende somente do número de amostras e não da dimensão delas.

Através dos resultados obtidos, pode-se ainda comprovar que utilizar núcleos não-lineares apresenta uma melhor performance do que o uso de núcleos lineares. Os modelos com núcleos lineares tiveram uma boa performance nas previsões de curto prazo (um e três meses), embora no longo prazo, os resultados foram pífios além de levarem mais tempo no treinamento do SVR. Assim, o SVR1 e SVR3 usando núcleos Gaussianos apresentaram melhores resultados comparados com SVRL1 e SVRL3 que utilizaram núcleos lineares constatando-se o ganho em incorporar a não-linearidade ao modelo.

PREVISÃO 1 PASSO À FRENTE

log_s30

	RMSE	MAE	DS	WDS	CP	CD
VAR1	0,003025	0,002978	60	0,870843	60	0
VAR3	0,002937	0,002701	40	0,865252	40	0
ECM1	0,00103	0,000882	60	1,444114	60	0
ECM3	0,003248	0,00287	40	0,675908	40	0
SVR1	0,006187	0,005337	60	0,627645	60	0
SVR3	0,005411	0,004391	100	0	80	20
SVRL1	0,003872	0,003574	80	0,105662	80	0
SVRL3	0,002622	0,002014	100	0	80	20

log_s60

	RMSE	MAE	DS	WDS	CP	CD
VAR1	0,003329	0,003279	80	0,353824	60	20
VAR3	0,003404	0,003043	40	0,786476	40	0
ECM1	0,001219	0,001098	80	0,595284	60	20
ECM3	0,003765	0,00322	40	0,556352	40	0
SVR1	0,004132	0,003647	100	0	80	20
SVR3	0,007393	0,007356	100	0	80	20
SVRL1	0,004025	0,003807	80	0,098148	80	0
SVRL3	0,001821	0,001736	100	0	80	20

log_s90

	RMSE	MAE	DS	WDS	CP	CD
VAR1	0,003712	0,003672	100	0	60	40
VAR3	0,003657	0,003049	60	0,050733	40	20
ECM1	0,001532	0,001469	100	0	60	40
ECM3	0,004214	0,003667	60	0,121542	40	20
SVR1	0,006206	0,005542	80	0,11488	60	20
SVR3	0,00485	0,004363	60	0,183155	60	0
SVRL1	0,002997	0,002572	80	0,067076	60	20
SVRL3	0,004154	0,00353	60	0,719383	40	20

log_s120

	RMSE	MAE	DS	WDS	CP	CD
VAR1	0,00318	0,003101	100	0	60	40
VAR3	0,003483	0,002957	60	0,102758	40	20
ECM1	0,001576	0,001435	100	0	60	40
ECM3	0,004391	0,003833	60	0,178746	40	20
SVR1	0,004145	0,003683	80	0,04904	60	20
SVR3	0,004066	0,003479	60	0,213284	60	0
SVRL1	0,002361	0,001962	80	0,031587	60	20
SVRL3	0,002289	0,001877	80	0,037504	60	20

<i>log_s180</i>						
	RMSE	MAE	DS	WDS	CP	CD
VAR1	0,003539	0,003397	80	0,133009	40	40
VAR3	0,003365	0,002954	40	1,335789	20	20
ECM1	0,002341	0,002062	80	0,075683	40	40
ECM3	0,004863	0,004035	40	1,441652	20	20
SVR1	0,004796	0,003988	60	0,195262	40	20
SVR3	0,004797	0,003889	80	0,066676	40	40
SVRL1	0,001717	0,001349	80	0,042318	40	40
SVRL3	0,003865	0,003553	60	0,621892	40	20
<i>log_s360</i>						
	RMSE	MAE	DS	WDS	CP	CD
VAR1	0,003851	0,00328	60	0,882269	20	40
VAR3	0,006138	0,005352	40	3,526784	20	20
ECM1	0,003803	0,003445	80	0,102103	40	40
ECM3	0,007854	0,005401	40	2,927699	20	20
SVR1	0,003749	0,003415	80	0,076044	40	40
SVR3	0,004339	0,004148	100	0	40	60
SVRL1	0,003173	0,002796	60	0,923533	20	40
SVRL3	0,006282	0,005739	60	0,728215	40	20
<i>log_s720</i>						
	RMSE	MAE	DS	WDS	CP	CD
VAR1	0,004338	0,003573	60	0,677911	20	40
VAR3	0,009825	0,008788	60	1,414752	40	20
ECM1	0,00549	0,004492	60	0,742189	20	40
ECM3	0,010832	0,007263	40	3,4119	20	20
SVR1	0,00997	0,009678	80	0,390124	40	40
SVR3	0,006189	0,005657	80	0,108004	20	60
SVRL1	0,004726	0,003806	40	4,497768	0	40
SVRL3	0,00449	0,003637	40	8,397764	20	20

Tabela 6.1: Resultados para previsões 1 passo à frente.

1 passo a frente	<i>log_s30</i>	<i>log_s60</i>	<i>log_s90</i>	<i>log_s120</i>	<i>log_s180</i>	<i>log_s360</i>	<i>log_s720</i>
SVR1/VAR(1)	2,045404	1,241336	1,671969	1,303541	1,355318	0,973624	2,298169
SVR1/VAR(3)	2,106098	1,213764	1,696906	1,190047	1,425325	0,610862	1,014708
SVR1/VEC1	6,004810	3,391022	4,052026	2,630836	2,049122	0,985882	1,816112
SVR1/VEC3	1,904690	1,097531	1,472797	0,944174	0,986284	0,477369	0,920434
SVR1/SVRL1	1,597868	1,026694	2,070458	1,755594	2,792677	1,181836	2,109364
SVR1/SVRL3	2,359477	2,269613	1,493876	1,811026	1,240895	0,596804	2,220283
SVR3/VAR(1)	1,788971	2,220864	1,306638	1,278691	1,355495	1,126624	1,426734
SVR3/VAR(3)	1,842055	2,171536	1,326126	1,167361	1,425511	0,706856	0,629944
SVR3/VEC1	5,251985	6,066848	3,166644	2,580683	2,049390	1,140809	1,127466
SVR3/VEC3	1,665898	1,963584	1,150986	0,926174	0,986413	0,552385	0,571418
SVR3/SVR1	0,874630	1,789091	0,781497	0,980937	1,000131	1,157145	0,620813
SVR3/SVRL1	1,397542	1,836849	1,618056	1,722126	2,793042	1,367555	1,309521
SVR3/SVRL3	2,063668	4,060545	1,167459	1,776501	1,241058	0,690588	1,378381

Tabela 6.2: Teste de Theil para previsões 1 passo à frente.

PREVISÃO 3 PASSOS À FRENTE

log_s30

	RMSE	MAE	DS	WDS	CP	CD
VAR1	0,012547	0,012271	60	0,470093	40	20
VAR3	0,004735	0,003289	60	0,120134	20	40
ECM1	0,004326	0,003557	60	0,178163	40	20
ECM3	0,005562	0,00478	40	0,493546	20	20
SVR1	0,013294	0,012539	40	1,72265	20	20
SVR3	0,015645	0,014006	60	0,410758	40	20
SVRL1	0,012894	0,01146	40	0,594815	40	0
SVRL3	0,008424	0,007124	80	0,10858	40	40

log_s60

	RMSE	MAE	DS	WDS	CP	CD
VAR1	0,012756	0,012485	80	0,192173	40	40
VAR3	0,004868	0,003518	40	0,885805	20	20
ECM1	0,004475	0,003511	80	0,102931	40	40
ECM3	0,00676	0,005807	40	0,622426	20	20
SVR1	0,011421	0,010739	60	0,465613	40	20
SVR3	0,015081	0,014884	100	0	40	60
SVRL1	0,014129	0,012227	40	0,516772	40	0
SVRL3	0,013907	0,011803	60	1,348012	20	40

log_s90

	RMSE	MAE	DS	WDS	CP	CD
VAR1	0,012617	0,012306	60	0,646722	20	40
VAR3	0,004846	0,003792	60	0,829642	20	40
ECM1	0,004766	0,003586	60	0,608433	20	40
ECM3	0,008137	0,007104	60	0,489442	20	40
SVR1	0,012102	0,011162	60	0,45148	20	40
SVR3	0,014685	0,013174	40	1,394195	20	20
SVRL1	0,010955	0,009148	20	20,03946	0	20
SVRL3	0,012875	0,011387	40	2,215827	0	40

log_s120

	RMSE	MAE	DS	WDS	CP	CD
VAR1	0,011619	0,011256	80	0,153403	20	60
VAR3	0,005155	0,004382	60	0,846251	20	40
ECM1	0,004835	0,003665	60	0,55057	20	40
ECM3	0,010303	0,009064	60	0,673176	20	40
SVR1	0,010253	0,00949	60	0,431458	20	40
SVR3	0,012105	0,010067	40	1,057406	20	20
SVRL1	0,010172	0,008942	20	16,84469	0	20
SVRL3	0,041062	0,030825	80	0,522948	20	60

<i>log_s180</i>						
	RMSE	MAE	DS	WDS	CP	CD
VAR1	0,010899	0,010263	60	0,6289	0	60
VAR3	0,006647	0,005494	20	1,93193	0	20
ECM1	0,006043	0,00479	40	5,01864	0	40
ECM3	0,013636	0,011654	40	3,513111	0	40
SVR1	0,00845	0,006916	20	13,36482	0	20
SVR3	0,010482	0,008113	40	3,080813	0	40
SVRL1	0,008117	0,007121	60	0,679034	0	60
SVRL3	0,056916	0,043658	60	3,394157	0	60

<i>log_s360</i>						
	RMSE	MAE	DS	WDS	CP	CD
VAR1	0,007783	0,006987	40	0,942643	0	40
VAR3	0,011031	0,009221	0	-	0	0
ECM1	0,007028	0,005639	20	3,82065	0	20
ECM3	0,020404	0,01706	20	2,66811	0	20
SVR1	0,003384	0,002643	40	16,23979	0	40
SVR3	0,006354	0,004942	40	1,919203	0	40
SVRL1	0,009652	0,007964	80	0,050088	20	60
SVRL3	0,028958	0,020596	40	5,418459	0	40

<i>log_s720</i>						
	RMSE	MAE	DS	WDS	CP	CD
VAR1	0,00479	0,004507	60	0,696748	0	60
VAR3	0,01518	0,013278	0	-	0	0
ECM1	0,008813	0,007316	40	0,823387	0	40
ECM3	0,025101	0,021429	20	2,564224	0	20
SVR1	0,015685	0,015343	40	1,611154	0	40
SVR3	0,006196	0,004621	40	2,504732	0	40
SVRL1	0,013562	0,011606	80	0,039508	20	60
SVRL3	0,006417	0,00471	20	7,530603	0	20

Tabela 6.3: Resultados para previsões 3 passos à frente.

3 passos a frente	<i>log_s30</i>	<i>log_s60</i>	<i>log_s90</i>	<i>log_s120</i>	<i>log_s180</i>	<i>log_s360</i>	<i>log_s720</i>
SVR1/VAR(1)	1,059546	0,895365	0,959183	0,882434	0,775279	0,434726	3,274710
SVR1/VAR(3)	2,807861	2,346074	2,497200	1,988822	1,271165	0,306723	1,033249
SVR1/VEC1	3,072785	2,551951	2,539226	2,120669	1,398204	0,481425	1,779808
SVR1/VEC3	2,390137	1,689480	1,487182	0,995175	0,619678	0,165834	0,624875
SVR1/SVRL1	1,030978	0,808333	1,104702	1,007969	1,041059	0,350556	1,156557
SVR1/SVRL3	1,578048	0,821246	0,939984	0,249697	0,148464	0,116846	2,444081
SVR3/VAR(1)	1,246922	1,182306	1,163883	1,041811	0,961707	0,816300	1,293535
SVR3/VAR(3)	3,304421	3,097929	3,030131	2,348025	1,576838	0,575945	0,408141
SVR3/VEC1	3,616196	3,369785	3,081126	2,503684	1,734426	0,903990	0,703037
SVR3/VEC3	2,812824	2,230914	1,804564	1,174914	0,768690	0,311393	0,246830
SVR3/SVR1	1,176846	1,320474	1,213411	1,180611	1,240467	1,877737	0,395007
SVR3/SVRL1	1,213303	1,067383	1,340458	1,190019	1,291399	0,658253	0,456849
SVR3/SVRL3	1,857120	1,084433	1,140588	0,294795	0,184165	0,219407	0,965430

Tabela 6.4: Teste de Theil para previsões 3 passos à frente.

PREVISÃO 6 PASSOS À FRENTE

log_s30

	RMSE	MAE	DS	WDS	CP	CD
VAR1	0,01499989	0,013177507	40	5,086776	0	40
VAR3	0,011483311	0,009883141	20	2,638136	0	20
ECM1	0,008449411	0,007316559	20	2,859991	0	20
ECM3	0,009613769	0,008682476	40	5,384587	0	40
SVR1	0,014819238	0,012956141	40	1,738183	0	40
SVR3	0,017554192	0,014697991	40	1,328053	0	40
SVRL1	0,03912583	0,03403454	0	-	0	0
SVRL3	0,096759387	0,072370867	40	0,173929	0	40

log_s60

	RMSE	MAE	DS	WDS	CP	CD
VAR1	0,013438106	0,011436381	40	6,656036	0	40
VAR3	0,011387639	0,009733114	20	2,469548	0	20
ECM1	0,008644234	0,007447949	40	0,452693	0	40
ECM3	0,012167598	0,011205936	40	4,276254	0	40
SVR1	0,011410549	0,009211651	40	1,569288	0	40
SVR3	0,013807585	0,012681214	60	1,335708	0	60
SVRL1	0,044026286	0,037807601	0	-	0	0
SVRL3	0,993033182	0,742509719	40	0,22442	0	40

log_s90

	RMSE	MAE	DS	WDS	CP	CD
VAR1	0,011480655	0,00927313	60	1,274853	0	60
VAR3	0,011274766	0,009486313	20	2,182511	0	20
ECM1	0,008863325	0,007689832	40	0,472478	0	40
ECM3	0,014992445	0,013985704	40	3,682913	0	40
SVR1	0,010515791	0,008521198	40	1,445854	0	40
SVR3	0,01418523	0,011725296	60	2,43046	0	60
SVRL1	0,03134233	0,025191098	40	1,338886	0	40
SVRL3	0,070626234	0,045715068	60	3,02781	0	60

log_s120

	RMSE	MAE	DS	WDS	CP	CD
VAR1	0,009257439	0,007052401	60	1,347498	0	60
VAR3	0,010277874	0,008653279	40	0,591554	0	40
ECM1	0,008546431	0,007398168	40	0,487772	0	40
ECM3	0,017707314	0,016472255	40	3,530507	0	40
SVR1	0,006853063	0,006065692	80	0,138956	0	80
SVR3	0,01134724	0,008723284	40	19,82217	0	40
SVRL1	0,030815872	0,027450586	60	0,254765	0	60
SVRL3	2,397508712	1,787129809	40	0,222336	0	40

<i>log_s180</i>						
	RMSE	MAE	DS	WDS	CP	CD
VAR1	0,006163078	0,005076522	60	0,574485	0	60
VAR3	0,010707125	0,009041719	40	0,476629	0	40
ECM1	0,008601478	0,00739104	40	0,508905	0	40
ECM3	0,021460332	0,019854831	40	3,450277	0	40
SVR1	0,004981086	0,004626384	80	0,039983	0	80
SVR3	0,007411916	0,005361739	60	6,096691	0	60
SVRL1	0,021761433	0,020565385	60	0,446848	0	60
SVRL3	3,235401305	2,458665178	40	0,233075	0	40
<i>log_s360</i>						
	RMSE	MAE	DS	WDS	CP	CD
VAR1	0,00406718	0,003342413	60	0,495901	0	60
VAR3	0,01436803	0,012516324	60	0,19004	20	40
ECM1	0,006961966	0,006109767	60	0,395363	20	40
ECM3	0,025370901	0,022522709	60	0,688154	20	40
SVR1	0,012008708	0,011880776	80	0,182764	0	80
SVR3	0,002226109	0,001657659	80	0,781978	20	60
SVRL1	0,045080402	0,037573219	40	0,484627	0	40
SVRL3	1,905383202	1,446053079	20	1,69552	0	20
<i>log_s720</i>						
	RMSE	MAE	DS	WDS	CP	CD
VAR1	0,010525625	0,010259766	80	0,22696	20	60
VAR3	0,018670251	0,015671851	80	0,047534	40	40
ECM1	0,006661323	0,006132999	60	0,743887	20	40
ECM3	0,027651991	0,02377458	80	0,189576	40	40
SVR1	0,024549501	0,024486349	60	0,650789	0	60
SVR3	0,009952337	0,008418459	80	0,335294	20	60
SVRL1	0,081032795	0,072230016	60	0,411719	20	40
SVRL3	0,279118629	0,21085696	0	-	0	0

Tabela 6.5: Resultados para previsões 6 passos à frente.

6 passos a frente	<i>log_s30</i>	<i>log_s60</i>	<i>log_s90</i>	<i>log_s120</i>	<i>log_s180</i>	<i>log_s360</i>	<i>log_s720</i>
SVR1/VAR(1)	0,987956	0,849119	0,915957	0,740276	0,808214	2,952588	2,332356
SVR1/VAR(3)	1,290502	1,002012	0,932684	0,666778	0,465212	0,835794	1,314899
SVR1/VEC1	1,753878	1,320019	1,186439	0,801863	0,579096	1,724902	3,685379
SVR1/VEC3	1,541460	0,937782	0,701406	0,387019	0,232107	0,473326	0,887802
SVR1/SVRL1	0,378758	0,259176	0,335514	0,222387	0,228895	0,266384	0,302958
SVR1/SVRL3	0,153156	0,011491	0,148894	0,002858	0,001540	0,006303	0,087954
SVR3/VAR(1)	1,170288	1,027495	1,235577	1,225743	1,202632	0,547335	0,945534
SVR3/VAR(3)	1,528670	1,212506	1,258140	1,104045	0,692241	0,154935	0,533059
SVR3/VEC1	2,077564	1,597317	1,600441	1,327717	0,861703	0,319753	1,494048
SVR3/VEC3	1,825943	1,134783	0,946159	0,640822	0,345378	0,087743	0,359914
SVR3/SVR1	1,184554	1,210072	1,348946	1,655791	1,488012	0,185375	0,405399
SVR3/SVRL1	0,448660	0,313621	0,452590	0,368227	0,340599	0,049381	0,122819
SVR3/SVRL3	0,181421	0,013904	0,200849	0,004733	0,002291	0,001168	0,035656

Tabela 6.6: Teste de Theil para previsões 6 passos à frente.

PREVISÃO 9 PASSOS À FRENTE

log_s30

	RMSE	MAE	DS	WDS	CP	CD
VAR1	0,008248	0,007106146	60	0,741385	0	60
VAR3	0,017091	0,012785026	20	2,797975	0	20
ECM1	0,016672	0,014465579	20	2,884369	0	20
ECM3	0,024625	0,023762432	40	2,375653	0	40
SVR1	0,007914	0,006747032	60	0,230655	0	60
SVR3	0,009817	0,007867215	40	3,758087	0	40
SVRL1	0,126617	0,108111008	20	6,756244	0	20
SVRL3	3,983688	2,866057293	60	5,076626	0	60

log_s60

	RMSE	MAE	DS	WDS	CP	CD
VAR1	0,007291	0,006698397	60	0,572626	0	60
VAR3	0,016467	0,012654918	20	3,070105	0	20
ECM1	0,017568	0,01527725	40	0,819614	0	40
ECM3	0,027321	0,026393933	40	2,398676	0	40
SVR1	0,007654	0,006592214	60	0,966196	0	60
SVR3	0,005923	0,004613831	60	0,863318	0	60
SVRL1	0,149145	0,128888504	20	1,800813	0	20
SVRL3	46,46501	33,96869349	60	4,630786	0	60

log_s90

	RMSE	MAE	DS	WDS	CP	CD
VAR1	0,008009	0,006734399	60	0,406426	0	60
VAR3	0,015939	0,012631763	20	3,644501	0	20
ECM1	0,018577	0,016080855	40	0,874571	0	40
ECM3	0,030026	0,029093651	40	2,393312	0	40
SVR1	0,00815	0,006629999	80	0,267748	0	80
SVR3	0,005526	0,004607619	80	0,226931	0	80
SVRL1	0,109113	0,09607977	40	1,225472	0	40
SVRL3	3,785733	2,491397027	40	0,252312	0	40

log_s120

	RMSE	MAE	DS	WDS	CP	CD
VAR1	0,009871	0,008509104	60	0,451043	0	60
VAR3	0,01531	0,012420726	20	4,28329	0	20
ECM1	0,018924	0,016613555	40	0,938381	0	40
ECM3	0,032163	0,031192355	40	2,37102	0	40
SVR1	0,011353	0,009833305	80	0,271769	0	80
SVR3	0,00608	0,005493404	80	0,028569	0	80
SVRL1	0,135333	0,128278588	60	0,378363	0	60
SVRL3	113,0356	82,50302966	60	4,695669	0	60

<i>log_s180</i>						
	RMSE	MAE	DS	WDS	CP	CD
VAR1	0,013497	0,012119471	80	0,310684	0	80
VAR3	0,015559	0,012800895	20	5,263087	0	20
ECM1	0,019187	0,016833271	40	0,96451	0	40
ECM3	0,03473	0,033725437	40	2,32192	0	40
SVR1	0,015142	0,014224337	80	0,275563	0	80
SVR3	0,00842	0,007629387	80	0,174785	0	80
SVRL1	0,116256	0,099446339	60	0,367707	0	60
SVRL3	152,6399	113,4602205	60	4,478686	0	60

<i>log_s360</i>						
	RMSE	MAE	DS	WDS	CP	CD
VAR1	0,020475	0,019926284	80	0,298939	0	80
VAR3	0,017403	0,0145326	20	4,674107	0	20
ECM1	0,017507	0,015008418	40	0,93298	0	40
ECM3	0,035414	0,034121651	40	2,434201	0	40
SVR1	0,027831	0,027585201	100	0	0	100
SVR3	0,014699	0,014227434	100	0	0	100
SVRL1	0,257033	0,233752947	60	0,466463	0	60
SVRL3	90,16503	66,9673666	60	4,532013	0	60

<i>log_s720</i>						
	RMSE	MAE	DS	WDS	CP	CD
VAR1	0,029702	0,029411898	80	0,303427	0	80
VAR3	0,018747	0,015998157	40	1,243181	0	40
ECM1	0,015341	0,013113207	60	0,62159	0	60
ECM3	0,034839	0,033341365	40	2,599079	0	40
SVR1	0,040488	0,040339731	100	0	0	100
SVR3	0,023929	0,022753931	100	0	0	100
SVRL1	0,413313	0,379139622	80	0,152845	0	80
SVRL3	12,6914	9,36849731	60	4,545322	0	60

Tabela 6.7: Resultados para previsões 9 passos à frente.

9 passos a frente	<i>log_s30</i>	<i>log_s60</i>	<i>log_s90</i>	<i>log_s120</i>	<i>log_s180</i>	<i>log_s360</i>	<i>log_s720</i>
SVR1/VAR(1)	0,959442	1,049762	1,017540	1,150201	1,121920	1,359242	1,363152
SVR1/VAR(3)	0,463034	0,464833	0,511309	0,741575	0,973206	1,599228	2,159693
SVR1/VEC1	0,474673	0,435690	0,438701	0,599935	0,789224	1,589752	2,639147
SVR1/VEC3	0,321360	0,280162	0,271423	0,352989	0,436010	0,785880	1,162144
SVR1/SVRL1	0,062500	0,051321	0,074690	0,083892	0,130251	0,108278	0,097961
SVR1/SVRL3	0,001986	0,000165	0,002153	0,000100	0,000099	0,000309	0,003190
SVR3/VAR(1)	1,190188	0,812297	0,690012	0,615959	0,623870	0,717901	0,805624
SVR3/VAR(3)	0,574393	0,359684	0,346727	0,397130	0,541174	0,844653	1,276381
SVR3/VEC1	0,588831	0,337133	0,297491	0,321279	0,438866	0,839649	1,559738
SVR3/VEC3	0,398647	0,216787	0,184057	0,189033	0,242454	0,415073	0,686828
SVR3/SVR1	1,240499	0,773792	0,678117	0,535523	0,556073	0,528163	0,591001
SVR3/SVRL1	0,077531	0,039712	0,050649	0,044926	0,072429	0,057189	0,057895
SVR3/SVRL3	0,002464	0,000127	0,001460	0,000054	0,000055	0,000163	0,001885

Tabela 6.8: Teste de Theil para previsões 9 passos à frente.

PREVISÃO 12 PASSOS À FRENTE

log_s30

	RMSE	MAE	DS	WDS	CP	CD
VAR1	0,012785	0,011104875	80	0,351243	0	80
VAR3	0,017542	0,014257534	20	7,614935	0	20
ECM1	0,027484	0,025764775	20	3,323593	0	20
ECM3	0,037204	0,036432708	40	2,201897	0	40
SVR1	0,011942	0,010133575	60	1,030474	0	60
SVR3	0,009159	0,007681361	60	0,618145	0	60
SVRL1	0,464962	0,386058677	40	1,052391	0	40
SVRL3	190,2517	134,4375952	40	0,185985	0	40

log_s60

	RMSE	MAE	DS	WDS	CP	CD
VAR1	0,015706	0,014612525	80	0,33614	0	80
VAR3	0,016218	0,013336954	20	11,76747	0	20
ECM1	0,028394	0,026975125	40	1,124791	0	40
ECM3	0,039589	0,038734625	40	2,180303	0	40
SVR1	0,017055	0,015888009	80	0,300803	0	80
SVR3	0,013867	0,013069159	80	0,146042	0	80
SVRL1	0,562653	0,478707079	60	0,185968	0	60
SVRL3	2208,811	1585,593226	40	0,205192	0	40

log_s90

	RMSE	MAE	DS	WDS	CP	CD
VAR1	0,019012	0,01835815	80	0,320327	0	80
VAR3	0,014874	0,012182103	20	22,45376	0	20
ECM1	0,029027	0,027885516	40	1,219133	0	40
ECM3	0,0419	0,040962016	40	2,154569	0	40
SVR1	0,018801	0,018153233	80	0,297072	0	80
SVR3	0,014295	0,013844683	60	0,630083	0	60
SVRL1	0,43703	0,395567504	60	0,281855	0	60
SVRL3	184,7589	118,502309	60	4,183234	0	60

log_s120

	RMSE	MAE	DS	WDS	CP	CD
VAR1	0,022286	0,021870045	80	0,311141	0	80
VAR3	0,013709	0,010934138	20	386,1503	0	20
ECM1	0,029094	0,028167895	40	1,314917	0	40
ECM3	0,043806	0,042743678	40	2,146863	0	40
SVR1	0,024216	0,023967195	80	0,289079	0	80
SVR3	0,01796	0,017718662	80	0,225278	0	80
SVRL1	0,637108	0,594543457	80	0,188917	0	80
SVRL3	5381,442	3857,031418	40	0,202355	0	40

<i>log_s180</i>						
	RMSE	MAE	DS	WDS	CP	CD
VAR1	0,026838	0,02661059	60	0,800211	0	60
VAR3	0,013059	0,010418503	0	-	0	0
ECM1	0,028469	0,027611857	20	4,132967	0	20
ECM3	0,045645	0,044431074	20	4,360891	0	20
SVR1	0,028363	0,02826394	60	0,748699	0	60
SVR3	0,022713	0,022500407	60	0,626448	0	60
SVRL1	0,636343	0,58747570	80	0,229213	2	60
SVRL3	7262,879	5290,566329	60	0,11493	20	40

<i>log_s360</i>						
	RMSE	MAE	DS	WDS	CP	CD
VAR1	0,034471	0,034191615	40	1,37623	0	40
VAR3	0,012087	0,009461557	20	1,038713	20	0
ECM1	0,024548	0,023634681	0	-	0	0
ECM3	0,045008	0,043024548	40	1,97638	20	20
SVR1	0,041457	0,041276548	60	0,549213	0	60
SVR3	0,030571	0,029342548	40	1,05479	0	40
SVRL1	1,378407	1,304860372	40	1,708295	0	40
SVRL3	4297,402	3126,842312	80	0,102282	40	40

<i>log_s720</i>						
	RMSE	MAE	DS	WDS	CP	CD
VAR1	0,04381	0,042774475	60	0,425162	20	40
VAR3	0,010222	0,008381673	60	0,227785	40	20
ECM1	0,01702	0,015803959	40	1,105296	20	20
ECM3	0,042342	0,038723992	60	0,698787	40	20
SVR1	0,052612	0,051888659	40	1,166367	0	40
SVR3	0,04127	0,037179809	20	2,581864	0	20
SVRL1	1,378407	1,304860372	40	1,708295	0	40
SVRL3	601,4188	435,3539169	60	0,780564	40	20

Tabela 6.9: Resultados para previsões 12 passos à frente.

12 passos a frente	<i>log_s30</i>	<i>log_s60</i>	<i>log_s90</i>	<i>log_s120</i>	<i>log_s180</i>	<i>log_s360</i>	<i>log_s720</i>
SVR1/VAR(1)	0,934123	1,085851	0,988903	1,086623	1,056799	1,202657	1,200922
SVR1/VAR(3)	0,680788	1,051605	1,263989	1,766461	2,171850	3,429892	5,146979
SVR1/VEC1	0,434524	0,600639	0,647701	0,832336	0,996256	1,688822	3,091295
SVR1/VEC3	0,320994	0,430798	0,448706	0,552805	0,621372	0,921099	1,242545
SVR1/SVRL1	0,025684	0,030311	0,043019	0,038010	0,044571	0,030076	0,038169
SVR1/SVRL3	0,000063	0,000008	0,000102	0,000004	0,000004	0,000010	0,000087
SVR3/VAR(1)	0,716389	0,882897	0,751886	0,805878	0,846299	0,886862	0,942022
SVR3/VAR(3)	0,522104	0,855052	0,961041	1,310070	1,739248	2,529266	4,037373
SVR3/VEC1	0,333242	0,488375	0,492462	0,617290	0,797815	1,245369	2,424861
SVR3/VEC3	0,246174	0,350278	0,341162	0,409980	0,497603	0,679235	0,974672
SVR3/SVR1	0,766911	0,813092	0,760324	0,741635	0,800814	0,737419	0,784416
SVR3/SVRL1	0,019698	0,024646	0,032708	0,028189	0,035693	0,022179	0,029940
SVR3/SVRL3	0,000048	0,000006	0,000077	0,000003	0,000003	0,000007	0,000069

Tabela 6.10: Teste de Theil para previsões 12 passos à frente.

7

Conclusão e trabalhos futuros

SVR é uma técnica alternativa promissora para a previsão das taxas de juros. Com relação aos outros métodos, o SVR é capaz de capturar relações não-lineares entre as variáveis macro-econômicas e variáveis da estrutura a termo e, suas previsões para médio (seis meses) e longo prazo (nove e doze meses) superam aquelas oriundas de modelos puramente lineares. Embora não tenham apresentado o mesmo desempenho no curto prazo (um e três meses), ainda assim aparecem como bons indicadores da direção de cada taxa permitindo uma estimação mais precisa da direção da curva de retornos como podemos visualizar nos gráficos (7.1-7.35).

Algumas questões podem ser melhor exploradas, constituindo um caminho para pesquisa futuras, dentre elas merecem destaque:

- Avaliar a conveniência de se utilizar outras funções de custo permitindo verificar o comportamento comparativo entre as mesmas;
- A necessidade de se empregar outros núcleos, os quais poderiam superar problemas para conseguir boas previsões em um e três passos à frente;
- A escolha dos parâmetros ε , γ e C . Estudos de Tao e Cao (25) na previsão de séries temporais financeiras mostraram que a cota C e o parâmetro do núcleo γ têm importância na performance do SVR e uma escolha inadequada desses parâmetros pode causar problemas de excesso (" *overfitting* ") e falta de ajuste (" *underfitting* "). Uma alternativa sugerida e aplicada com sucesso pelos autores é o uso de parâmetros adaptáveis, ou seja, modificações nos parâmetros C e ε são feitas a fim de colocar mais peso nos pontos dos dados de treinamento mais recentes e menos peso nos dados de treino mais distantes, pois nas séries temporais financeiras não-estacionárias os pontos de dados mais recentes podem oferecer informações mais úteis que os pontos de dados de treinamento distantes.
- A habilidade de generalização do SVR também deveria ser testada aplicando o método para outras séries temporais.

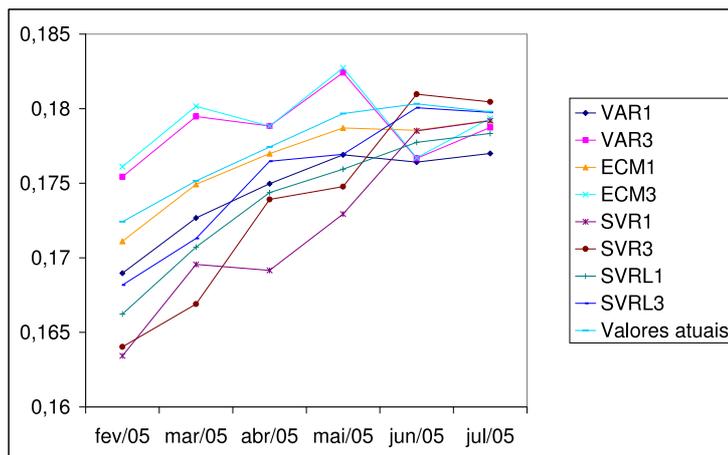


Figura 7.1: Valores previstos nos dados de teste de $\log_s 30$ (1 mês).

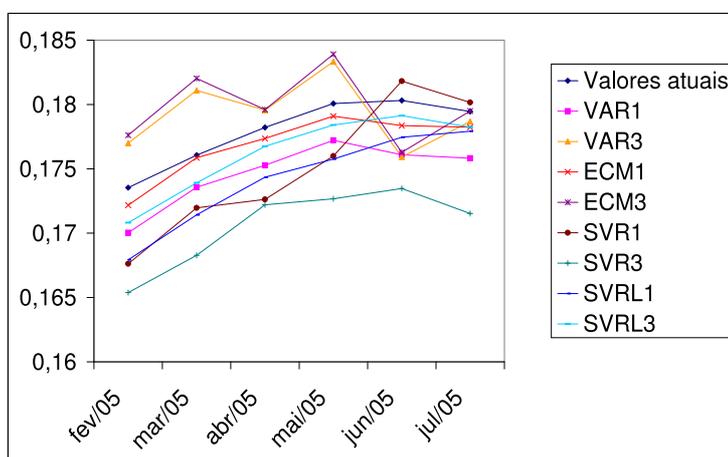


Figura 7.2: Valores previstos nos dados de teste de $\log_s 60$ (1 mês).

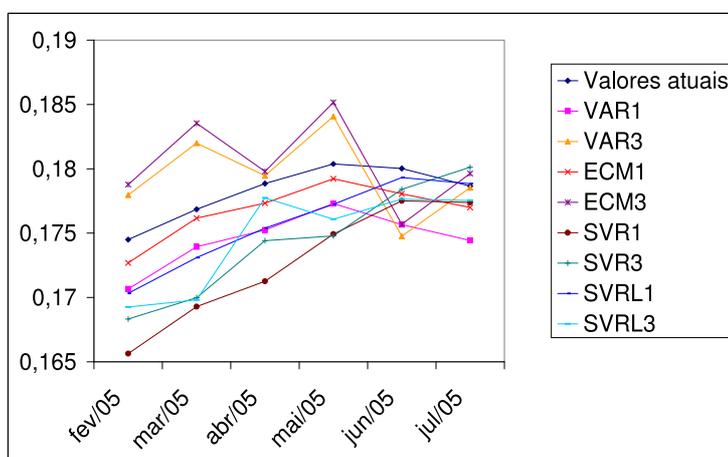


Figura 7.3: Valores previstos nos dados de teste de $\log_s 90$ (1 mês).

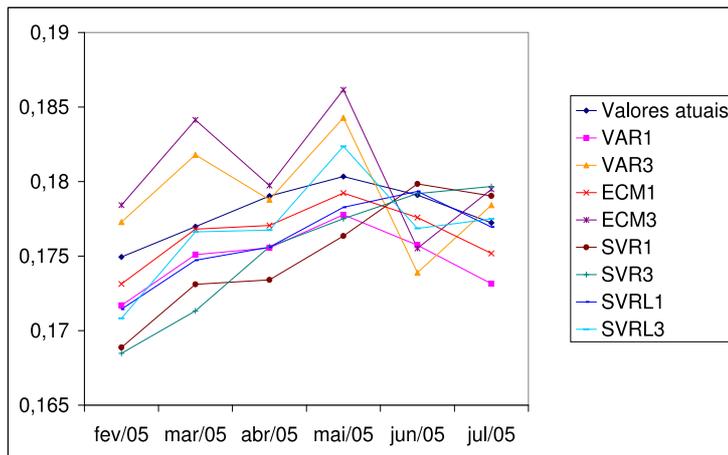


Figura 7.4: Valores previstos nos dados de teste de $\log_s 120$ (1 mês).

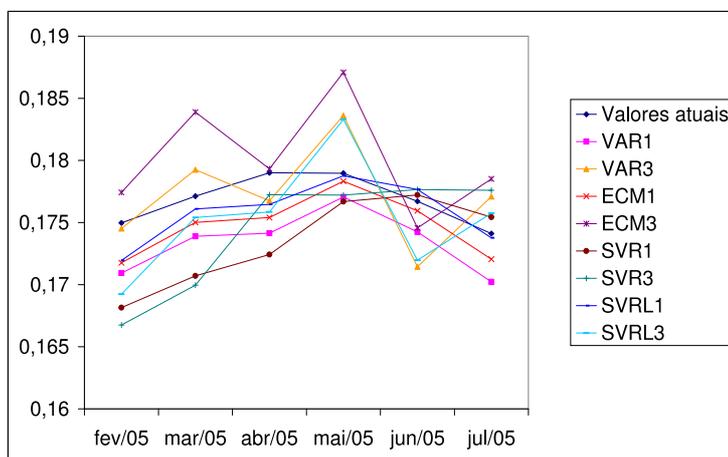


Figura 7.5: Valores previstos nos dados de teste de $\log_s 180$ (1 mês).

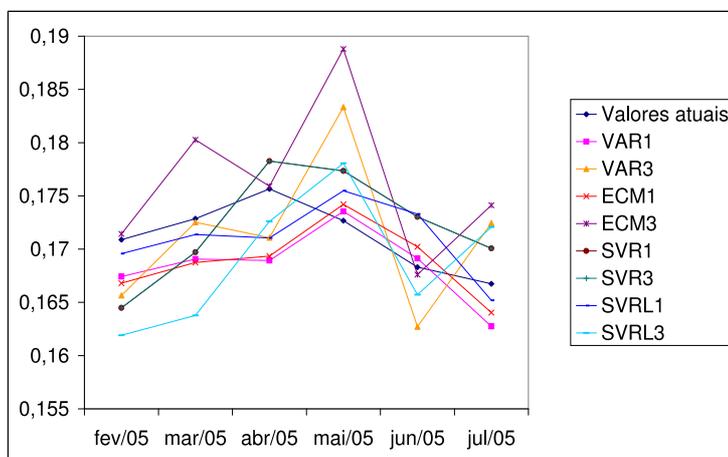


Figura 7.6: Valores previstos nos dados de teste de $\log_s 360$ (1 mês).

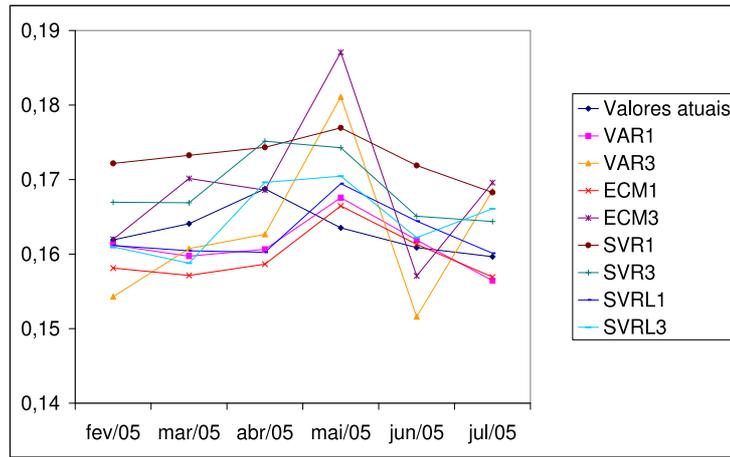


Figura 7.7: Valores previstos nos dados de teste de $\log_s 720$ (1 mês).

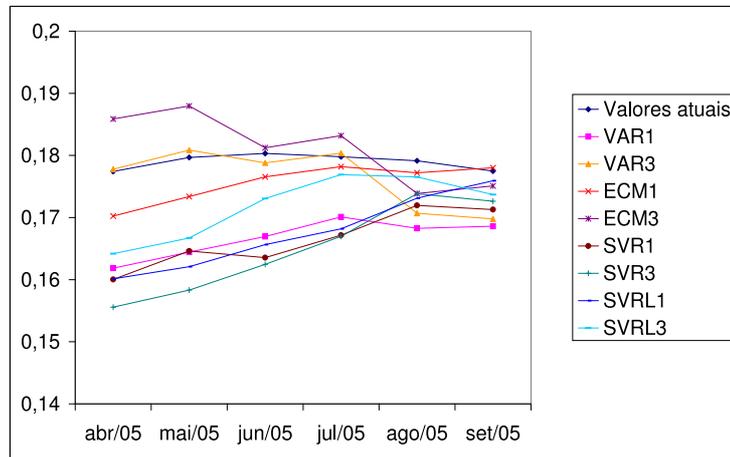


Figura 7.8: Valores previstos nos dados de teste de $\log_s 30$ (3 meses).

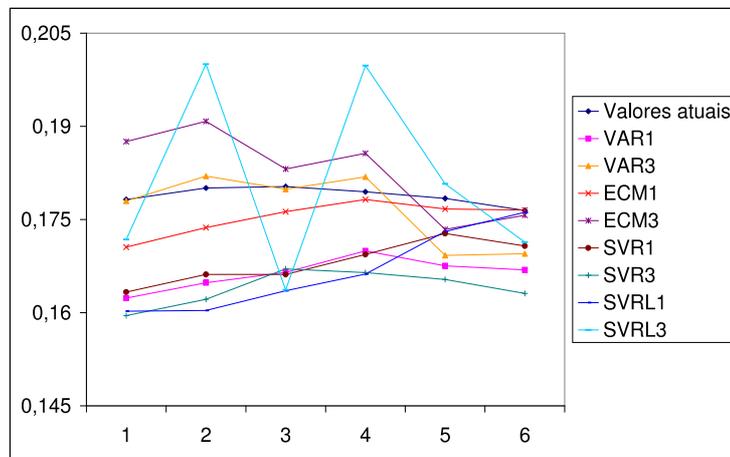


Figura 7.9: Valores previstos nos dados de teste de $\log_s 60$ (3 meses).

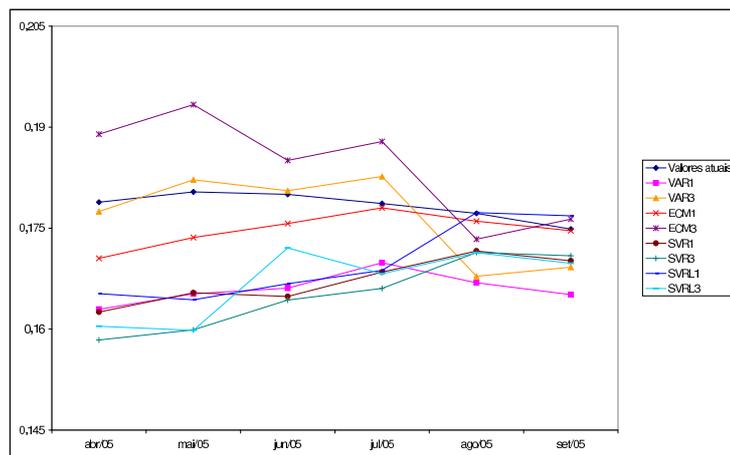


Figura 7.10: Valores previstos nos dados de teste de $\log_s 90$ (3 meses).

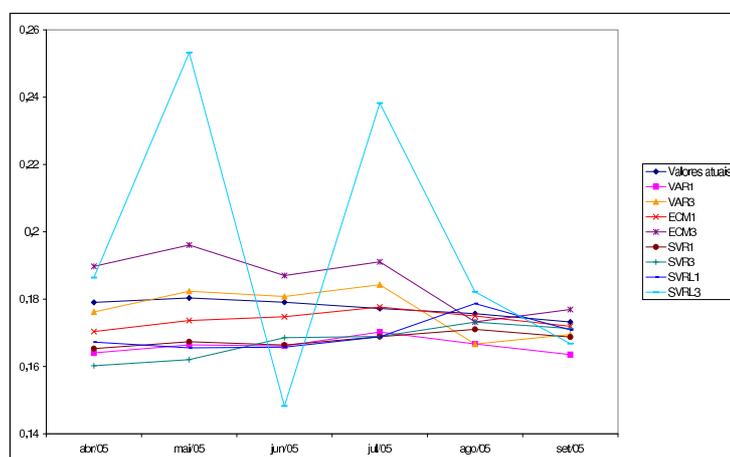


Figura 7.11: Valores previstos nos dados de teste de $\log_s 120$ (3 meses).

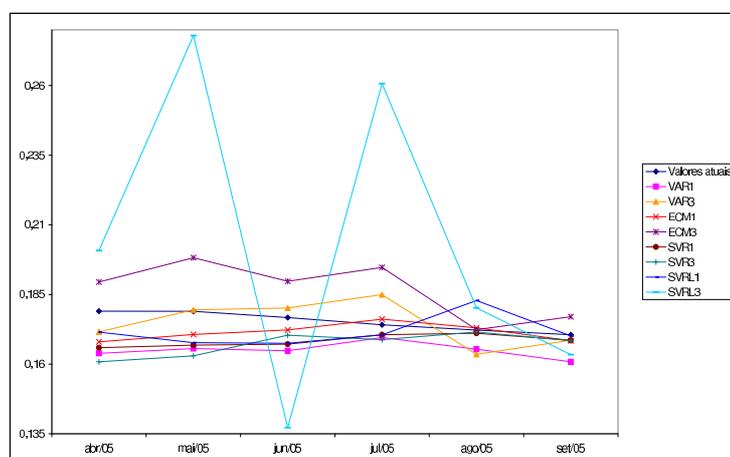


Figura 7.12: Valores previstos nos dados de teste de $\log_s 180$ (3 meses).

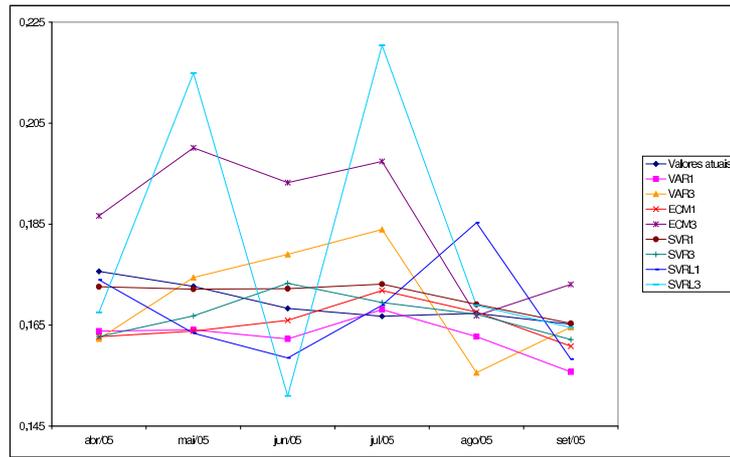


Figura 7.13: Valores previstos nos dados de teste de $\log_s 360$ (3 meses).

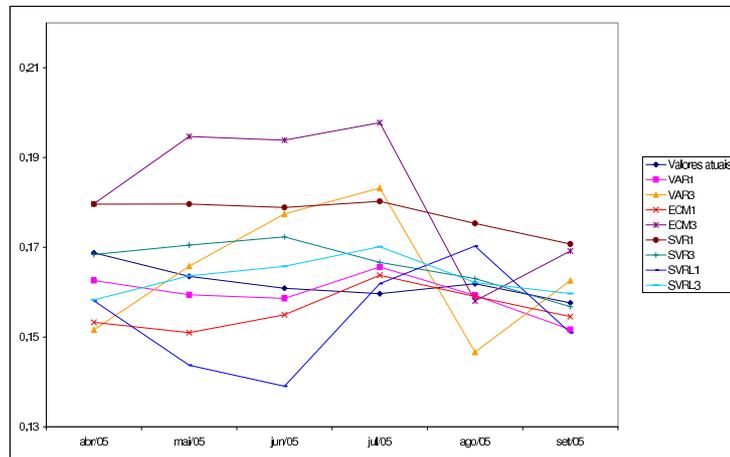


Figura 7.14: Valores previstos nos dados de teste de $\log_s 720$ (3 meses).

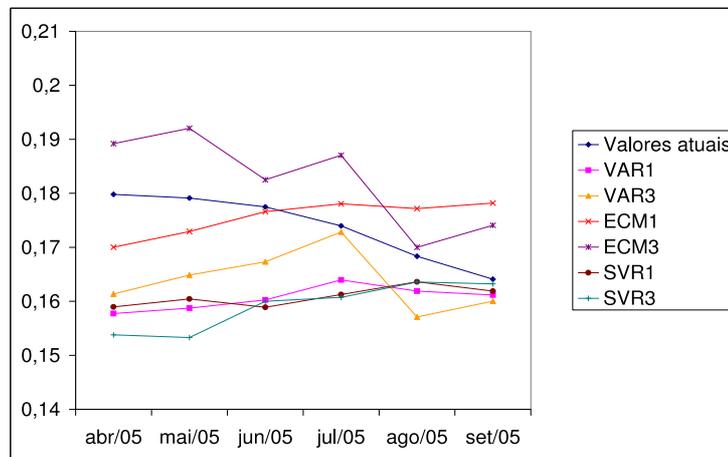


Figura 7.15: Valores previstos nos dados de teste de $\log_s 30$ (6 meses).

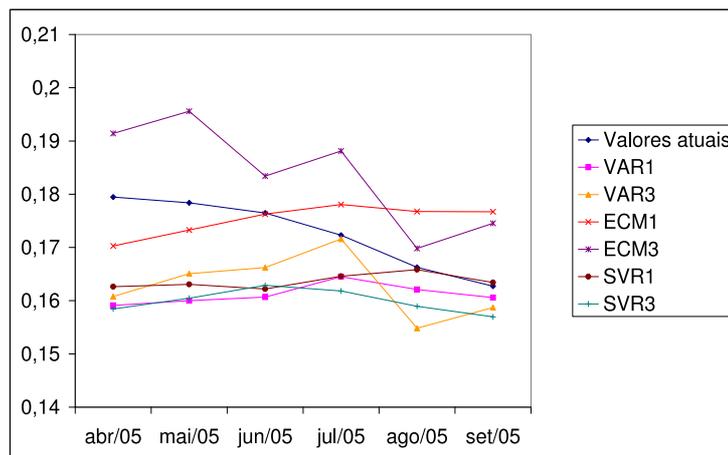


Figura 7.16: Valores previstos nos dados de teste de $\log_s 60$ (6 meses).

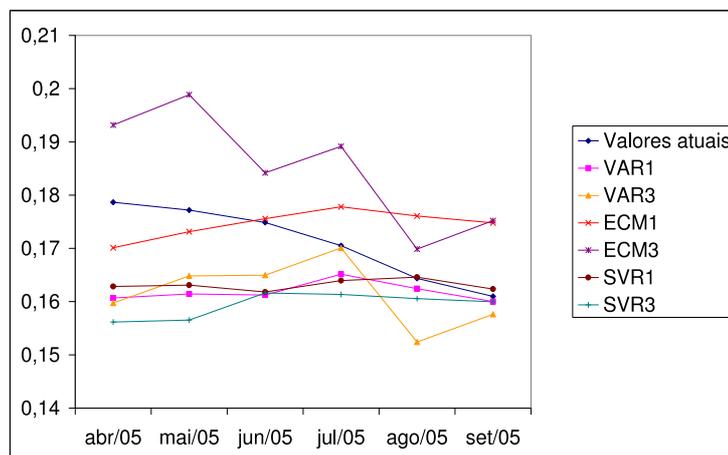


Figura 7.17: Valores previstos nos dados de teste de $\log_s 90$ (6 meses).

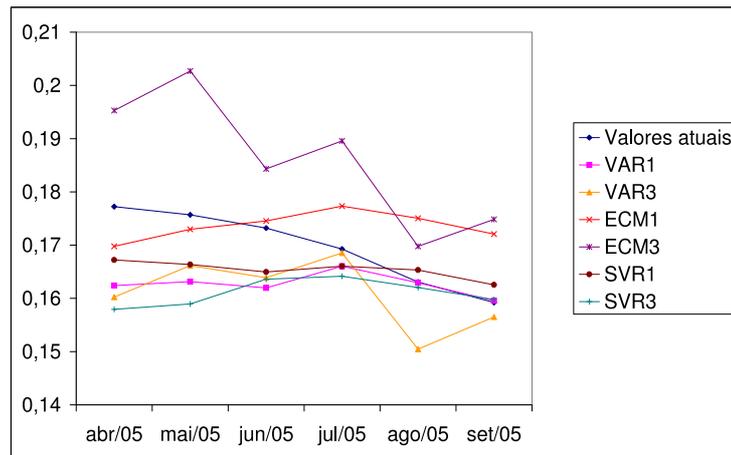


Figura 7.18: Valores previstos nos dados de teste de $\log_s 120$ (6 meses).

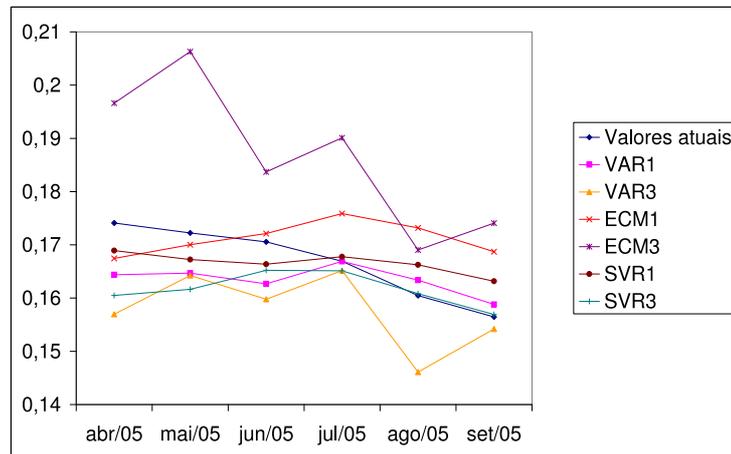


Figura 7.19: Valores previstos nos dados de teste de $\log_s 180$ (6 meses).

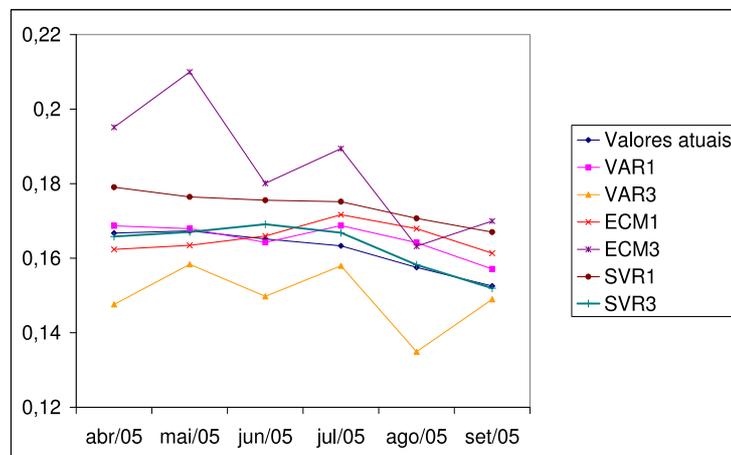


Figura 7.20: Valores previstos nos dados de teste de $\log_s 360$ (6 meses).

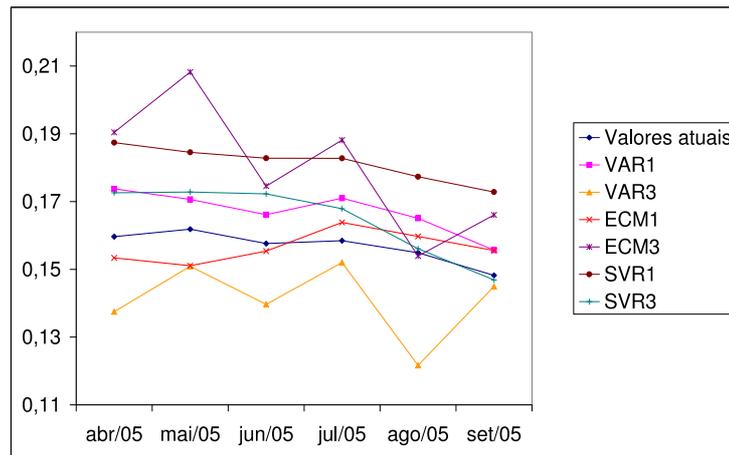


Figura 7.21: Valores previstos nos dados de teste de $\log_s 720$ (6 meses).

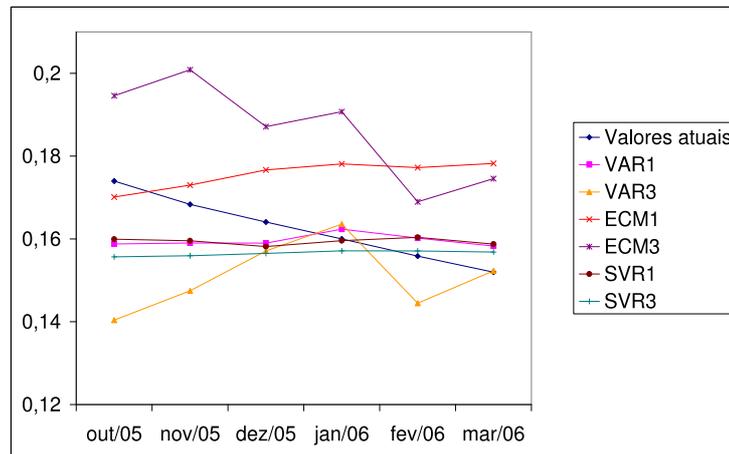


Figura 7.22: Valores previstos nos dados de teste de $\log_s 30$ (9 meses).

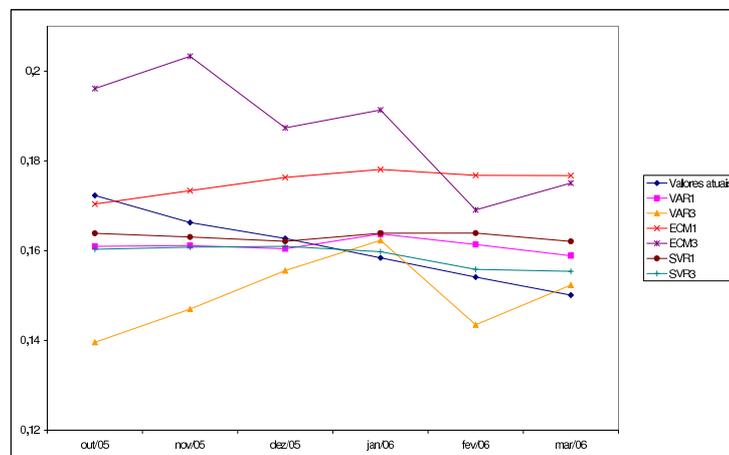


Figura 7.23: Valores previstos nos dados de teste de $\log_s 60$ (9 meses).

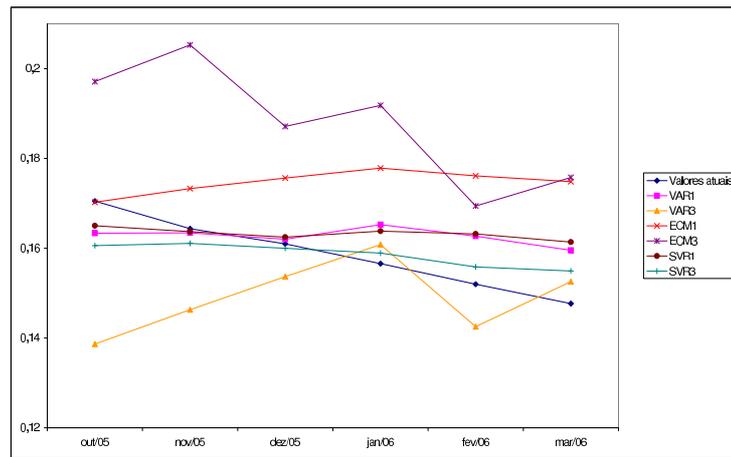


Figura 7.24: Valores previstos nos dados de teste de $\log_s 90$ (9 meses).

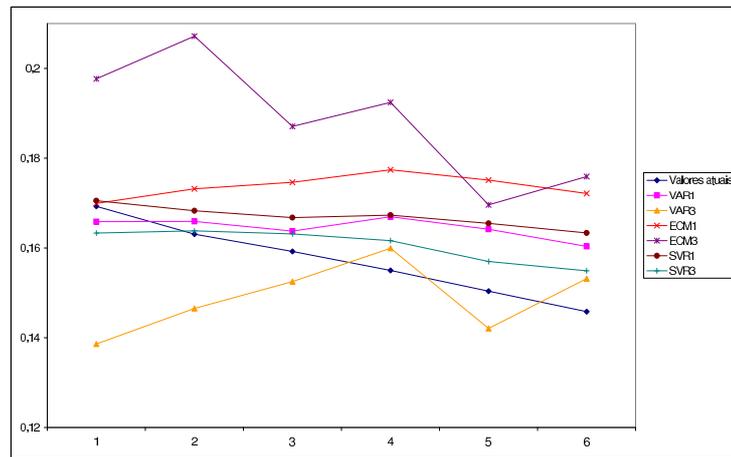


Figura 7.25: Valores previstos nos dados de teste de $\log_s 120$ (9 meses).

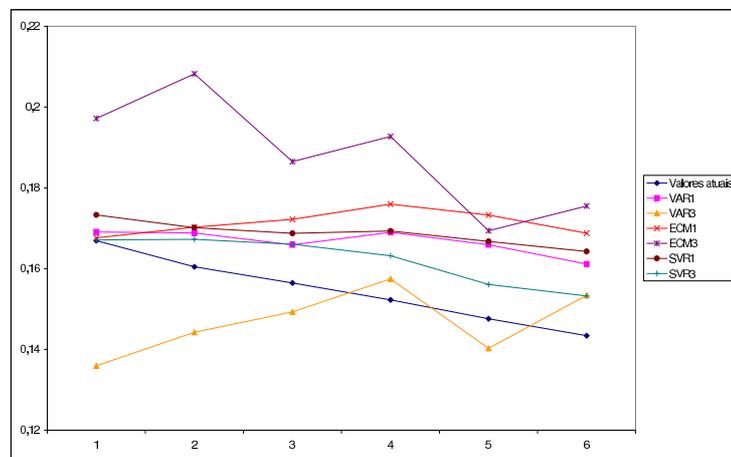


Figura 7.26: Valores previstos nos dados de teste de $\log_s 180$ (9 meses).

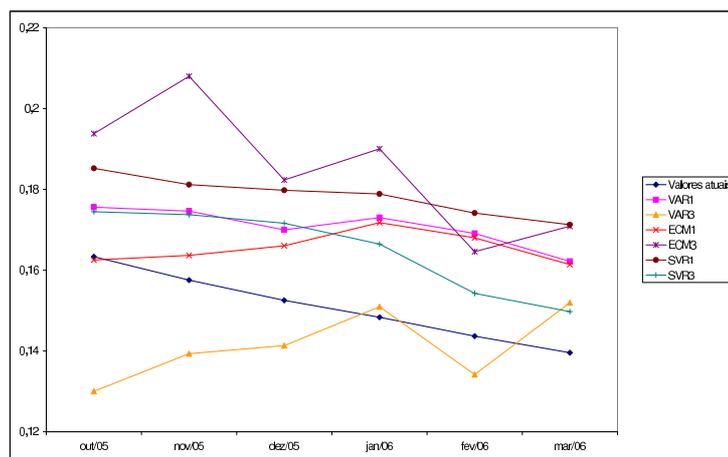


Figura 7.27: Valores previstos nos dados de teste de $\log_s 360$ (9 meses).

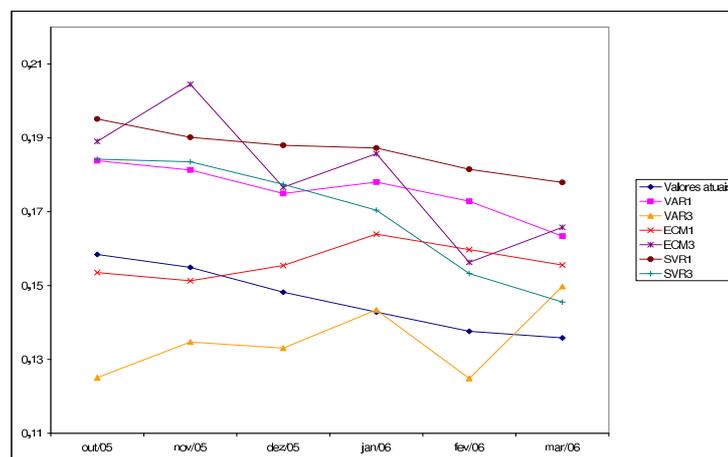


Figura 7.28: Valores previstos nos dados de teste de $\log_s 720$ (9 meses).

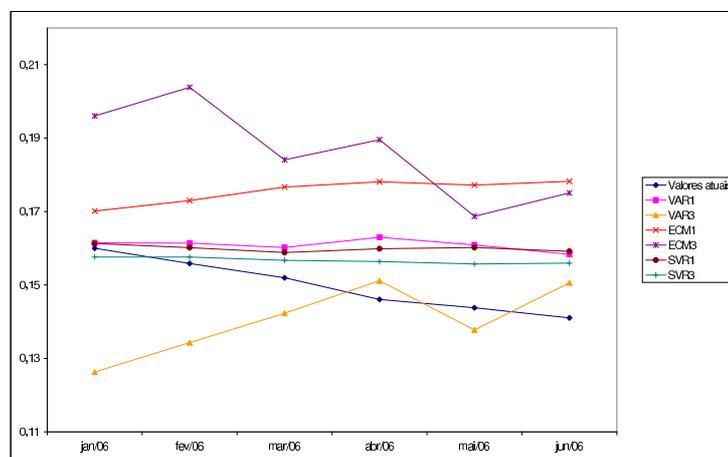


Figura 7.29: Valores previstos nos dados de teste de $\log_s 30$ (12 meses).

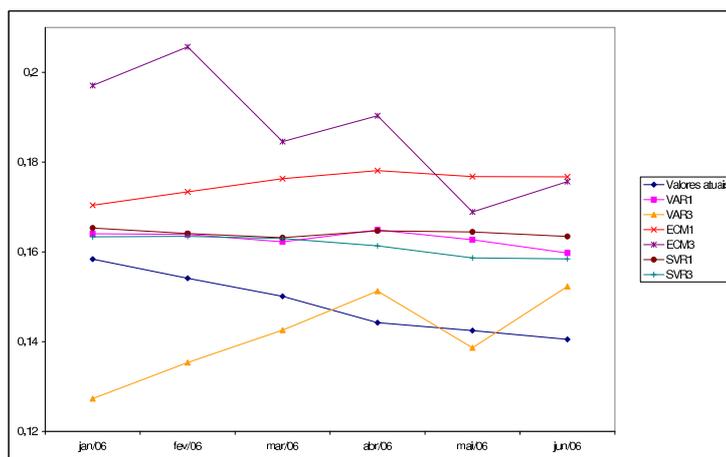


Figura 7.30: Valores previstos nos dados de teste de $\log_s 60$ (12 meses).

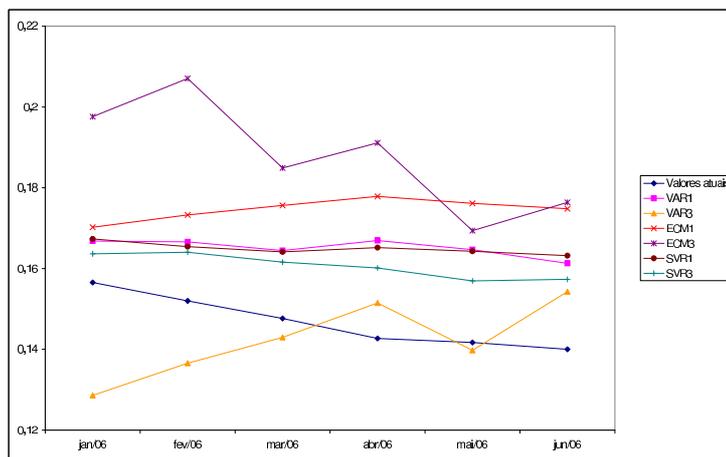


Figura 7.31: Valores previstos nos dados de teste de $\log_s 90$ (12 meses).

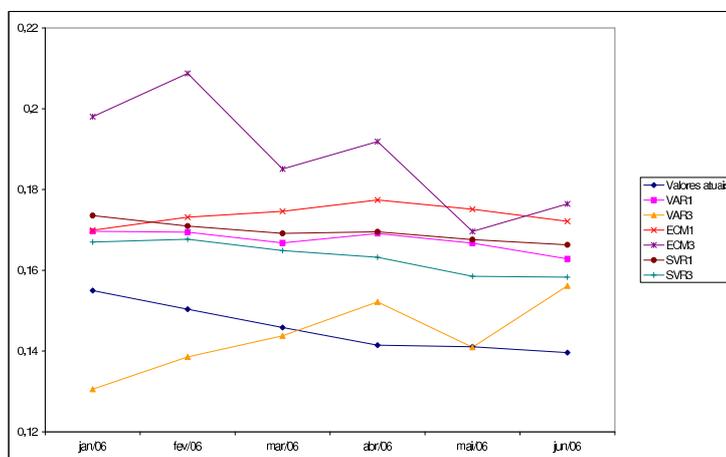


Figura 7.32: Valores previstos nos dados de teste de $\log_s 120$ (12 meses).

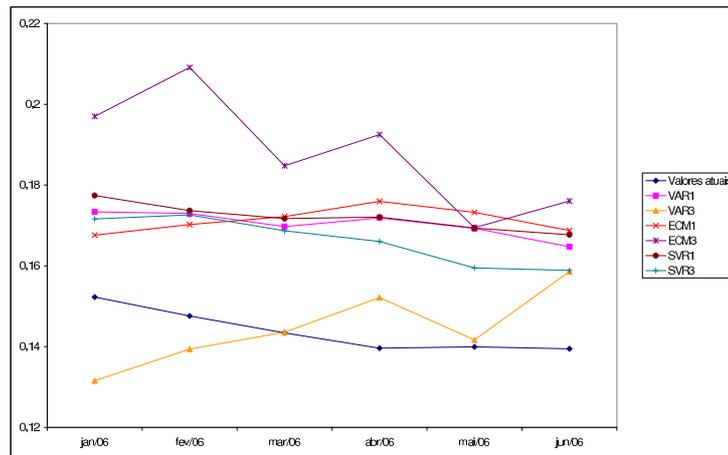


Figura 7.33: Valores previstos nos dados de teste de $\log_s 180$ (12 meses).

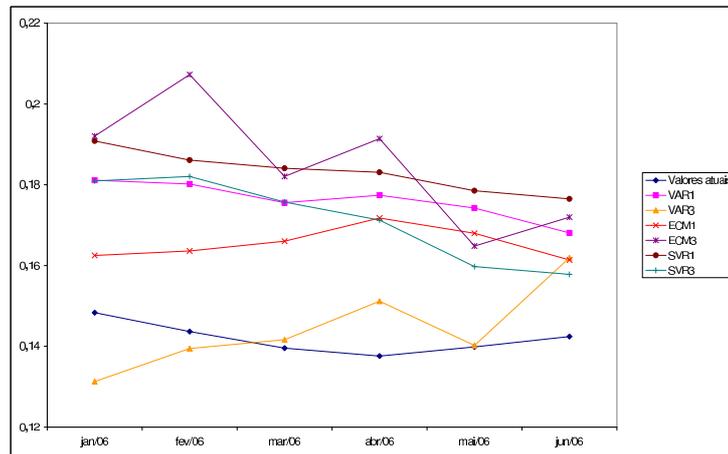


Figura 7.34: Valores previstos nos dados de teste de $\log_s 360$ (12 meses).

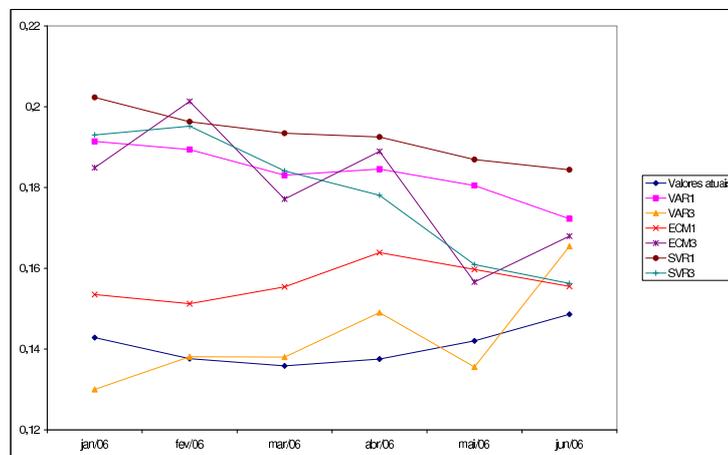


Figura 7.35: Valores previstos nos dados de teste de $\log_s 720$ (12 meses).

Referências Bibliográficas

- [1] BOENDER, G. C. E.; P. V. AALST, F. H.. **Modelling and management of assets and liabilities of pension plans in the netherlands.** Worldwide Asset and Liability Modeling, 1998. 1.1
- [2] BOSER, B. E.; GUYON, I. M. ; VAPNIK, V. N.. **A training algorithm for optimal margin classifiers.** Annual Workshop on Computational Learning Theory, p. 144–152, 1992. 3.2
- [3] BURGESS, C. J. C.. **A tutorial on support vector machines for pattern recognition.** Data Mining and Knowledge Discovery, 2:121–167, 1998. (document), 3.2, 3.2
- [4] CHANG, C. C.; LIN, C. J.. **Libsvm : a library for support vector machines, software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.** 2001. 3.7
- [5] CHONG, E. K. P.; ZAK, S. H.. **An Introduction to Optimization.** John Wiley & Sons, 2nd. edition, 2001. 2.1
- [6] CORTES, C.; VAPNIK, V.. **Support vector networks.** Machine Learning, 20(3):273–297, 1995. 3.2
- [7] CRISTIANINI, N.; SHAWE-TAYLOR, J.. **An Introduction to Support Vector Machines and Other Kernel-based Learning Methods.** Cambridge University Press, 2000. (document), 2, 3.1.1, 3.2, 3.5, 3.5.1
- [8] DERT, C.. **A dynamic model for asset liability management for defined benefit pension funds.** Worldwide Asset and Liability Modeling, 1998. 1.1
- [9] FUKUDA, R.; LOPES, H. ; VEREDA, L.. **Estimating var models for the term structure of interest rates.** IME 2006 Conference at Leuven, 2006. 1
- [10] GAO, J. B.; GUNN, S. R. ; HARRIS, C. J.. **Mean field method for the support vector machine regression.** Neurocomputing, 2001. 1

- [11] KIM, K.. **Financial time series forecasting using support vector machines.** *Neurocomputing*, 55:307–319, 2003. 1, 5.1
- [12] KUZMAN, H.. **A support vector machine for avo interpretation.** 2003 SEG Abstracts, p. 181–184, 2003. 1
- [13] LI, J.. **Multiatributes pattern recognition for reservoir prediction.** *Innovation: 2005 CSEG National Convention*, p. 205–208, 2005. 1
- [14] MARTIN, C.. **Prognóstico de demanda de potência elétrica para planejamento e operação de sistemas elétricos.** Universidade Federal de Santa Catarina. Tese apresentada ao Curso de Pós-Graduação em Engenharia de Produção, 2005. 3.2
- [15] MULLER, K. R.; SMOLA, A. J.; RATSCH, G.; SCHÖLKOPF, B.; KOHLMORGEN, J. ; VAPNIK, V.. **Predicting time series with support vector machines.** In W. Gerstner, A. Germond, M. Hasler, and J.D. Nicoud, editors, *Artificial Neural Networks ICANN'97*, 1327:999–1004, 1997. 3.3.2
- [16] **R project url (2007) .** <http://www.r-project.org/>. 3.7
- [17] SCHÖLKOPF, B.. **The kernel trick for distances.** Technical report, Microsoft Research, 2000. 3.5.1
- [18] SCHÖLKOPF, B.; SMOLA, A. J.. **Learning with kernels: Support Vector Machines, Regularization, Optimization, and Beyond.** The MIT Press, 2001. (document), 3.3.1, 3.3, 3.4, 3.5.1, 3.7
- [19] SCHÖLKOPFT, B.; BURGES, C. J. C. ; SMOLA, A. J.. **Advances in kernel methods.** 1998. 1
- [20] SCHÖLKOPFT, B.; BURGES, C. J. C. ; VAPNIK, V.. **Extracting support data for a given task.** In: U. M. Fayyad and R. Uthurusamy (eds):*Proceedings, First International Conference and Knowledge Discovery and Data Mining*, p. 252–257, 1995. 1
- [21] SILVA, A. P. A.; FERREIRA, V. H.. **On eletric load forecasting and neural networks.** 4th International Institute of Forecasters' Workshop, 2007. 3.2
- [22] SMOLA, A. J.. **Regression estimation with support vector learning machines.** Master's thesis. Physik Department, Technische Universität München, 1996. (document), 3.6

- [23] SMOLA, A. J.; SCHÖLKOPF, B.. **A tutorial on support vector regression.** *Statistics and Computing*, 14(3):199–222, 2003. 3.2
- [24] S.RÜPING. **Svm kernels for time series analysis.** CS Departament, AI Unit, University of Dortmund, 2001. 3.5.1
- [25] TAY, F. E. H.; CAO, L.. **Application of support vector machines in financial time series forecasting.** *Omega*, 29:309–317, 2001. 1, 5.1, 7
- [26] TIKHONOV, A. N.. **On solving ill-posed problem and method of regularization.** *Dokl. Akad. Nauk USSR*, 153:501–504, 1963. 3.3.1
- [27] VAPNIK, V.; CHERVONENKIS, A.. **A note on one class of perceptrons.** *Automation and Remote Control*, 25, 1964. 3.2
- [28] VAPNIK, V.; LERNER, A.. **Pattern recognition using generalized portrait method.** *Automation and Remote Control*, 24:774–780, 1963. 3.2
- [29] V.VAPNIK. **Statistical Learning Theory.** Wiley, 1998. (document), 2, 3.2, 3.1, 3.3.1, 3.5.1
- [30] V.VAPNIK. **The Nature of Statistical Learning Theory.** Springer-Verlag, 2nd. edition, 2000. (document), 1, 1.1, 1, 3.2, 3.3.1
- [31] ZHAO, B, Z. H.. **Nonlinear classification of avo attributes using svm.** *Nonlinear Geophysics*, 25, 2005. 1

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)