



## Construção de Previsões Intervalares de Séries Temporais via Metodologia Support Vector Machines e Técnicas Bootstrap

Hiago Henrique Gomes de Araujo<sup>a</sup>

[hiago.araujo.00@gmail.com](mailto:hiago.araujo.00@gmail.com)

Isis Didier Lins<sup>a</sup>

[isis.lins@ceerma.org](mailto:isis.lins@ceerma.org)

Márcio das Chagas Moura<sup>a</sup>

[marcio@ceerma.org](mailto:marcio@ceerma.org)

Enrique López Droguett<sup>a,b</sup>

[elopezdroguett@ing.uchile.cl](mailto:elopezdroguett@ing.uchile.cl)

<sup>a</sup>Centro de Estudos e Ensaio em Risco e Modelagem Ambiental – Departamento de Engenharia de Produção – Universidade Federal de Pernambuco  
Av. da Arquitetura, s/n - Cidade Universitária, Recife - PE, 50740-550

<sup>b</sup>Departamento de Engenharia Mecânica – Universidade do Chile  
Av. Beauchef 851, Santiago, Chile

### RESUMO

As previsões de séries temporais usualmente são pontuais, no entanto, quando esses pontos são acompanhados de um intervalo, torna-se possível avaliar a precisão dessa medida e apoiar decisões com mais eficiência. Este artigo traz uma metodologia para a construção de estimativas intervalares para previsões de séries temporais, que foi realizada via regressão por *Support Vector Machines* (SVM) e construção de intervalos de confiança via métodos *bootstrap*. Nessa pesquisa, utilizou-se o bootstrap estacionário e o bootstrap da máxima entropia em quatro séries temporais com comportamentos distintos: Temperatura da superfície do mar, Venda de uma família de produtos laticínios, cotação do dólar e aceleração em um rolamento. Foi possível comparar os resultados obtidos e concluir que o bootstrap de máxima entropia oferece previsões intervalares com qualidade superior ao bootstrap estacionário.

**PALAVRAS CHAVE.** Bootstrap. Support Vector Machines. Estimativas intervalares.

**Estatística**

### ABSTRACT

Time series forecasts are usually point estimates, but when they are provided with an interval, it becomes possible to evaluate the precision of that measure and to support decisions more effectively. This paper presents a methodology to build confidence intervals of time series forecasting. The prediction is performed through regression by Support Vector Machines (SVM) and construction of confidence intervals are obtained via bootstrap techniques. In this research, we used the stationary bootstrap and bootstrap of maximum entropy in four time series with different behaviors: sea surface temperature, sale of a family of dairy products, dollar sale value and acceleration in a bearing. By comparing the obtained results, we conclude that the maximum entropy bootstrap gives intervals with superior quality than the stationary bootstrap.

**KEYWORDS.** Bootstrap. Support Vector Machines. Interval Predict.

**Statistics**



## 1. Introdução

Uma série temporal é “qualquer conjunto de observações ordenadas no tempo” [Morettin e Toloi, 1981]. No âmbito da análise de séries temporais encontram-se diversos modelos de regressão e previsão que possuem aplicações em diferentes áreas do conhecimento, como economia, gestão da produção, engenharia de manutenção, estudos climáticos, entre outras.

*Support Vector Machines* (SVM), também conhecido como *Kernel Machines*, é um algoritmo de aprendizado estatístico que pode ser usado tanto para realizar classificação (identificação de padrões) quanto para realizar modelos regressivos. Seu modelo é embasado na teoria do aprendizado estatístico e na metodologia de *Structural Risk Minimization*. O SVM é um método não-paramétrico, o que significa que o modelo não exige hipóteses sobre a distribuição dos dados originais [KECMAN, 2005]. É possível encontrar aplicações deste algoritmo em modelos de regressões para séries financeiras [CAO e TAY 2003], séries de temperatura do ar [RADHIKA e SHASHI, 2009], temperatura da superfície do oceano [LINS et al. 2013].

O modelo de regressão via SVM realiza previsões pontuais. No entanto as previsões em formatos de intervalos de confiança são mais informativas para a tomada de decisão, já que indicam a precisão associada às estimativas pontuais. Desta forma, os métodos bootstrap se tornaram bastante populares como ferramentas de construção de intervalos no âmbito das previsões via *Kernel Machines* [DE BRABANTER et al.]. O principal atrativo do bootstrap é seu caráter não-paramétrico que possibilita a construção de intervalos precisos sem exigir hipóteses sobre a distribuição do conjunto de dados.

É possível encontrar modelos de SVM de regressão (SVR) em conjunto com o método bootstrap sendo aplicado por diversos autores [LINS et al, 2015; LENDASSE et al, 2004; ANGUITA et al, 2000]. No entanto, essas aplicações não envolvem realizar a amostragem para modelar séries temporais, que não podem ser replicadas pelas técnicas mais comuns de bootstrap que pressupõem independência dos dados amostrais. Nesse contexto, o bootstrap de blocos estacionário que se adapta aos modelos de séries temporais, porém exige estacionaridade da série foi proposto [POLITIS e ROMANO 1994]. No entanto, o bootstrap de máxima entropia é desenvolvido para realizar a replicação de séries temporais sem exigir essa hipótese.

Este trabalho propõe dois métodos: um baseado em SVM e bootstrap estacionário e o outro é composto por SVM e bootstrap de máxima entropia, os métodos são aplicados à quatro séries temporais com comportamentos distintos: a temperatura da superfície do oceano em determinada região, o valor de venda do dólar ao longo de um ano, o volume de vendas de produtos laticínios e dados da aceleração de rolamento em uma máquina.

O artigo está organizado da seguinte maneira: a Seção 2 mostra a base teórica necessária para realizar as previsões, a Seção 3 contém uma explicação da metodologia desenvolvida, a Seção 4 apresenta os resultados obtidos, e a Seção 5 expõe as conclusões do trabalho.

## 2. Fundamentação Teórica

### 2.1 Support Vector Machines

O SVM foi desenvolvido partindo de um embasamento teórico forte e posteriormente seguindo para aplicações. Devido a isso, a ferramenta, documentada inicialmente em 1964, só começou a ganhar visibilidade em 1992, quando a metodologia passou a ser identificada como relevante para aplicações na prática. Atualmente, o SVM apresenta maior desempenho do que outros algoritmos de aprendizado estatísticos, como Redes Neurais, para diversos problemas conhecidos [KECMAN, 2005].

Assim como o algoritmo das Redes Neurais, o SVM é capaz de fazer aproximações para qualquer função multivariada em qualquer grau de precisão. Dessa forma, esse algoritmo é extremamente recomendado para modelar sistemas com funções não-lineares, desconhecidas ou parcialmente conhecidas [KECMAN, 2005].

Nesta pesquisa, usa-se o *Support Vector Machines* para regressão (SVR). No entanto, o SVM pode ser aplicado para resolver problemas de classificação, nos quais são procuradas funções que melhor dividam um conjunto de dados, de acordo com cada categoria. Para maiores



detalhes em classificação via SVM, consultar [CHIH-WEI HSU, CHIH-CHUNG CHANG, 2008].

### 2.1.1 Support Vector Machines de Regressão

Em uma regressão, busca-se encontrar a dependência funcional de uma variável  $y$  com um vetor de variáveis independentes  $x$ . Dessa forma, o algoritmo do SVM de regressão recebe como entrada um conjunto de dados (conjunto de treinamento) relativos às variáveis dependentes e independentes. Após processar esses valores, é possível encontrar uma aproximação que representa a relação entre  $y$  e  $x$ , e calcular valores de  $y$  correspondentes a um novo conjunto de variáveis  $x$  (conjunto de teste). Para associar os valores de  $y$  e  $x$  o algoritmo se baseia na resolução do seguinte problema de otimização:

$$\min_{w,b,\xi,\xi^*} \frac{1}{2} \|w^2\| + C \cdot \sum_l (\xi_l + \xi_l^*), \quad l = 1, 2, \dots, L \quad (1)$$

$$s. a. \quad y_l - w^T \phi(x_l) - b \leq \varepsilon + \xi_l, \quad \forall l \quad (2)$$

$$-y_l + w^T \phi(x_l) + b \leq \varepsilon + \xi_l^*, \quad \forall l \quad (3)$$

$$\xi_l, \xi_l^* > 0. \quad (4)$$

A função objetivo é formada por duas partes: a primeira representa um indicador de capacidade do SVM e, quanto menor, mais generalista é a função encontrada; a segunda parte representa um custo associado aos erros entre os valores reais e previstos da variável dependente na etapa de treinamento. Esses custos são computados para cada ponto  $l$  por  $\xi_l$  e  $\xi_l^*$ , que representam variáveis de folga. O parâmetro  $C$  representa uma medida de *trade-off* entre as duas partes da função objetivo.

As restrições dizem respeito aos limites (superior e inferior) determinados por uma função em formato de um tubo de tamanho  $2\varepsilon$ , que envolve os valores da variável dependente. Os valores de  $w$  e  $b$  são a resposta do problema de otimização (1) - (4) e são os coeficientes estimados da equação de regressão dada pela Eq. (5) [KECMAN, 2005]:

$$y = f(x) = w^T \phi(x) + b \quad (5)$$

Ao se utilizar o modelo mostrado anteriormente para realizar um SVM quando a dimensão do vetor  $x$  é muito elevada, encontramos dois principais problemas [KECMAN, 2005]:

- A escolha de um mapeamento apropriado para  $\Phi(x)$  pode se tornar bastante complexa;
- O produto escalar  $\Phi^T(x)\Phi(x)$  pode exigir um grande esforço computacional caso a dimensão do vetor  $\Phi(x)$  seja muito grande. Esse problema se relaciona ao fenômeno da *maldição da dimensionalidade*.

Esse problema pode ser contornado por meio da introdução de kernels,  $K(x_i, x_j) = \Phi^T(x_i)\Phi(x_j)$ . A vantagem básica do uso dessas funções é evitar um mapeamento direto  $\Phi(x)$  e calcular o produto interno com pouco esforço computacional. A *Gaussian Radial Basis Function* (RBF) é a função kernel (6) mais utilizada, uma de suas vantagens é a necessidade de somente um parâmetro.

$$K(x_i, x_j) = \exp \left[ -\frac{(x_i - x_j)^2}{2\sigma^2} \right]^d \quad (6)$$

Para a resolução do modelo de programação matemática, utiliza-se a sua versão dual, a qual pode ser obtida através das condições de Karush-Kuhn-Tucker. Encontra-se o seguinte modelo:

$$\max_{\alpha, \alpha^*} L_D = -\frac{1}{2} \sum_i \sum_j (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) x_j^T x_j - \sum_i (\varepsilon - y_i) \alpha_i - \sum_i (\varepsilon + y_i) \alpha_i^*, \quad i, j = 1, 2, \dots, l \quad (7)$$



$$s. a. \quad \sum_i (\alpha_i - \alpha_i^*) = 0 \quad (8)$$

$$0 \leq \alpha_i \leq C \quad (9)$$

$$0 \leq \alpha_i^* \leq C \quad (10)$$

em que  $\alpha$  e  $\alpha^*$  representam os vetores de dimensão  $l$  compostos pelos multiplicadores de Lagrange para o problema. Esses valores são encontrados a partir da resolução do modelo, e são utilizados na construção do modelo de regressão (11), no qual o índice 0 indica que o valor faz parte da solução ótima.

$$f(x) = w_0^T \Phi(x) + b_0 = \sum_{i=1}^l (\alpha_{0i} - \alpha_{0i}^*) K(x_i, x) + b_0. \quad (11)$$

### 2.1.1.1 Previsão de série temporal

O modelo do SVR vem sendo amplamente utilizado no âmbito da previsão de séries temporais e tem apresentado resultados excelentes quando comparado a outros modelos, como Redes Neurais [LINS, 2012]. O modelo autorregressivo para a previsão de séries temporais é apresentado a seguir:

$$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-p}), \quad (12)$$

Em que  $y_t$  é a observação da série no ponto  $t$ . Esse valor é função das  $p$  medidas que o antecedem. Desta forma, para previsão de um passo à frente, o  $x$  será correspondente aos próprios valores da série temporal, sendo composto por  $t - p$  vetores de dimensão  $p$  (tabela 1).

Tabela 1 – Entradas e Saídas de SVR para séries temporais

i	Entrada				Saída
1	$y_1$	$y_2$	...	$y_p$	$y_{p+1}$
2	$y_2$	$y_3$	...	$y_{p+1}$	$y_{p+2}$
...	...	...	...	...	...
$t - p$	$y_{t-p}$	$y_{t-p+1}$	...	$y_{t-1}$	$y_t$

Fonte: Lins, 2012 (adaptado)

Após a construção e resolução do modelo de otimização torna-se possível calcular a previsão de um passo da série temporal (13):

$$y_{t+1} = f(y_t, y_{t-1}, \dots, y_{t-p}) = \sum_{i=1}^l (\alpha_{0i} - \alpha_{0i}^*) K(x_i, x) + b_0 \quad (13)$$

## 2.2 Bootstrap

Os métodos bootstrap se dão a partir de técnicas de replicação de amostras, e possuem grande versatilidade em inferência estatística. São exemplos de aplicação: estimativa de erro padrão, estimativa de viés e construção de intervalos de confiança. Um grande atrativo para o uso desses algoritmos como método de análise de incerteza é a possibilidade de aplicá-los sem assumir hipóteses quanto à distribuição da população.

O processo de reamostragem consiste em extrair diversas amostras de um conjunto de dados inicial. Assuma uma amostra inicial; quando se aplica o bootstrap nesses valores, obtém-se uma grande quantidade de amostras diferentes, mas com comportamento semelhante ao da amostra original  $X^1, X^2, \dots, X^B$ . A estatística  $s(X)$  de interesse será calculada para cada uma



dessas amostras geradas e, desta forma, obtém-se um conjunto de valores da estatística de interesse. [EFRON e TIBISHINARI, 1993].

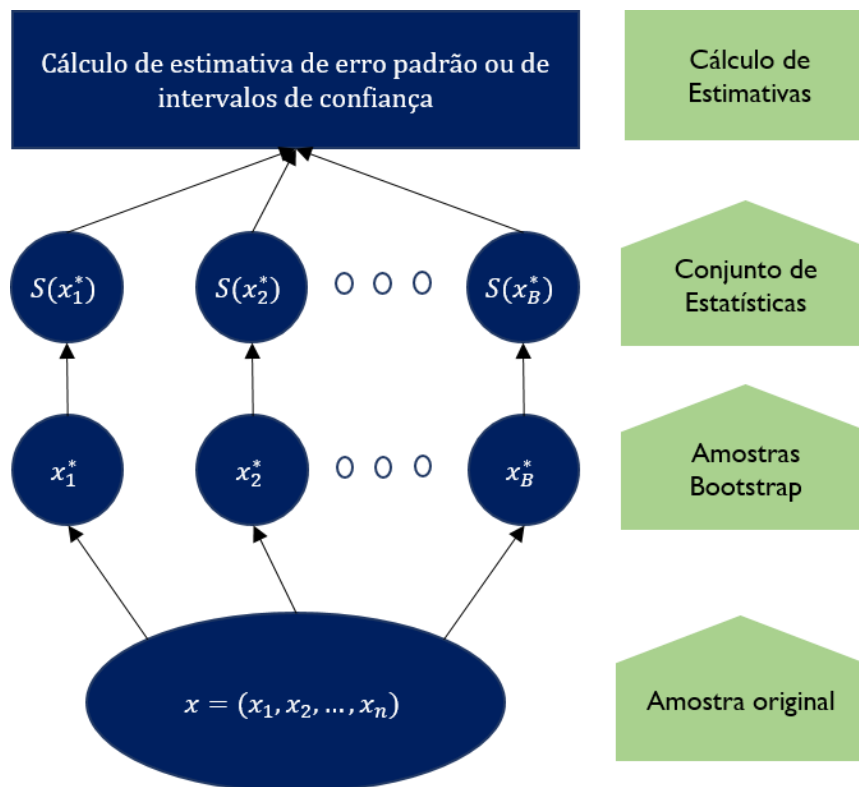


Figura 1 – Esquema de Bootstrap

A forma mais básica do bootstrap realiza as replicações a partir de reamostragem com reposição. Ela assume que os dados originais são independentes e identicamente distribuídos (i.i.d.), e pode ser definido pelo seguinte passo-a-passo:

1. Supõe-se que o conjunto de dados replicados pode ser indicado por  $X = \{x_1, x_2, \dots, x_n\}$
2. Simula-se um conjunto de variáveis aleatórias discretas com distribuição uniforme entre 0 e  $n$ :  $I = \{I_1, I_2, I_3, \dots, I_n\}$ .
3. Constrói-se a amostra bootstrap a partir dos valores obtidos anteriormente:  $x^* = \{x_{I_1}, x_{I_2}, \dots, x_{I_n}\}$ ;
4. Repetem-se as etapas 2 e 3 diversas vezes para obter uma quantidade ideal de amostras bootstrap.

### 2.2.1 Bootstrap Estacionário

O bootstrap básico não se mostra eficiente quando os dados analisados não são independentes e identicamente distribuídos. Dessa forma, essa técnica de reamostragem não é eficaz para a análise de séries temporais. Contudo, existem diversos métodos de replicação desenvolvidos para esse caso.

O bootstrap estacionário foi introduzido em 1994 e assume a hipótese de que a série analisada é estacionária e fracamente dependente. Esse algoritmo busca formar réplicas da série original, das quais é possível recalculer uma estatística desejada. Um fator importante do bootstrap estacionário é que ele garante que as séries replicadas serão estacionárias [POLITIS e ROMANO, 1994].

De forma resumida, a metodologia consiste em dividir a série temporal original em conjuntos de dados, chamados de blocos. O tamanho desses blocos deve seguir uma distribuição geométrica. Os blocos são reamostrados e unidos, buscando formar uma nova série temporal. O processo se repete diversas vezes, obtendo-se diversas séries das quais será calculada uma estatística.



Buscando entender melhor os procedimentos do algoritmo, suponha que  $b$  seja um valor aleatório de distribuição geométrica, com média  $l$ , e que  $i$  representa uma posição na série temporal. Dessa forma, um bloco de tamanho  $b$ , iniciado a partir da medida posicionada em  $i$  ( $X_i$ ), é representado pelo seguinte conjunto:

$$B_{i,b} = \{X_i, X_{i+1}, \dots, X_{i+b-1}\} \quad (14)$$

Descreve-se como obter replicações a partir do bootstrap estacionário. O processo pode ser resumido nos seguintes passos [POLITIS E ROMANO, 1994]:

1. Assuma que a série da qual se deseja extrair as replicações é apresentada como:  $X = \{X_1, X_2, \dots, X_N\}$ ;
2. Assuma que  $L_1, L_2, L_3 \dots$  é uma sequência de medidas independentes e identicamente distribuídas com distribuição geométrica;
3. Assuma que  $I_1, I_2, I_3 \dots$  é uma sequência de números independentes e identicamente distribuídos, seguindo uma distribuição uniforme discreta limitada entre 1 e  $N$ ;
4. Constrói-se uma sequência de blocos a partir dos valores anteriores:  $B_{I_1, L_1}, B_{I_2, L_2}, B_{I_3, L_3}, \dots$ ;
5. A série replicada será construída pela junção dos blocos na ordem desta sequência, até que a série replicada possua, ao menos,  $N$  valores;
6. O processo é repetido para se obter uma quantidade ideal de amostras replicadas para a análise, no caso desta pesquisa,  $R = 1000$ .

## 2.2.2 Bootstrap de Máxima Entropia

Muitas situações do mundo real envolvem séries não estacionárias ou séries nas quais a hipótese de estacionariedade é de difícil verificação. Visto isso bootstrap de máxima entropia (ME Bootstrap) foi apresentado e é aplicável para o caso de séries temporais desse tipo [VINOD, 2004]. Seu diferencial está relacionado à possibilidade de análises, com mais precisão do que no uso de testes de hipóteses baseados em teoria assintótica. Além disso, como este método é válido para séries não estacionárias, ele evita o uso de transformações que destroem o formato da série.

Podemos ver aplicações dessa técnica em diversas pesquisas com temas distintos, por exemplo: análise séries temporais relacionadas à economia asiática [VINOD 2006], estimativas quanto à relação entre consumo de energia e saúde econômica da Turquia [YALTA 2011], análise de risco sobre instituições financeiras [ADAMS 2014], ou cálculo de estimativas sobre a temperatura do ar na Europa central [BARBOSA et al. 2011].

Sendo o conjunto  $\Omega$  formado pela população na qual a série analisada está inserida, o Bootstrap de máxima entropia constrói diversas réplicas da série como elementos de  $\Omega$  a partir de um algoritmo de 7 passos, que foram desenvolvidos para que o formato e dependência com o tempo da função autocorrelação sejam passados da série original para as replicadas [VINOD, 2013].

O passo-a-passo para a replicação de uma série de tamanho  $T$  pode ser visualizado a seguir [VINOD e LOPEZ-DE-LACALLE, 2009]:

1. Ordene os dados originais em ordem crescente, obtendo uma nova série  $x_{(t)}$ , e armazene a posição destes valores antes da transformação;
2. Calcule os pontos intermediários entre os valores da série já ordenados  $z_t = \frac{(x_{(t)} - x_{(t-1)})}{2}$ ,  $t = 2, 3, \dots, T - 1$ ;
3. Calcule o valor da média truncada ( $m_{trm}$ ) dos desvios  $x_{(t)} - x_{(t-1)}$ . Calcule os valores de  $z_1$  e de  $z_T$  utilizando as relações  $z_1 = x_{(1)} - m_{trm}$  e  $z_T = x_{(T)} - m_{trm}$ ;
4. Construa uma função densidade de máxima entropia utilizando os valores de  $z$  como extremidades dos intervalos. A distribuição consiste na união de intervalos de distribuições uniformes com probabilidades iguais (Figura 2). Esta distribuição garante que a média dentro de cada intervalo satisfaça a restrição da preservação da média (do





- inglês: “*mean-preserving constraint*”), essa restrição é desenvolvida para garantir que o teorema ergódico seja satisfeito [VINOD, 2008].
5. Simule, a partir do método da transformada inversa (gerando um valor uniforme entre 0 e 1 e buscando, na distribuição, o quantil referente à medida simulada),  $T$  valores da distribuição construída, ordenando-os de forma crescente;
  6. Reordene os valores originais utilizando a posição armazenada no primeiro passo. Dessa forma, garante-se o formato da série temporal; os valores sempre estarão posicionados de acordo com a sua dimensão. Por exemplo, os valores de maiores dimensões nas séries replicadas e na série original possuem a mesmo índice de posição;
  7. Repete-se o processo entre os passos 2 e 6 para obter diversas réplicas e permitir a análise.

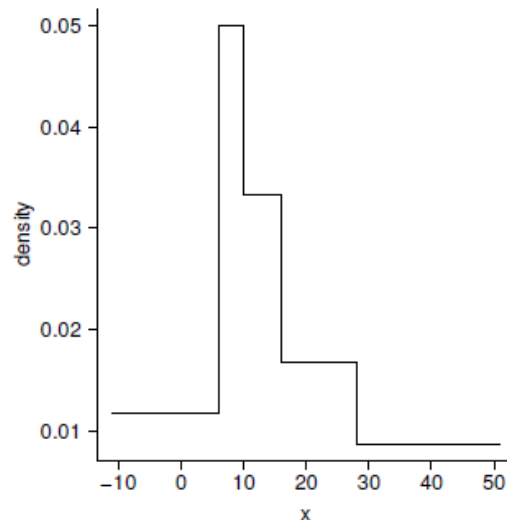


Figura 2 – Exemplo de Função Densidade de Máxima Entropia (Fonte: Vinod e Lopes-de-Lacalle, 2009)

### 2.2.3 Intervalos de confiança via bootstrap pelo método dos percentis

define-se diversos métodos para encontrar um intervalo de confiança utilizando reamostragem [EFRON e TIBISHINARI, 1993]. Dentre eles, encontra-se o método dos percentis, no qual são extraídos quantis de acordo com um valor de significância adotado  $\alpha$ . Esse método possui grande popularidade devido a sua simplicidade [CARPENTER, 2010]. Além disso, a ferramenta não exige hipóteses sobre os parâmetros da distribuição. O intervalo é obtido pelos quantis do conjunto de medidas extraído das réplicas ( $\theta^*$ ):

$$[\theta_{inf}, \theta_{sup}] = [\theta_{\frac{1-\alpha}{2}}^*, \theta_{\frac{\alpha}{2}}^*] \quad (15)$$

### 3. Métodos propostos

Para a construção do intervalo de confiança, os dois bootstraps foram utilizados para a replicação dos dados originais e a previsão foi realizada via SVR para cada replicação. O conjunto de todas as previsões realizadas é utilizado para a construção de um intervalo de confiança (Figura 3). Para cada um dos bootstraps foi necessário replicar o conjunto original de dados 1000 vezes, pois esse valor é ideal para construção de intervalos via método do percentil [CARPENTER, 2010]. O modelo foi implementado de forma computacional através da linguagem R, sendo utilizadas as bibliotecas boot, meboot e kernlab.

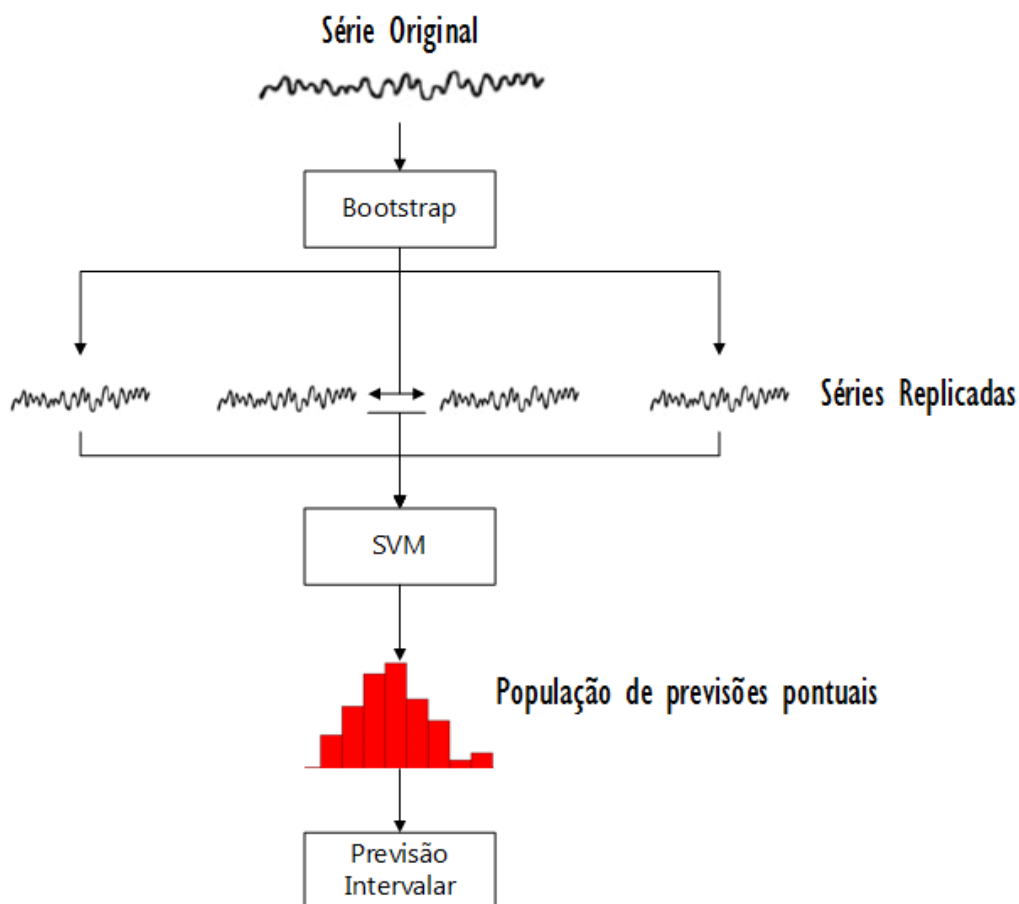


Figura 3 – Metodologia de SVM e bootstrap

Para a construção do modelo de regressão foi necessário estimar os parâmetros ideais do SVM, para isso utilizou-se o método de otimização chamado *particle swarm optimization* (PSO), a metodologia PSO em conjunto com o SVR implica em retirar um conjunto de dados da série analisada (no caso, 10%), e a partir da otimização, buscar o conjunto de parâmetros cuja regressão é melhor ajustada para os valores de teste [LINS, 2009].

#### 4. Aplicações

O SVM combinado ao bootstrap estacionário e ao bootstrap de máxima entropia foi aplicado a quatro séries temporais de diferentes naturezas, a saber:

- Temperatura da superfície do Oceano (SST): Temperatura superficial diária do oceano em graus Celsius (nas coordenadas: 10°S 10°W) referente ao período entre 8 de outubro de 2006 e 8 de outubro de 2007, os dados são originados do projeto PIRATA, e podem ser coletados a partir do endereço (<http://www.pmel.noaa.gov/pirata/>) após a validação.
- Vendas: A venda total de uma família de laticínios. O conjunto de dados passou por uma transformação linear devido à sua confidencialidade.
- Cotação do dólar: O valor de compra do dólar coletado nos dias úteis entre os dias 16 de março de 2016 e 17 de março de 2017, os dados foram retirados do site: <https://economia.uol.com.br/cotacoes/cambio/dolar-comercial-estados-unidos/?historico>.
- Aceleração de rolamento em máquina: Dados de vibração de um rolamento até a quebra fornecidos em termos da aceleração durante um intervalo de tempo de 10s a cada medição. Os dados são referentes ao desafio de prognóstico [IEEE PHM, 2012], para a viabilidade do uso dos dados foi necessário realizar um tratamento que reduziu drasticamente a quantidade de dados.



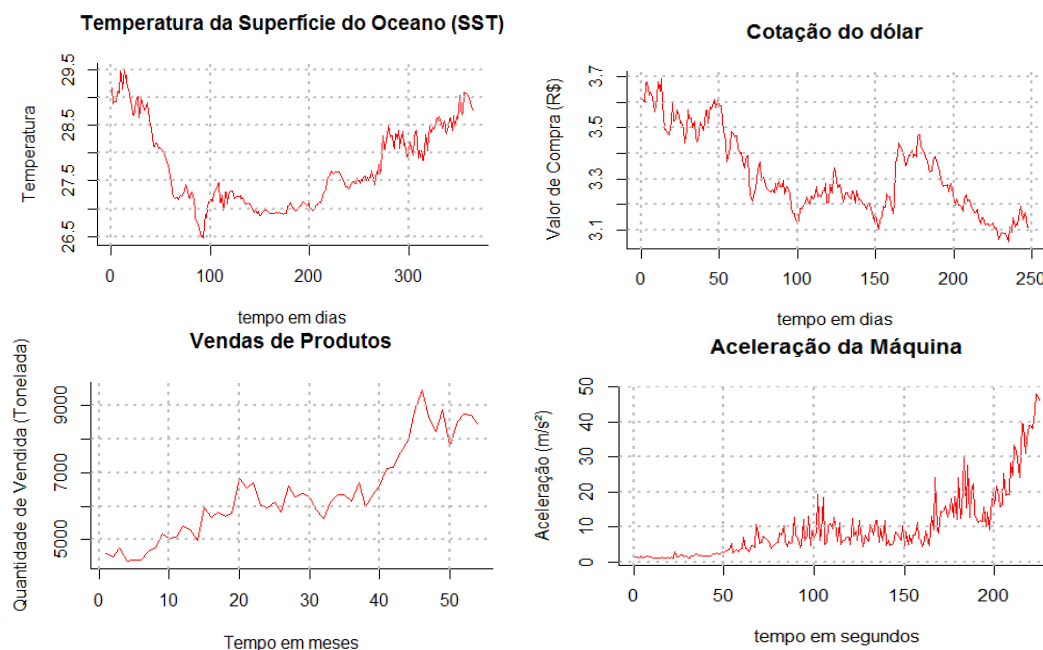


Figura 4 – Séries Temporais Avaliadas

Após aplicação do SVM+PSO a cada uma das séries, foram encontrados os valores adequados para os parâmetros do SVM (Tabela 2).

Tabela 2 – Valores dos parâmetros de SVM para cada uma das séries

	C	$\sigma$	$\epsilon$
SST	917,1542	0,1057	0,1685
Cotação do dólar	1103,8661	0,6555	0,02742
Vendas	15,6191	2,2358	0,0345
Aceleração de rolamento	190,5738	0,2353	0,0012

Após a escolha dos parâmetros do SVM, foram realizadas as replicações via bootstrap estacionário e bootstrap de máxima entropia, seguindo os algoritmos apresentados anteriormente. Resolveu-se o modelo de otimização do SVR e calculou-se o valor da previsão de um passo à frente para cada uma das séries replicadas utilizando a relação de regressão (12).

Com o conjunto de previsões de cada série, foram construídos intervalos por meio do método do percentil. Os resultados estão presentes nas Tabelas 3 e 4 e nas Figuras 5 e 6. Os histogramas facilitam a observação do formato da distribuição das previsões fornecidas pelo SVM para as réplicas bootstrap. Com esses resultados é possível comparar a qualidade dos dois tipos distintos de bootstrap em estimativas relacionadas à previsão de um passo dessas séries temporais. As linhas verticais nos histogramas representam os limites do intervalo, da média das previsões e da previsão pontual fornecida pelo SVM a partir da amostra original

Tabela 3 – Resultados para bootstrap estacionário

Série Analisada	$\theta_{inf}$	$\theta_{sup}$	Amplitude do IC	Média das Previsões	Previsão SVM na série original	Valor Original	Amplitude da série
Série de SST	26,871	29,182	2,31	27,7902	28,7826	28,77	4,29
Cotação do dólar	3,094	3,604	0,510	3,313	3,118	3,100	0,6386
Venda de laticínios	4225	8738	4483	6296	8785	8606	5966,19
Aceleração do rolamento	1,52	38,97	37,45	9,23	47,281	46,250	47,1592



Tabela 4 – Resultados para Bootstrap de Máxima Entropia

Série Analisada	$\theta_{inf}$	$\theta_{sup}$	Amplitude do IC	Média das Previsões	Previsão SVM na série original	Valor Original	Amplitude da série
Serie de SST	28,7033	28,8954	0,1921	28,8000	28,7826	28,77	4,29
Cotação do dólar	3,097	3,127	0,030	3,113	3,118	3,100	0,6386
Venda de Laticínios	8408	9995	1597	8894	8785	8606	5966,19
Aceleração do Rolamento	40,2871	53,8705	13,5834	46,3430	47,2891	46,250	47,1592

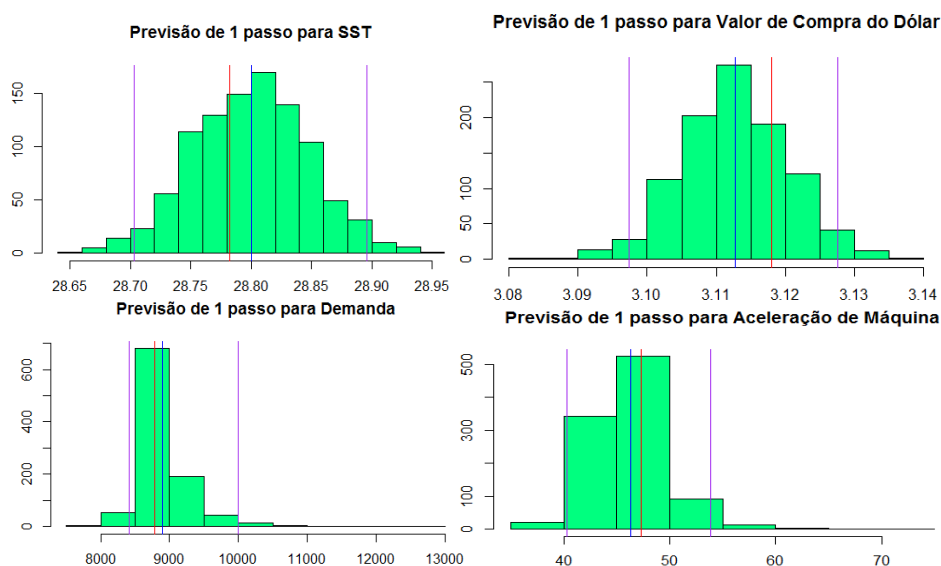


Figura 5 – Histogramas de previsões pontuais em séries replicadas para bootstrap de máxima entropia

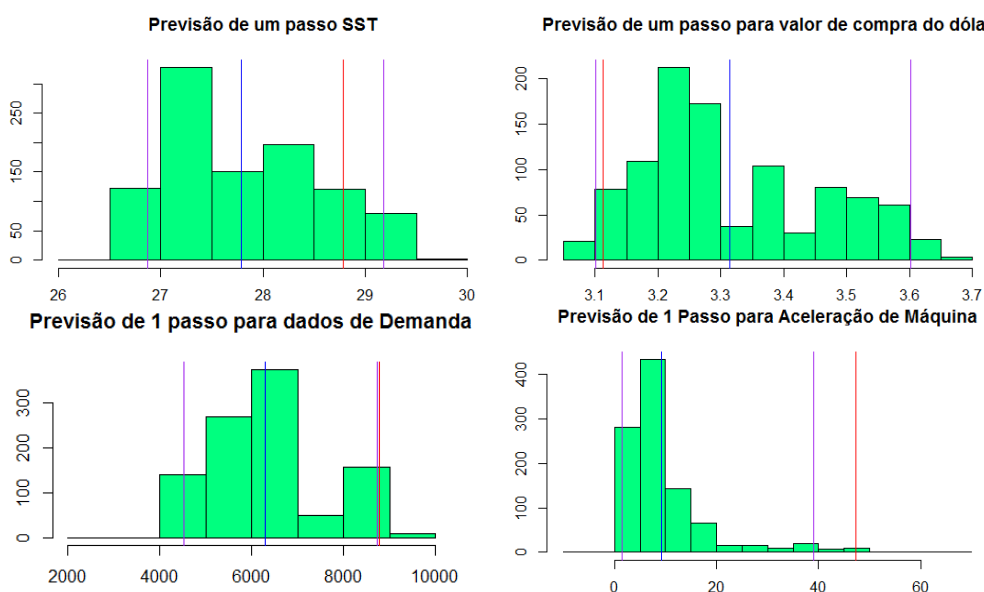


Figura 6 – Histogramas de previsões pontuais em séries replicadas para bootstrap estacionário



Observa-se, inicialmente, uma discrepância em relação ao tamanho dos intervalos obtidos com relação aos dois tipos de bootstrap. Para o bootstrap estacionário, os intervalos são muito grandes em relação à amplitude da série. Por exemplo, para o caso da série da temperatura da superfície do oceano, o intervalo de confiança obtido pelo bootstrap estacionário possui amplitude com o valor de 53,84% da amplitude de todo conjunto de dados representando um intervalo de péssima qualidade quando comparado aos 4,47% obtido para o intervalo referente ao bootstrap de máxima entropia.

O bootstrap estacionário destrói a dependência dos valores com o tempo no conjunto de dados e isso leva a previsões que englobam quase que completamente a série analisada, em séries com tendência (dólar, demanda e vibração de rolamento) os valores os quais buscava-se acertar a previsão se localizaram na extremidade ou fora do intervalo construído.

O bootstrap de máxima entropia mantém a ordem e a dependência no tempo da série original e com isso é possível acertar todos os pontos previstos para as quatro séries temporais analisadas. Observa-se que no caso em que ambos os intervalos acertaram a previsão a média da amostra obtida por máxima entropia está evidentemente mais próxima do valor previsto do que para o caso estacionário. Por exemplo, para o caso da temperatura da superfície do oceano, o valor da média dista do original em 0,01 para o bootstrap de máxima entropia e em 0,21 para o bootstrap estacionário.

#### 4. Conclusão

Neste artigo foi proposto um método diferenciado para a construção previsões intervalares de séries temporais, uniu-se a regressão do SVM a dois tipos de bootstrap específicos para esse tipo de conjunto de dados. As metodologias foram aplicadas para quatro séries temporais distintas e foi possível verificar qual das duas técnicas de replicações utilizadas fornece intervalos com melhor precisão.

É necessário encontrar a técnica de reamostragem que melhor se adapta ao conjunto de dados analisados. Com a designação dos intervalos de confiança da previsão, foi possível concluir que, para as séries analisadas, os resultados são mais favoráveis quando se utiliza o bootstrap de máxima entropia, pois utilizando o bootstrap estacionário foram obtidos intervalos de confiança que englobam praticamente todos os valores presentes na série.

Com essa pesquisa, foi possível implementar a construção de intervalos de confiança de previsões SVM através de análises bootstrap, foi possível identificar a maior viabilidade, para o tipo de séries estudada, de uma técnica de bootstrap sobre a outra. Isso se dá devido à dependência dos dados no tempo, que é destruída quando ocorre a divisão da série em blocos.

#### 5. Referências

ANGUITA, D.; BONI, A.; RIDELLA, S. Evaluating the Generalization Ability of Support Vector Machines through the Bootstrap. **Neural Processing Letters**, v. 11, n. 1, p. 51–58, 2000.

CAO, L. J.; TAY, F. E. H. Support vector machine with adaptive parameters in financial time series forecasting. **IEEE Transactions on Neural Networks**, v. 14, n. 6, p. 1506–1518, nov. 2003.

CARPENTER, J.; BITHELL, J. Bootstrap confidence intervals: When, which, what? A practical guide for medical statisticians. **Statistics in Medicine**, v. 19, n. 9, p. 1141–1164, 2000.

CHIH-WEI HSU, CHIH-CHUNG CHANG, AND C.-J. L. A Practical Guide to Support Vector Classification. **BJU international**, v. 101, n. 1, p. 1396–400, 2008.



DE BRABANTER, K. et al. Approximate confidence and prediction intervals for least squares support vector regression. **IEEE Transactions on Neural Networks**, v. 22, n. 1, p. 110–120, 2011.

EFRON, B.; TIBSHIRANI, R. J. An introduction to the bootstrap. New York: Chapman e all/CRC. 1993.

IEEE PHM 2012 Prognostic Challenge. Scoring of results and application procedure. Web page. <http://www.femto-st.fr/f/d/IEEE-Challenge-Appli.pdf>. Acessado: 2016-08-15

KECMAN, V. “Support vector machines: an introduction. In Studies in Fuzziness and Soft Computing, Wang, L. (ed.), v. 177, p. 1-47. Berlin Heidelberg: Springer-Verlag. 2005

LENDASSE, A. et al. **Fast bootstrap applied to LS-SVM for long term prediction of time series**. 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541). **Anais...IEEE**, [s.d.]Disponível em: <<http://ieeexplore.ieee.org/document/1380003/>>. Acesso em: 14 abr. 2017

LINS, Isis Didier. SUPPORT VECTOR MACHINES AND PARTICLE SWARM OPTIMIZATION APPLIED TO RELIABILITY PREDICTION. 2009. 95 f. Dissertação (Mestrado) - Curso de Engenharia de Produção, Universidade Federal de Pernambuco, Recife, 2009.

LINS, I. D. et al. Computational Statistics and Data Analysis Prediction of sea surface temperature in the tropical Atlantic by support vector machines. **Computational Statistics and Data Analysis**, v. 61, n. February 2015, p. 187–198, 2013.

LINS, I. D. et al. Computing confidence and prediction intervals of industrial equipment degradation by bootstrapped support vector regression. **Reliability Engineering and System Safety**, v. 137, p. 120–128, 2015.

MORETTIN, Pedro Alberto; DE CASTRO TOLOI, Clélia Maria. **Modelos para previsão de séries temporais**. Instituto de matematica pura e aplicada, 1981.

POLITIS, Dimitris N.; ROMANO, Joseph P. The stationary bootstrap. **Journal of the American Statistical association**, v. 89, n. 428, p. 1303-1313, 1994..

RADHIKA, Y.; SHASHI, M. Atmospheric Temperature Prediction using Support Vector Machines. **International Journal of Computer Theory and Engineering**, v. 1, n. 1, p. 55–58, 2009.

VINOD, H. D. Ranking mutual funds using unconventional utility theory and stochastic dominance. **Journal of Empirical Finance**, v. 11, n. 3, p. 353–377, 2004.

VINOD, H. D. Maximum entropy ensembles for time series inference in economics. **Journal of Asian Economics**, v. 17, n. 6, p. 955–978, 2006.

VINOD, H. D.; LOPEZ-DE-LACALLE, J. Maximum Entropy Bootstrap for Time Series: The meboot R Package. **Journal of Statistical Software**, v. 29, n. 5, p. 1–19, 2009.

YALTA, A. T. Analyzing energy consumption and GDP nexus using maximum entropy bootstrap: The case of Turkey. **Energy Economics**, v. 33, n. 3, p. 453–460, 2011.