

Uma Introdução às Séries Temporais

Hugo, Jeandro, Rhudney, Tiago



UNIVERSIDADE
FEDERAL
DE PERNAMBUCO



19 de Outubro de 2015



1 Introdução

2 Estratégias de Previsão

3 Exercícios

4 Referências

Séries Temporais

O que é

- Conjunto de observações de um fenômeno ordenadas no tempo;
- Processo estocástico de tempo discreto;

Exemplos:

- O consumo **mensal** de energia elétrica de uma casa registrado durante os **últimos 2 anos**.
- Temperatura de uma reação química a cada **segundo** em **dois minutos consecutivos**.
- Índice pluviométrico **anual** da região Nordeste no **período de 1988 a 2001**.

Análise de Séries Temporais

- Por quê:
 - Crescente aumento do armazenamento de informações temporais (ex.: Data Warehouse, Cloud).
 - Necessidade de descobrir novas informações a partir dos dados armazenados.
- Objetivos
 - Sumário dos dados (ex.: construção de histogramas, existencia de variações sazonais, tendências, . . .)
 - Estabelecimento de causalidade (ex.: série de vendas está relacionado com série de propagandas)
 - Classificação (ex.: eletrocardiograma classificado como normal ou anormal)
 - **Previsão de Séries Temporais**

Principais métodos para previsão de Séries Temporais

- Modelos Lineares e não Lineares: AR, Holt Winters, ARIMA, SARIMA ...
- Redes Neurais, SVM (*Support Vector Machine*), ...



Figura: Guerra de Algoritmos

Livro Texto

- Teoria Séries Temporais,
- Características do R,
- Exercícios

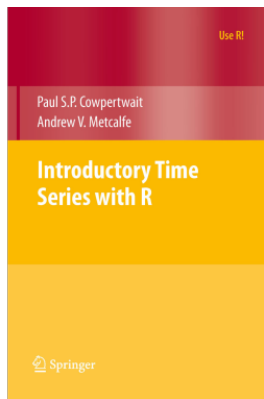


Figura: Livro Texto [1]

Estratégias de Previsão

- Estimar valores da série em tempos futuros, a partir de informações já disponíveis.
- Diminuir incerteza, otimizar ações, diminuir perdas.



Exemplo (Empresa Naval)

A leading international marine paint company uses statistics available in the public domain to forecast the numbers, types, and sizes of ships to be built over the next three years. One source of such information is World Shipyard Monitor, which gives brief details of orders in over 300 shipyards. The paint company has set up a database of ship types and sizes from which it can forecast the areas to be painted and hence the likely demand for paint. The company monitors its market share closely and uses the forecasts for planning production and setting prices.

Correlação Cruzada

- Suponha modelos de séries temporais com variáveis x e y (estacionárias na média e variância)
- As variáveis devem ser serialmente correlacionadas, e correlacionadas entre si em diferentes *lags*
- O modelo combinado é estacionário em segunda ordem se todas essas correlações dependem somente do lag

Cross Covariance Function (ccvf)

$$\bullet \gamma_k(x, y) = E[(x_{t+k} - \mu_x)(y_t - \mu_y)]$$

Correlação Cruzada

Cross Covariance Function (ccvf)

- $\gamma_k(x, y) = \gamma - k_{y,x}$

A função ccf, $\rho_k(x, y)$

- $\rho_k(x, y) = \frac{\gamma_k(x, y)}{\sigma_x \sigma_y}$

A função ccvf e ccf podem ser estimadas de uma série temporal:

- $c_k(x, y) = \frac{1}{n} \sum_{t=1}^{n-k} (x_{t+k} - \bar{x}) - (y_t - \bar{y})$

A amostra acf é definida por:

- $r_k(x, y) = \frac{c_k(x, y)}{\sqrt{co(x, x)co(y, y)}}$

Alisamento Exponencial (AE)

- Fundamentado na suposição de que as últimas observações contêm mais informações sobre o futuro e, portanto, são mais relevantes para a previsão;

O método de AE é definido através da equação:

$$a_t = \alpha x_t + (1 - \alpha)a_{t-1}, \quad t = 1, \dots, N$$

onde a_t é o valor exponencialmente (suavizado) e α é a constante de alisamento, $0 < \alpha < 1$.

Alisamento Exponencial (AE)

Os erros de predição são dados através da equação:

- $e_t = x_t - \hat{x}_{t|t-1} = x_t - a_{t-1}$

Por padrão o R obtém o valor de α minimizando a soma dos erros quadrados. *SS1PE (Sum of Squared one-step-ahead Prediction Errors)*

- $SS1PE = \sum_{t=2}^n e_t^2 = e_2^2 + e_3^2 + \dots + e_n^2 \quad a_1 = x_1$

Exemplo

The number of letters of complaint received each month by a motoring organisation over the four years 1996 to 1999 are available on the website. At the beginning of the year 2000, the organisation wishes to estimate the current level of complaints and investigate any trend in the level of complaints. We should first plot the data, and, even though there are only four years of data, we should check for any marked systematic trend or seasonal effects.

Método de Holt-Winters

Introduz um novo parâmetro: tendência

$$a_t = \alpha(x_t - s_{t-p}) + (1 - \alpha)(a_{t-1} + b_{t-1})$$

$$b_t = \beta(a_t - a_{t-1}) + (1 - \beta)b_{t-1}$$

$$s_t = \gamma(x_t - a_t) + (1 - \gamma)s_{t-p}$$

Onde a_t , b_t e s_t são estimados pelas médias do efeito sazonal no tempo t e α , β e γ são os parâmetros de alisamento.

Exemplo: Four-year-ahead forecasts for the air passenger data

The seasonal effect for the air passenger data appeared to increase with the trend, which suggests that a 'multiplicative' seasonal component be used in the Holt-Winters procedure. The predict function in R can be used with the fitted model to make forecasts into the future.

Exercícios


- 1 Describe the association and calculate the ccf between x and y for k equal to 1, 10, and 100.

Exercícios

The data in the file wine.dat are monthly sales of Australian wine by category, in thousands of litres, from January 1980 until July 1995. The categories are fortified white, dry white, sweet white, red, rose, and sparkling. The sweet white wine time series is plotted in Figure 3.9, and there is a dramatic increase in sales in the second half of the 1980s followed by a reduction to a level well above the starting values. The seasonal variation looks as though it would be better modelled as multiplicative, and comparison of the SS1PE for the fitted models confirms this.

- Plote o gráfico a série do vinho branco doce
- Quais os parâmetros de alizamento ótimos ?
- Plote o gráfico com os valores ajustados.

Referências I

-  Paul SP Cowpertwait and Andrew V Metcalfe.
Introductory time series with R.
Springer Science & Business Media, 2009.