Universidade Presbiteriana
Mackenzie

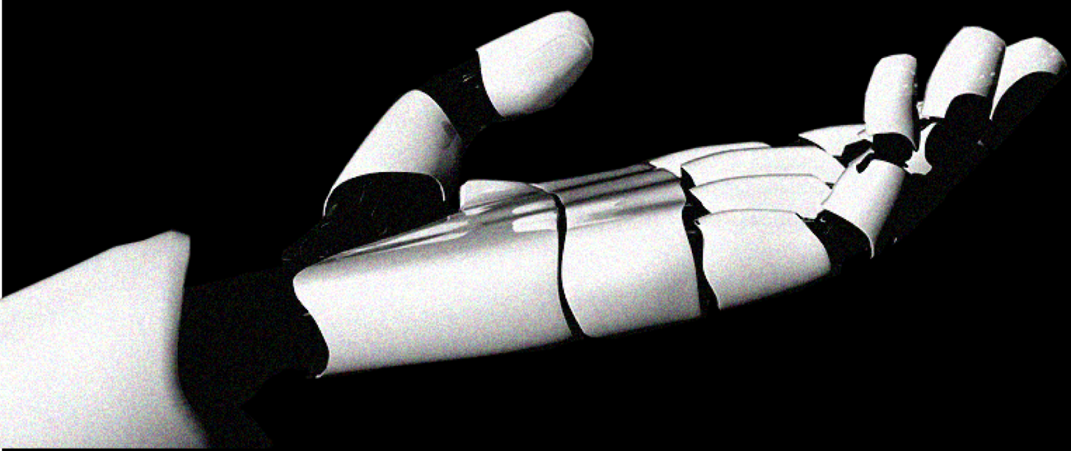
Inteligência Artificial

módulo 4

Introdução à Previsão de Séries Temporais e à Aprendizagem por Reforço

Universidade Presbiteriana
Mackenzie

Trilha 4
Regressão para Séries Temporais e Suavização



Sumário

Introdução ao Estudo da Trilha de Aprendizagem

O objetivo é discutir os principais conceitos referentes ao estudo de regressão e de suavização para séries temporais, tendo como base os seguintes tópicos:

- introduzir os principais conceitos de regressão para séries temporais;
- apresentar a suavização para séries temporais.

Regressão para séries temporais

Uma das formas de se estudar uma série temporal é aplicando os modelos de regressão, sendo estes muito utilizados para análises em diversas áreas.

!

O uso da regressão tem como finalidade obter uma equação que consiga apresentar e explicar, da melhor forma possível, a relação existente entre uma variável dependente com uma ou mais variáveis independentes, e assim possibilitando estimar os valores futuros da variável dependente (MORETTIN & TOLOI, 2006).

Quando nos referenciamos ao uso dos métodos de regressão para analisar séries temporais, estamos considerando usar métodos estatísticos a fim de proporcionar uma melhor compreensão do comportamento da série temporal.

Coeficiente de correlação

Para verificar como uma variável influencia em outra variável, precisamos quantificar o **grau de associação entre estas duas variáveis** e isto é feito utilizando o coeficiente de correlação. O coeficiente de correlação mede este grau de associação com um único número que varia entre -1 a 1 (BUSSAB & MORETTIN, 2017).

Uma das formas de se **calcular este coeficiente é o de Pearson**, que utiliza o número de observações, as quais devem ser iguais para as duas variáveis X e Y. Nesta aplicação, a correlação, conforme o valor do coeficiente é mais próximo de 1 ou -1, indica que é mais forte onde o sinal indica se a correlação é negativa ou positiva.

Quando o coeficiente de correlação de Pearson é elevado ao quadrado, temos o coeficiente de determinação. Através deste coeficiente obtemos a quantidade de variação de uma variável devido à variação de outra variável.

$$R^2 = r^2$$

Existem dois métodos principais para realizar a análise de uma série temporal: regressão linear simples e regressão linear múltipla. Como ambos os modelos de regressão descrevem a relação entre duas variáveis utilizando uma equação matemática, utilizamos o coeficiente de correlação para calcular este número.

Uma das formas de se obter os coeficientes de regressão é pelo **Método dos Mínimos Quadrados (MMQ)**, o qual busca definir a curva ideal que minimiza o somatório dos erros. No entanto, como os erros podem resultar em valores positivos ou negativos, elevamos os mesmos ao quadrado para solucionar esse problema (SILVA & MATTOS, 2017).

! **O Método dos Mínimos Quadrados é um dos métodos mais utilizados para estimar os parâmetros da equação de uma regressão linear. Isto porque o MMQ necessita que algumas suposições mínimas sejam atendidas para ser aplicado, entre elas temos que variável X é controlada, pois não sofre influência das variações aleatórias, e que os erros da equação são não correlacionados.**

A seguir, veremos um pouco sobre cada um destes métodos: regressão linear simples e múltipla.

Regressão linear Simples

O objetivo de se realizar uma análise de regressão é o de fazer uma análise estatística e assim obter uma equação que permita explicar as alterações encontradas na variável dependente causadas pela variação das variáveis independentes.

!

Em outras palavras, a Regressão Linear Simples visa fornecer uma equação que descreve o comportamento de uma variável em função do comportamento da outra. Tendo que o gráfico da equação de regressão é uma reta, o objetivo é apresentar a melhor reta que se ajusta ao conjunto de dados observados das variáveis analisadas.

Quando temos uma relação linear que possibilite sumarizar a dependência encontrada entre as variáveis dependentes e independentes, podemos então utilizar a seguinte equação para descrever esta relação (BUSSAB & MORETTIN, 2017):

$$Y_i = \alpha + \beta X_i$$

Onde temos que:

- Y_i é o valor encontrado para a variável dependente Y no i-ésimo nível da variável independente X;
- α é a constante de regressão que representa a interceptação da reta com o eixo Y;
- β é o coeficiente de regressão que representa o quanto a variável Y varia em média para um aumento de uma unidade da variável X;
- X_i é o i-ésimo nível da variável independente Xi (i = 1,2,...n).

Esses parâmetros estão representados na Figura 1.

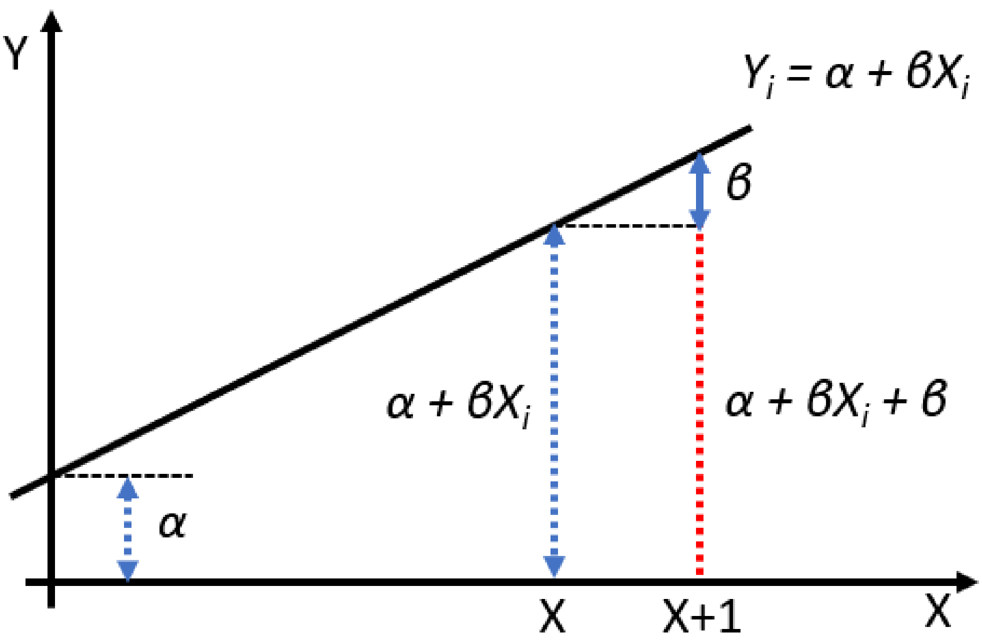


Figura 1 - Representação da equação $Y_i = \alpha + \beta X_i$. Fonte: Elaborado pelo autor.

Esta equação para a relação linear entre X e Y é determinística, indicando que todos os pontos das variáveis se encontram exatamente na reta de regressão. Porém, isto não costuma acontecer, normalmente os valores observados são encontrados fora da linha reta.

Teremos então uma diferença entre o valor encontrado e o valor fornecido pela equação. O erro é como é chamada esta diferença, que podemos representar por σ, sendo uma variável aleatória que pode ser devido ao efeito de variáveis que não foram consideradas ou a erros de medição dos dados.

O erro irá quantificar a falha obtida pelo modelo em se ajustar aos dados. Ao adicionar esse erro à equação de regressão linear, temos:

$$Y_i = \alpha + \beta X_i + \sigma_i$$

Onde temos as mesmas informações da equação anterior acrescidas por:

- σ_i é o erro associado à distância entre o valor observado Yi e o valor correspondente na curva para o mesmo nível i de X.

No método da regressão linear simples, temos que a variável independente X é uma variável controlada pelo responsável pela análise, enquanto a variável dependente Y é uma variável aleatória, ou seja, para cada valor que X pode assumir, temos uma distribuição de probabilidade para Y. Na figura 2, vemos uma representação da regressão linear simples.

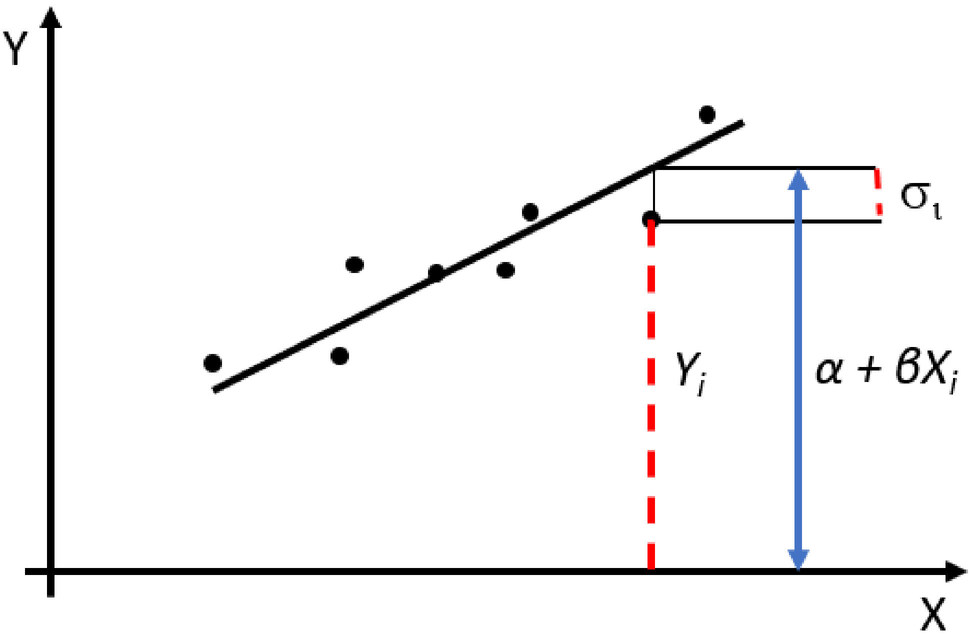


Figura 2 - Representação da regressão linear simples. Fonte: Elaborado pelo autor.

Conforme o gráfico na figura 2, o valor Y_i representa o valor observado do dado; $\alpha + \beta X_i$ representa o valor apresentado pela equação e σ_i é o erro entre o valor esperado e o encontrado.

Regressão linear múltipla

Outro método de regressão aplicado para séries temporais é a regressão linear múltipla. Esta regressão tem como diferencial, em relação à regressão linear simples, três ou mais variáveis; uma única variável dependente (Y) e duas ou mais variáveis independentes (X_i, i = 1, 2, ...).

!

Assim como na regressão linear simples, o objetivo da análise com a regressão múltipla é encontrar uma equação que possa ser usada para estimar os valores de Y conforme os valores apresentados pelas variáveis independentes.

Uma característica deste método, em comparação com a regressão linear simples, é que, **com o acréscimo das variáveis independentes, possibilita-se melhorar a capacidade de predição**. Porém a técnica de cálculo se torna mais complexa. A seguir, vemos a equação que representa este método:

$$Y_i = \alpha + \beta_1 X_{1j} + \beta_2 X_{2j} + \ldots + \beta_k X_{kj} + \upsilon_j, \quad j = 1, \ldots, n$$

Onde temos que:

- Y_i é a variável dependente obtida em função das variáveis independentes;
- X_{ij} (i = 1 a k) são as variáveis independentes fixas;
- β_i (i = 1 a k) são os coeficientes de regressão parciais;
- υ_j são os erros associado à distância entre o valor observado Y_i e o valor correspondente na curva para o mesmo nível i de X.

Suavização

Quando realizamos uma análise de uma série temporal, considera-se que as observações dos dados do passado possuem informações que indicam o padrão de comportamento da série. Assim, os métodos vistos ao longo das trilhas deste componente apresentam formas de se identificar o padrão e separá-lo dos ruídos ou variáveis aleatórias.

!

É neste ponto que encontramos os modelos de suavização. A suavização considera que os valores extremos de uma série temporal representam a aleatoriedade. Utilizando, então, a suavização, é possível suavizar estes pontos extremos, permitindo identificar um padrão, e com isto podendo utilizar estes pontos na série temporal para prever os valores futuros (MORETTIN & TOLOI, 2006).

O procedimento de suavização é chamado de **Modelo de suavização exponencial**, este modelo possui uma simplicidade muito grande e uma eficácia computacional satisfatória, além de apresentar uma razoável precisão.

Para utilizar o modelo de suavização exponencial, devemos obter os dados da série temporal e separar em duas partes, sendo uma para a modelagem e outra para a testagem.

Então deve-se definir qual método de suavização deve ser utilizado para a série temporal, verificando se existe tendência, sazonalidade e a forma da sazonalidade. Os principais modelos existentes são:

- Suavização exponencial simples;
- Suavização linear de Holt;
- Suavização Holt-Winters aditivo e multiplicativo.

A seguir, veremos cada um destes modelos.

Suavização exponencial simples

A suavização exponencial simples é indicada quando temos **modelos sem tendência**. Desta forma, os dados são representados pela média da série e podem ser representados pela equação:

$$Y_t = \mu + \varepsilon_t \, , \; t = 1, 2, \ldots, T$$

Onde temos que:

- Y_t é a variável a ser estimada;
- μ é um nível médio constante;
- ε_t é a variável aleatória no instante t.

Neste caso, é preciso estimar o μ do modelo. Podemos estimar o μ utilizando as médias móveis. As médias móveis são utilizadas para descrever o procedimento em que uma nova média é calculada trocando o dado observado mais antigo e incluindo o dado mais recente.

Assim, a média será a previsão para o próximo período e com isto podem-se estimar os dados futuros. Dentre os tipos de médias móveis temos as médias móveis ponderadas, em que se atribui um peso para cada valor observado na série.

Ao se adotarem as médias móveis ponderadas, a observação mais recente apresenta uma melhor informação para o dado futuro e, por isto, quanto mais antigo for o dado, menor deve ser o seu peso atribuído.

Este procedimento deve ser feito de forma gradual, assim, para os dados mais recentes, devemos atribuir um peso maior, e quanto mais antigo, menor o peso. Para saber mais sobre as médias móveis, ver a trilha 3, “Médias móveis”, deste mesmo componente de aprendizagem.

A estimativa do μ pode ser realizada através da **Suavização Exponencial Simples (SES)**. Este método considera que as observações da série temporal mais recentes têm mais importância para a definição do modelo.

Assim temos a equação:

$$\hat{u} = \hat{Y}_t = \hat{Y}_{t-1} + \alpha\,(Y_{t-1} - \hat{Y}_{t-1}) = \alpha Y_{t-1} + (1-\alpha)\hat{Y}_{t-1},\ t = 1, 2, \ldots, T$$

Na figura 3, vemos um exemplo da aplicação da suavização exponencial simples com α igual a 0,9 e α igual a 0,2. Também está sendo considerado que trabalharemos com números inteiros.

Figura 3 - Suavização exponencial simples. Fonte: Elaborado pelo autor.

mês/ano	t	valor	α = 0,9	α = 0,2
jan/18	0	102		
fev/18	1	100	102	102
mar/18	2	110	100	102
abr/18	3	143	109	103
mai/18	4	100	140	111
jun/18	5	98	104	109
jul/18	6	107	99	107
ago/18	7	111	106	107
set/18	8	102	111	108
out/18	9	130	103	107
nov/18	10	115	127	111
dez/18	11	112	116	112
jan/19	12			

Na aplicação da fórmula, o primeiro valor $\hat{Y}_t$ terá o último valor da série temporal. Assim, no caso, o valor de janeiro de 2018 é 102, então o $\hat{Y}_t$ irá ter o valor 102. O cálculo segue a equação conforme o exemplo para o mês de dezembro de 2018, em que se substituiu o mês e ano pelo t: 0, 1, 2 e assim por diante:

$$\hat{Y}_{11} = \alpha Y_{10} + (1-\alpha)\hat{Y}_{10} = (0{,}9)(115) + (0{,}1)(127) = 116$$

Agora com α igual a 02, temos:

$$\hat{Y}_{11} = \alpha Y_{10} + (1- \alpha)\hat{Y}_{10} = (0,2)(115) + (0,8)(111) = 112$$

A seleção do α é muito importante, pois está relacionado com a previsão do valor a ser obtido. Neste caso, quanto menor for o α, mais estável será a previsão, já que quando se utilizam valores baixos para α, temos que pesos maiores serão destinados aos dados do passado, mais antigos.

Na figura 4, vemos um gráfico comparando o valor e os dois resultados apresentados na figura 3.

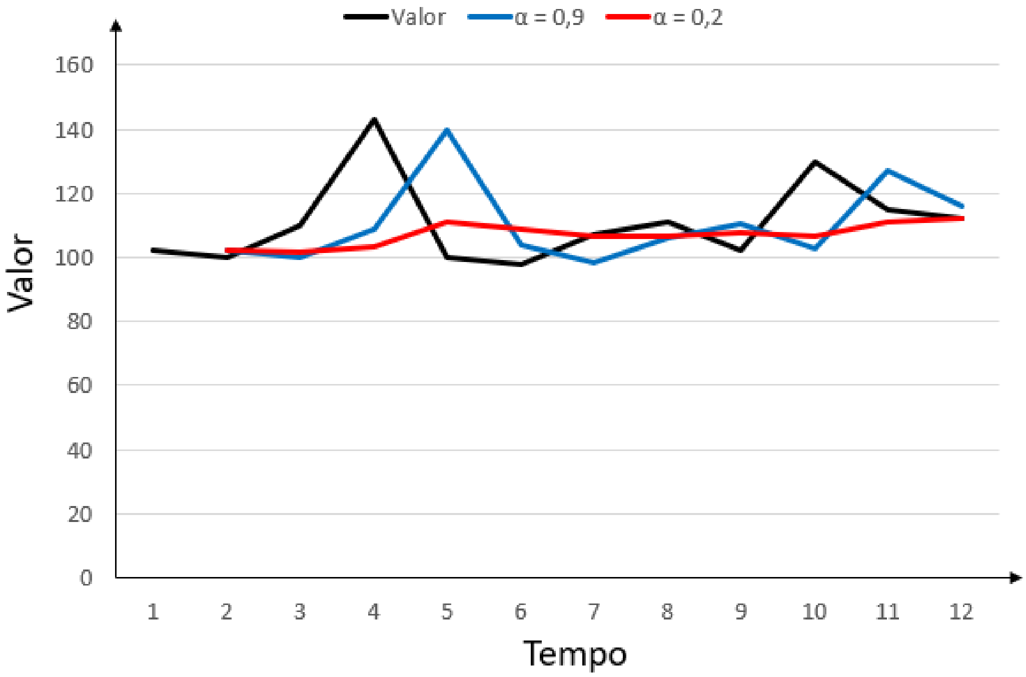


Figura 4 - Suavização exponencial simples. Fonte: Elaborado pelo autor.

Como o valor de α é livre, para determinar o melhor valor possível para cada série temporal podemos utilizar algumas formas de comparação. A média dos quadrados dos erros (MSE – *Mean Square Error*) é a forma mais utilizada.

A aplicação deste procedimento inicia selecionando um valor inicial para α, as previsões serão geradas a partir deste valor inicial. Então se comparam os valores observados dos dados com os valores previstos obtidos e calcula-se a média dos quadrados das diferenças entre eles.

Por fim, o valor de α que minimiza a média é o a ser adotado na suavização exponencial simples.

Suavização linear de Holt

A suavização linear de Holt foi apresentada em 1957, quando Holt estendeu o método da suavização exponencial simples (SES) com o objetivo de obter previsões para séries com tendência.

! **O motivo principal é que, quando o método SES é aplicado em uma série temporal que tem tendência linear, este método apresentava resultados que eram superestimados ou subestimados com relação aos valores reais da série.**

A equação que representa este modelo é:

$$Y_t = \mu + T_t + \varepsilon_t \text{ , } t = 1, 2, \dots, T$$

Onde temos que:

- Y_t é a variável a ser estimada;
- μ é um nível médio constante;
- T_t é o componente da tendência no período t;
- ε_t é a variável aleatória no instante t.

O modelo de suavização linear de Holt pode ser estimado utilizando regressão linear simples, assim obtemos o modelo de Holt:

$$\hat{Y}_t = \alpha\hat{Y}_t + \beta\hat{Y}_{t-1} + \varepsilon_t,\ t = 1, 2, \dots, T$$

Onde temos que:

- α é uma constante para suavizar o nível médio;
- β é uma constante de suavização para modelar a tendência da série.

Para se obter o nível médio e a tendência da série, devemos calcular L_t e T_t, respectivamente, através das equações:

$$L_t = \alpha Y_t + (1 - \alpha)(L_{t-1} + T_{t-1}) \qquad 0 < \alpha < 1$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \qquad 0 < \beta < 1$$

Então teremos que as previsões serão obtidas por:

$$\hat{Y}_t(h) = L_t + hT_t \qquad h > 0$$

Para iniciar o procedimento da aplicação da suavização exponencial linear de Holt, devemos inicialmente obter uma estimativa para os primeiros valores de L_1 e T_1. Para isto, podemos inicialmente igualar o valor de L_1 com Y_1 e, no caso do T_1, igualar a diferença entre Y_2 e Y_1.

Também é preciso verificar quais serão os valores iniciais de α e β. Para isto, podemos fazer da mesma forma que ao se usar o SES, procurando pela menor média de quadrados dos erros através da variação dos α e β.

Suavização Holt-Winters aditivo e multiplicativo

A suavização Holt-Winter é indicada **quando a série temporal apresenta tanto a tendência quanto a sazonalidade**. Em meados de 1960, o modelo de suavização linear de Holt foi estendido por Winters para tratar da sazonalidade.

A suavização de Holt-Winters pode ser aplicada quando o componente sazonal apresenta comportamento aditivo ou multiplicativo e é baseado em três equações de suavização: uma para o nível, outra para a tendência e, por fim, uma para a sazonalidade. Sendo que esta última pode ser aditiva ou multiplicativa.

Para se obter o nível médio e a tendência da série, devemos calcular Lt e Tt, respectivamente, através das equações:

$$L_t = \alpha(Y_t - S_{t-s}) + (1 - \alpha)(L_{t-1} + T_{t-1}) \qquad 0 < \alpha < 1$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \qquad 0 < \beta < 1$$

$$S_t = \gamma(Y_t - L_t) + (1 - \gamma)S_{t-s} \qquad 0 < \gamma < 1$$

E as previsões serão obtidas por:

$$\hat{Y}_t(h) = L_t + hT_t + S_{t-s+h} \qquad h = 1,2, \ldots, s$$

Onde γ é o coeficiente de suavização da componente sazonal e S é um período completo da sazonalidade, isto é, se estamos trabalhando com dados mensais, $S = 12$, com dados trimestrais, $S = 4$.

Assim também ocorre para a suavização de Holt-Winters multiplicativo, ou seja, quando o componente sazonal é multiplicativo. No caso, as equações serão:

$$L_t = \alpha(Y_t / S_{t-s}) + (1 - \alpha)(L_{t-1} + T_{t-1}) \qquad 0 < \alpha < 1$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} \qquad 0 < \beta < 1$$

$$S_t = \gamma(Y_t / L_t) + (1 - \gamma)S_{t-s} \qquad 0 < \gamma < 1$$

E as previsões serão obtidas por:

$$\hat{Y}_t(h) = (L_t + hT_t)\, S_{t-s+h} \qquad h = 1,2, \ldots, s$$

Síntese

Neste *e-book*, foram apresentados os principais conceitos sobre regressão para séries temporais, percorrendo desde coeficiente de correlação até os principais tipos, que são a regressão linear simples e a regressão linear múltipla. Por fim, para encerrar o tema sobre séries temporais, foi apresentada a suavização, e também seus conceitos e principais modelos, que são a suavização exponencial simples, a suavização linear de Holt e a suavização Holt-Winters aditiva e multiplicativa.

Continuem consumindo os outros materiais da trilha, como o audioblog e o vídeo explicativo. Cada material foi elaborado para melhorar a compreensão do assunto tratado nesta trilha.

Referências

AGGARWAL, C. C. ***Neural Networks and Deep Learning****: A textbook*. Berlin: Springer, 2018.

BISGAARD, S.; KULAHCI, M. ***Time Series Analysis and Forecasting by Example***. Wiley, 2011.

BUSSAB, W. O.; MORETTIN, P.A. **Estatística Básica**. 9ª ed. São Paulo: Saraiva, 2017.

GÉRON, A. ***Hands-On Machine Learning with Scikit-Learn and TensorFlow***. Sebastopol: O'Reilly, 2017.

GRAESSER, L.; KENG, W. L. ***Foundations of Deep Reinforcement Learning****: Theory and Practice in Python*. Boston: Addison-Wesley, 2019.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de séries temporais**. 2ª ed. São Paulo: Edgard Blücher/ ABE- Projeto Fisher, 2006.

SILVA, V. M.; MATTOS, V. L. D. O método dos mínimos quadrados no ajuste de um modelo polinomial. **Scientia Plena**, 13(4), p. 01-08, 2017.

SUTTON, R. S.; BARTO, A. G. ***Reinforcement Learning***. 2ª ed. Cambridge: The MIT Press, 2018.

Universidade Presbiteriana
Mackenzie