Universidade Presbiteriana Mackenzie

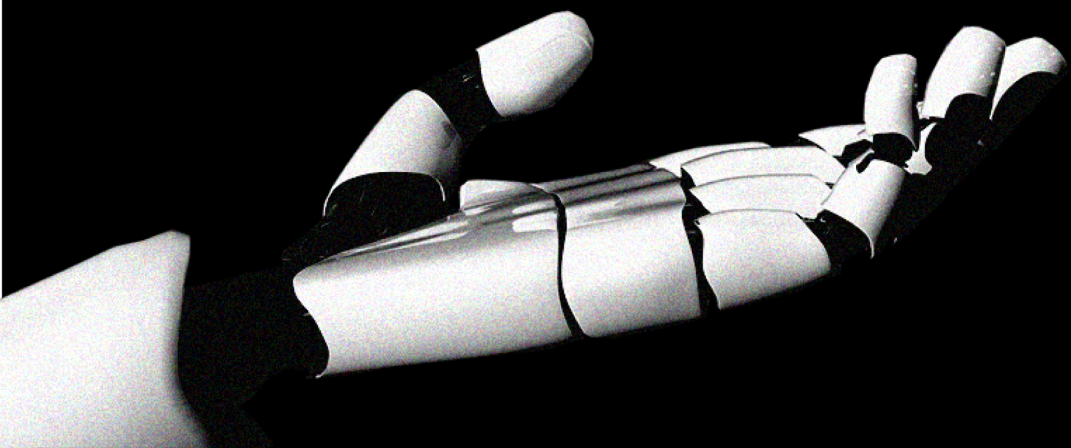
Inteligência Artificial

módulo 4

Introdução à Previsão de Séries Temporais e à Aprendizagem por Reforço

Universidade Presbiteriana Mackenzie

Trilha 2
Tendência e sazonalidade



Sumário

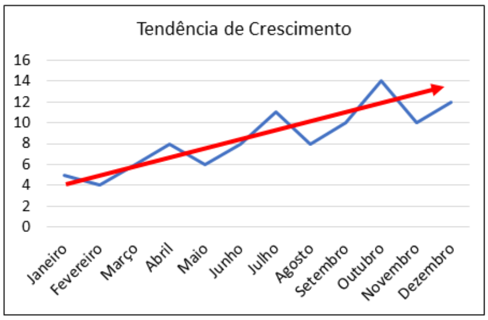
Introdução ao Estudo da Trilha de Aprendizagem

O objetivo é apresentar os conceitos e finalidades de tendência e sazonalidade relacionados às séries temporais, divididos nos seguintes tópicos:

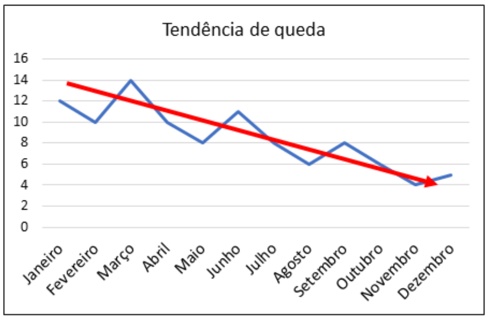
- apresentar os conceitos de tendência e seus usos nas séries temporais;
- apresentar os conceitos de sazonalidade e seus usos nas séries temporais.

Tendência nas séries temporais

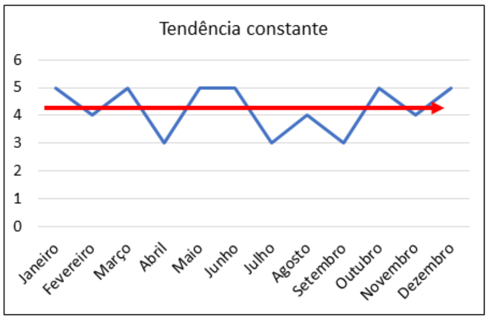
Na trilha 1, vimos que na decomposição clássica de uma série temporal temos o componente da tendência, este componente indica se a série temporal ao longo do tempo possui tendência de crescimento, queda ou se mantêm estável, constante. Na figura 1, vemos as representações das tendências:



a) Série temporal com tendência de crescimento



b) Série temporal com tendência de queda



c) Série temporal com tendência constante

Figura 1 - Tipos de tendências. Fonte: Elaborado pelo autor.

Assumindo que a fórmula que representa uma decomposição clássica de uma série temporal é a seguinte:

$$X_t = T_t + S_t + a_t,\ t = 1, 2, \ldots N$$

Onde T_t é o componente de tendência, S_t é o componente cíclico ou sazonal e a_t é o componente aleatório ou ruído branco. Podemos expressar a tendência linear da seguinte forma:

$$T_t = \alpha + \beta t,$$

Nesta equação, α e β são as constantes a serem estimadas.

!

Uma das finalidades de se isolar o componente da tendência é a de identificá-la e assim conseguir realizar o processo de previsão, através da análise de regressão. Outra finalidade é a de remover a tendência para que seja possível reconhecer os demais componentes da série temporal. Neste caso, utilizando as médias móveis. Na próxima seção, sazonalidade, temos um exemplo de como se calcular a média móvel, item necessário para se obter o componente da sazonalidade.

Quando estamos verificando a existência do componente de tendência em uma série temporal, precisamos identificar se os dados observados são independentes. Normalmente os dados de séries temporais não são independentes, possuindo uma dependência das observações, ainda mais nos casos em que os intervalos analisados são curtos.

A informação sobre a dependência dos dados é importante, pois a maioria dos testes de tendência consideram que existe esta dependência. **Os principais testes para identificar as tendências são: testes de Wald-Wolfowitz, o teste de Cox-Stuart, o teste de Mann-Kendall e o teste de Spearman**.

Dentre estes testes, podemos mencionar o teste de Mann-Kendall. Este é um teste não paramétrico, em que a hipótese é que temos uma característica constante da série temporal em que os valores estão sempre constantes e sem alteração na distribuição da probabilidade (SNEYERS, 1975).

Os métodos clássicos são bem eficientes para se trabalhar com a série temporal quando temos uma tendência linear e/ou sazonalidade regular. Contudo, quando a tendência e/ou a sazonalidade mudam constantemente ao longo do tempo ou quando temos uma correlação entre os valores sucessivos das flutuações irregulares, então eles não funcionam bem (CHAFTFIELD, 2000).

Nas situações em que a série temporal tiver mais do que um componente não aleatório, é sugerido que se realize o teste da existência da tendência após a remoção do outro componente (MORETTIN & TOLOI, 2006).

Existem diversos métodos para estimar a tendência T_t, a maioria utiliza regressão e ajuste exponencial de dados. Segundo Morettin e Toloi (2006), os métodos mais utilizados são:

1. ajustar uma função temporal, como um polinômio, uma exponencial ou alguma função suave de t;
2. realizar a suavização ou filtro nos valores da série temporal ao redor de um ponto específico, assim estimando a tendência daquele ponto;
3. suavizar os valores da série temporal, realizando ajustes sucessivos de retas de mínimos quadrados, sendo esta uma técnica de regressão local chamada "lowess".

Obtendo a estimativa da tendência através de T'_t, torna-se possível obter a série temporal ajustada para a tendência ou a série com a tendência removida:

$$Y_t = X_t + T'_t$$

onde Y_t é a nova série temporal ajustada, X_t é a série temporal original e T'_t é a tendência.

Métodos de Ajustes

! **O método de ajuste linear tem como objetivo encontrar os coeficientes da equação da reta mais ajustados aos dados analisados. Este método consiste na utilização do método dos mínimos quadrados.**

Segundo Barroso (1997), no método de ajuste linear, temos que a fórmula da tendência se apresenta na forma mais básica, relacionando duas variáveis qualquer X e Y, sendo a equação:

$$Y = a + bX$$

Onde Y é o valor da tendência, X é o valor do tempo, a é o coeficiente linear da reta e b é o coeficiente angular da reta (se positivo indica tendência crescente, se negativo, a tendência é decrescente).

Para obtermos a tendência por mínimos quadrados, temos que considerar que a variável independente sempre será o tempo. Por exemplo, em uma série temporal registrada diariamente durante 120 dias, a variável independente assume os valores dos dias, indo de 1 a 120.

Normalmente a tendência é modelada utilizando uma função polinomial, já que possui um número grande de parâmetros. A função polinomial contém tanto a equação da reta quanto um ajuste linear múltiplo. Este método é mais adequado quando temos tendências que são crescentes ou decrescentes.

As equações a serem utilizadas para obter as tendências podem ser a linear, polinomial, logarítmica ou exponencial, por exemplo. Para aplicá-las, podem ser utilizados aplicativos computacionais nos quais se usa como entrada a série temporal e se obtêm a tendência como saída.

Uma **linha de tendência linear** é utilizada com séries temporais simples, composta por dados lineares. Os dados analisados serão lineares caso o padrão em seus pontos parecer com uma linha. Quando temos este tipo de linha de tendência, verificamos que as informações da série temporal estão crescendo ou diminuindo de forma constante. Na figura 1, vimos representações da tendência linear.

Quando temos uma linha de tendência polinomial, normalmente vemos uma **linha curva** indicando uma variação dos dados. Nas situações em que temos uma série temporal bem longa, é interessante usar esta tendência para poder analisar as subidas e decidas dos dados.

Para determinar a ordem do polinomial, podemos utilizar o número de flutuações nos dados ou através da quantidade de curvas (vales ou picos) que aparecem na linha de tendência. Por exemplo, em uma linha de tendência polinomial de ordem 2, temos apenas uma curva, já na ordem 3, podemos ter um ou dois picos ou vales. Na figura 2, vemos um exemplo de gráficos de tendência polinomial:

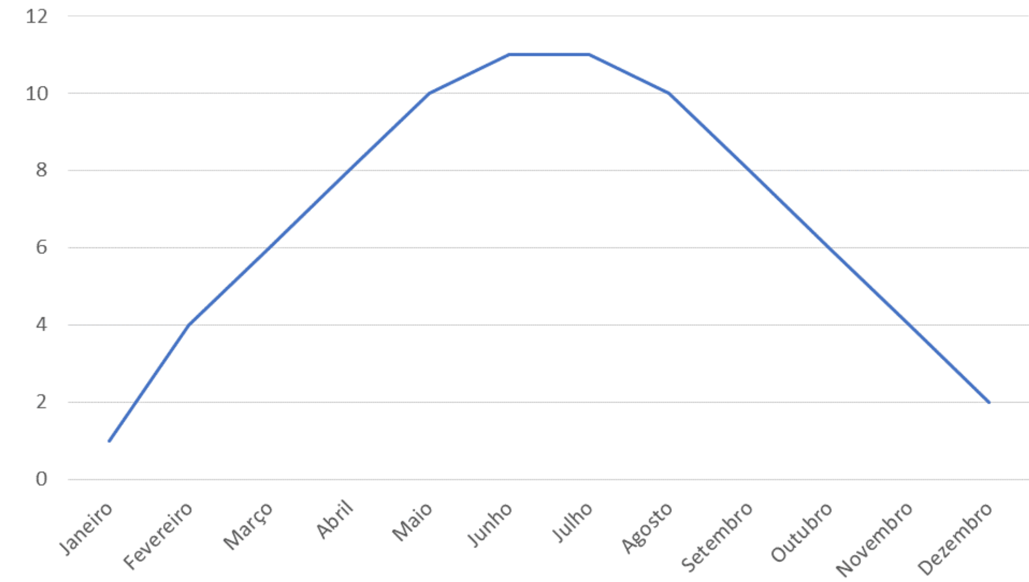


Figura 2 - Linha de tendência polinomial. Fonte: Elaborado pelo autor.

A linha de tendência logarítmica é uma linha curva muito prática quando temos uma série temporal em que a taxa de alteração nos dados aumenta ou diminui rapidamente em um curto espaço de tempo. Neste caso, podemos ter valores negativos e/ou positivos. Na figura 3, temos um exemplo de tendência exponencial:

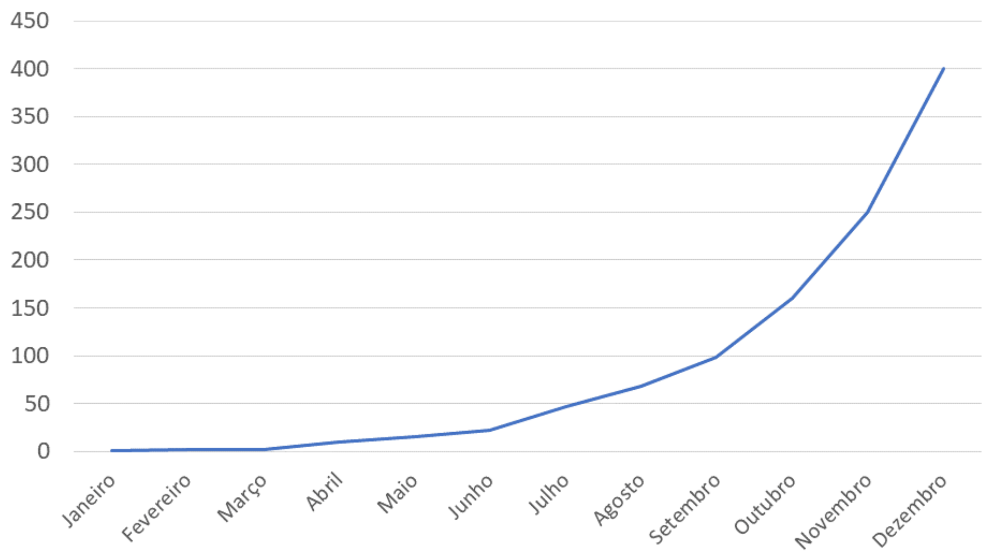


Figura 3 - Linha de tendência logarítmica. Fonte: Elaborado pelo autor.

O método de Regressão Local ou “Lowess” (*locally weighted regression scatter plot smoothing*) é indicado quando temos um ponto de interesse e desejamos estimar uma função na vizinhança deste ponto e, para isto, utilizamos a média ponderada entre as observações (MORETTIN & TOLOI, 2006).

“Lowess” é a realização de uma suavização por meio de sucessivas médias ponderadas em subconjuntos de dados.

Sazonalidade nas séries temporais

!

A sazonalidade é definida quando, em uma série temporal, encontramos oscilações de subida e de descida que ocorrem sempre em um determinado período, como ano, mês ou dia da semana. Como o componente da sazonalidade possui movimentos que ocorrem em intervalos regulares de tempo, torna-se mais fácil de prevê-los.

Temos como exemplos da sazonalidade o padrão encontrado na área de vendas de brinquedos, que é maior em outubro e em dezembro. Também podemos citar as medições das variações de temperatura em um determinado local ao longo dos anos, ou o número de matrículas de alunos nas escolas, que normalmente acontecem no começo do ano.

Morettin e Toloi (2006) reproduzem um trecho extraído de Pierce (1980) que se refere à importância de efetuar o ajuste da sazonalidade em uma série temporal que é muito relevante:

!

“Tem havido no passado um interesse em se ter dados disponíveis sobre fenômenos importantes, sociais e econômicos, para os quais a variação sazonal foi removida. As razões relacionam-se, geralmente, com a ideia de que nossa habilidade em reconhecer, interpretar ou reagir a movimentos importantes não sazonais numa série (tais como pontos de mudança e outros eventos cíclicos, novos padrões emergentes, ocorrências não esperadas para as quais causas possíveis são procuradas) é perturbada pela presença dos movimentos sazonais”.

A sazonalidade pode ser observada normalmente das seguintes formas (MORETTIN & TOLOI, 2006):

- entre observações para meses sucessivos em um determinado ano;
- entre as observações para o mesmo mês em anos sucessivos.

Considerando a equação para o modelo de séries temporais Xt:

$$X_t = T_t + S_t + a_t,\ t = 1,2,\ldots N$$

onde T_t é o componente de tendência, S_t é o componente cíclico ou sazonal e a_t é o componente aleatório ou ruído branco. Devemos obter a série temporal sem o efeito da sazonalidade, para isto, então, obter:

$$X^{SA}_t = X_t - S'_t$$

Quando nos referenciamos à decomposição da série temporal no componente da sazonalidade, devemos considerar qual será o método da decomposição, sendo este Aditivo ou Multiplicativo.

O modelo apresentado até agora para série temporal se refere ao modelo aditivo, onde temos:

$$X_t = T_t + S_t + a_t$$

Já o modelo de decomposição multiplicativo assume que a série temporal é o resultado do produto dos componentes:

$$X_t = T_t \times S_t \times a_t$$

Quando o modelo for Aditivo, inicialmente precisaremos remover o componente da tendência da série temporal, deixando a série somente com a sazonalidade e o ruído branco:

$$X_t - T_t = S_t + a_t$$

Em seguida, para se estimar a sazonalidade, precisamos calcular a diferença entre o valor observado (X_t) e as médias móveis centradas em cada período da série, por exemplo, para cada mês do ano ou para cada dia da semana.

As **médias móveis** podem ser de qualquer ordem, em que a média móvel de ordem “k” é a obtenção da média de “x” valores observados antes do valor atual até “x” valores depois do valor atual. Considerando que o número de períodos (n) é chamado de ordem da série temporal.

Já a **centralização** é a obtenção da média aritmética de dois valores sucessivos obtidos pelas médias móveis. Na figura 4, vemos um exemplo da média móvel não centrada e da média móvel centrada.

Neste exemplo da figura 4, estamos considerando um período de 12 meses, então devemos realizar o cálculo considerando os 6 meses anteriores ao valor desejado e os 6 meses posteriores. Depois, na sequência, devemos ir mantendo este período, porém sempre acrescentando o próximo valor e desprezando o primeiro do grupo anterior, apresentando o resultado no período central, no exemplo, no sexto mês de cada grupo de valores:

Mês/Ano	Número série temporal	Média móvel de 12 meses	
		Não centrada	Centrada
jan/20	2		
fev/20	4		
mar/20	3		
abr/20	5		
mai/20	6		
jun/20	2		
jul/20	7	4,84	4,795
ago/20	7	4,75	
set/20	8		
out/20	3		
nov/20	5		
dez/20	6		
jan/21	1		
fev/21	2		

Figura 4 - Obtenção da média móvel e média móvel centrada. Fonte: Elaborado pelo autor.

No exemplo da figura 4, para se obter a primeira média móvel, calcula-se a média dos valores obtidos entre janeiro de 2020 e dezembro de 2020, inclusive, obtendo-se 4,84. Realiza-se a mesma ação para o próximo ciclo, indo de fevereiro de 2020 até janeiro de 2021, obtendo então 4,75. Com isto, obtemos as médias móveis não centradas. Para obter a média móvel centrada, basta fazer a média aritmética dos dois valores: 4,84 e 4,75. Obtendo, então, 4,795.

A média móvel centrada representa a tendência da série temporal, assim, para se obter a série temporal sem a tendência, basta subtrair a média móvel centrada do valor original da série temporal. Na figura 5, vemos o cálculo da série temporal sem a tendência:

		Média móvel de 12 meses		
Mês/Ano	**Número série temporal**	**Não centrada**	**Centrada**	**Série sem tendência**
jan/20	2			
fev/20	4			
mar/20	3			
abr/20	5			
mai/20	6			
jun/20	2			
jul/20	7	4,84	4,795	2,205
ago/20	7	4,75		
set/20	8			
out/20	3			
nov/20	5			
dez/20	6			
jan/21	1			
fev/21	2			

7 – 4,795 = 2,205

Figura 5 - Obtenção da média móvel e média móvel centrada. Fonte: Elaborado pelo autor.

A sazonalidade é obtida agrupando-se os valores obtidos da série sem tendência, isto para cada período da série temporal a ser analisada. Por exemplo, na figura 5, obtivemos o valor 2.205 para o mês de julho. Para se obter a sazonalidade, devemos calcular a média de todos os valores obtidos sem a tendência para o mês de julho. Assim obteremos a média para cada mês.

Quando estamos considerando o modelo multiplicativo, o cálculo para se obter a sazonalidade é diferente. Também precisaremos remover a tendência da série temporal, porém o cálculo é mais complexo. Teremos:

$$X_t/T_t = (T_t \times S_t \times a_t)/Tt = S_t \times a_t$$

A sazonalidade é obtida pela técnica da razão de média móvel. A execução desta técnica se baseia em calcular a razão entre o valor observado na série temporal e o valor da média móvel.

Na figura 6, vemos a representação do cálculo para este modelo:

Mês/Ano	Número série temporal	Média móvel de 12 meses: Não centrada	Média móvel de 12 meses: Centrada	Série sem tendência
jan/20	2			
fev/20	4			
mar/20	3			
abr/20	5			
mai/20	6			
jun/20	2			
jul/20	$\boxed{7}$	4,84	$\boxed{4,795}$	$\boxed{1,459854}$
ago/20	7	4,75	⋮	
set/20	8	⋮	⋮	
out/20	3	⋮	⋮	
nov/20	5	⋮	⋮	
dez/20	6	⋮		
jan/21	1	⋮		
fev/21	2			

7/ 4,795 = 1,459854

Figura 6 - Obtenção do índice de sazonalidade pelo modelo da multiplicidade. Fonte: Elaborado pelo autor.

Quando o índice da sazonalidade obtido é superior a 1, temos a indicação de uma influência sazonal de crescimento, já quando o valor é menor que 1, temos uma influência sazonal de decrescimento. Este índice também pode ser multiplicado por 100% para facilitar na interpretação dos valores obtidos.

Por fim, para obter o valor do ruído no modelo aditivo, precisamos apenas subtrair o valor de tendência e de sazonalidade do valor observado na série temporal:

$$a_t = X_t - T_t - S_t$$

Para o modelo multiplicativo, o cálculo do ruído é obtido pela razão do valor da série temporal dividido pelos valores dos demais componentes:

$$a_t = X_t / (T_t \times S_t)$$

Síntese

Neste *e-book*, foram apresentados os principais conceitos sobre tendência, indicando os tipos que podem ser: crescente, queda e constante. Também vimos os conceitos e a importância de se analisar o componente da sazonalidade nas séries temporais, apresentando os diferentes modelos e análises, além de um exemplo de como se calcular a tendência e a sazonalidade para uma série temporal utilizando as médias móveis centradas. Por fim, apresentamos como o ruído é obtido utilizando os demais componentes da série temporal: tendência e sazonalidade.

Esta trilha possui uma atividade de aprofundamento que estará disponível para entrega até o final do componente. Continuem consumindo os outros materiais da trilha, como o audioblog e o vídeo explicativo. O conjunto de todos estes materiais possibilitarão uma compreensão melhor do que é a tendência e sazonalidade.

Referências

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

BARROSO, L. C. *et al*. **Cálculo numérico (com aplicações)**. São Paulo: Harbra, v. 384, p. 2a, 1987.

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

CHAFTFIELD, C. ***Time Series Forecasting***. Chapman & Hall/ CRC, 2000.

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

MONTGOMERY, D. C.; JENNINGS, C. L.; KULAHCI, M. ***Introduction to Time Series Analysis and Forecasting***. 2ª ed. Wiley, 2015.

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

PIERCE, D. A. *A survey of recente developments is seasonal adjustment*. **The American Statistician**, 34(3), p. 125-134, 1980.

SNEYERS, R. ***Sobre el análisis estadístico de las series de observaciones***. [S.l.]: Secretariado de la Organización Meteorológica Mundial, 1975.

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

Universidade Presbiteriana
Mackenzie