Introdução à Predição de Séries Temporais

Prof. Jefferson T. Oliva

Aprendizado de Máquina e Reconhecimento de Padrões (AM28CP)
Engenharia de Computação
Departamento Acadêmico de Informática (Dainf)
Universidade Tecnológica Federal do Paraná (UTFPR)
Campus Pato Branco

Sumário

- Séries Temporais
- Comparação de Séries Temporais
- Redução de Dimensionalidade
- Representação da Série Temporal no Domínio de Frequência
- Representação da Série Temporal no Domínio de Tempo-Frequência
- Predição de Séries Temporais
- Avaliação de Modelos de Predição de Séries Temporais

Introdução

- Diversos fenômenos podem ser observados e representados ao longo do tempo, tais como:
  - Sinais biológicos (*e.g.* batimentos cardíacos)
  - Negócios (*e.g.* demanda de produtos)
  - Sensoriamento (*e.g.* monitoramento sísmico)
  - Entre outros
- A análise desses dados tem diversas aplicações, tais como:
  - Compreensão do comportamento temporal
  - Classificação do comportamento temporal
  - Detecção de novidades
  - Predição

Introdução

- Os dados temporais são analisados em diversos campos:
  - Engenharia
  - Governo
  - Mercado financeiro
  - Meteorologia
  - Negócios
  - Saúde
- Esses dados são representados por **séries temporais**

Sumário

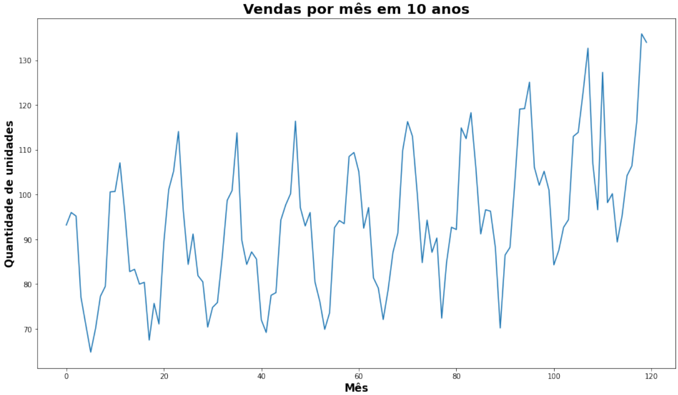
Séries Temporais

- Uma série temporal é um conjunto de observações ordenadas no tempo
  - $S = \{S_0, S_1, S_2, ..., S_n\}$, onde $S_t \in \mathbb{R}$ representa uma observação S_i em um instante de tempo i

- Uma série temporal pode ser:
  - Contínua: os dados são coletados de forma ininterrupta, isto é, continuamente no tempo

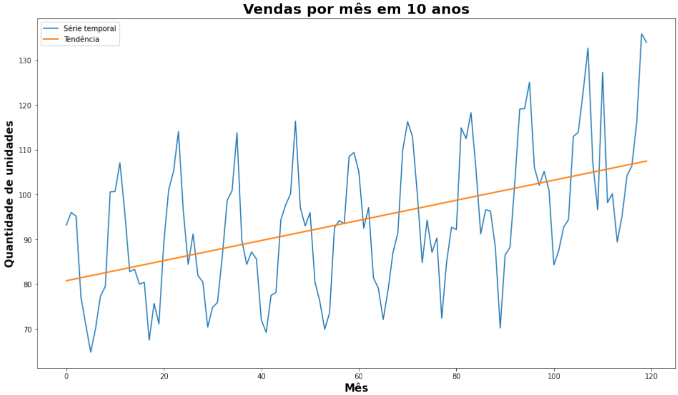
  - Discreta: os dados são obtidos em tempos específicos, ou seja, com interrupções
    - Uma série temporal discreta, geralmente, é um segmento de uma série temporal contínua

- Exemplo de série temporal

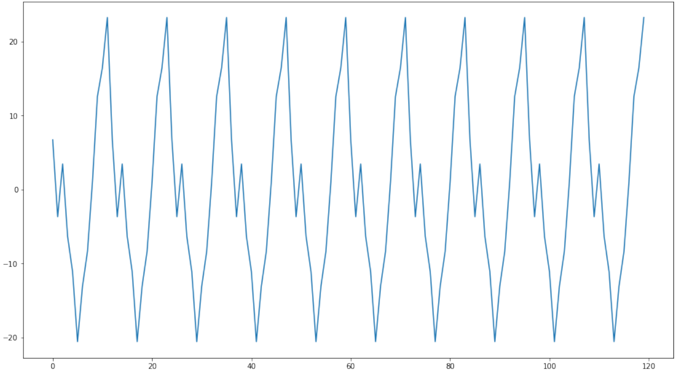


- Uma série temporal também pode ser representada pelos seguintes componentes:
  - Tendência: movimentos regulares desenvolvidos em intervalo de tempo
    - Esse componente possui comportamento unidirecional (crescente ou decrescente)
  - Sazonalidade: possui movimentos periódicos similares e ocorre de forma regular em um período fixo
    - São denominados como padrões
  - Resíduo: movimentos que não pertencentes à sazonalidade e à tendência
    - Comportamento aleatório ou irregular
    - Geralmente são considerados ruídos

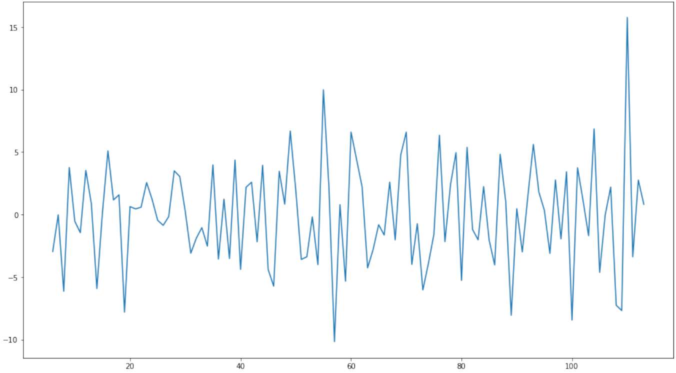
- Tendência


Vendas por mês em 10 anos
Série temporal
Tendência
Quantidade de unidades
Mês
130
120
110
100
90
80
70
0
20
40
60
80
100
120

- Sazonalidade



- Resíduo
  - Exemplo: ruído branco (sinal aleatório), que contém valores aleatórios



- Um componente de série temporal no instante t pode ser representado por meio da seguinte equação:
  - $S_t = X_t + Y_t + Z_t$
  - X: tendência
  - Y: sazonalidade
  - Z: resíduo

- Estacionariedade
  - Uma série temporal é estacionária se os seus dados oscilam em uma média e variância constantes, ou seja, o comportamento da série não é alterado no decorrer do tempo
  - Em algumas aplicações, a propriedade de estacionariedade é indispensável
  - Caso a série temporal não seja estacionária, pode ser aplicada a função denominada primeira diferença para torná-la estacionária:

    $$S'_t = S_t - S_{t-1}$$

  - Alguns testes estatísticos foram propostos para verificar se uma série temporal é estacionária ou não
    - Exemplo: teste de Dickley-Fuller

Séries Temporais

- Séries temporais também podem ser utilizadas em processos relacionados à mineração de dados
  - Agrupamento: definição de grupos de séries temporais de acordo com grau de semelhança
  - Classificação: determinação o grupo em que a série temporal pertence
  - Regressão: relação entre variáveis e predição de valores de séries temporais
  - Regras de associação: descoberta de relações relevantes em bases dados e geração de regras de associação a partir de padrões relevantes de séries temporais

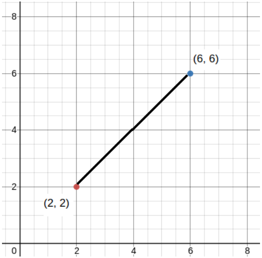
Comparação de Séries Temporais

Comparação de Séries Temporais

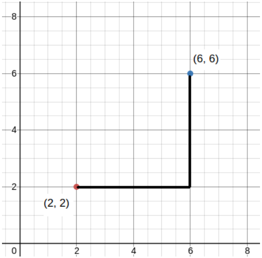
- Uma das tarefas fundamentais para o processamento de séries temporais é a determinação de critérios para determinar quão similar são esses dados

- Diversas medidas de similaridade podem ser aplicadas em séries temporais, tais como:
  - Distância euclidiana
  - Distância de Manhattan
  - Distância de Minkowski (norma L_p)
  - Correlação-cruzada
  - *Dynamic time warping* (DTW)

- Distância euclidiana
  - Uma das técnicas de medida de similaridade mais utilizada
  - Determina o comprimento de dois pontos em linha reta
  - $D_e(S1, S2) = \sum_{t=0}^{n-1} \sqrt{(S1_t - S2_t)^2}$



- Distância de Manhattan
  - Também conhecida como geometria do taxi
  - Versão simplificada da distância euclidiana
  - Comumente utilizada em aplicações de tempo real
  - $D_m(S1, S2) = \sum_{t=0}^{n-1} |S1_t - S2_t|$



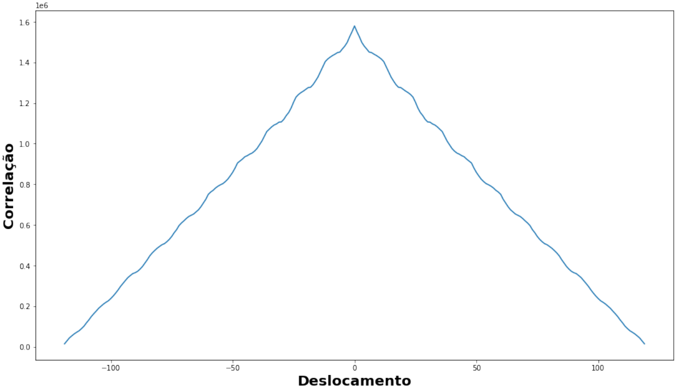
Comparação de Séries Temporais

- Distância de Minkowski (norma L_p)
  - $D_L(S1, S2) = \sum_{t=0}^{n-1} (|S1_t - S2_t|^p)^{\frac{1}{p}}$
  - $p = 1$: distância de Manhattan
  - $p = 2$: distância euclidiana
  - $p = \infty$: distância suprema

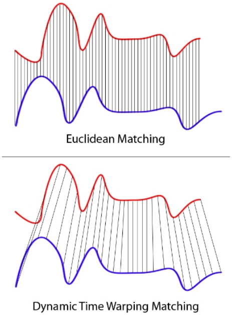
Comparação de Séries Temporais

- Correlação-cruzada
  - Para o cálculo de similaridade, na correlação-cruzada é um parâmetro de deslocamento $\Phi = \{-n, ..., -1, 0, 1, ..., n\}$, onde n é o comprimento de uma das séries temporais
  - $CC(S1, S2, \Phi) = \begin{cases} \sum_{t=0}^{n-\Phi-1} S1_{i+\Phi} * S2_i & \Phi \geq 0 \\ CC(S2, S1, -\Phi) & \Phi < 0 \end{cases}$

- Correlação-cruzada
  - Exemplo de correlograma resultante da comparação entre duas séries temporais de tamanho 120 (correlação-cruzada para todos os parâmetros de deslocamento)



- *Dynamic time warping* (DTW)
  - Utilizado para mensurar a semelhança entre séries temporais independentemente do tamanho e da variação do tempo
  - É realizada uma busca do melhor alinhamento ponto-a-ponto entre séries temporais



Fonte: `https://www.sflscientific.com/data-science-blog/2016/6/3/dynamic-time-warping-time-series-analysis-ii`

- *Dynamic time warping* (DTW)
  - Dada duas séries temporais $S1$ e $S2$ de tamanho $n1$ e $n2$, respectivamente, o DTW é aplicado nas seguintes etapas:
    1. Geração de uma matriz $n1 \times n2$ de distância (*e.g.* euclidiana)
    2. Busca pela rota W que minimize o custo de distância entre duas séries temporais

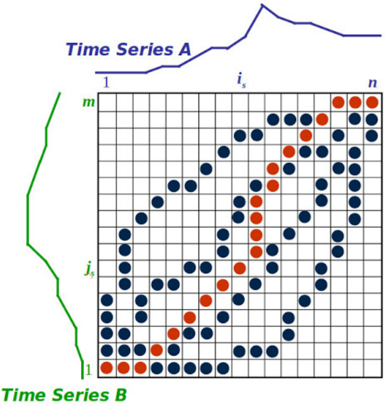
$$W = \{W_1, W_2, ..., W_L\} \tag{1}$$

$$DTW = min_W \left\{ \frac{1}{L} * \sqrt{\sum_{i=1}^{L} W_i} \right\} \tag{2}$$

onde L é o tamanho da rota e $max(n1, n2) \leq L \leq n1 + n2 - 1$

Comparação de Séries Temporais

- *Dynamic time warping* (DTW)
  - Exemplo de matriz de distâncias, onde os pontos de cor vermelha pertencem à solução ótima



Fonte: `http://ros-developer.com/2017/11/17/ros-packages-for-dynamic-time-warping/`

Redução de Dimensionalidade

Redução de Dimensionalidade

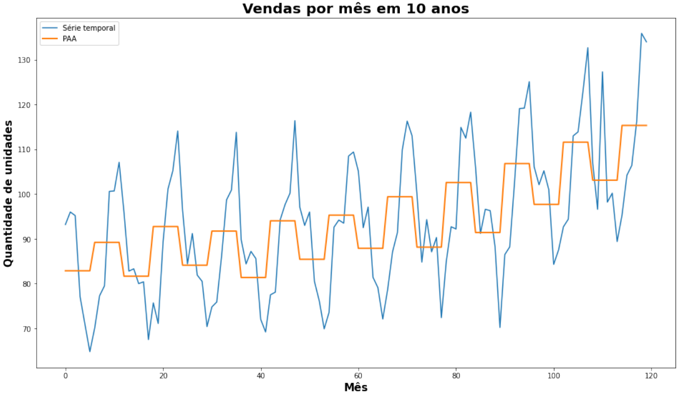
- Séries temporais são comumente caracterizadas pela sua alta dimensionalidade
- Grande quantidade de dados pode acarretar em custo elevado para armazenamento e processamento
- Séries temporais com alta dimensionalidade é um dos grandes obstáculos para a eficiência de vários algoritmos
- A redução da dimensionalidade das séries temporais pode ser desejável em várias aplicações
  - Por mais que os dados temporais sejam simplificados, a complexidade da análise é reduzida e as principais características são mantidas

- As principais vantagens da redução de dimensionalidade de séries temporais são:
  - Redução do tempo para o processamento dos dados
  - Diminuição das dimensões irrelevantes e redundantes
  - Possibilidade de Melhora no desempenho na detecção de padrões
- Diversos métodos de redução de dimensionalidade foram propostos, dos quais, o *piecewise aggregate approximation* (PAA) é um dos mais conhecidos
  - Várias outras técnicas são baseadas em PAA

- *Piecewise aggregate approximation* (PAA)
  - Reduz a dimensionalidade de uma série temporal de tamanho n para uma outra equivalente de tamanho n'
  - O método reduz a dimensionalidade por meio da separação da série temporal em segmentos de mesmo tamanho
  - Em cada segmento é calculada a média aritmética

$$S'_i = \frac{n'}{n} \sum_{j=\frac{n}{n'}*(i-1)+1}^{\frac{n}{n'}*i} S_j \tag{3}$$

- *Piecewise aggregate approximation* (PAA)
  - Exemplo de série temporal e o respectivo resultado da aplicação do PAA



Sumário

Representação da Série Temporal no Domínio de Frequência

Representação da Série Temporal no Domínio de Frequência

- Séries temporais podem conter informações relevantes que não são detectáveis no decorrer do tempo
- Uma alternativa é a representação desses dados temporais no domínio de frequência, no qual é possível analisar a proporção do sinal para cada faixa de frequência
- Para a análise de séries temporais no domínio de frequência, as mesmas deve ser convertidas por meio de métodos específicos
  - A transformada de Fourier é uma das técnicas comumente utilizadas para a conversão de dados no domínio do tempo para o domínio de frequência

- Existem diversos tipos de transformada de Fourier, os quais são utilizados de acordo com as características dos dados

- Para séries temporais discretas, por exemplo, pode ser utilizada a transformada discreta de Fourier (TDF), que pode ser computada por meio da seguinte equação:

$$X_k = \sum_{j=1}^{n} S_j e^{-i2\pi k \frac{j}{n}}$$

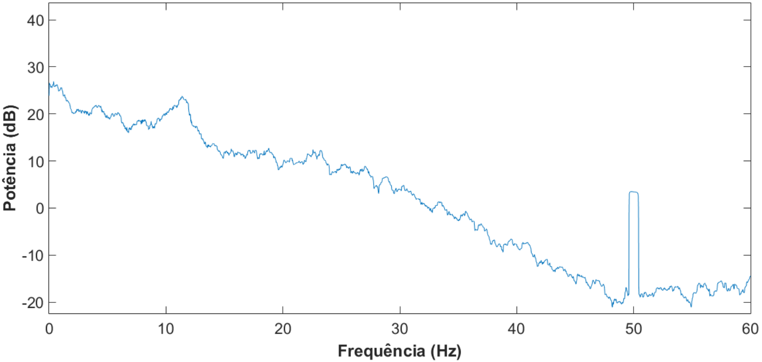
  onde:
  - X_k é o k-ésimo componente de Fourier
  - S é a série temporal
  - $2\pi k$ é a frequência angular
  - i é a unidade imaginária
  - $e^{-i2\pi k \frac{j}{n}} = \cos(-2\pi k \frac{j}{n}) + i \sin(-2\pi k \frac{j}{n})$ é a fórmula de Euller

  - A função acima é equivalente à operação de correlação-cruzada entre uma série temporal e uma senoide de frequência k

- A transformada de Fourier resulta em números complexos denominados como coeficientes de Fourier
  - Para evitar o uso de números complexos em processos computacionais, os componentes de Fourier podem ser utilizados para a geração do espectro de potência
  - O espectro de potência representa a distribuição de energia do sinal em componentes de frequência
  - O *i*-ésimo valor de um espectro de potência pode ser obtido pela seguinte equação:

$$P_i = |X_i|^2$$

- Exemplo de espectro de potência



Representação da Série Temporal no Domínio de Tempo-Frequência

Representação da Série Temporal no Domínio de Tempo-Frequência

- A principal desvantagem do espectro de potência é a ausência de resolução temporal
- Uma solução para problema da resolução temporal pode ser a divisão da série temporal em diversas partes
  - Para cada parte pode ser aplicada a transformada de Fourier
  - Para evitar o problema conhecido como vazamento (*leakage*), que pode ser definido como "manchas" em imagens, pode ser utilizada uma função janela deslizante
- Com a aplicação do procedimento acima, uma série temporal pode ser representada no domínio tempo-frequência, onde é possível verificar os componentes de frequência para cada instante de tempo

Representação da Série Temporal no Domínio de Tempo-Frequência

- Para a representação de dados temporais no domínio de tempo-frequência pode ser aplicada a transformada de Fourier de curto termo
  - Para séries temporais discretas: transformada discreta de Fourier de curto termo (TDFC)

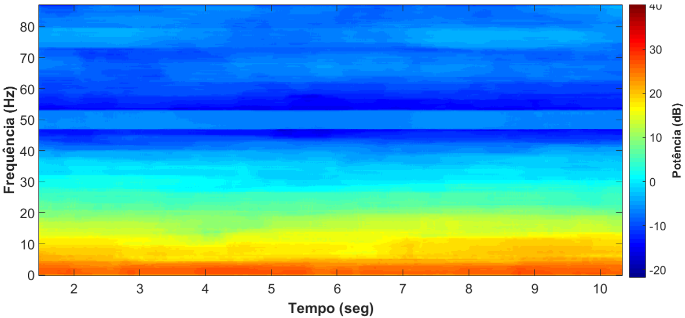
$$X_{t,k} = \sum_{j=t}^{l} S_j W_{(j-t)} e^{-i2\pi k \frac{j}{n}}$$

  onde:
    - W é a função janela deslizante
    - l é o comprimento da janela
    - $X_{t,k}$ é o coeficiente de Fourier para a frequência k e instante de tempo t

Representação da Série Temporal no Domínio de Tempo-Frequência

- A partir dos coeficientes $X_{t,k}$ pode ser gerado um espectrograma:
  - $E_{t,k} = |X_{t,k}|^2$

- Exemplo de espectrograma:



Predição de Séries Temporais

Predição de Séries Temporais

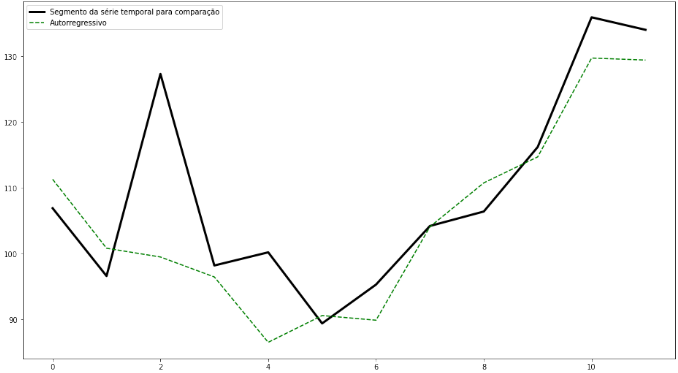
- Métodos para construção de modelos preditivos a partir de séries temporais podem ser divididos em:
    - Paramétricos: esses métodos assumem que uma série temporal pode ser descrita utilizando um conjunto limitado de parâmetros (*e.g.* média, desvio-padrão)
    - Não-paramétricos: os parâmetros são estimados sem a consideração de que a série temporal tem alguma estrutura particular
- Exemplos de métodos paramétricos: modelo autorregressivo, médias móveis, ARMA (*autoregressive-moving-average*), ARIMA (*autoregressive integrated moving average*)
- Exemplos de métodos não-paramétricos: máquina de vetores de suporte, redes neurais artificiais, vizinhos mais próximos

- Modelo autorregressivo (AR)
  - A partir de um instante t, esse modelo especifica que a saída depende linearmente dos valores anteriores
    - Por exemplo, em um modelo autorregressivo de ordem 1 ($p = 1$), o valor de S_t depende de S_{t-1}
  - Um modelo autorregressivo de ordem p, ou $AR(p)$, pode ser definido pela seguinte equação:

$$S_t = c + \epsilon_t + \sum_{i=1}^{p} (S_{t-i} * \varphi_i)$$

onde c é uma constante, ϵ é um número aleatório (representação do ruído branco), e φ_i é o i-ésimo parâmetro

- Modelo autorregressivo
  - Comparação entre um segmento da série temporal e o resultado de predição pelo modelo autorregressivo



- Médias móveis simples (MM)
  - Esse modelo utiliza os q últimos dados da série temporal para a predição do próximo valor
  - O valor de q também define a ordem do modelo, que pode ser definido por $MM(q)$
  - Um modelo de médias móveis simples pode ser definido pela seguinte equação:

$$S_t = \frac{1}{q} \sum_{i=t-q}^{t-1} S_i$$

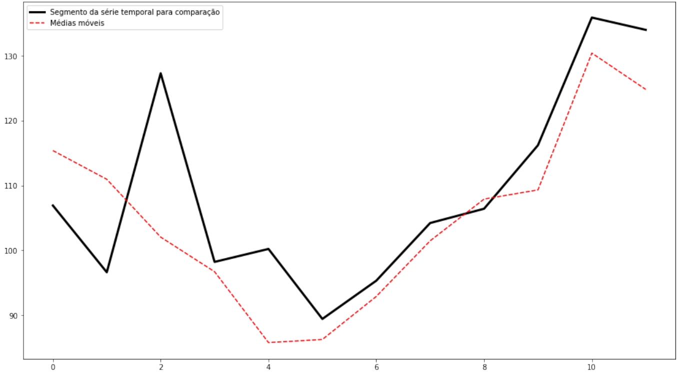
- Médias móveis simples (MA – *moving average*)
  - Outra equação para a construção de modelos por médias móveis simples:

$$S_t = \mu + \epsilon_t + \sum_{i=1}^{q} (\epsilon_{t-i} * \theta_i)$$

onde μ é o valor esperado (geralmente é atribuído um valor igual a zero) de X_t e θ é o i-ésimo parâmetro do modelo

- Essa equação é utilizada na implementação dos dois próximos métodos de predição

- Médias móveis simples (MA – *moving average*)
  - Comparação entre um segmento da série temporal e o resultado de predição pelo modelo médias móveis

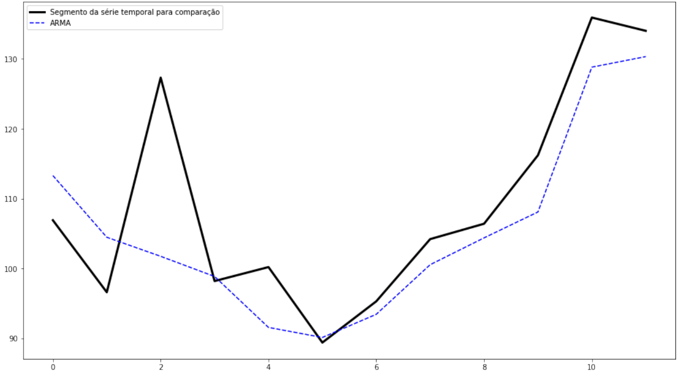


- ARMA (*autoregressive-moving-average*)
  - Modelo auto-regressivo de médias móveis
  - Esse modelo consiste na combinação entre autorregressão e médias móveis
  - Desse modo, ARMA é composto por dois parâmetros: p (ordem da autorregressão) e q (ordem da médias móveis)
  - O modelo ARMA de ordem p e q, ou $ARMA(p, q)$ pode ser definido pela seguinte equação:

$$S_t = c + \epsilon_t + \sum_{i=1}^{p}(S_{t-i} * \varphi_i) + \sum_{i=1}^{q}(\epsilon_{t-i} * \theta_i)$$

- ARMA (*autoregressive-moving-average*)
  - Para $ARMA(p, q)$, temos as seguintes relações:
    - $ARMA(p, 0) = AR(p)$
    - $ARMA(0, q) = MM(p)$
  - No ARMA, os componentes AR e MM se complementam para a geração de um modelo preditivo
  - A principal vantagem do ARMA é a possibilidade de geração de modelos ajustáveis à série temporal utilizando menor quantidade de parâmetros em relação aos métodos AR e MM

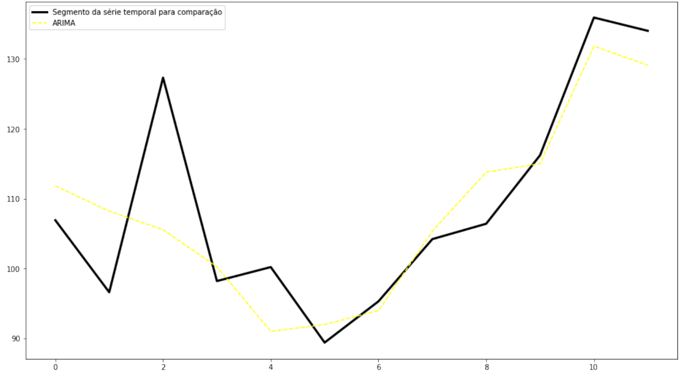
- ARMA (*autoregressive-moving-average*)
  - Comparação entre um segmento da série temporal e o resultado de predição pelo modelo ARIMA



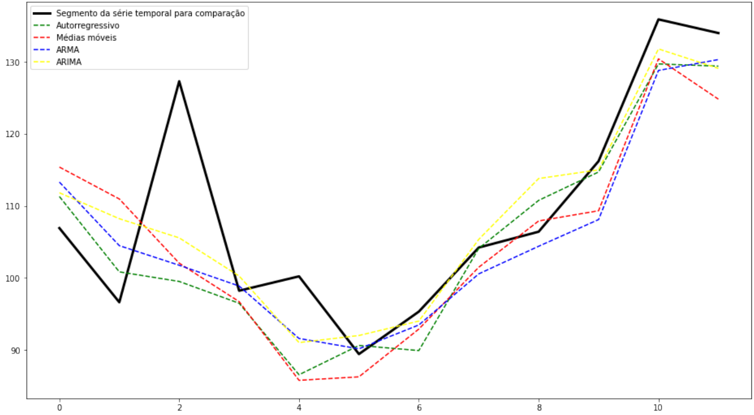
- ARIMA (*autoregressive integrated moving average*)
  - Modelo auto-regressivo integrado de médias móveis
  - Generalização do modelo ARMA
  - Um modelo ARIMA é resultado da combinação entre autorregressão de ordem p, integração (operação de diferenciação sucessiva) de ordem d, médias móveis de ordem q
  - Exemplo operação de integração de ordem d:
    $I_t(d) = (S_t - S_{t-1}) - (S_{t-1} - S_{t-2}) - ... - (S_{t-d+1} - S_{t-d})$
  - Desse modo, um modelo $ARIMA(p, q, d)$ pode ser estimado pela seguinte equação:
    $$S_t = c + \epsilon_t + \sum_{i=1}^{p}(I_{t-i}(d) * \varphi_i) + \sum_{i=1}^{q}(\epsilon_{t-i} * \theta_i)$$

- ARIMA (*autoregressive integrated moving average*)
  - Para $ARIMA(p, d, q)$, temos as seguintes relações:
    - $ARIMA(p, 0, 0) = AR(p)$
    - $ARIMA(0, d, 0) = I(d)$
    - $ARIMA(0, 0, q) = MM(q)$
    - $ARIMA(p, 0, q) = ARMA(p, q)$
  - É importante ressaltar que $AR(p)$, $MM(q)$ e $ARMA(p, q)$ são apropriados para séries temporais estacionárias
  - Para lidar com séries não estacionárias, no ARIMA é utilizado um procedimento de integração para assegurar a propriedade de estacionariedade dos dados temporais

- ARIMA (*autoregressive integrated moving average*)
  - Comparação entre um segmento da série temporal e o resultado de predição pelo modelo ARIMA



- Comparações entre um segmento de série temporal e resultados de modelos preditivos



Avaliação de Modelos de Predição de Séries Temporais

Avaliação de Modelos de Predição de Séries Temporais

- Diversas medidas são utilizadas para a avaliação de modelos de predição de séries temporais
- Exemplos de métricas:
  - Erro quadrático médio (EQM)

    $$EQM = \frac{1}{N}\sum_{i}^{N}(\overline{\mathrm{Y}}_i - Y_i)^2$$

  - Raiz quadrada do EQM (*root-mean-square error* – RMSE)

    $$RMSE = \sqrt{EQM}$$

  - Erro médio absoluto (EMA)

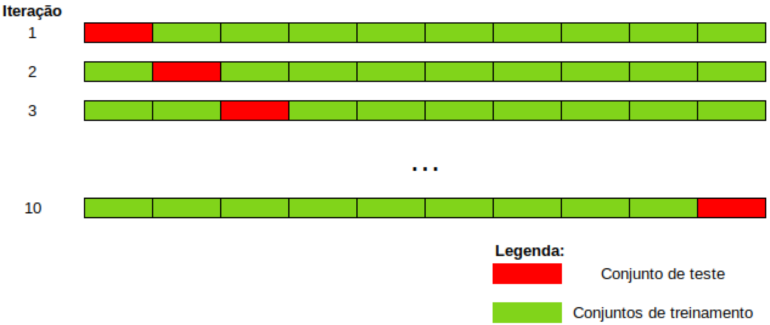
    $$EMA = \frac{1}{N}\sum_{i}^{N}(\overline{\mathrm{Y}}_i - Y_i)$$

  onde Y_i é o i-ésimo dado temporal, $\overline{\mathrm{Y}}_i$ é o i-ésimo resultado da predição e N é a quantidade dados utilizados na predição

- Os exemplos de medidas apresentados no slide anterior também são comumente utilizadas para a avaliação de modelos construídos para problemas de regressão

- Para a avaliação de modelos de predição de séries temporais, também pode ser aplicada a validação cruzada

- Na validação cruzada, os dados são divididos em conjuntos de teste e de treinamento
  - O *k-fold*, por exemplo, divide o conjunto de dados em k conjuntos, sendo um para teste e os $k-1$ restantes para treinamento do modelo em um processo que é repetido k vezes
  - Ilustração de um exemplo para $k = 10$ (*10-fold cross-validation*)



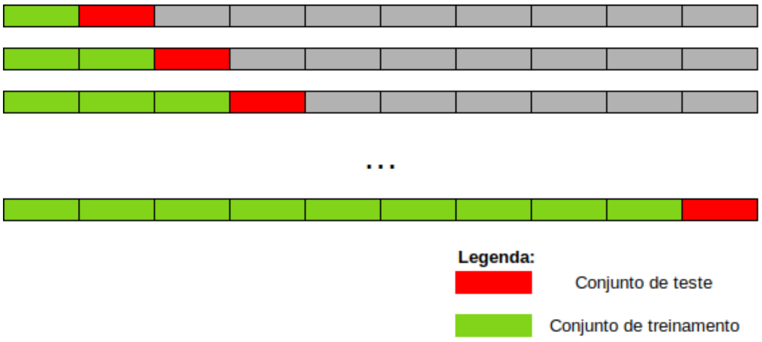
- Entretanto, durante a aplicação da validação cruzada, geralmente não é considerada a ordem dos dados, os quais são amostrados aleatoriamente em conjuntos de teste e de treinamento
  - Essa abordagem não recomendada para algumas aplicações em dados temporais, como a predição de séries temporais

- As séries temporais não devem ter os seus valores amostrados aleatoriamente, pois não faria sentido para a predição de valores futuros, que é dependente de séries históricas (dados sequenciais)

- Em outras palavras, para a predição de valores futuros, há a dependência temporal entre os dados

- A validação cruzada pode ser aplicada em modelos de forma incremental
    - O conjunto de dados é dividido em k partições, mas mantendo a ordem temporal entre as observações
    - Na primeira etapa (iteração), o primeiro conjunto é utilizado para treinamento e o segundo, para teste
    - Na segunda etapa, os dois primeiros conjuntos são usados para treinamento e o terceiro, para teste
    - Na última etapa, os $k - 1$ primeiros conjuntos são utilizados para treinamento e o k-ésimo para teste

- Ilustração de exemplo validação cruzada de 10 *folds* para séries temporais



- Outras abordagens de validação cruzada também podem ser aplicadas em séries temporais
  - Por exemplo: validação cruzada "bloqueada" (*blocked cross-validation*)
    - O conjunto de dados é dividido em k *folds*, onde, para cada *fold*, uma porcentagem dos dados é utilizado para treinamento e o restante, para teste
    - No entanto, essa abordagem pode acarretar na discrepância ("leakage") de resultados

C. Chatfield.
The analysis of time series: an introduction.
Boca Raton: Taylor & Francis, 2003.

D. C. Montgomery; C. L. Jennings; M. Kulahci.
Introduction to time series analysis and forecasting.
Nova Jérsei: John Wiley & Sons, 2015.

E. Galdino.
Análise e previsão de séries temporais.
Acesso em 11 de janeiro de 2021. Disponível em:
https://www.cin.ufpe.br/~psgmn/Series%20Temporais/Aula_01.pdf

G. E. P. Box, G. M. Jenkins.
Time series analysis: forecasting and control.
São Francisco: Holden-Day, 2015.

- J. T. Oliva.
  Geração automática de laudos médicos para o diagnóstico de epilepsia por meio do processamento de eletroencefalogramas utilizando aprendizado de máquina.
  Tese de Doutorado. Universidade de São Paulo.
- P. A. Morettin; C. Toloi.
  Análise de séries temporais: Modelos lineares univariados.
  São Paulo, SP: Blucher, 2018.
- P. J. Brockwell; R. A. Davis.
  Introduction to time series and forecasting.
  Nova Iorque: Springer-Verlag, 2002.
- R. J. Hyndman.
  Cross Validation for Time Series.
  Disponível em:
  `https://robjhyndman.com/hyndsight/tscv/`, 2016.

S. Shrivastava.
Cross Validation in Time Series.
Disponível em: https://medium.com/@soumyachess1496/cross-validation-in-time-series-566ae4981ce4, 2020.

Apêndice

- Exemplo de definição de parâmetros: modelo autorregressivo
  - Os parâmetros podem ser definidos através de diversas abordagens
    - Por exemplo: mínimos quadrados na equação de Yule-Walker

$$\begin{bmatrix} 1 & a_1 & a_2 & a_3 & \dots & a_{n-1} \\ a_1 & 1 & a_1 & a_2 & \dots & a_{n-2} \\ a_2 & a_1 & 1 & a_1 & \dots & a_{n-3} \\ . & . & . & . & & . \\ . & . & . & . & & . \\ . & . & . & . & & . \\ a_{n-1} & a_{n-2} & a_{n-3} & a_{n-4} & \dots & 1 \end{bmatrix} \begin{bmatrix} \varphi_1 \\ \varphi_2 \\ \varphi_3 \\ . \\ . \\ . \\ \varphi_n \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ . \\ . \\ . \\ a_n \end{bmatrix} \quad (4)$$

onde a_d é o coeficiente de autocorrelação com parâmetro de deslocamento d

- Para médias móveis, no ARIMA, os parâmetros também são ajustados por meio da regressão por mínimos quadrados