

ANA RAQUEL

POSTECH

DATA ANALYTICS
ANÁLISE DE SÉRIES TEMPORAIS

AULA 05

SUMÁRIO

O QUE VEM POR AÍ?	3
CONHEÇA SOBRE O ASSUNTO	4
HANDS ON	12
O QUE VOCÊ VIU NESTA AULA?	13
REFERÊNCIAS.....	14
PALAVRAS-CHAVE	15

EMANIP

O QUE VEM POR AÍ?

Nessa aula, você irá aprender sobre modelos estatísticos lineares para séries temporais. Esses modelos estão relacionados com o modelo de regressão linear, mas representam as correlações entre pontos de dados na mesma série temporal, diferente dos métodos tradicionais aplicados a dados transversais, onde cada ponto é independente um do outro na amostra dos dados. Vamos lá?

EMAN

CONHEÇA SOBRE O ASSUNTO

Diferença entre regressão linear e modelos forecasting de séries temporais

Você aprendeu sobre regressão linear nas aulas passadas, porém, consegue imaginar qual é a diferença de um modelo de regressão linear com um modelo estatístico de forecasting para séries temporais?

Uma regressão linear presume que você tenha dados independentes entre si e distribuídos, mas isso não ocorre em séries temporais, onde os pontos (dados) próximos no tempo costumam estar fortemente correlacionados uns com os outros. Nessa aula, vamos conhecer os modelos **autorregressivos (AR)**, modelos de **média móvel (MA)** e modelos **autorregressivos integrados de média móvel (ARIMA)**.

Modelos autorregressivos (AR)

O modelo autorregressivo (AR) toma como base a intuição de que o passado prediz o futuro, onde o valor em um ponto no tempo (t) é uma função dos valores da série em pontos anteriores ao tempo.

É importante ressaltar que esse modelo pode funcionar bem quando temos como pré-requisito uma série temporal estacionária (quando a média e a variância dos dados permanecem constante ao longo do tempo). A estacionariedade não é apenas uma questão gráfica para representar um modelo, mas sim um conceito matemático que pode ser determinado em relação às especificidades de qualquer modelo estatístico.

Para séries não estacionárias, é possível transformá-las em séries estacionárias usando técnicas como diferenciação ou decomposição sazonal (para remover o componente de tendência), antes de aplicar um modelo AR.

Modelos de média móvel (MA)

Um modelo de média móvel (MA) se baseia em um processo em que o valor em cada ponto no tempo é uma função dos termos de “erro” do valor do passado recente, independente dos outros. Um modelo de média móvel pode ser expresso de forma semelhante a um modelo autorregressivo, exceto que os termos incluídos na equação linear se referem a termos de erros presentes e passados, em vez de valores presentes e passados do próprio processo.

Por natureza, os modelos MA têm estacionaridade fraca, sem a necessidade de impor restrições aos seus parâmetros.

Modelos Autorregressivos Integrados de Média Móvel (ARIMA)

Depois de conhecer, de forma separada, os modelos AR e MA, chegou o momento de aprender sobre o modelo ARIMA (modelo autorregressivo integrado de média móvel). Esse modelo combina as técnicas dos modelos AR e MA, porém também leva em conta a diferenciação (uma forma de remover tendências e tornar a série temporal estacionária).



Figura 1 - Modelo ARIMA

Fonte: Elaborado pela autora (2023)

O que é a diferenciação?

É a conversão de uma série temporal em uma série temporal de mudanças nos valores ao longo do tempo (Remover um componente de tendência polinomial consiste em diferenciar a série até que ela se torne estacionária).

E por que falamos tanto em tornar a série estacionária? O modelo ARIMA não funciona bem quando temos uma série não estacionária, por isso, um dos pré-requisitos para a construção do modelo é garantir que a série seja estacionária.

Hiperparâmetros do algoritmo:

- **P:** Número de lags que devem ser incluídos no modelo (Lags são valores auto correlacionados que devem ser levados em consideração, ou seja, os valores históricos que serão utilizados para prever o futuro).

- **D:** Número de vezes que as observações serão diferenciadas (Objetivo de tornar a série estacionária).

- **Q:** Ordem de média móvel, tamanho de uma janela (lag) de média móvel (Para calcular o erro e encontrar a melhor previsão para calcular os resíduos e fazer o forecast).

Passos para aplicar o ARIMA:

1. Visualize os dados da série temporal;
2. Identifique se a data é estacionária;
3. Plote os gráficos de correlação e autocorrelação;
4. Construa o modelo ARIMA com base nos dados.

Como garantir a estacionariedade da série temporal?

Para analisar a estacionariedade da série, você precisa visualizar os dados. Ao plotar a média móvel da sua série, é possível identificar se a série possui um componente de tendência.

```
#Plotando a média móvel
orig = plt.plot(indexedDataset, color='blue', label='Original')
mean = plt.plot(rolmean, color='red', label='Rolling Mean')
std = plt.plot(rolstd, color='black', label='Rolling Std')
plt.legend(loc='best')
plt.title('Rolling Mean & Standard Deviation')
plt.show(block=False)
```

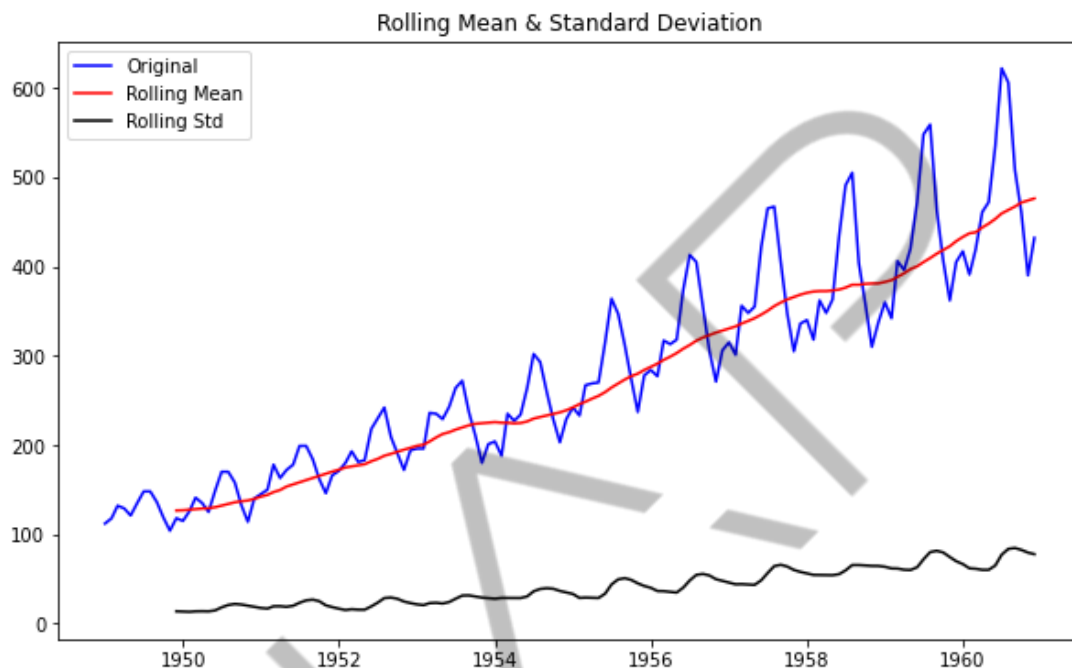


Figura 2 - Média Móvel

Fonte: Elaborado pela autora (2023)

No gráfico da figura 2 – Média Móvel, vemos que a própria média móvel tem um componente de tendência, embora o desvio padrão contínuo seja constante com o tempo. Para que a série temporal seja estacionária, é preciso garantir que ambas as estatísticas tenham uma tendência. Para aumentar ainda a hipótese de que a série temporal não é estacionária, você pode realizar o teste ADCF (Dickey–Fuller test).

O teste de Dickey-Fuller permite saber se há presença significativa de tendência nas séries temporais das variáveis por meio de um teste de hipótese.

```
#Perform Augmented Dickey-Fuller test:
print('Results of Dickey Fuller Test:')
dfctest = adfuller(indexedDataset['#Passengers'], autolag='AIC')

dfoutput = pd.Series(dfctest[0:4], index=['Test Statistic', 'p-value', '#Lags Used', 'Number of Observations Used'])
for key,value in dfctest[4].items():
    dfoutput['Critical Value (%)'%key] = value

print(dfoutput)
```

Results of Dickey Fuller Test:

Test Statistic	0.815369
p-value	0.991880
#Lags Used	13.000000
Number of Observations Used	130.000000
Critical Value (1%)	-3.481682
Critical Value (5%)	-2.884042
Critical Value (10%)	-2.578770

dtype: float64

Figura 3 - ADF Teste

Fonte: Criado pela autora (2023)

Podemos observar, de acordo com o resultado do teste, os seguintes tópicos:

- Valor p alto.
- Os valores críticos em intervalos de confiança de 1%, 5%, 10% devem ser o mais próximo possível das Estatísticas de Teste.

Portanto, podemos dizer com segurança que essa série temporal não é estacionária. Mas o que pode ser realizado para torná-la estacionária?

Você pode, por exemplo, realizar algumas técnicas sobre a série temporal, tal como transformação logarítmica ou decaimento exponencial sobre os dados, e validar novamente com o teste ADF.

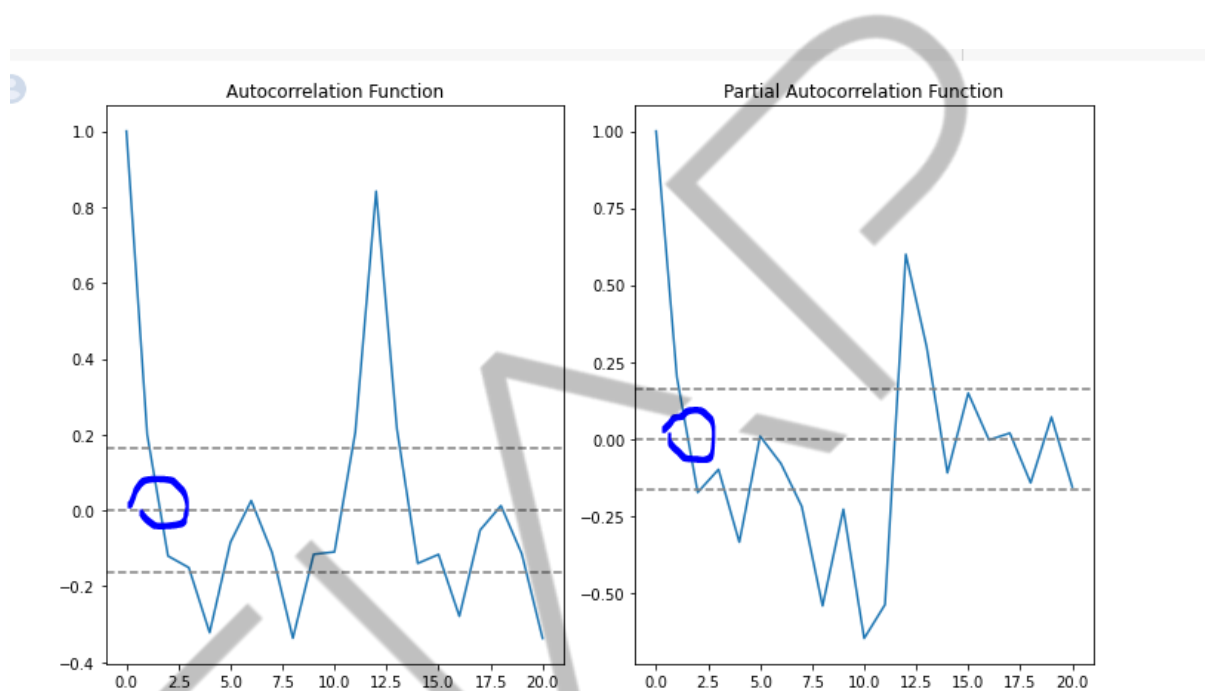
Hiperparâmetros do modelo ARIMA

O modelo ARIMA é especificado em termos de parâmetros (P, D, Q). Você já aprendeu a como identificar uma série temporal estacionária, mas como definir numa primeira tentativa o parâmetro P e o parâmetro Q?

Você pode utilizar os gráficos de ACF (para 'Q') e o gráfico de PACF (para 'P').

Vamos encontrar em qual ponto cada gráfico passa em zero e este ponto será o valor de P e Q inicial (talvez, em alguns casos, pequenas alterações nos parâmetros do ARIMA possam melhorar/piorar os resultados, vale a pena alterar os valores um pouco positivamente e negativamente para observar o desempenho).

Observe, na figura 4 – Encontrando Q e P, o que ocorre quando aplicamos os plots ao nosso dataset de entrada (lags é o número de amostras).



- A partir do gráfico ACF, vemos que a curva toca a linha $y=0,0$ em $x=2$ ($Q = 2$);
- Do gráfico PACF, vemos que a curva toca a linha $y=0,0$ em $x=2$ ($P = 2$).

Figura 4 - Encontrando Q e P

Fonte: Elaborado pela autora (2023)

O parâmetro D representa o número de vezes em que a série temporal precisa ser diferenciada (integrada) para torná-la estacionária, ou seja, para remover tendências e sazonalidades da série.

Após identificar os parâmetros, podemos seguir com a construção do modelo ARIMA, conforme podemos observar na figura 5 – Configurando o modelo ARIMA.

```
# AR+I+MA = ARIMA model
model = ARIMA(indexedDataset_logScale, order=(2,1,2)) #(p,d,q)
results_ARIMA = model.fit(dispatch=-1)
plt.plot(datasetLogDiffShifting)
plt.plot(results_ARIMA.fittedvalues, color='red')
plt.title('RSS: %.4f'%sum((results_ARIMA.fittedvalues - datasetLogDiffShifting['#Passengers'])**2))
print('Plotting ARIMA model')
```

Plotting ARIMA model

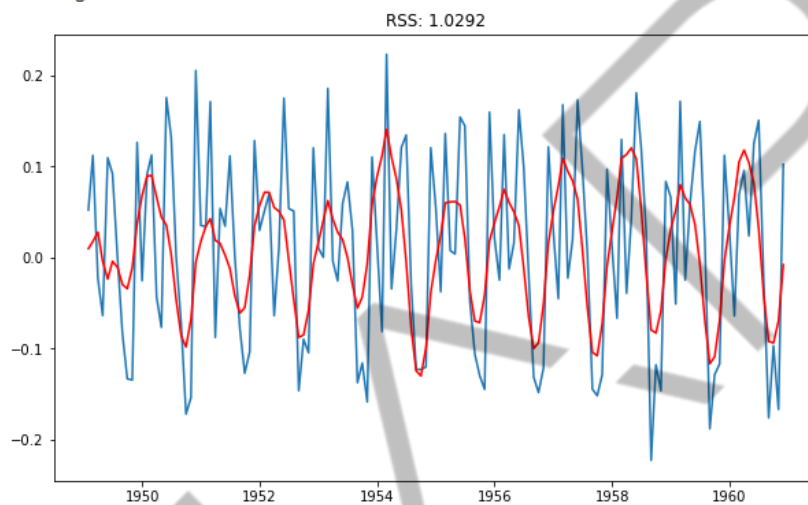


Figura 5 - Configurando o modelo ARIMA
Fonte: Elaborado pela autora (2023)

Modelo AutoARIMA

O modelo AutoARIMA (Autoregressive Integrated Moving Average) é uma extensão do modelo ARIMA, que inclui a seleção automática dos parâmetros P, D e Q através de um algoritmo de busca, com o objetivo de realizar a combinação de parâmetros que minimiza um critério de avaliação estatística.

Métricas de desempenho de modelos de séries temporais

Você aprendeu alguns modelos de forecasting de séries temporais, mas como podemos avaliar a performance desses modelos? Será que apenas gerando o forecasting já é o suficiente para concluir a previsão dos valores?

Bom, vamos te apresentar uma das principais métricas.

A medida de erro, normalmente utilizada para avaliar a qualidade do ajuste de um modelo, é a chamada raiz quadrada do erro médio quadrático (RMSE). Esta é a raiz quadrada do MSE (Erro médio quadrático, que é a média das diferenças ao quadrado entre as previsões do modelo e os valores reais), e é uma medida mais interpretável do erro, pois tem as mesmas unidades que os dados originais. Nesse caso, um modelo bom é aquele que possui o menor erro quadrático médio. RMSE é uma boa medida, porque geralmente ela representa explicitamente o que vários métodos tendem a minimizar.

O Coeficiente de determinação (R^2) mede a proporção da variância dos dados que é explicada pelo modelo.

HANDS ON

Agora, chegou o momento de ver, na prática, como aplicar modelos de forecasting no Python, utilizando os conhecimentos adquiridos na aula de hoje.

EMSE

O QUE VOCÊ VIU NESTA AULA?

Como criar modelos de forecasting com Python.

Daqui em diante, é importante que você replique os conhecimentos adquiridos para fortalecer mais suas bases e conhecimentos.

IMPORTANTE: não esqueça de praticar com o desafio da disciplina, para que assim você possa aprimorar os seus conhecimentos!

Você não está sozinho(a) nesta jornada! Te esperamos no Discord e nas *lives* com os(as) professores(as) especialistas, onde você poderá tirar dúvidas, compartilhar conhecimentos e estabelecer conexões!

REFERÊNCIAS

NIELSEN, Aileen. **Análise Prática de Séries Temporais: Predição com Estatística e Aprendizado de Máquina.** [s.l.]: O'Reilly Media, Inc., 2021.

EMAP

PALAVRAS-CHAVE

Palavras-Chave: Arima, Autoarima, Média Móvel.

EMSE

The background is a dark blue gradient with abstract, wavy lines in shades of teal, yellow, and red. Scattered throughout are small, light blue dots and various geometric shapes, including a circle with the number 7, a circle with an 'x', a circle with a '0', and a hexagon.

POSTECH