

# Profissão: Cientista de Dados



# GLOSSÁRIO



# Regressão II



Dica: para encontrar rapidamente a palavra que procura aperte o comando CTRL+F e digite o termo que deseja achar.

- **Diferencie previsão de explicação**
- **Faça inferência sobre os parâmetros**
- **Diferencie qualidade de complexidade**
- **Selecione modelos**
- **Regularização**



# Diferencie previsão de explicação



# Diferencie previsão de explicação

## ● Explicação

Em estatística, refere-se à identificação e interpretação de padrões de associação entre variáveis.

## ● Previsão

Em aprendizado de máquina, refere-se à descoberta de padrões em dados existentes para prever novas observações.

## ● Intervalo de confiança

Faixa de valores dentro da qual se espera que um parâmetro populacional desconhecido caia, com um certo nível de confiança.

## ● Redução de erro

Foco da previsão, onde o objetivo é minimizar a diferença entre os valores previstos e os valores reais.



# Faça inferência sobre os parâmetros



# Faça inferência sobre os parâmetros

## • Valor p

A probabilidade de obter um resultado pelo menos tão extremo quanto o observado, assumindo que a hipótese nula é verdadeira. Se o valor p for pequeno, rejeitamos a hipótese nula.

## • Variância Homogênea

Uma condição para a distribuição conhecida do beta, onde a variância (a medida da dispersão dos dados) é a mesma para todas as observações.



# Diferencie qualidade de complexidade





# Diferencie qualidade de complexidade

## • **AIC (Critério de Informação de Akaike)**

É uma medida de qualidade de um modelo estatístico. É uma função de perda que busca equilibrar a complexidade do modelo com a sua capacidade de explicar os dados. Quanto menor o AIC, melhor o modelo.

## • **Intercepto**

Em um modelo de regressão, o intercepto é o valor esperado da variável dependente quando todas as variáveis independentes são iguais a zero.

## • **Correlação Espúria**

É uma correlação aparente entre duas variáveis que na verdade não estão relacionadas. Isso pode ocorrer quando uma terceira variável não considerada no modelo está influenciando ambas.

## • **Mínimos Quadrados**

É um método de otimização que busca minimizar a soma dos quadrados das diferenças entre os valores observados e os valores previstos por um modelo.



# Diferencie qualidade de complexidade

## ● Navalha de Occam

Princípio filosófico que sugere que a explicação mais simples (ou seja, aquela que requer o menor número de suposições) é geralmente a melhor.

## ● Overfitting

Acontece quando um modelo é excessivamente complexo e se ajusta muito bem aos dados de treinamento, mas não generaliza bem para novos dados.

## ● R Quadrado

É uma medida de quão bem as previsões de um modelo se ajustam aos dados reais. Varia de 0 a 1, onde 1 indica que o modelo explica toda a variabilidade dos dados.



# Selezione modelos



# Selecione modelos

## • Algoritmo Stepwise

É um método de seleção de variáveis que combina os métodos forward e backward. Ele adiciona e remove variáveis conforme necessário, até que nenhuma alteração seja feita no modelo.

## • Forward Selection

É um método de seleção de variáveis que começa com um modelo vazio e adiciona variáveis da menos relevante para a mais relevante até que nenhuma variável seja estatisticamente significativa.

## • Backward Elimination

É um método de seleção de variáveis que começa com todas as variáveis possíveis e remove as menos relevantes até que todas as variáveis restantes sejam estatisticamente significativas.



# Regularização



# Regularização

## Elastic Net

É uma técnica de regularização que combina as penalidades L1 e L2. É útil quando há várias características correlacionadas.

## Estimador de Mínimos Quadrados

É um método de otimização que busca minimizar a soma dos quadrados das diferenças entre o valor observado e o valor previsto pelo modelo.

## Regularização

É uma classe de algoritmos aplicados à regressão que visam reduzir a complexidade do modelo. O foco é mais na otimização do que na teoria da informação ou na relevância da variável.



# Bons estudos!

