escola britânica de artes criativas & tecnologia

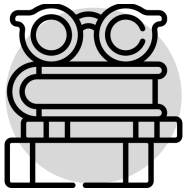
Profissão: Cientista de Dados

escola
britânica de
artes criativas
& tecnologia

GLOSSÁRIO

Regressão I

- **Conheça a equação do modelo de regressão**
- **Utilize statsmodels para Regressão**
- **Estime mínimos quadrados**
- **Avalie a qualidade do modelo**
- **Realize a pós-poda**



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

Conheça a equação do modelo de regressão

Conheça a equação do modelo de regressão

- **Alfa (α)**

  Em um modelo de regressão, alfa é a constante que representa o valor da variável dependente (Y) quando a variável independente (X) é zero.

- **Epson (ε)**

  Em um modelo de regressão, Epson é um elemento que representa a variabilidade dos dados que não pode ser explicada pelo modelo.

- **Beta (β)**

  No contexto de um modelo de regressão, beta é a inclinação da linha de tendência. Representa a mudança na variável dependente (Y) para cada mudança unitária na variável independente (X).

Conheça a equação do modelo de regressão

- **Variável Dependente (Y)**
  Em um modelo de regressão, a variável dependente é a variável que está sendo prevista ou estimada. No exemplo da aula, a variável dependente era a gorjeta.

- **Sigma quadrado**
  É a variância de Y, um parâmetro de dispersão dos pontos em torno do valor esperado, que é dado pela reta de regressão.

- **Variável Independente (X)**
  Em um modelo de regressão, a variável independente é a variável que é usada para prever ou estimar a variável dependente.

Utilize statsmodels para Regressão

Utilize statsmodels para Regressão

- **Erro quadrado médio**
  Medida que quantifica a diferença entre os valores estimados e os valores reais. É um dos atributos disponíveis no objeto 'reg' do statsmodels.

- **Intervalo de confiança**
  Faixa de valores dentro da qual se espera que um parâmetro populacional desconhecido esteja, com um certo nível de confiança.

- **Intercepto**
  Parâmetro estimado em um modelo de regressão que representa o valor esperado da variável dependente quando todas as variáveis independentes são iguais a zero.

- **Método 'predict'**
  Método do objeto 'reg' do statsmodels usado para fazer previsões para novas observações com base no modelo de regressão ajustado.

Utilize statsmodels para Regressão

- **Objeto 'reg'**
  Objeto criado usando a função OLS do statsmodels para definir e rodar um modelo de regressão.

- **Statsmodels**
  Biblioteca Python usada para estimar modelos estatísticos e realizar testes estatísticos.

- **Ordinary Least Squares (OLS)**
  Método usado para estimar os parâmetros desconhecidos em um modelo de regressão linear.

Estime mínimos quadrados

Estime mínimos quadrados

Distribuição dos Estimadores

Refere-se à distribuição probabilística dos estimadores de parâmetros em um modelo estatístico. Saber a distribuição dos estimadores permite fazer inferências sobre eles, como estimar intervalos de confiança e realizar testes de hipóteses.

Método de Mínimos Quadrados

É um método matemático para encontrar os melhores valores de parâmetros que minimizam a soma dos quadrados dos resíduos. É comumente usado em regressão linear para estimar os coeficientes do modelo.

Soma dos Quadrados dos Resíduos

É a soma dos quadrados das diferenças entre os valores observados e os valores previstos por um modelo. O objetivo do método de mínimos quadrados é encontrar os valores dos coeficientes do modelo que minimizam essa soma.

Estime mínimos quadrados

Testes de Hipóteses

São procedimentos estatísticos que permitem tomar decisões sobre a população com base em dados amostrais. Em regressão linear, os testes de hipóteses podem ser usados para testar a significância dos coeficientes do modelo.

Valor Observado

É o valor real de uma variável dependente que é observado nos dados. Em regressão linear, a diferença entre o valor observado e o valor previsto é o resíduo.

Valor Previsto

É o valor de uma variável dependente previsto por um modelo com base em valores específicos das variáveis independentes. Em regressão linear, o valor previsto é uma função linear dos coeficientes do modelo e das variáveis independentes.

Avalie a qualidade do modelo

Avalie a qualidade do modelo

Coeficiente de Correlação

É a raiz quadrada do R quadrado. Mede a força e a direção da relação linear entre duas variáveis.

Coeficiente de Determinação (R quadrado)

É a soma de quadrados do modelo dividida pela soma de quadrados total. É interpretado como a proporção da variância explicada pelo modelo. Quanto maior o R quadrado, melhor o modelo.

Realize a pós-poda

Realize a pós-poda

Alfa

É um parâmetro que é variado de pequeno para grande para ver como a impureza muda em uma árvore de decisão.

Base de Testes

É um conjunto de dados separado usado para avaliar o desempenho de um modelo de aprendizado de máquina.

Realize a pós-poda

Base de Treinamento

É o conjunto de dados usado para treinar um modelo de aprendizado de máquina.

C, C p, ou Parâmetro de Custo de Complexidade

É uma ferramenta usada na pós-poda de árvores de decisão. Um custo alto atribuído à complexidade da árvore resulta em uma árvore mais enxuta, com menos profundidade e quebras.

Realize a pós-poda

Gradient Boosting

É uma técnica de aprendizado de máquina que usa árvores de decisão e é baseada no princípio de melhorar os erros de previsão de um modelo anterior.

Random Forests

É uma técnica de aprendizado de máquina que usa múltiplas árvores de decisão para fazer previsões.

Realize a pós-poda

Ruído

É a variabilidade específica da base de treinamento que não deve ser aprendida por um modelo de aprendizado de máquina.

Variabilidade

É a medida de quanto os dados em um conjunto de dados variam.

escola britânica de artes criativas & tecnologia

Bons estudos!