

Profissão: Cientista de Dados



GLOSSÁRIO



Regressão Logística I



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

- **Conheça a regressão logística**
- **Bivariada**
- **Regressão logística**
- **Use mais de uma variável**
- **Qualidade do modelo**



Conheça a regressão logística



Conheça a regressão logística

Evento

No contexto da regressão logística, um evento é a ocorrência de um resultado específico que a variável resposta está tentando prever.

Linha de regressão

Linha que melhor se ajusta aos dados em um gráfico de dispersão, representando a relação entre as variáveis.

Função de regressão

Uma função matemática que descreve a relação entre a variável dependente e as variáveis independentes em um modelo de regressão.

Regressão logística

Modelo de regressão usado para prever a probabilidade de uma variável dependente categórica, geralmente binária.



Conheça a regressão logística

● Variável binária

Variável que pode assumir apenas dois valores, geralmente 0 e 1.

● Variável independente

Variável que é manipulada ou controlada em um estudo ou experimento para testar seus efeitos na variável dependente.

● Variável dependente

Variável que é o resultado de interesse em um estudo ou experimento.

● Variável resposta

Outro nome para a variável dependente em um modelo de regressão.



Bivariada



Bivariada

● **Análise Bivariada**

É uma análise estatística que envolve a comparação de duas variáveis para determinar a relação entre elas.

● **Tabela Cruzada**

É uma tabela que mostra a frequência com que diferentes combinações de variáveis ocorrem.



Regressão logística



Regressão logística

• Função de ligação

Em estatística, uma função de ligação é uma função que estabelece a relação entre a média de uma variável dependente e os parâmetros de um modelo estatístico. No caso da regressão logística, a função de ligação é a função logística.

• Log da chance

O log da chance de um evento é o log da probabilidade do evento dividida pela probabilidade do evento não ocorrer. É usado na regressão logística para transformar as probabilidades em uma escala logarítmica.

• Função logística

A função logística é uma função que mapeia qualquer valor real para um valor entre 0 e 1. É usada na regressão logística para garantir que as probabilidades previstas estejam sempre entre 0 e 1.

• Método predict

Em ciência de dados, o método predict é usado para obter as previsões de um modelo para um conjunto de dados. No caso da regressão logística, ele retorna as probabilidades previstas para cada observação.



Regressão logística

• Modelo linear generalizado

Um modelo linear generalizado é uma generalização do modelo linear que permite a resposta variar de acordo com uma distribuição de probabilidade diferente da distribuição normal. A regressão logística é um exemplo de um modelo linear generalizado.

• Probabilidade de um evento

A probabilidade de um evento é a chance de o evento ocorrer. Na regressão logística, a probabilidade de um evento é modelada como uma função dos preditores.

• Variável de resposta binária

Uma variável de resposta binária é uma variável que tem apenas dois possíveis resultados, como sobreviver ou não sobreviver a um desastre. A regressão logística é usada quando a variável de resposta é binária.



Use mais de uma variável



Use mais de uma variável

● Parâmetros do modelo

Os coeficientes que o modelo de regressão logística aprende durante o treinamento. Eles indicam a relação entre as variáveis independentes (por exemplo, sexo, classe de viagem) e a variável dependente (sobrevivência).

● Razão de chances

Uma medida estatística que compara a probabilidade de um evento ocorrer em um grupo em relação à probabilidade de ocorrer em outro grupo.



Qualidade do modelo



Qualidade do modelo

● Calibragem

Refere-se à precisão das estimativas de probabilidade do modelo em relação à realidade. Um modelo bem calibrado terá suas previsões correspondendo à frequência observada dos eventos.

● Discriminação

Capacidade do modelo de distinguir entre eventos e não eventos. Um modelo que discrimina bem será capaz de diferenciar corretamente entre diferentes categorias ou classes.

● Curva ROC (Receiver Operating Characteristic)

Ferramenta gráfica para avaliar a capacidade de discriminação de um modelo. Traça a taxa de verdadeiros positivos contra a taxa de falsos positivos para diferentes pontos de corte. A área sob a curva ROC é um indicador da qualidade do modelo.

● KS (Kolmogorov-Smirnov)

Estatística usada para comparar duas distribuições de probabilidade. Mede a maior distância entre as curvas de distribuição acumulada de dois grupos, com uma maior distância indicando uma melhor discriminação.



Bons estudos!

