**Regressão Linear**

Regressão linear é uma técnica estatística usada para modelar a relação entre uma variável dependente (também chamada de variável resposta) e uma ou mais variáveis independentes (também chamadas de variáveis explicativas ou preditoras). É frequentemente usada em análises preditivas e modelagem de dados para prever valores futuros da variável dependente com base nas informações conhecidas das variáveis independentes.

Para entender como funciona a regressão linear, primeiro precisamos entender os conceitos básicos:

* Variável dependente: é a variável que estamos tentando prever. É representada por y na equação da regressão.
* Variáveis independentes: são as variáveis que usamos para prever a variável dependente. São representadas por x na equação da regressão.
* Coeficiente de regressão: é uma medida que indica a mudança na variável dependente (y) para uma unidade de mudança na variável independente (x). É representado por beta (β) na equação da regressão.
* Intercepto: é o valor de y quando x é igual a zero. É representado por alpha (α) na equação da regressão.

A equação da regressão linear é dada por:

y = α + βx + ε

Onde ε é o erro aleatório, que é a diferença entre o valor real de y e o valor previsto de y. O objetivo da regressão linear é minimizar o erro aleatório e encontrar os valores de α e β que fornecem a melhor previsão de y com base em x.

Existem dois tipos principais de regressão linear:

* Regressão linear simples: envolve uma variável independente e uma variável dependente. A equação da regressão linear simples é dada por:

y = α + βx + ε

* Regressão linear múltipla: envolve duas ou mais variáveis independentes e uma variável dependente. A equação da regressão linear múltipla é dada por:

y = α + β1x1 + β2x2 + ... + βnxn + ε

Onde x1, x2, ..., xn são as variáveis independentes e β1, β2, ..., βn são os coeficientes de regressão para cada variável independente.

Existem várias métricas que podem ser usadas para avaliar a precisão da regressão linear, como o coeficiente de determinação (R²), que mede a proporção da variação na variável dependente que é explicada pelas variáveis independentes, e o erro padrão residual, que mede a dispersão dos valores residuais em relação aos valores ajustados.

A regressão linear é uma ferramenta poderosa para prever valores futuros com base em dados históricos. No entanto, é importante ter cuidado ao interpretar os resultados da regressão linear e lembrar que a correlação não implica causalidade. É sempre importante considerar outras variáveis e fatores que possam influenciar a variável dependente além das variáveis independentes incluídas no modelo de regressão linear.

Neste exemplo, estamos usando uma biblioteca **numpy** para criar dois arrays: **x** e **y**. O array **x** é uma matriz com uma única coluna, que representa nossos dados de entrada. O array **y** é um array unidimensional que representa nossos dados de saída.

Em seguida, importamos a classe **Linear Regression** da biblioteca **sklearn.linear\_model**. Esta classe nos permite criar um modelo de regressão linear e treiná-lo com nossos dados de entrada e saída.

Depois, criamos uma instância do modelo **Linear Regression**e seguido o método **fit** passando os dados **x** e **y**. Isso treina o modelo com nossos dados de entrada e saída.

Finalmente, usamos o método **predict** para prever o valor de saída **y** para um novo valor de entrada **x\_new**. Em seguida, imprimimos o resultado da previsão. Neste exemplo, a saída será um array com um único valor: **array ([5.2])**, que é a previsão para **x\_new = 6**.

Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente