**Adjusted R²**

**Descrição**

O R² é uma métrica usada para avaliar a qualidade de um modelo de regressão. Ele mede a proporção da variação na variável dependente que pode ser explicada pela variável independente. No entanto, o R² não leva em consideração o número de variáveis independentes no modelo, o que pode levar a uma interpretação incorreta da qualidade do modelo.

O Adjusted R², por outro lado, leva em consideração o número de variáveis independentes no modelo e é uma medida mais precisa da qualidade do modelo. Ele penaliza a inclusão de variáveis independentes desnecessárias e aumenta quando variáveis independentes relevantes são adicionadas ao modelo.

**Fórmula**

A fórmula para o Adjusted R² é:

Onde:

* R² é o R² padrão do modelo.
* N é o número de observações no modelo.
* K é o número de variáveis independentes no modelo.

**Interpretando o Adjusted R²**

O Adjusted R² varia de 0 a 1 e quanto mais próximo de 1, melhor é o ajuste do modelo. No entanto, é importante lembrar que o Adjusted R² não é uma medida absoluta da qualidade do modelo e deve ser usado em conjunto com outras métricas para avaliar o modelo de regressão.

**Exemplo**

Suponha que estamos construindo um modelo de regressão linear para prever o preço de casas com base em variáveis como tamanho, número de quartos, localização, etc.

Podemos calcular o R² do modelo e obter um valor de 0.75, o que indica que 75% da variação nos preços das casas é explicada pelas variáveis independentes no modelo.

No entanto, também sabemos que existem 4 variáveis independentes no modelo. Para calcular o Adjusted R² e assumindo que temos 100 observações, podemos usar a fórmula:

Ou seja, o Adjusted R² é de 0.735, o que indica que o modelo explica 73.5% da variação nos preços das casas, levando em consideração o número de variáveis independentes no modelo.

Com isso, podemos concluir que embora o R² de 0.75 seja bom, o Adjusted R² de 0.735 fornece uma medida mais precisa da qualidade do modelo, levando em conta o número de variáveis independentes no modelo.

**Aplicações**

O Adjusted R² é particularmente útil quando se compara modelos de regressão que têm um número diferente de variáveis independentes.

Aqui estão algumas aplicações do R² ajustado:

1. Seleção de variáveis: O R² ajustado pode ser usado para selecionar as variáveis mais relevantes em um modelo de regressão. Comparando o R² ajustado de diferentes modelos com diferentes conjuntos de variáveis independentes, é possível identificar o modelo que melhor se ajusta aos dados, levando em consideração o número de variáveis utilizadas. Um modelo com um R² ajustado mais alto indica que uma porcentagem maior da variação na variável dependente é explicada pelas variáveis independentes incluídas no modelo.
2. Comparação de modelos: O R² ajustado também é útil para comparar diferentes modelos de regressão que têm um número diferente de variáveis independentes. O R² tradicional pode aumentar à medida que mais variáveis são adicionadas ao modelo, mesmo que essas variáveis não sejam realmente relevantes. O R² ajustado, por outro lado, penaliza a inclusão de variáveis não sejam realmente relevantes. O R² ajustado, por outro lado, penaliza a inclusão de variáveis irrelevantes, proporcionando uma medida mais justa e comparável da qualidade do modelo.
3. Avaliação da qualidade do modelo: O R² ajustado pode ser usado como uma métrica para avaliar a qualidade global do modelo. Um valor mais alto de R² ajustado indica que o modelo é capaz de explicar uma maior porcentagem da variação na variável dependente, levando em consideração o número de variáveis independentes. No entanto, é importante lembrar que o R² ajustado deve ser interpretado em conjunto com outras métricas e considerações, pois o ajuste do modelo não é o único critério para avaliar sua qualidade.
4. Validação cruzada: O R² ajustado pode ser usado em conjunção com técnicas de validação cruzada para avaliar a capacidade de generalização de um modelo de regressão. A validação cruzada envolve dividir os dados em conjuntos de treinamento e teste e ajustar o modelo aos dados de treinamento. Em seguida, o modelo é avaliado usando os dados de teste. O R² ajustado pode ser calculado no s dados de teste para avaliar a capacidade do modelo de se ajustar a novos dados não vistos anteriormente.

Em resumo, o R² ajustado é uma métrica útil para selecionar variáveis relevantes, comparar modelos de regressão com diferentes números de variáveis independentes, avaliar a qualidade global do modelo e avaliar sua capacidade de generalização. No entanto, é importante considerar o R² ajustado em conjunto com outras métricas e considerações relevantes para uma análise completa e adequada do modelo.

**Limitações**

Embora o Adjusted R² seja uma métrica útil para avaliar a qualidade do modelo regressão, ele também tem algumas limitações, incluindo:

1. Interpretação inadequada: O Adjusted R² pode ser interpretado de maneira inadequada se for usado como a única métrica para avaliar a qualidade do modelo. Ele não considera a qualidade dos dados ou a precisão das previsões do modelo. Portanto, é importante usar outras métricas de avaliação do modelo, como o erro quadrático médio (MSE) ou o erro absoluto médio (MAE).
2. Suposições de linearidade: O Adjusted R² assume que o relacionamento entre as variáveis independentes e dependentes é linear. Se o relacionamento for não linear, o Adjusted R² pode subestimar ou superestimar a qualidade do modelo.
3. Sensibilidade ao tamanho da amostra: O Adjusted R² é sensível ao tamanho da amostra e pode ser afetado por amostras grandes ou pequenas. Em geral, quanto maior o tamanho da amostra, maior será o valor do Adjusted R², o que pode levar a uma interpretação excessivamente positiva da qualidade do modelo.
4. Sensibilidade ao número de variáveis independentes: O Adjusted R² pode ser afetado pelo número de variáveis independentes incluídas no modelo. Quanto mais variáveis independentes forem incluídas, menor será o valor do Adjusted R², o que pode levar a uma interpretação excessivamente negativa da qualidade do modelo.

Em resumo, o Adjusted R² é uma métrica útil para avaliar a qualidade do modelo de regressão, mas deve ser usado em conjunto com outras métricas de avaliação do modelo. Ele também tem algumas limitações, incluindo suposições de linearidade, sensibilidade ao tamanho da amostra e número de variáveis independentes.

**Conclusão**

Em conclusão, o Adjusted R² é uma métrica valiosa para avaliar a qualidade e um modelo de regressão, considerando o número de variáveis independentes. Ele oferece uma medida mais precisa e justa do ajuste do modelo, penalizando a inclusão de variáveis irrelevantes. O Adjusted R² é aplicado na seleção de variáveis, comparação de modelos, avaliação da qualidade global do modelo e validação cruzada. No entanto, é importante interpretar o Adjusted R² em conjunto com outras métricas e considerar suas limitações, como a interpretação inadequada quando usada isoladamente, suposições de linearidade, sensibilidade ao tamanho da amostra e número de variáveis independentes. Ao utilizar o Adjusted R² e outras métricas apropriadas, os pesquisadores e analistas podem obter uma visão abrangente da qualidade e capacidade de generalização do modelo de regressão.