



A Decision Tree Regressor - Teoria

☰ Ciclo	Ciclo 06: Algoritmos baseado em árvores
# Aula	46
🕒 Created	@March 10, 2023 4:55 PM
☑ Done	☑
☑ Ready	☑

Objetivo da Aula:

- ☐ A Decision Tree Regression
- ☐ A impureza de uma árvore
- ☐ Resumo
- ☐ Próxima aula

Conteúdo:

▼ 1. A Decision Tree Regression

O algoritmo Decision Tree Regressor é utilizado para problemas de regressão. O processo de treinamento é semelhante ao de problemas de classificação, porém a escolha do atributo na separação é feita de forma a minimizar o erro entre os pontos da folha e o seu valor médio.

O algoritmo Decision Tree Regressor possui seis passos para o seu treinamento, que consistem em escolher um atributo do conjunto de dados e, para cada possível valor do atributo selecionado, encontrar o valor da impureza da

separação, repetindo os passos para todas as combinações de atributo e valores a fim de encontrar a combinação atributo-valor que retorne o menor valor da função custo da Decision Tree. Uma vez definido o par atributo-valor, faz-se a separação do conjunto de dados em dois nós filhos. Repete-se o processo até que os valores dos parâmetros sejam atendidos.

▼ 2. A impureza de uma árvore

O algoritmo Decision Tree também pode ser utilizado para problemas de regressão. Nesse caso, a impureza dos nós é medida utilizando o erro quadrático médio (MSE) ou o erro absoluto médio (MAE) ao invés dos critérios de impureza utilizados em problemas de classificação.

▼ 2.1 O critério MSE

A fórmula do erro quadrático médio (MSE) para o cálculo da impureza de uma folha da Decision Tree é:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y}_i)^2$$

onde:

- n é o número total de exemplos
- y é o valor real do i -ésimo exemplo
- \bar{y} é o valor previsto para o i -ésimo exemplo

▼ 2.2 O critério MAE

A fórmula do erro absoluto médio (MAE) para o cálculo da impureza de uma folha da Decision Tree é:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

onde:

- n é o número total de exemplos
- y é o valor real do i -ésimo exemplo
- \hat{y} é o valor previsto para o i -ésimo exemplo

▼ 2.3 O critério de Friedman MSE

A fórmula do erro quadrático médio de Friedman (Friedman MSE) para o cálculo da impureza de uma folha da Decision Tree é:

$$Friedman\ MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \frac{p}{n} \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

onde:

- n é o número total de exemplos
- y é o valor real do i -ésimo exemplo
- y_hat é o valor previsto para o i -ésimo exemplo
- p é o número total de variáveis preditoras
- y_bar é a média dos valores reais de y

▼ 3. Os 6 passos do treinamento da Decision Tree Regressor

No algoritmo Decision Tree para problemas de regressão, a impureza dos nós é medida utilizando o erro quadrático médio (MSE). A redução da impureza acontece da seguinte forma:

1. Primeiro, a árvore avalia qual é o melhor atributo para fazer a divisão dos dados em subconjuntos. Essa escolha é feita com base em qual atributo resultaria na maior redução do erro de predição.
2. Após encontrar o melhor atributo, a árvore faz a divisão dos dados em subconjuntos com base nos valores desse atributo.
3. Para cada subconjunto, é calculado o erro quadrático médio (MSE) entre os valores reais e os valores preditos.
4. Em seguida, é calculada a impureza desse nó utilizando a média desses erros quadráticos médios.
5. A árvore continua fazendo a divisão dos subconjuntos recursivamente até que os subconjuntos resultantes sejam puros ou até que um critério de parada seja atingido.

6. O processo de divisão continua até que não seja mais possível reduzir a impureza ou até que um critério de parada seja atingido, como por exemplo, limite máximo de profundidade da árvore ou número mínimo de amostras em um nó.

▼ 4. Resumo

1. O processo de treinamento da Decision Tree Regressor se resume a subdividir os dados originais em nós filhos, a fim de minimizar a impureza das folhas.
2. A impureza de uma Decision Tree Regressor pode ser medida utilizando os critérios MSE, Friedman MSE e MAE.
3. A Decision Tree possui 6 passos para o seu treinamento.

▼ 5. Próxima aula

A Decision Tree Regressor - Prática