

12

의사결정나무 개념

[의사결정나무 개념]

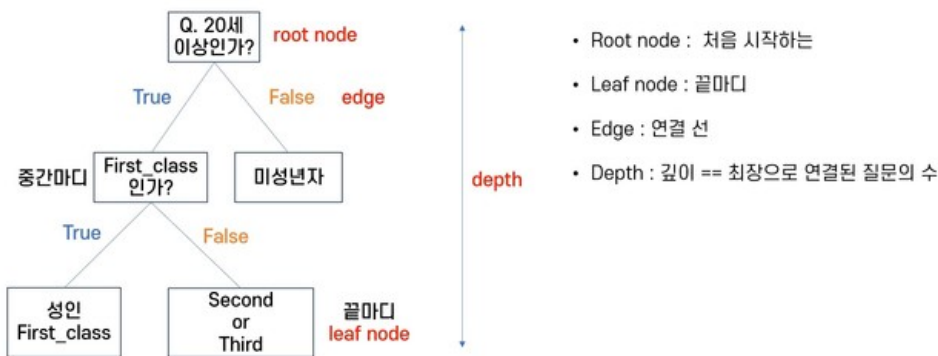
대부분의 앙상블 모델들이 의사결정나무를 기반으로

이루어져 있습니다.

개념과 사용하는 용어들에 익숙해지시기 바랍니다.

00:00

의사결정나무 개념



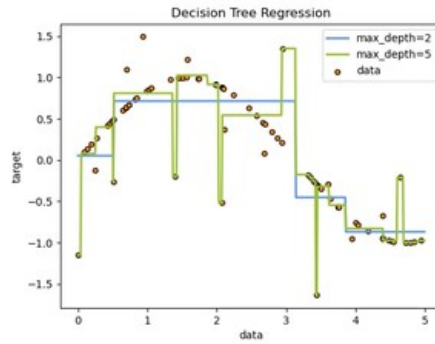
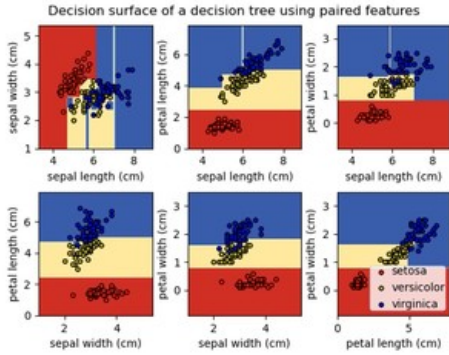
[의사결정나무 회귀와 분류 설명]

의사결정 나무는 회귀와 분류분석 모두 사용할 수 있습니다.

- 매개변수를 조절하지 않으면, train 데이터의 score가 1이 나온다는 점이 특이한 점입니다. (과적합)
- 오버피팅(과적합)의 문제를 가지치기로 해결해 주어야 합니다.

02:31

의사결정나무 개념



[의사결정나무 장단점]

모든 모델에서 장단점을 알고 계셔야 합니다.

의사결정나무 개념

• 장점

- 1) 데이터 스케일에 영향을 받지 않는다.
- 2) 특히, 데이터 내에 이진 특성과 연속적 특성이 혼합되어 있어도 무관함
- 3) 결정트리 모델이 어떻게 훈련되었는지 경로로 해석이 가능하다. 즉, 설명력이 좋다.

• 단점

- 1) 가지치기를 하더라도 과대적합이 되는 경향이 있다.
- 2) 모델의 일반화 성능이 좋지 않다.

• 매개변수 튜닝

- 1) max_depth
트리의 최대 깊이
- 2) min_samples_split
내부 노드를 분할하는 데 필요한 최소 샘플 수
- 3) min_samples_leaf
리프노드에 있어야 하는 최소 샘플 수