HW #6

**의사결정트리**

각 데이터의 속성을 예측 가능한 규칙들의 조합으로 나타내어 **분류**를 수행할 수 있도록 하는 지도학습 모델. 적절한 **분리기준**과 **정지규칙**을 지정하여 의사결정나무 모델을 생성한다.

분류와 회귀가 모두 가능하며, 분류의 경우 끝마디에서 새로운 데이터를 분류, 회귀의 경우 끝마디의 종속변수 평균을 예측값으로 반환.

결정트리 모델은 데이터에 대한 어떤 가정도 하지 않기 때문에, 비파라미터 모델이다.

**가지치기 =** 의사결정트리에서의 과적합 방지 과정

**CART 알고리즘**

각 노드에서 **비용함수**를 최소화하는 특성과 특성의 임계값을 결정함

지니 불순도와 엔트로피 방식에는 큰 차이가 없으나, 지니 불순도보다 훈련이 빠른 엔트로피가 기본값으로 사용됨.

결정트리 단점

훈련셋 회전 민감도 – 항상 축에 수직인 분할을 사용하기 때문

높은 분산 – 훈련 데이터의 작은 변화에도 매우 민감함