

6장 결정 트리 답지_김찬호

6.1 max_depth는 모델의 무엇을 설정하는 매개변수 인가.

↳ 결정 트리의 최대 깊이

6.2 한 노드의 질문의 대답이 세가지 이상일 경우 즉, 자식 노드가 세개 이상일 경우 사용해야 하는 알고리즘의 예시.

↳ ID3

6.3 한 샘플이 특정 클래스에 속할 확률은 그 샘플이 분류된 리프 노드에 있는 클래스의 훈련 샘플의 비율과 같다. (T)

6.4 CART 알고리즘은 자식 노드의 개수가 2개 이상인 경우 사용한다. (F)

-> 3개 이상인 경우에는 사용하지 못한다.

6.5 훈련 세트가 클수록 훈련 속도가 느려지는데 이때 어떤 매개 변수를 어떻게 지정해야 훈련 세트가 작을 경우의 훈련 속도를 높일 수 있는지.

↳ presort = True

6.6 지니 불순도와 엔트로피가 각각 0일 때 어떻게 다른지 서술하시오.

↳ 지니 불순도는 한 노드의 모든 샘플이 같은 클래스에 속해 있을 때 0이 되며, 엔트로피는 어떤 세트가 한 클래스의 샘플만 담고 있을 때 0이 된다.

6.7 결정 트리의 형태를 제한하는 매개변수가 있는데 이때 min_으로 시작하는 매개변수와 max_로 시작하는 매개변수의 차이점을 서술하시오.

↳ min_으로 시작하는 매개변수는 증가시킬수록 모델에 규제가 커져 과대적합이 방지된다. Max_로 시작하는 매개변수는 감소시킬수록 모델에 규제가 커져 과대적합이 방지된다.

6.8 샘플의 타깃값을 예측할 때 어떤 값이 예측값이 되는가.

↳ 샘플에 해당하는 리프 노드에 있는 모든 훈련 샘플의 평균 타깃값

6.9 결정 트리는 훈련 세트의 회전에 민감하다는 특징을 가지고 있는데, 이때 훈련 세트의 적절한 회전 방향을 결정하는 기법은?

↳ PCA 기법