

preview2. DA팀 박보영

결정트리

규칙노드에서 분할하여 리프노드로.

깊이가 깊어질수록 결정트리의 예측 성능이 저하될 가능성 높음.

결정노드는 정보 균일도가 높은 데이터 세트를 먼저 선택할 수 있도록 규칙을 만듦(서브 데이터 세트로 쪼개어 내려감)

정보이득=1-엔트로피 / 지니계수 = 불평등계수

경계시각화를 통해서 이상치 확인

앙상블학습=여러개의 분류기를 생성하고 그 예측을 결합함으로써 보다 정확한 최종 예측을 도출하는 기법 \Rightarrow 보팅, 배깅, 부스팅

보팅, 하드-다수결, 소프트-평균

배깅-랜덤포레스트=여러 개의 결정트리 분류기가 전체 데이터에서 배깅 방식으로 각자의 데이터를 샘플링해 개별적으로 학습을 수행한 뒤 최종적으로 모든 분류기가 보팅을 통해 예측결정을 하게 됨. 일부가 중첩되게 샘플링된 데이터 세트