

Random Forest

2022년 6월 3일 금요일 오후 8:47

1. Decision Tree 단점

- 중간 node error 발생 시 다음 단계로 전파
- 데이터의 미세한 변동에 크게 영향 받음
- depth가 커지면 overfitting 가능성 높아짐

2. Ensemble

- 여러 모델의 예측을 통합하여 하나의 모델로 만듦(다수결, 평균 방법 사용)
- 여러 모델의 장점 결합
- Base 모델 독립, $\text{predict} > 0.5$
- Random Forest에서 다수의 DT를 ensemble

3. Bagging(Bootstrap + Aggregating)

- random sampling with replacement
- 복원 추출로 원본데이터와 같은 분포를 유지