**랜덤포레스트 & 배깅(Bagging) & 보팅(Voting)**

**앙상블 학습**

**Theory**

앙상블 학습(Ensemble)을 통한 분류는 여러 개의 분류기(Classifier)를 생성하고 그 예측을 결합함으로써 보다 정확한 최종 예측을 도출하는 기법을 말합니다. 대부분의 정형 데이터 분류 시 앙상블 기반 알고리즘이 뛰어난 성능을 나타내고 있습니다. **(RandomForest, Gradient Boost, XGBoost)**

앙상블 학습의 유형은 전통적으로 보팅(Voting), 배깅(Bagging), 부스팅(Boosting)의 세 가지로 나눌 수 있으며, 이외에도 스태킹을 포함한 다양한 앙상블 방법이 있습니다.

그 중에서도 보팅과 배깅은 여러 개의 분류기가 투표를 통해 최종 예측 결과를 결정하는 방식입니다. 보팅과 배깅의 다른 점은 보팅의 경우 서로 다른 알고리즘을 가진 분류기를 결합하는 것이고, 배깅은 동일한 분류기를 데이터 샘플링을 서로 다르게 가져가면서 학습을 수행하는 방법입니다. **( LinearRegression, KnearestNeighbor, SVM )**

**보팅(Voting Classifier)**

**Theory**

보팅 방법(Voting Classifier)은 앞에서 말했듯, 서로 다른 여러 개의 분류기를 학습한 후 투표를 통해 최종 예측결과를 결정하는 방식입니다. 보팅 방법에는 하드 보팅 방법과 소프트 보팅 방법 이렇게 두 가지가 있습니다. 하드 보팅은 다수의 분류기가 예측한 결괏값 중 다수의 분류기가 결정한 결괏값으로 최종 결과를 선정하는 방식이고, 소프트 보팅은 분류기들의 레이블 당 결정 확률의 평균을 구하여 가장 높은 확률의 레이블을 선택하는 방식입니다. 일반적으로 소프트 보팅 방식이 많이 사용됩니다.

**▷ 결측치, 이상치 처리 -> 개별 학습기 하이퍼 파라미터 튜닝 -> 보팅 분류기 학습**

**배깅(Bagging Classifier)**

**Theory**

배깅 방법(Bagging Classifier)은 앞에서도 말했듯, 동일한 분류기를 데이터 샘플링을 서로 다르게 가져가면서 학습을 수행하고, 해당 분류기들의 결괏값 투표를 통해 최조 예측 결과를 결정하는 방법입니다.

**▷ 결측치, 이상치 처리 -> 학습기 하이퍼 파라미터 튜닝 -> 배깅 분류기 학습**

**랜덤포레스트(RandomForest)**

**Theory**

랜덤 포레스트는 배깅의 대표적인 알고리즘입니다. 랜덤 포레스트는 여러 개의 결정 트리 분류기가 전체 데이터에서 배깅 방식으로 각자의 데이터를 샘플링해 개별적으로 학습을 수행한 뒤 최종적으로 모든 분류기가 보팅을 통해 예측을 결정하게 됩니다.

**▷ 결측치, 이상치 처리 -> 랜덤포레스트 하이퍼 파라미터 튜닝**