

## 간단한 앙상블 기법

1. Voting - 각각의 단일 모델들이 예측한 분류 중 가장 많은 비율을 차지한 레이블을 최종 결과로 예측하는 방법

- 1) Hard Voting (각 모델들의 softmax 적용 값에서 가장 큰 값만 참고해 비율 계산)

ex) A, B, C 모델의 softmax 적용 값이 다음과 같다면

A 모델 = {1일 확률: 0.7, 2일 확률: 0.2, 3일 확률: 0.1}

B 모델 = {1일 확률: 0.4, 2일 확률: 0.3, 3일 확률: 0.3}

C 모델 = {1일 확률: 0.0, 2일 확률: 0.9, 3일 확률: 0.1}

각 모델의 max 분류만 참고하여

“A번 모델 = 1이라 예측”

“B번 모델 = 1이라 예측”

“C번 모델 = 2이라 예측”

와 같이 생각하여 가장 비율이 큰 1을 최종 분류로 결정하는 방식으로 voting을 진행

- 2) Soft Voting ( softmax 적용 값을 모두 참고하여 비율 계산)

ex) A, B, C 모델의 softmax 적용 값이 다음과 같다면

A 모델 = {1일 확률: 0.7, 2일 확률: 0.2, 3일 확률: 0.1}

B 모델 = {1일 확률: 0.4, 2일 확률: 0.3, 3일 확률: 0.3}

C 모델 = {1일 확률: 0.0, 2일 확률: 0.9, 3일 확률: 0.1}

각 모델의 softmax 값을 모두 참고하여

“1이라 예측한 softmax 값 총합 =  $0.7(A) + 0.4(B) + 0.0(C) = 1.1$ ”

“2이라 예측한 softmax 값 총합 =  $0.2(A) + 0.3(B) + 0.9(C) = 1.4$ ”

“3이라 예측한 softmax 값 총합 =  $0.1(A) + 0.3(B) + 0.1(C) = 0.5$ ”

와 같이 생각하여 가장 총합이 큰 2를 최종 분류로 결정하는 방식으로 voting을 진행

## 2. Bagging - (Boostrapping + Aggregating)

Boostrapping - 전체 집합에서 무작위 복원추출을 통해 여러 부분집합을 만드는 행위

Aggregating - 평균이나 최빈값을 도출하는 동작

Boostrapping을 통해 생성된 각기 다른 데이터셋으로부터 학습한 여러 모델들의 아웃풋을 집계하여 최종된 예측값/분류값 도출

### Prediction

**Averaging : Bagging**을 통해 얻은 각 모델들의 출력값의 평균을 최종 출력으로 사용한다.

if

Model\_1  $\Rightarrow$  5.5

Model\_2  $\Rightarrow$  7

Model\_3  $\Rightarrow$  4.5

Final\_output =  $(5.5 + 5 + 4.5)/3 = 5.0$

### Classification

데이터 성격에 맞는 **Voting** 방식을 택한다.

FYI, Random Forest 모델은 각 의사결정나무의 출력을 Hard-Voting을 통해 집계한다.