

Part.04

Ensemble Learning

I Ensemble Learning2

FASTCAMPUS
ONLINE

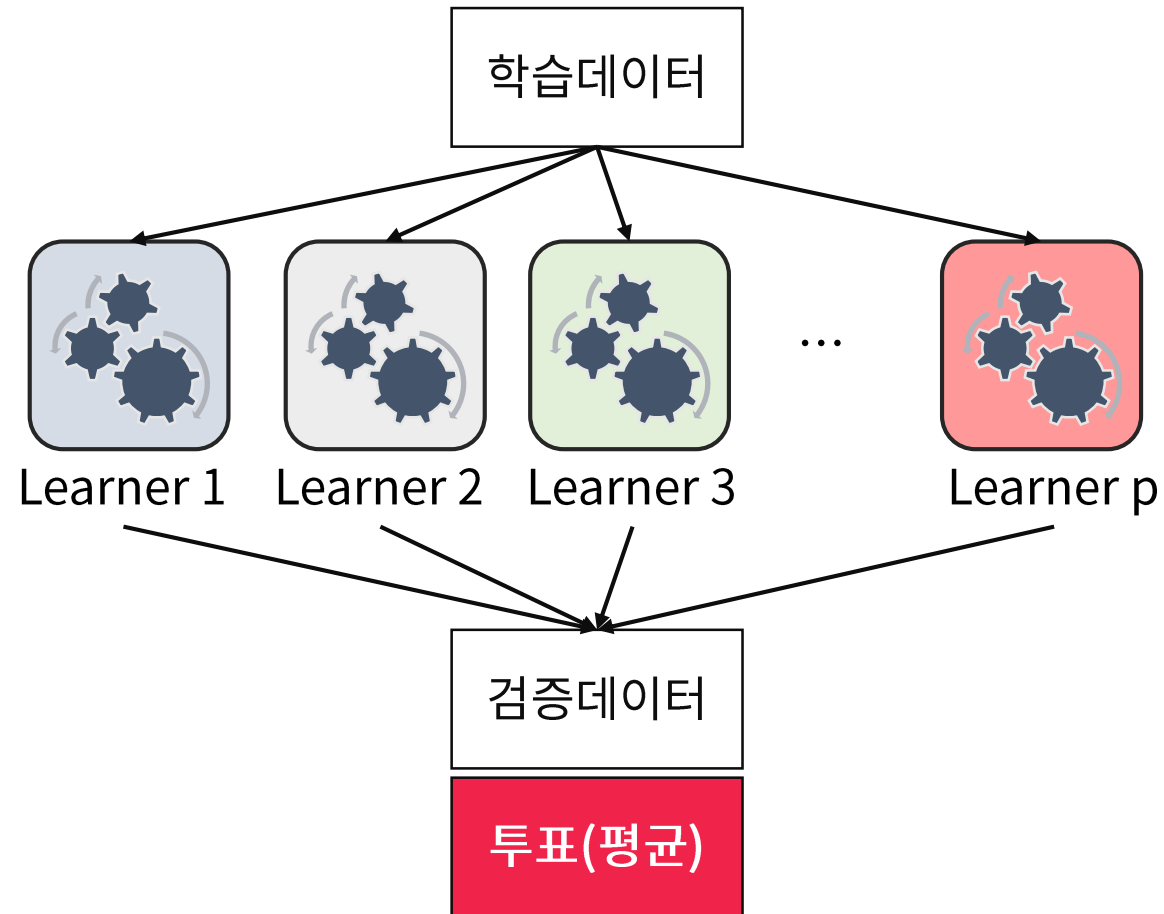
머신러닝과 데이터분석 A-ZI

강사. 이경택

I Ensemble Learning이란

■ Ensemble Learning

- 여러 개의 기본 모델을 활용하여 하나의 새로운 모델을 만들어내는 개념



I Ensemble Learning이란

■ Ensemble Learning

L classifier with p probability of wrong,

$$P_{ensemble}(K; L, p) = \binom{L}{K} p^K (1 - p)^{L-K}$$

$p < 1/2 \rightarrow L$ 이 증가함에 따라 기댓값은 낮아짐

$p > 1/2 \rightarrow L$ 이 증가함에 따라 기댓값이 올라감 낮아짐

- 이진 분류 문제에서 각각의 base learner들은 0.5 이상의 정확도를 가져야함
- 각각의 분류기는 독립이어야함
- 총 분류기의 수는 이론적으로 무한대이어야 함

I Ensemble Learning이란

■ Ensemble Learning

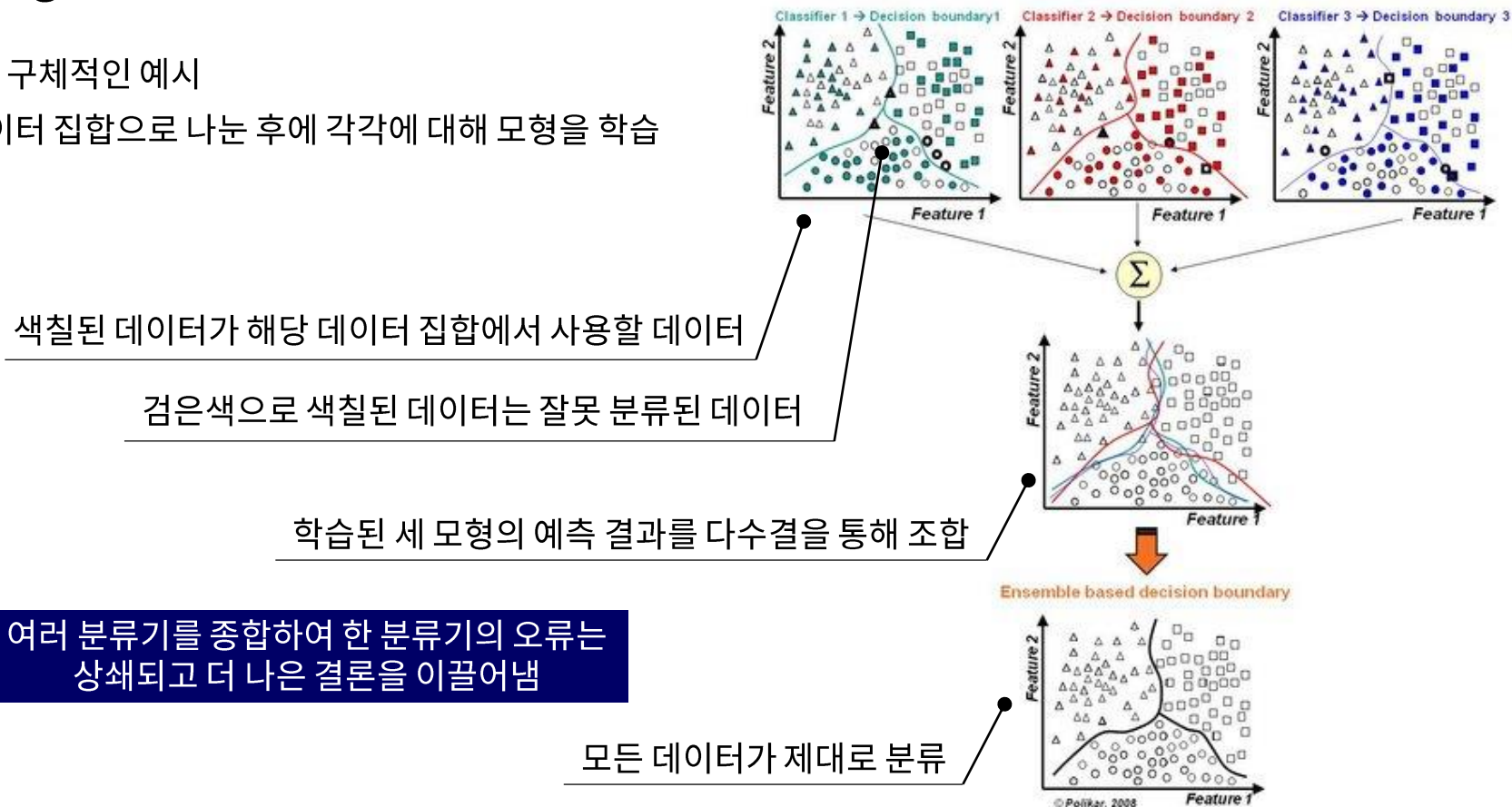
Ensemble learning 예제

Label	Learner1	Learner2	Learner3	voting
1	<u>1</u>	0	0	0
0	1	<u>0</u>	<u>0</u>	<u>0</u>
1	0	<u>1</u>	<u>1</u>	<u>1</u>
1	0	0	0	0
1	<u>1</u>	<u>1</u>	<u>1</u>	<u>1</u>
0	1	1	1	1
0	<u>0</u>	<u>0</u>	1	<u>0</u>
0	<u>0</u>	1	<u>0</u>	<u>0</u>
정확도	0.5	0.5	0.5	<u>0.625</u>

I Ensemble Learning이란

■ Ensemble Learning

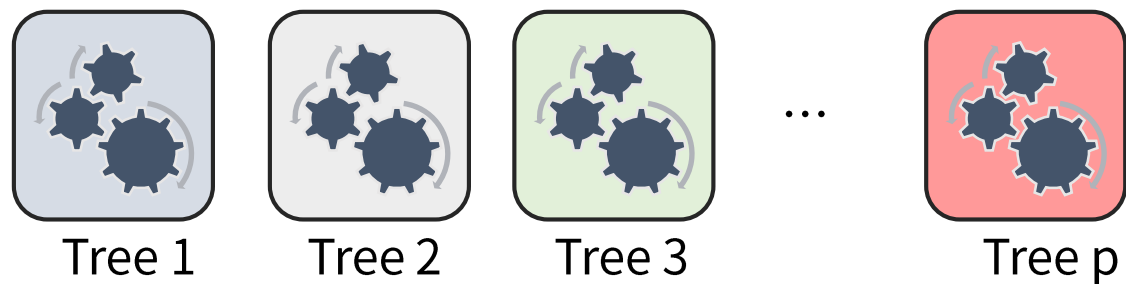
- 오른쪽 그림은 앙상블의 구체적인 예시
- 전체 데이터를 총 세 데이터 집합으로 나눈 후에 각각에 대해 모델을 학습



I Ensemble Learning이란

■ Ensemble Learning

- Test 데이터에 대해 다양한 의견(예측값)을 수렴하기 위해 overfitting이 잘되는 모델을 기본적으로 사용



- Ensemble 개념 자체는 여러 모델의 조합을 뜻하기 때문에 Tree가 아닌 다른 모델을 사용해도 무방
- 가장 많이 쓰이는 RandomForest, Boosting은 이 Tree 기반 모델

Part.04

Ensemble Learning

Ensemble의 종류

FASTCAMPUS
ONLINE

머신러닝과 데이터분석 A-Z

강사. 이경택