

Part.04

Ensemble Learning

# | Gradient Boosting의 종류

FASTCAMPUS  
ONLINE

머신러닝과 데이터분석 A-Z

강사. 이경택

# I Gradient Boosting의 종류

## ■ Gradient Boosting

- $x$ 를 입력 받아  $y$ 를 예측하는 모델  $h_0$ 가 있다고 하자.

$$y = h_0(x) + \text{error}$$

- Error가 예측 불가능한 랜덤 노이즈가 아닌 경우, 예측 성능을 올리는 가장 직관적인 방법은 error를 제거하는 것.
- 그렇다면 어떻게 error를 제거할 수 있을까?

$$\text{error} = h_1(x) + \text{error2}$$

$$\text{error2} = h_2(x) + \text{error3}$$

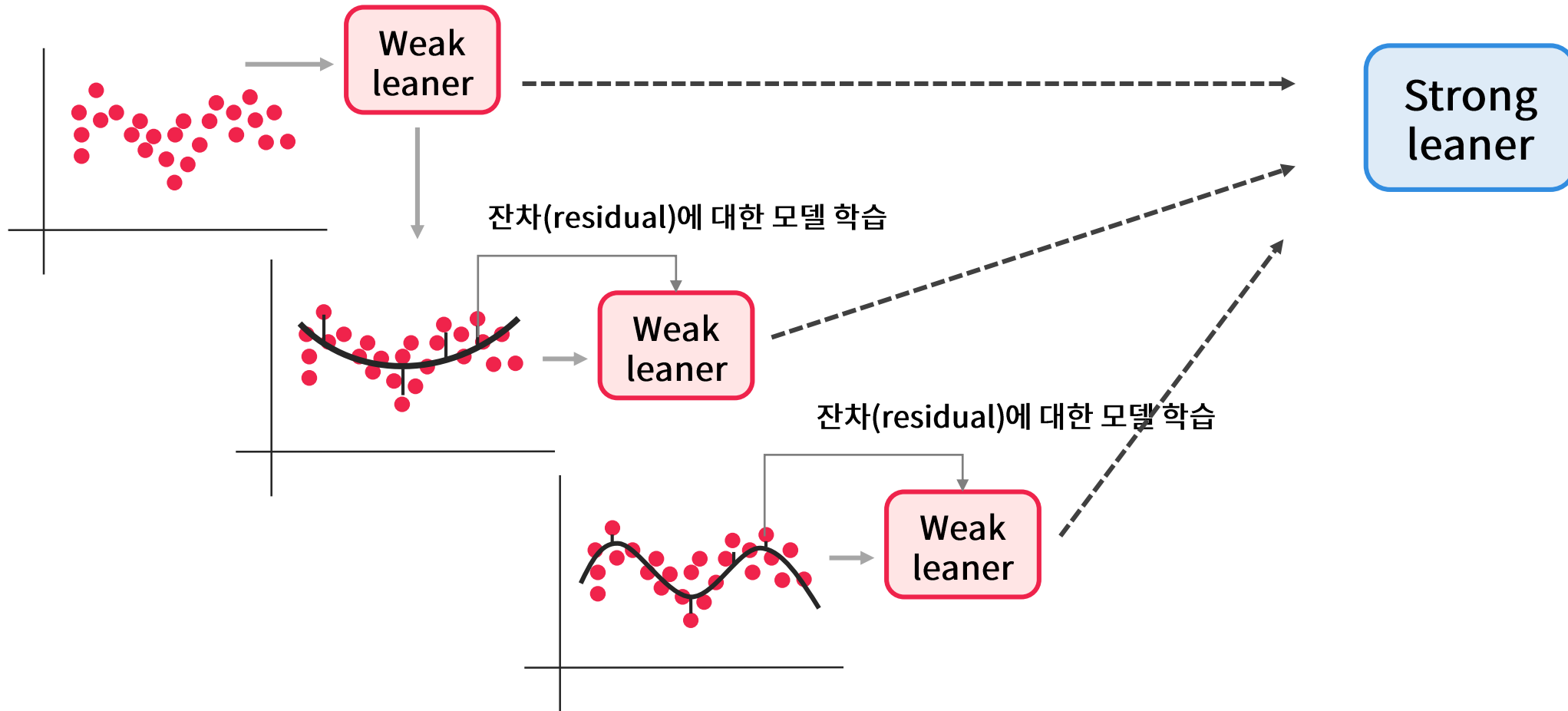
⋮



$$y = h_0(x) + h_1(x) + h_2(x) + \cdots + \text{small error}$$

# I Gradient Boosting의 종류

## ■ Gradient Boosting



# I Gradient Boosting의 종류

## ■ XGBoost

- 머신러닝 알고리즘 대회인 Kaggle competition과 KDD Cup에서 우수한 성적을 낸 팀들이 XGBoost를 많이 활용한 것이 알려지면서 주목을 받음.
- XGBoost는 Gradient Boosting 개념을 의사 결정 나무에 도입한 알고리즘으로 데이터 별 오류를 다음 Round 학습에 반영 시킨다는 측면에서 기존 Gradient Boosting과 큰 차이는 없음.
- 다만, XGBoost는 Gradient Boosting과는 달리 학습을 위한 목적식에 Regularization term이 추가되어 모델이 과적합 (overfitting) 되는 것을 방지해줌.
- Regularization term을 통해 XGBoost는 복잡한 모델에 페널티를 부여함.

Loss function

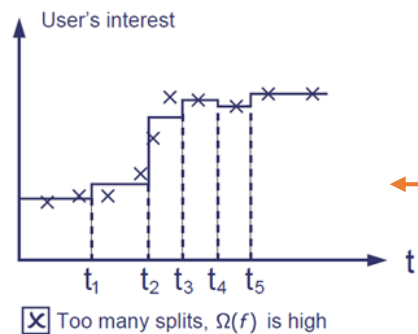
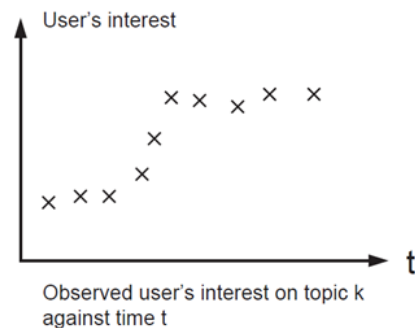
Regularization

$$obj^{(t)} = \sum_{i=1}^n l(Y_i, \hat{Y}_i^{(t)}) + \sum_{i=1}^t \Omega(f_i)$$

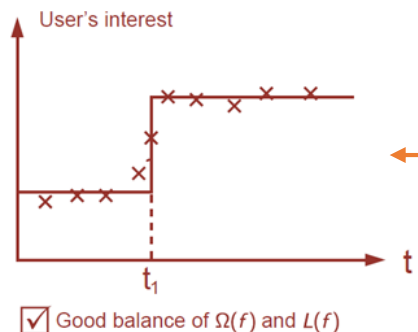
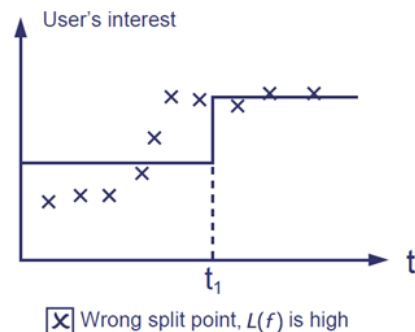
# I Gradient Boosting의 종류

## ■ XGBoost 예시

- 아래는 특정 주제에 대한 유저의 흥미 변화를 모델링하는 문제임.
- 오류를 최소화하기 위해 약한 분류기를 너무 많이 학습할 위험이 있는 Gradient Boosting은 과적합에 취약함.



Gradient Boosting

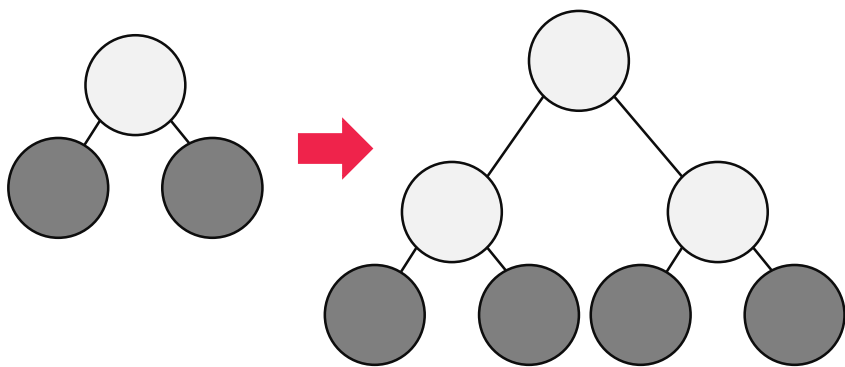


XGBoost

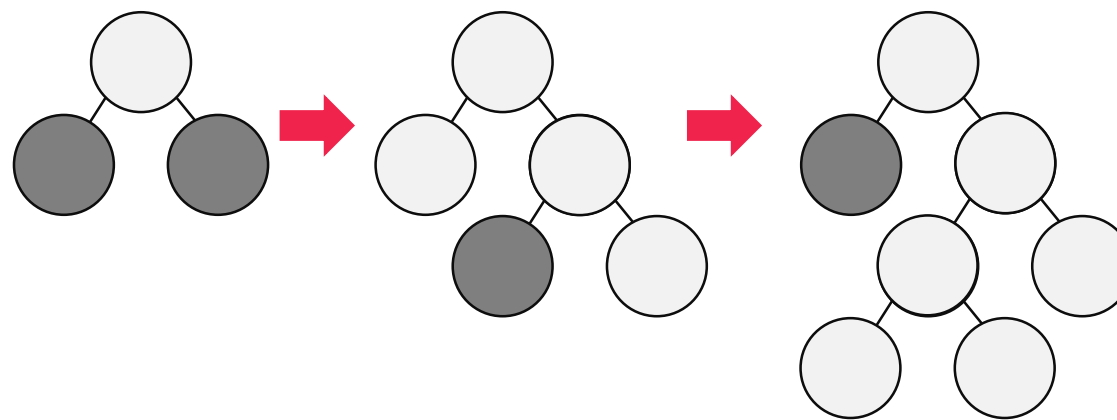
# I Gradient Boosting의 종류

## ■ LightGBM

- Xgboost와 다르게 leaf-wise loss사용(loss를 더 줄일 수 있음)
- Xgboost 대비 2배 이상 빠른 속도(동일 파라미터 기준), GPU지원
- Overfitting에 민감하여, 대량의 학습데이터를 필요로 함



Level-wise growth



Leaf-wise growth

# I Gradient Boosting의 종류

- Catboost(unbiased boosting with categorical features)
  - 잔차 추정의 분산을 최소화 하면서 bias를 피하는 boosting기법
  - 관측치를 포함한 채로 boosting하지 말고, 관측치를 뺀 채로 학습해서 그 관측치에 대한 unbiased residual을 구하고 학습하자는 아이디어
  - Categorical features가 많은 경우 잘 맞는다고 알려져 있음
  - Categorical feature를 one-hot encoding방식이 아니라, 수치형으로 변환하는 방법 제안
  - 논문에서 제시하는 실험결과로는 (Catboost > LightGBM > Xgboost)



Part.04

Ensemble Learning

# I Stacking

FASTCAMPUS  
ONLINE

머신러닝과 데이터분석 A-Z

강사. 이경택