Statistical Machine Learning

7주차

담당: 17기 이서연

1. What is Ensemble?

2. Ensemble Methods

3. Ensemble Models

1. What is Ensemble Learning?

Ensemble

Wisdom of the crowd

Ensemble

Ensemble learning

- 다수의 기본 분류 모델(base classifier, weak classifier)의 예측 결과를 종합하여, 정확한 예측 성능을 얻도록 하는 방법론

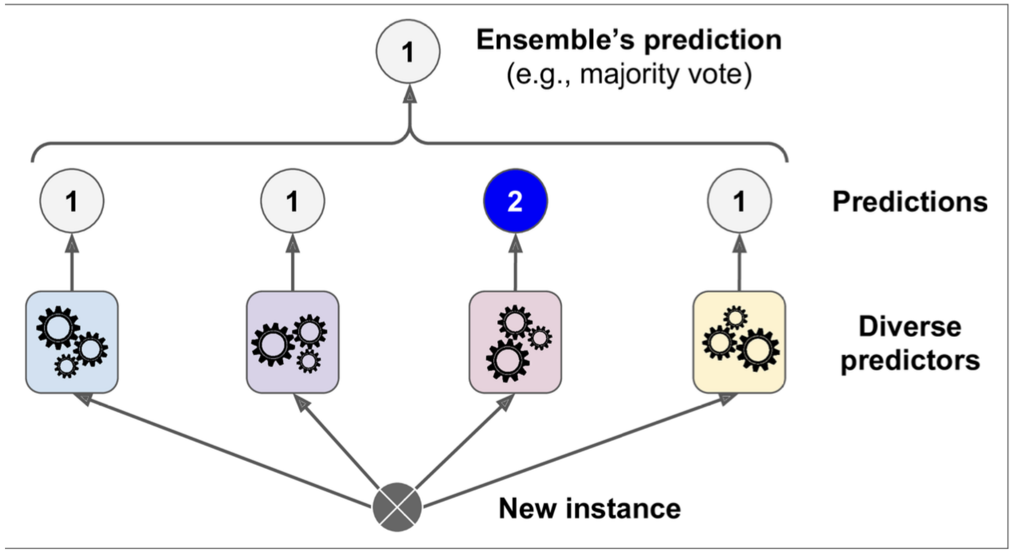
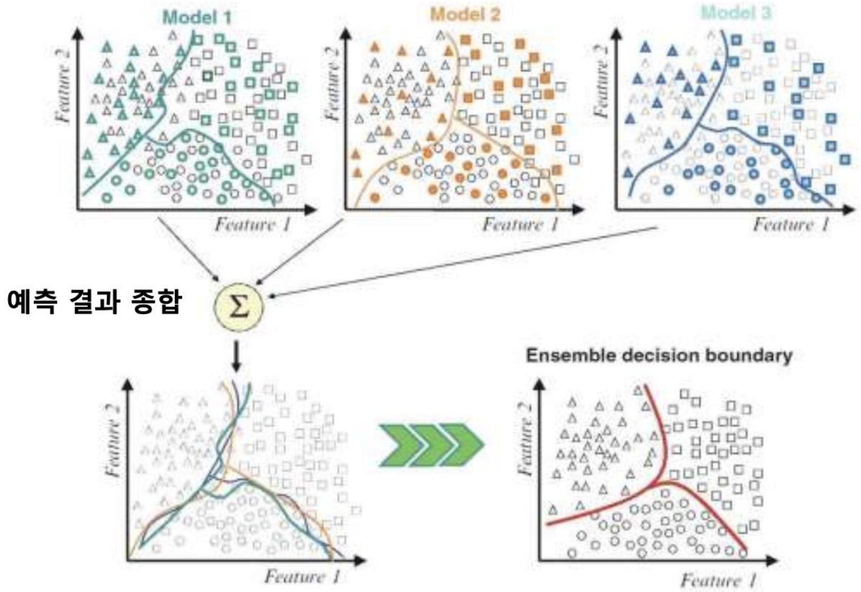


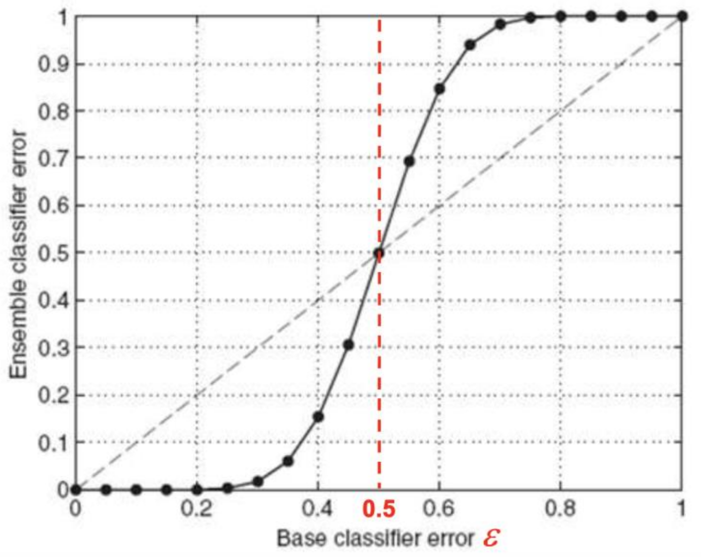
Figure 7-2. Hard voting classifier predictions



Ensemble

Example

- 25 base classifiers
- Error rate $\varepsilon = 0.35$
- Each independent
- Ensemble classifier : Majority vote

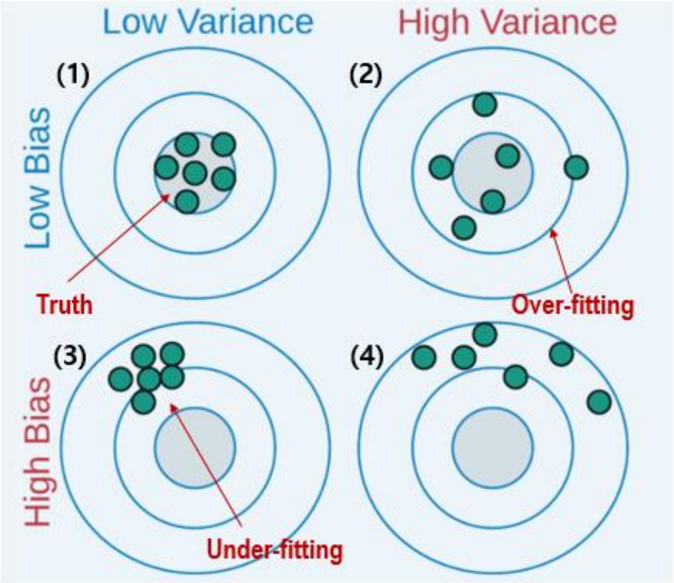


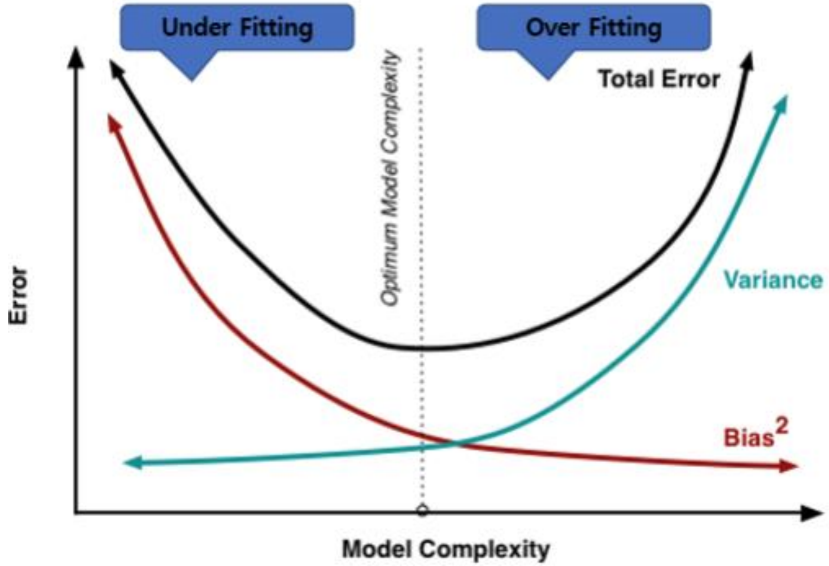
$$e_{ensemble} \sim Binomial(25, 0.35)$$

$$P(incorrect\ classifier\ \geq 13) = e_{ensemble} = \sum_{i=13}^{25} \binom{25}{i} \varepsilon^{i} (1-\varepsilon)^{25-i} = 0.06$$

Probability that the ensemble classifier makes wrong prediction

Bias and Variance





$Y = \mathrm{f(x) + e}$ (Y: predictions, x: covariates)

$$\mathrm{Error(x)} = \mathrm{E}\left[(Y - f'(x))^2\right] = \mathrm{E}\left[((f(x) - f'(x)) + e)^2\right]$$

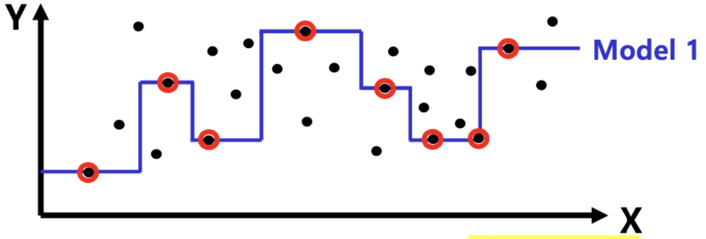
$$= (\underbrace{E[f'(x) - f(x)]}_{\text{Bias}})^2 + \underbrace{\mathrm{E}\left[(f'(x) - E[f'(x)])^2\right]}_{\text{Variance}} + \sigma_e^2$$

$$\mathrm{Error(x)} = Bias^2 + Variance + Irreducible\ Error$$

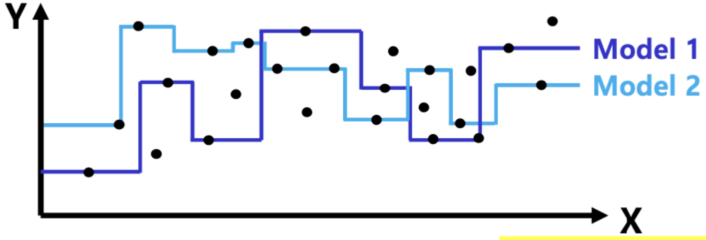
Ensemble

Ensemble Learning

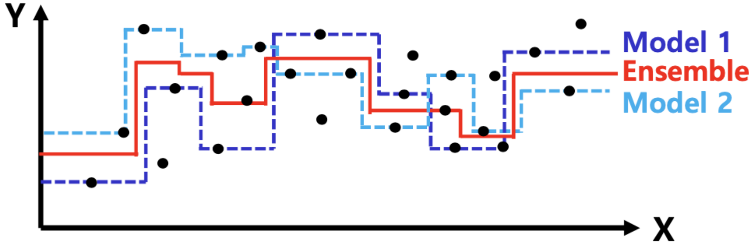
- Reduce Learning error
- Reduce Bias
- Reduce Variance



Sample 데이터에 대해 잘 예측 " Low bias "



Model 1과 Model 2 예측 값이 서로 다름 " High variance "



Model 1과 Model 2 예측 값의 평균 사용: Ensemble

"Low variance"

Ensemble

Voting

Hard Voting

Soft Voting

Weighted Voting

Stacking

Meta level Learning

Blending

Bagging

Bootstrap + Aggregating

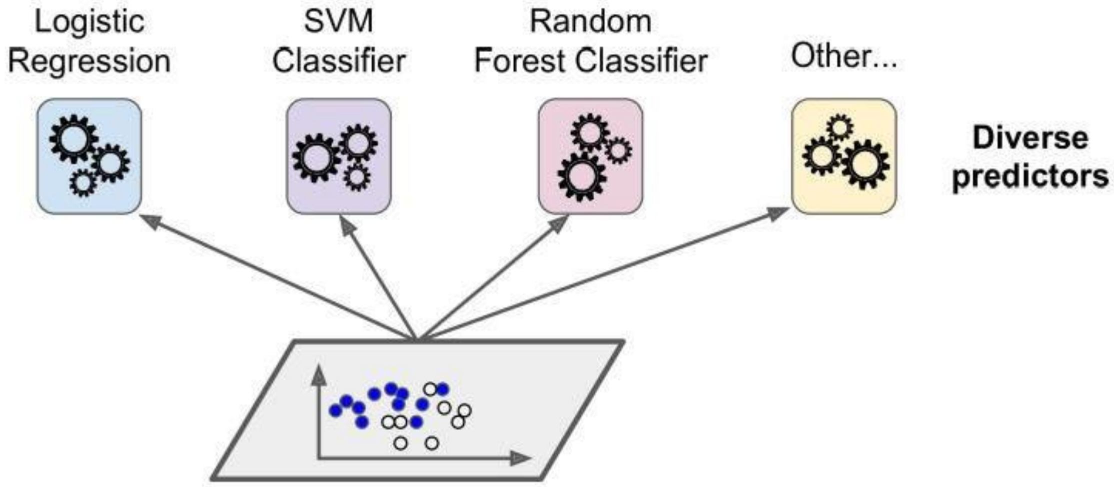
Boosting

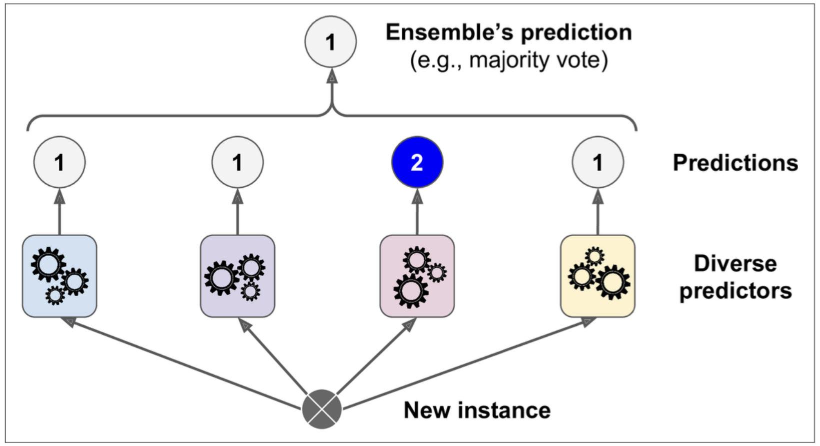
Error learner

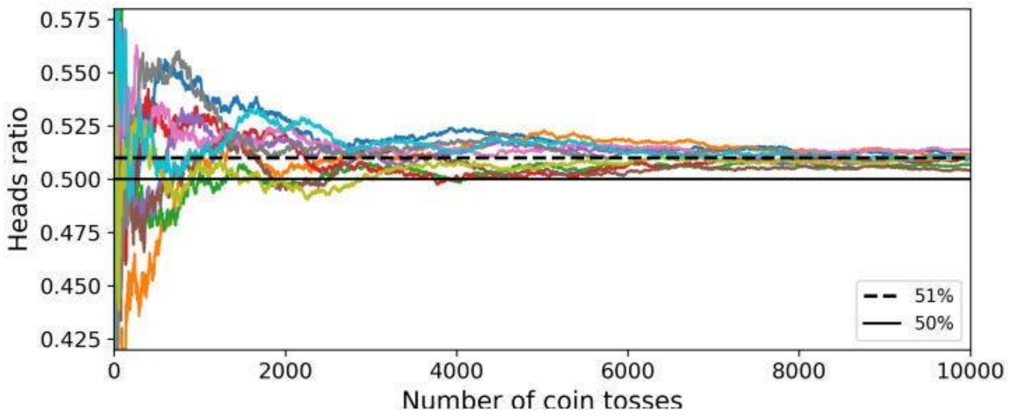
2. Ensemble Methods

Voting

Hard Voting : Majority Voting



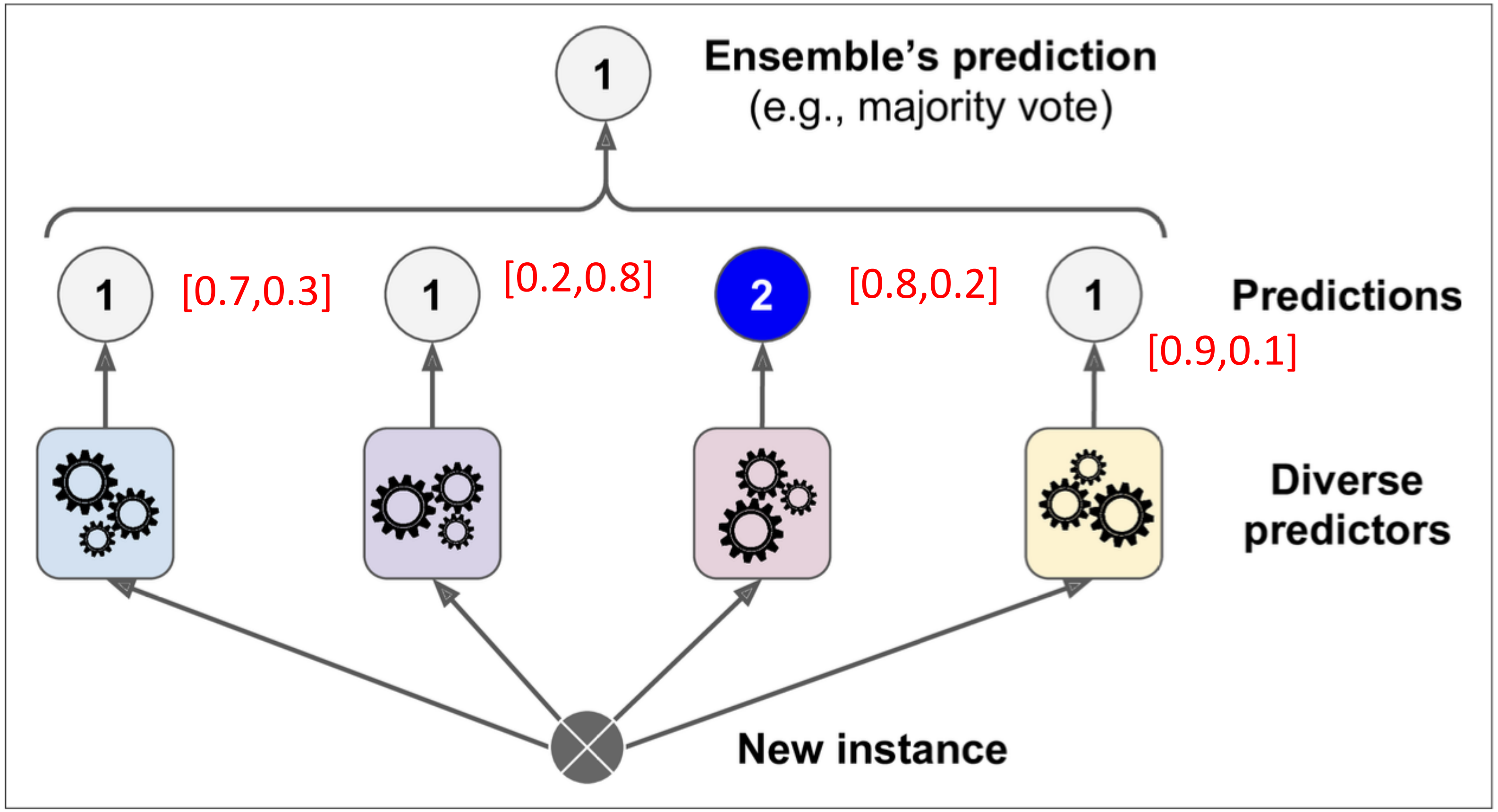




Voting

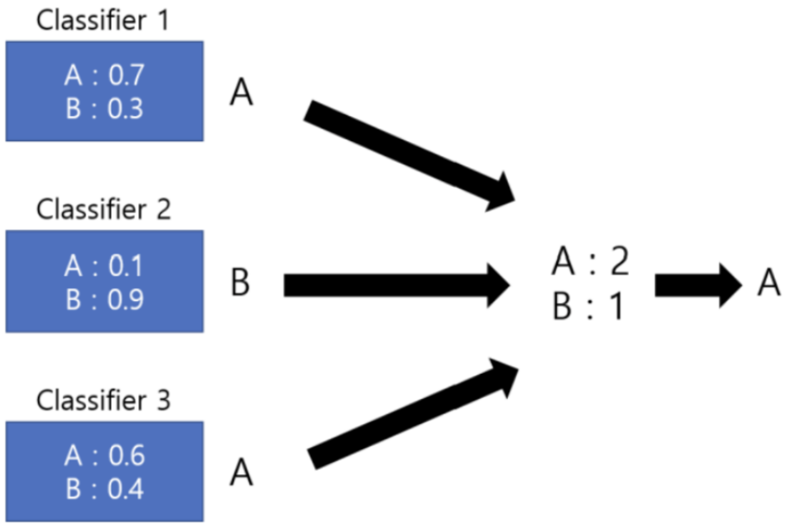
Soft Voting : Average Voting

[0.65,0.3]

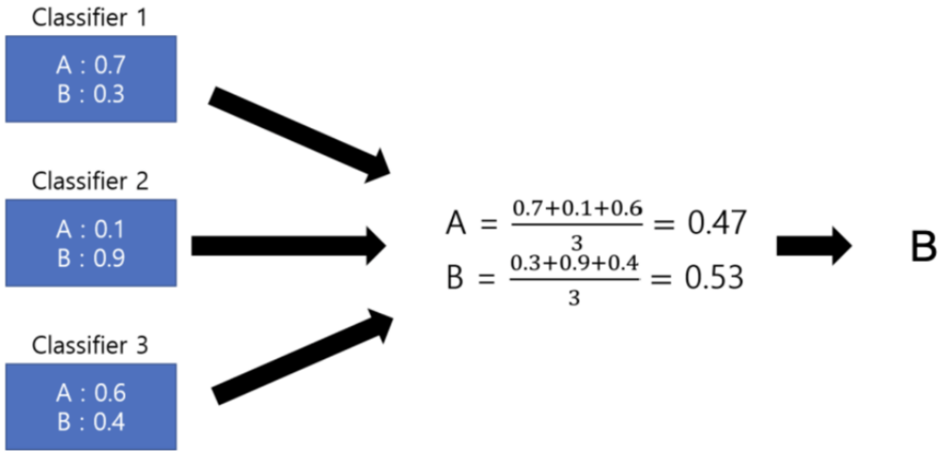


Voting

Hard Voting

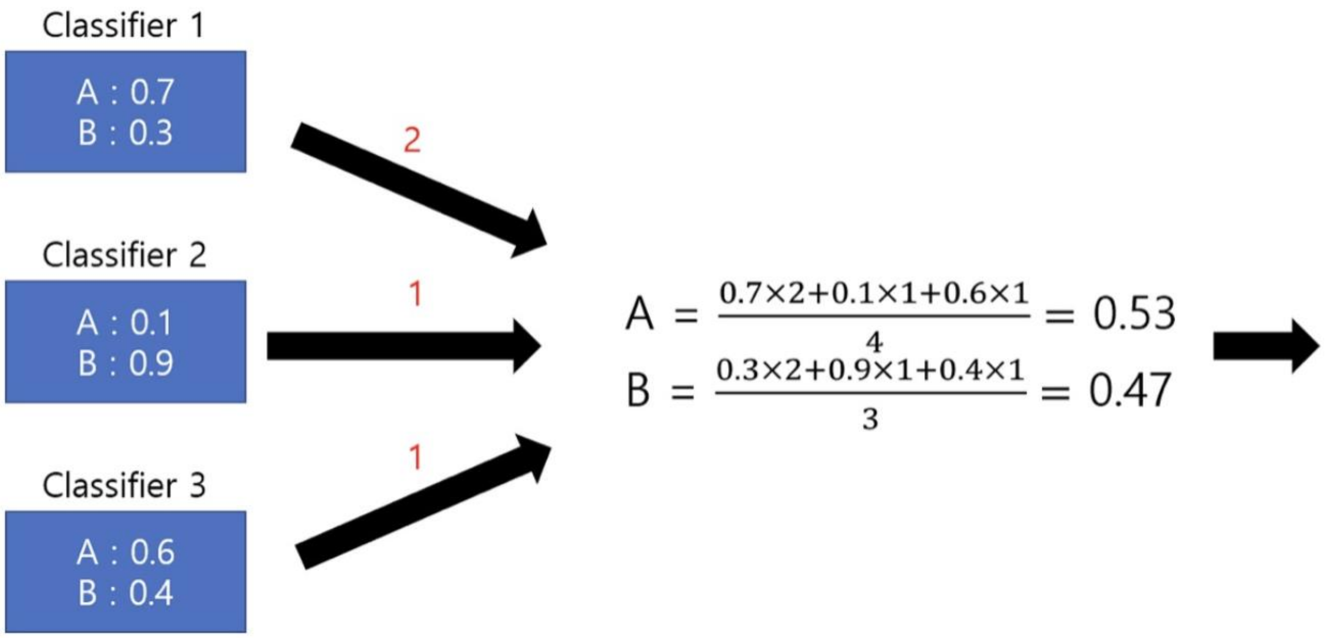


Soft Voting



Voting

Soft + Weighted



Bagging

Bagging = Bootstrap + Aggregating(Average)

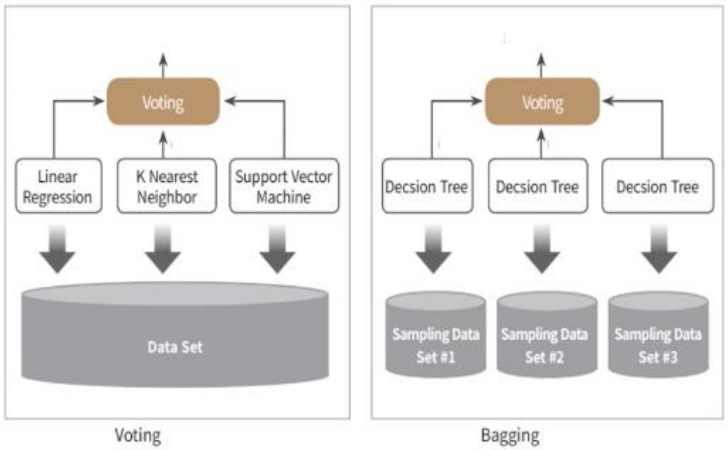
Bootstrap : sampling with Replacement → Variance 개선

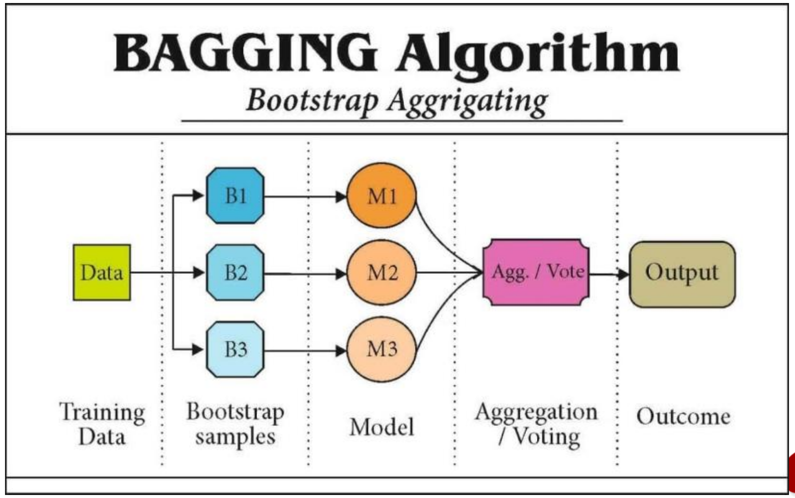
Probability that one sample is not chosen by bootstrap
(N records, N sample size)

$= (1 - \frac{1}{N})^N$

If N is large enough, then $\lim_{N \to \infty} \left(1 - \frac{1}{N}\right)^N = e^{-1} = 0.3678$
36.7% of original train dataset

Aggregating : Majority Voting, Weighting, Soft voting



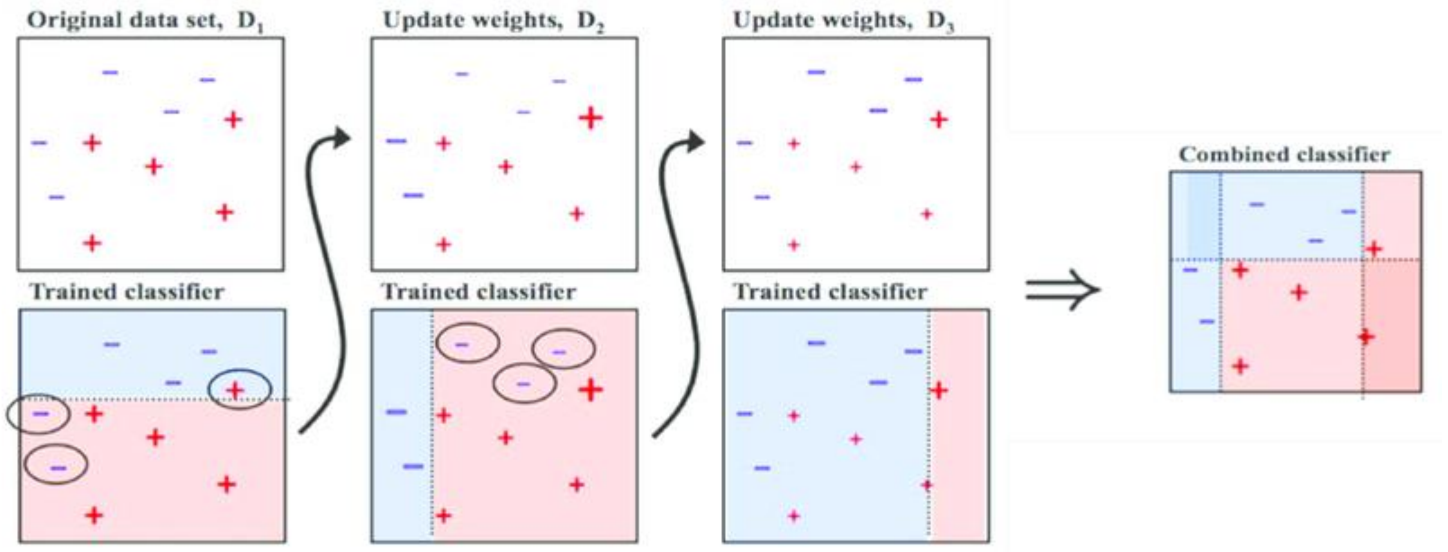


Boosting

Boosting

- 오분류된 샘플에 더 많은 가중치 부여 -> 오답을 다시 학습
- 예측이 틀린 데이터가 다시 뽑힐 가중치가 높아진다.
- 이전 모델이 잘못 예측한 부분을 집중적으로 학습

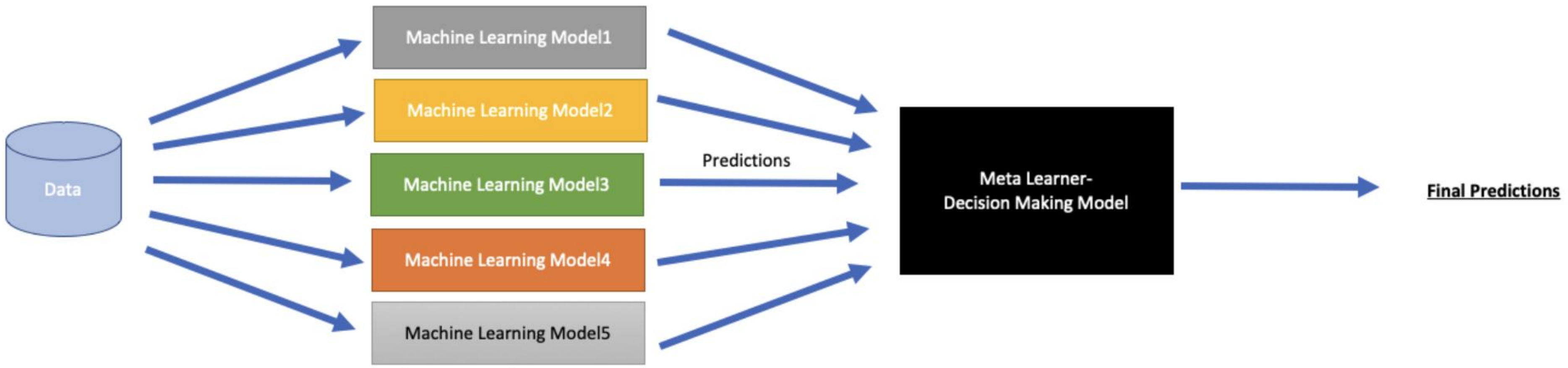
→ Bias 개선



Stacking

Stacking Generalization

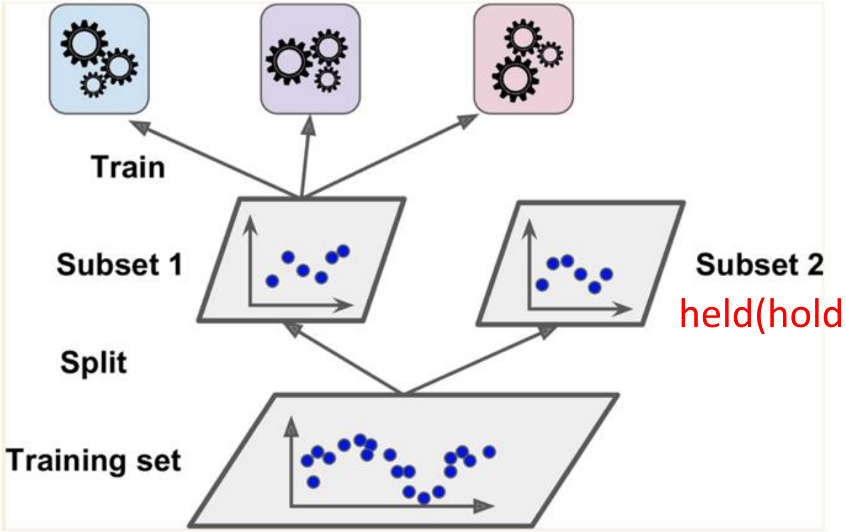
- Meta-learning model
- 여러 가지 모델들의 예측값을 최종 모델의 학습 데이터로 사용
- K-fold cv
- Step 0 : 각 weak model에 k-fold cv를 적용하여 예측 데이터를 형성
- Step 1 : step 0에서 만든 예측 데이터를 stack.하여 meta-model을 train 및 예측



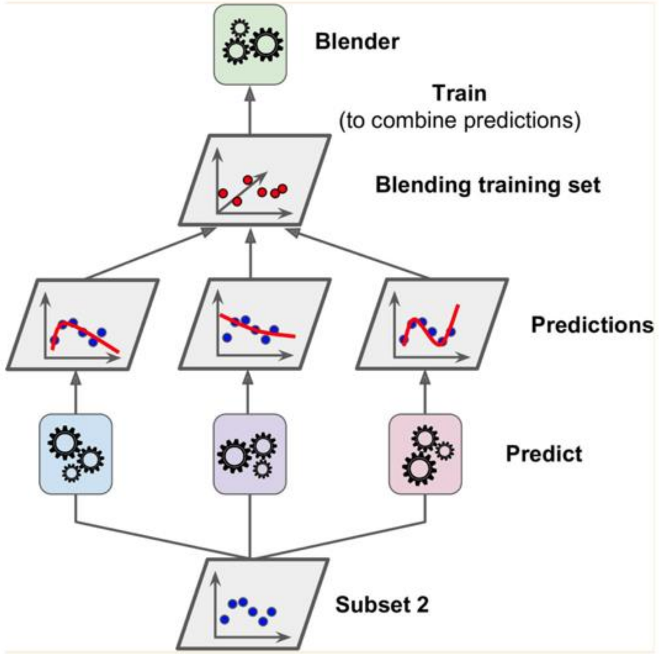
Blending

Blending Generalization

- Meta-learning model
- 개별 모델의 예측값을 다시 input으로 사용
- Use hold-out set



held(hold) out set



3. Ensemble Models

RandomForest

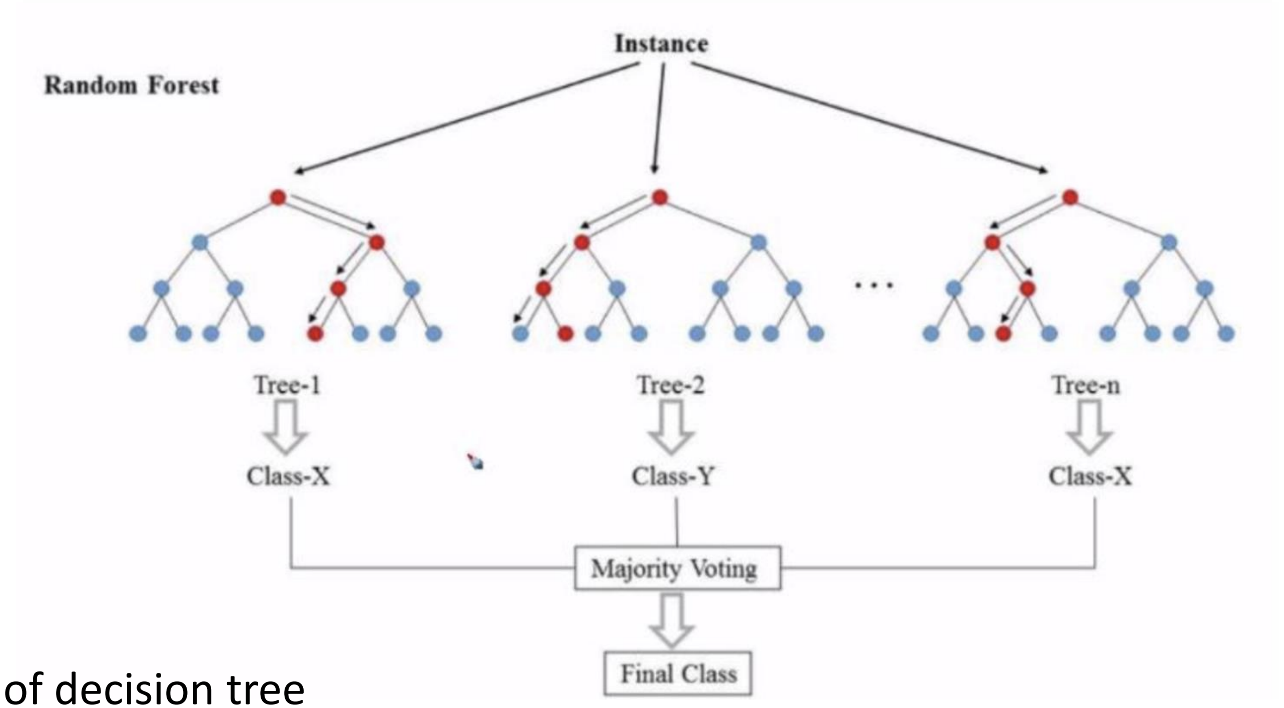
Feature Bagging → RandomForest

<u>RandomForest Decision Tree Generation</u>

- **Forest-RI(random input)**
  randomly select F features
  to split each node of decision tree

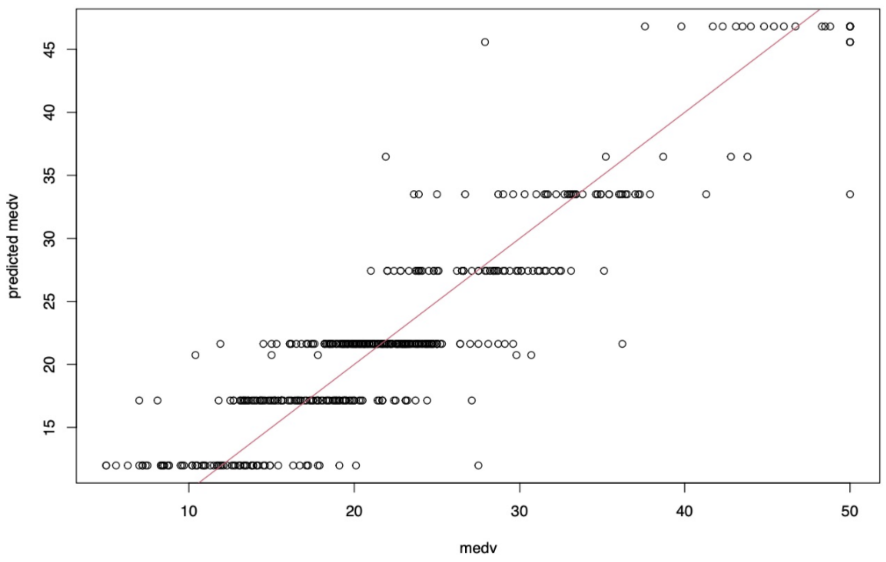
- **Forest-RC(randomly combined)**
  F randomly combined new features
  (F linear combination)

- **Randomly select**
  one of the F best splits at each node of decision tree

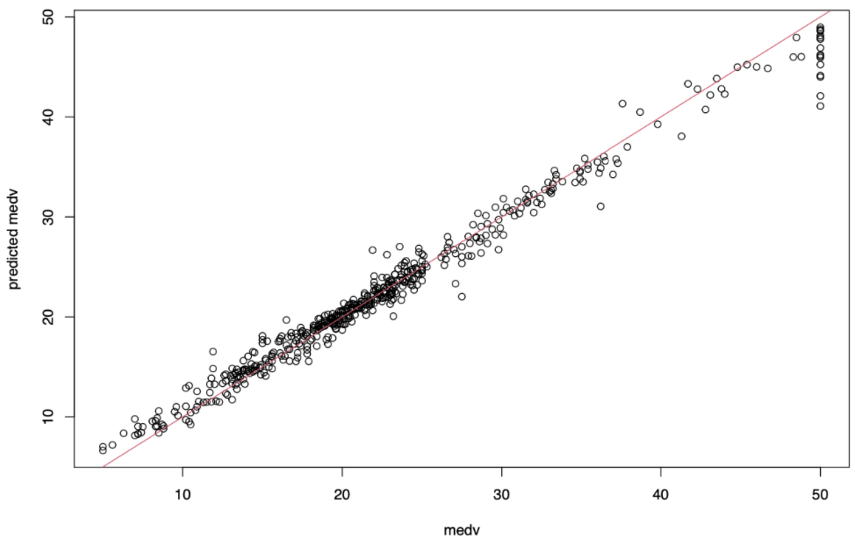


RandomForest

Single Tree



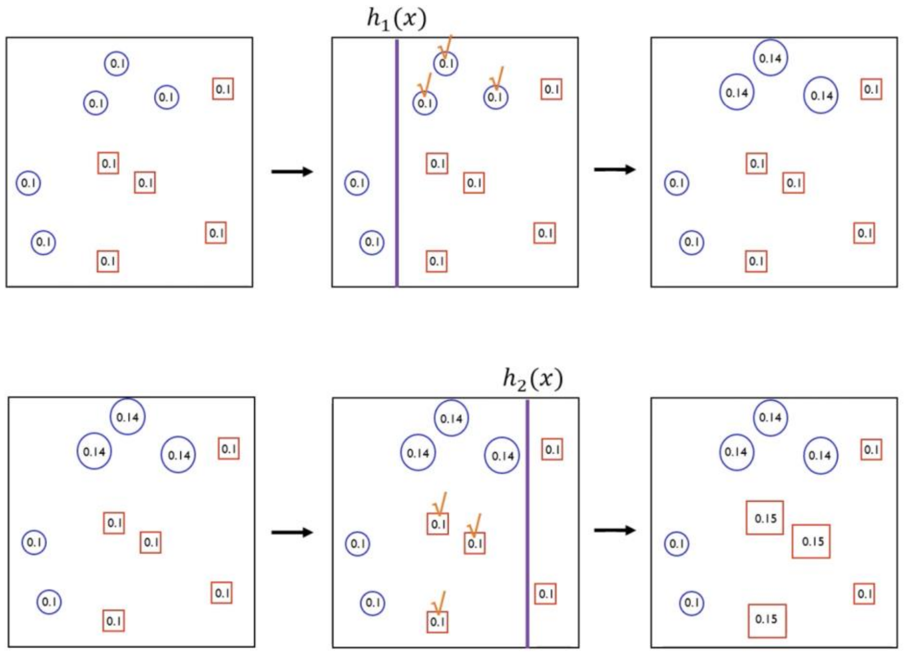
Random Forest



Adaboost

Adaboost : Adaptive + Boosting

- Adaptive : 이전 모델이 잘못 분류한 데이터의 가중치를 adaptive하게 변경
- Boosting : 이전 모델이 잘못 분류한 데이터들을 중심으로 학습



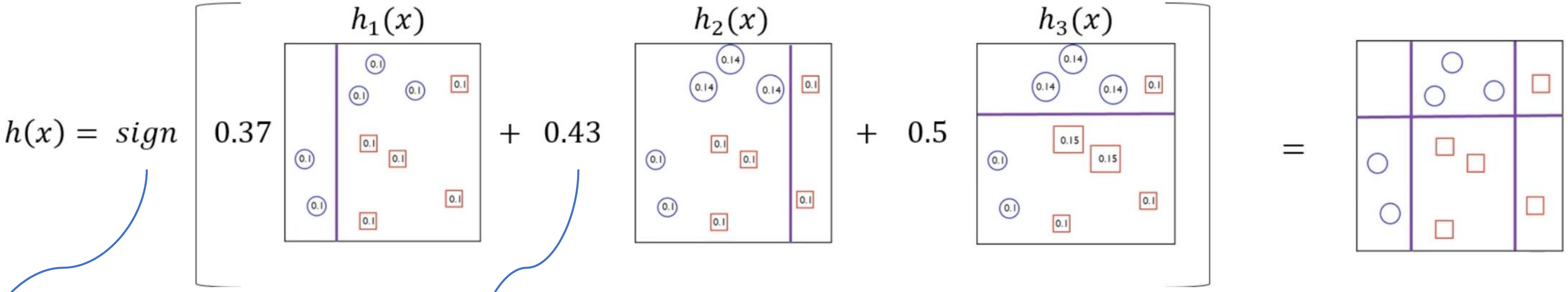
정분류 sample : 그대로

오분류 sample : 가중치 ↑

Adaboost

$$W_i = \frac{1}{n} \quad L_j = \frac{\sum_{i=1}^{n} W_i \, I(y_i \neq h_i(x))}{\sum_{i=1}^{n} W_i} . \, \alpha_j = \log\left(\frac{1 - L_j}{L_j}\right)$$

$$h(x) = sign\left[\sum_{i=1}^{m=3} \alpha_j h_j(x)\right]$$



aggregating

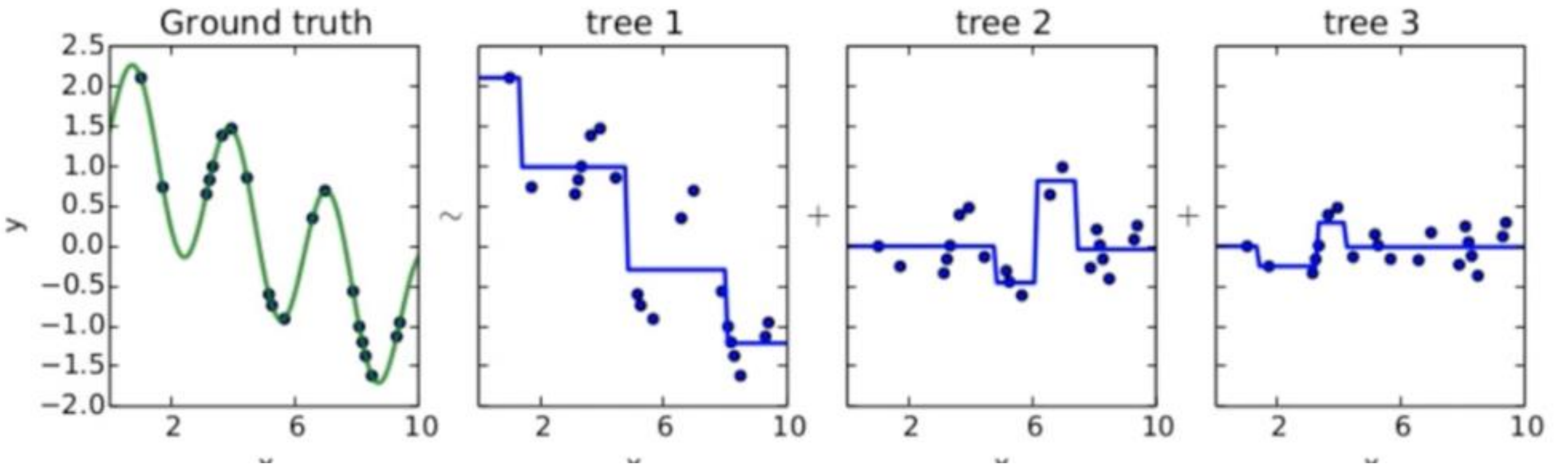
Model importance based on error rate

Gradient Boosting(GBM)

- Gradient boosting = Boosting with gradient decent
- Tree 1을 통해 Y를 예측하고 residua로 tree2 다시 학습
- 점차 residual(실제값과 예측값의 차이 작아짐
- Gradient boosting model = tree1 + tree2 + tree3

$$loss\ function : (y, f(x)) = \frac{1}{2}(y - f(x))^2$$

$$negative\ gradient : \frac{\partial(y, f(x))}{\partial f(x)} = -\frac{\partial\left[\frac{1}{2}(y - f(x))^2\right]}{\partial f(x)} = -(f(x) - y) = y - f(x)$$



Ensemble models

- RandomForest
- ExtraTrees
- Adaboost
- GradientBoost
  - XGBoost
  - LightGBM
  - CatBoost

수고하셨습니다!

해당 세션자료는 KUBIG Github에서 보실 수 있습니다!