

Unit 2

Adaboost & Gradient Boosting

11주차. 앙상블(Ensemble)

학습 내용

- Adaboost
- Gradient Boosting

학습 목표

- Adaboost 알고리즘을 이해할 수 있다.
- Gradient Boosting 알고리즘을 이해할 수 있다.

Adaboost Adaptive Boosting

Schapire & Freund, 2012 「Boosting: Foundations and Algorithms」

Weak learner를 Re-weighted
트레이닝 데이터 위에서 여러 번 학습

학습된 classifiers
또는 regressors의 결과를 결합

Strong Learner 앙상블 모형 완성

- Classifiers 결합 : Weighted Majority Vote
- Regressors 결합 : 평균

Adaboost

✓ Boosting

Weak learner가 앙상블 내에서 생성될 때

c_t 와 c_{t+1} 은 연관성을 가지고 생성

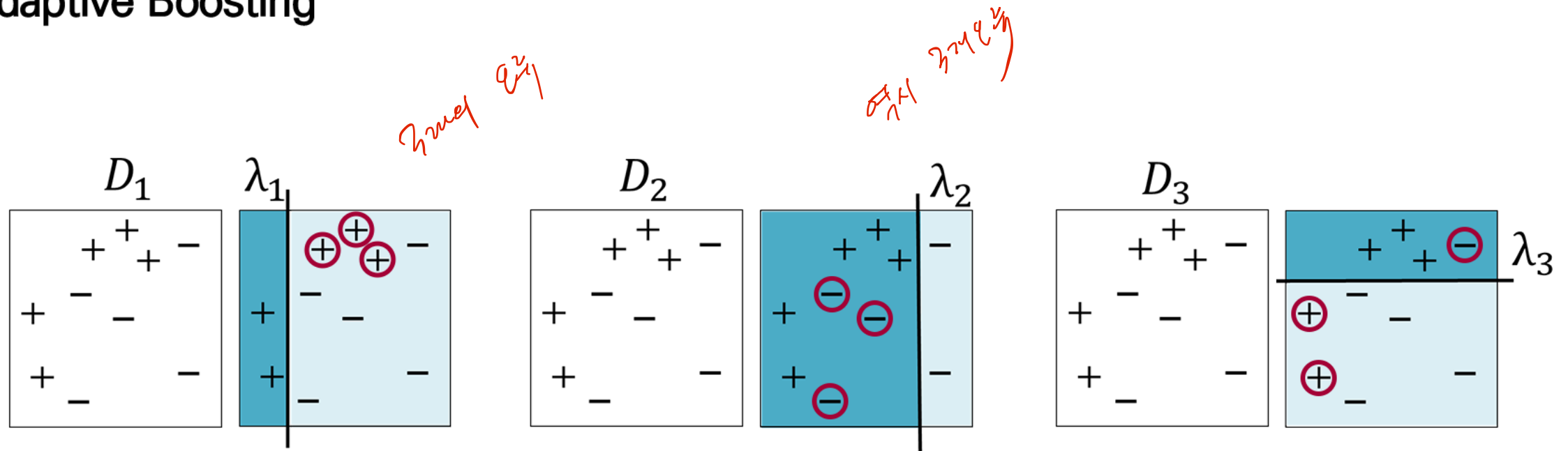
- » x 의 샘플을 c_t 가 옳게 분류하는 것과 틀리게 분류하는 것으로 분류
- » 옳게 분류하는 샘플은 인식이 가능해졌으므로 가중치를 낮춤
- » 틀리게 분류하는 샘플은 아직 "까다로운" 상태이므로 가중치를 높임
- » c_{t+1} 을 학습시키기 위한 샘플 집합의 resampling 과정에서, 가중치가 높은 샘플이 뽑힐 확률이 높아지는 정책 사용



Resampling할 때는 Replacement가 허용되지 않는다!

Adaboost

✓ Adaptive Boosting



Adaboost

✓ Adaptive Boosting

$$H(x) = \sum_t \rho_t h_t(x)$$

$$H = \text{sign} \left(0.42 \begin{array}{|c|c|} \hline \text{light blue} & \text{light blue} \\ \hline \end{array} + 0.65 \begin{array}{|c|c|} \hline \text{dark blue} & \text{light blue} \\ \hline \end{array} + 0.92 \begin{array}{|c|c|} \hline \text{dark blue} & \text{light blue} \\ \hline \end{array} \right)$$

$$= \begin{array}{|c|c|c|c|} \hline \text{dark blue} & \text{dark blue} & \text{dark blue} & \text{light blue} \\ \hline \text{dark blue} & \text{dark blue} & \text{dark blue} & \text{light blue} \\ \hline \text{light blue} & \text{light blue} & \text{light blue} & \text{light blue} \\ \hline \text{light blue} & \text{light blue} & \text{light blue} & \text{light blue} \\ \hline \end{array}$$

“ Weak Learner의
앙상블을 통해 정확히 분류 ”

Adaboost

Adaboost 샘플링

[동일 가중값]

[illegible]

학습진행

가중치가 다르게 샘플링

[illegible]

Unit 2 Adaboost & Gradient Boosting

Gradient Boosting

✓ Gradient Boosting 역사

Gradient Boosting

Gradient Descent + Boosting

Adaboost가 Boosting의
알고리즘 이라고
Adaboost 소개

Adaboost 를 특별한
Loss Function을 써서
Gradient Descent 소개

다양한 Loss Function에
대해서 Gradient
Boosting 방법을 활용

1996 Freund et al.
1997 Freund and Schapire

1998 Breiman et al.
1999 Breiman

2000 Friedman et al.
2001 Friedman

☑ Gradient Boosting 소개

Gradient Boosting

» MART : Multiple Additive Regression Tree

» GBRT : Gradient Boosted Regression Tree

➡ 약한 분류기를 반복적으로 학습하며
예측이 틀린 관측값에 집중

차별점

임의의 미분 가능한 손실 함수를 최적화문제로 일반화

→ 함수 재설정 point.

☑ Stochastic Gradient Descent

SGD(Stochastic Gradient Descent)

» 모델(=함수, F) 을 optimization 할 때, **모델의 구조는 고정**

» **매개변수(parameter)에 대해 편미분**

- θ s of Logistic regression
- θ s of Linear regression

최적화 문제

$$F(x|\Theta) = \min_{\Theta} \text{Loss}(y, F(x|\Theta))$$

☑ Gradient Boosting 차이점

Gradient Boosting

- » 모델의 구조가 고정되어 있다고 가정하지 않음
- » 데이터를 가장 잘 설명하는 함수를 찾는 것이 목적

➡ Gradient Boosting은 Loss 함수를
파라미터뿐만 아니라 함수 자체에 대해서 최소화

Gradient Boosting

☑ 가법 모형

손실 함수를 최소화하기 위해
Weak Learner 추가



가법 모형(Additive Model)

→ 함수 최적화 문제

기존 함수



새로운
Tree

최적 파라미터와 최적 Function을 동시에 찾는 것이 목표!
최적화 문제 해결이 조금 복잡!

Gradient Boosting

☑ GB와 SGD의 차이

SGD(Stochastic Gradient Descent)

- 1개의 복잡한 모델을 학습

GB(Gradient Boosting)

- 여러 개의 간단한 모델을 학습

✓ Gradient Boosting 알고리즘

▶ 새로운 모형

초기화 : $f_0(x) = \underset{\gamma}{\operatorname{argmin}} \sum_{i=1}^N L(y_i, \gamma)$

For : $m=1$ to M

■ $i=1, \dots, N$ 에 대해, 다음을 계산 $r_{im} = - \left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)} \right]_{f=f_{m-1}}$ → 항목에 대해 편미분.

■ 타겟이 r_{im} 인 회귀 트리를 학습하여 terminal regions : $R_{jm}, j = 1, 2, \dots, J_m$ 을 생성

■ $j = 1, \dots, J_m$ 에 대해, 다음을 계산 $\gamma_{jm} = \underset{\gamma}{\operatorname{argmin}} \sum_{x_i \in R_{jm}} L(y_i, f_{m-1}(x_i) + \gamma)$ ▶ Additive model

■ Update : $f_m(x) = f_{m-1}(x) + \sum_{j=1}^{J_m} \gamma_{jm} I(x \in R_{jm})$ 회보올려주는 값에 update

출력 : $\hat{f}(x) = f_M(x)$

☑ Adaboost와 GB의 공통점

Adaboost와 Gradient Boosting **공통점**

1

Boosting Approach

2

Weak Learner를 기본으로 **반복적으로**
예측하기 힘든 관측값에 더 집중해서
성능을 부스트

3

여러 약한 학습자가 합쳐져 하나의
강한 학습자 생성

☑ Adaboost와 GB의 차이점

Adaboost와 Gradient Boosting **차이점**

Adaboost	Gradient Boosting
이전에 잘못 분류된 관측값에 더 높은 가중치 부여	이전 반복에서 큰 잔차를 남긴 관측값 측정
높은 <u>가중값 데이터 포인트</u> 에 의해 단점 검색	<u>기울기</u> 를 통해 단점을 찾아냄
지수 손실 함수를 통해 적합화 정도가 나쁜 표본에 관해 더 많은 가중값 부여	기울기 Boosting이 오차항을 더 분해해서 줄여 나감 → 가법 모형

☑ Adaboost와 GB의 차이점

Adaboost와 Gradient Boosting **차이점**

Adaboost	Gradient Boosting
기울기 부스팅의 특별한 경우로 지수 손실함수 사용	기울기의 개념이 보다 일반적

앙상블 모형은 **여러 가지 방법이 존재**하기 때문에
다양한 Approach로 **새로운 모형이 개발**되고 있는 상황