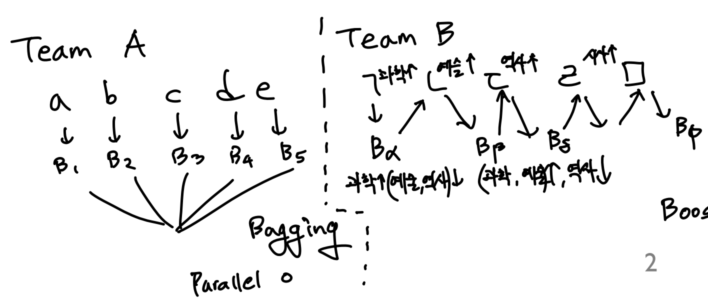
AdaBoost / GradientBoost

Ensemble Learning:
Adaptive Boosting (AdaBoost)

Pilsung Kang
School of Industrial Management Engineering
Korea University

Boosting: AdaBoost

- AdaBoosting: Idea
  - ✓ Strong model vs. Weak model → better than random Guessing
    - A weak model, performing only slightly better than random guessing, could be boosted in to arbitrarily accurate strong model
  - ✓ New classifiers should focus on difficult cases
    - Examine the learning set
    - Get some rule of thumb → Weak Model
    - Reweight the examples of the training set, concentrate on hard cases for the previous rule
    - Derive the next rule of thumb
    - …
    - Build a single, accurate predictor by combining the rules of thumb



Boosting → 일반적으로 속도↑

Parallel X

Boosting: AdaBoost

- AdaBoosting: Idea

*Classification Case

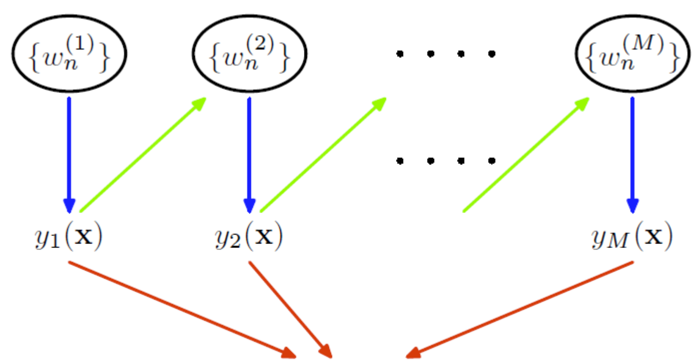
  ✓ Strong model vs. Weak model

    - A weak model, performing only slightly better than random guessing, could be boosted in to arbitrarily accurate strong model

  ✓ Train models sequentially, with a new model training at each round

  ✓ At the end of each round, misclassified examples are identified and have their emphasis increased in a new training set which is then fed back into the next round

  ✓ Large errors made by earlier models can be compensated by the subsequent models



$$Y_M(\mathbf{x}) = \operatorname{sign}\left(\sum_{m}^{M} \alpha_m y_m(\mathbf{x})\right)$$

Boosting: AdaBoost

- AdaBoosting: Algorithm

Algorithm 2 Adaboost

Input: Required ensemble size T

Input: Training set $S = \{(x_1, y_1), (x_2, y_2), ..., (x_N, y_N)\}$, where $y_i \in \{-1, +1\}$

Define a uniform distribution $D_1(i)$ over elements of S.

첫번째 데이터셋에 example i가 선택될 확률

for $t = 1$ to T **do**

<Stump Tree>

한번에 그어서 분리

↓

Weak Model

Train a model h_t using distribution D_t.

Calculate $\epsilon_t = P_{D_t}(h_t(x) \neq y)$

If $\epsilon_t \geq 0.5$ break

오분류율

random guessing 보다는 나아야한다

Set $\alpha_t = \frac{1}{2} \ln \left(\frac{1-\epsilon_t}{\epsilon_t} \right)$

$\alpha_t \begin{cases} \epsilon_t \approx 0.5 \approx \frac{1}{2}\ln\left(\frac{0.5}{0.5}\right) = 0 \\ \epsilon_t = 0 \approx \frac{1}{2}\ln\left(\frac{1}{0}\right) \approx \infty \end{cases}$

Update $D_{t+1}(i) = \frac{D_t(i) \exp(-\alpha_t y_i h_t(x_i))}{Z_t}$

가중치 업데이트

where Z_t is a normalization factor so that D_{t+1} is a valid distribution.

$D_t(i) \rightarrow$ t시점의 선택확률 기준점

$\alpha_t \rightarrow h_t$가 정확하면 Sampling Rate의 민동성↑

end for

For a new testing point (x', y'),

$H(x') = sign\left(\sum_{t=1}^{T} \alpha_t h_t(x') \right)$

Test Data에 대한 가중치

$y_i = h_t(x_i) \Rightarrow 1$

1	1
-1	-1

$y_i \neq h_t(x_i) \Rightarrow -1$

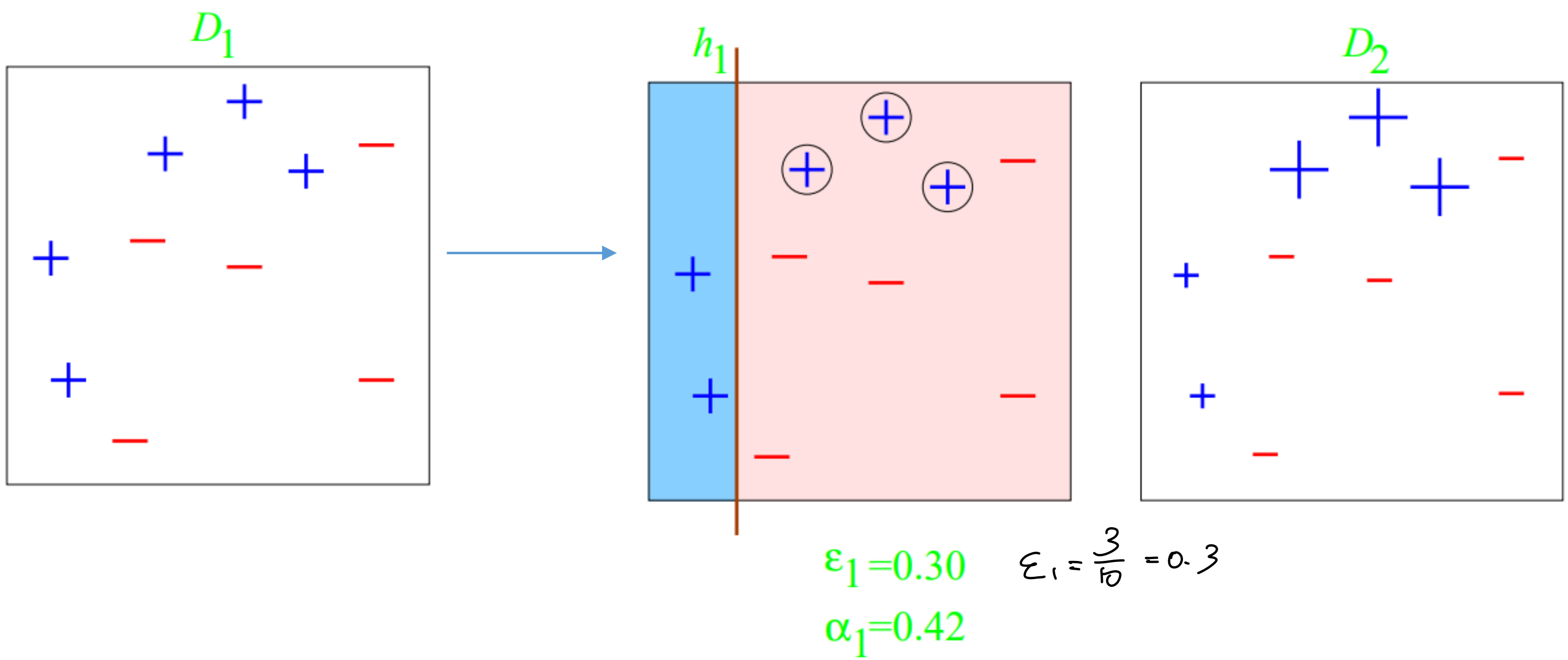
1	-1
-1	1

* h_t가 i의 정답 $D_{t+1}(i)\downarrow$

* h_t가 i의 오답 $D_{t+1}(i)\uparrow$

Boosting: AdaBoost

- Illustrative example 1
  - ✓ Round 1



- 3 misclassifications out of 10: $\epsilon_i = 0.30$
- Model confidence: $\alpha_i = \frac{1}{2} \log \left(\frac{1-\epsilon_i}{\epsilon_i} \right) = \frac{1}{2} \log \frac{1-0.3}{0.3} = 0.42$

https://www.cse.buffalo.edu/~jcorso/t/CSE455/files/lecture_boosting.pdf

Boosting: AdaBoost

- AdaBoost Example

  ✓ The selection probability of x_i for the next training dataset

$$D_{t+1}(i) = \frac{D_t(i)\exp(-\alpha_t y_i h_t(x_i))}{Z_t}$$

정분류된 데이터는 선택될 확률↓

↕

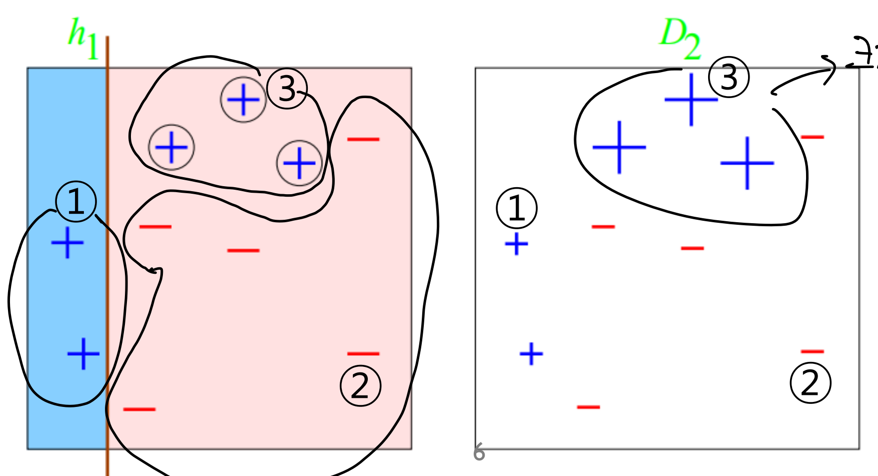
오분류는 증가

  ✓ Case 1: $y_i = 1, h_t(x_i) = 1 \rightarrow y_i h_t(x_i) = 1 \rightarrow -\alpha_t y_i h_t(x_i) < 0$ → decrease p

  ✓ Case 2: $y_i = -1, h_t(x_i) = -1 \rightarrow y_i h_t(x_i) = 1 \rightarrow -\alpha_t y_i h_t(x_i) < 0$ → decrease p

  ✓ Case 3: $y_i = 1, h_t(x_i) = -1 \rightarrow y_i h_t(x_i) = -1 \rightarrow -\alpha_t y_i h_t(x_i) > 0$ → increase p

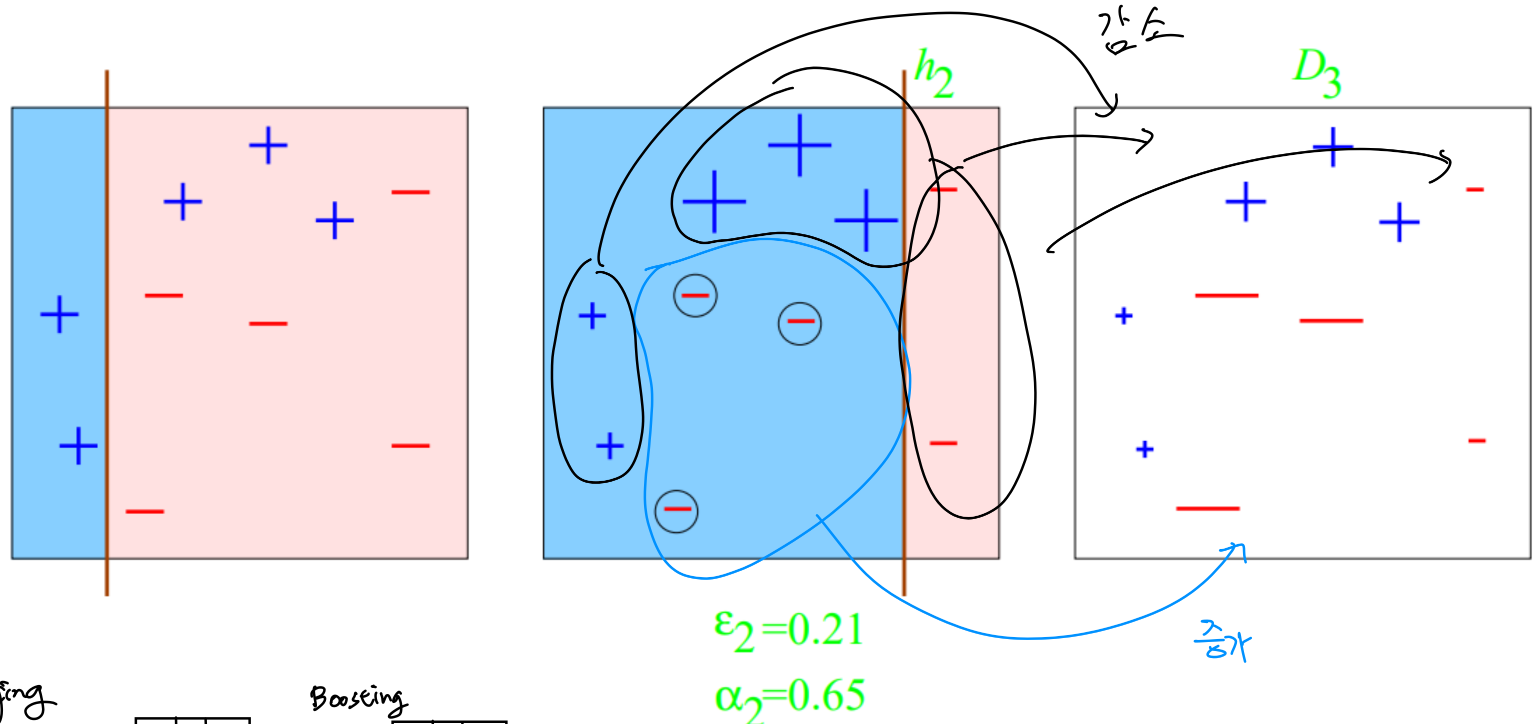
  ✓ α_t is the confidence of the current model that controls the magnitude of change



크기↑ = 선택될 확률↑

Boosting: AdaBoost

- Illustrative example 1
  - ✓ Round 2



Bagging

B_1	0.3	0.3	0.3
B_2	0.3	0.3	0.3
B_3	0.3	0.3	0.3

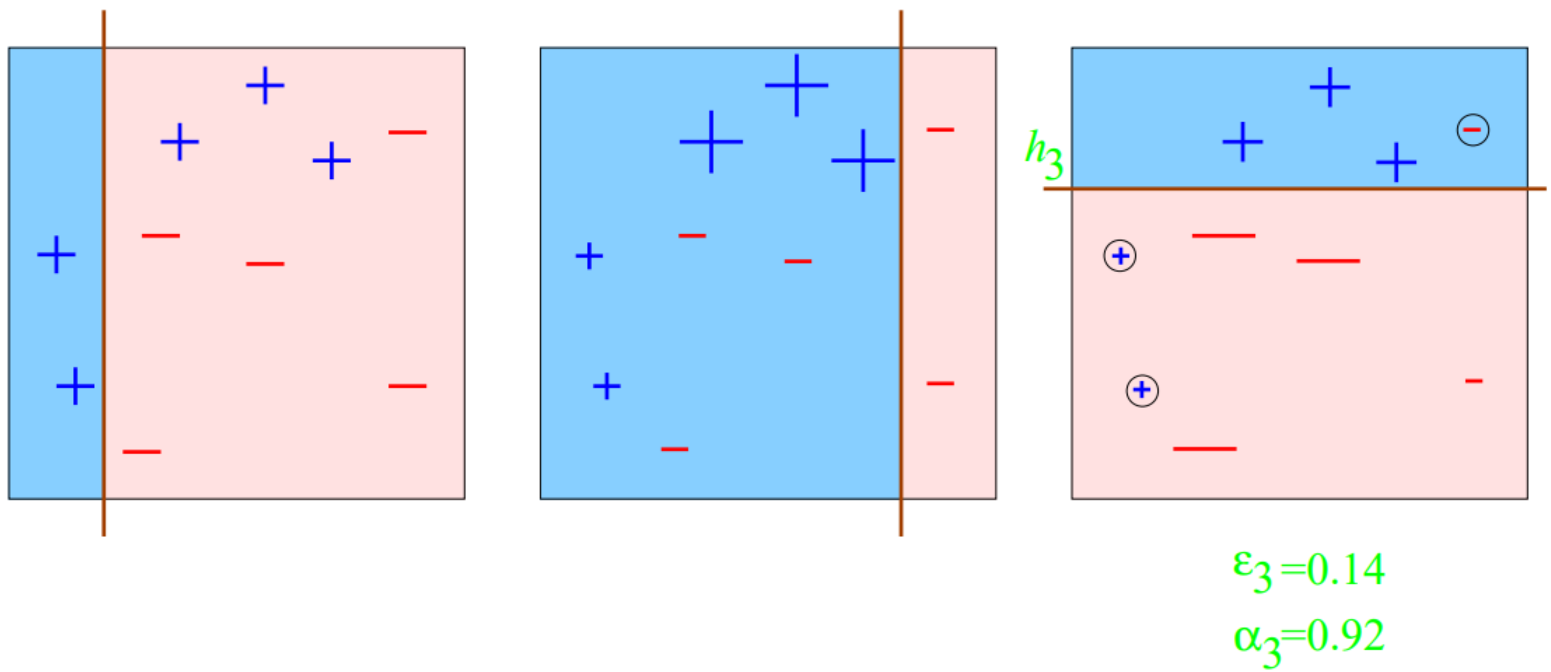
Boosting

0.3	0.3	0.3
0.5	0.3	0.2
0.4	0.4	0.2

https://www.cse.buffalo.edu/~jcorso/t/CSE455/files/lecture_boosting.pdf

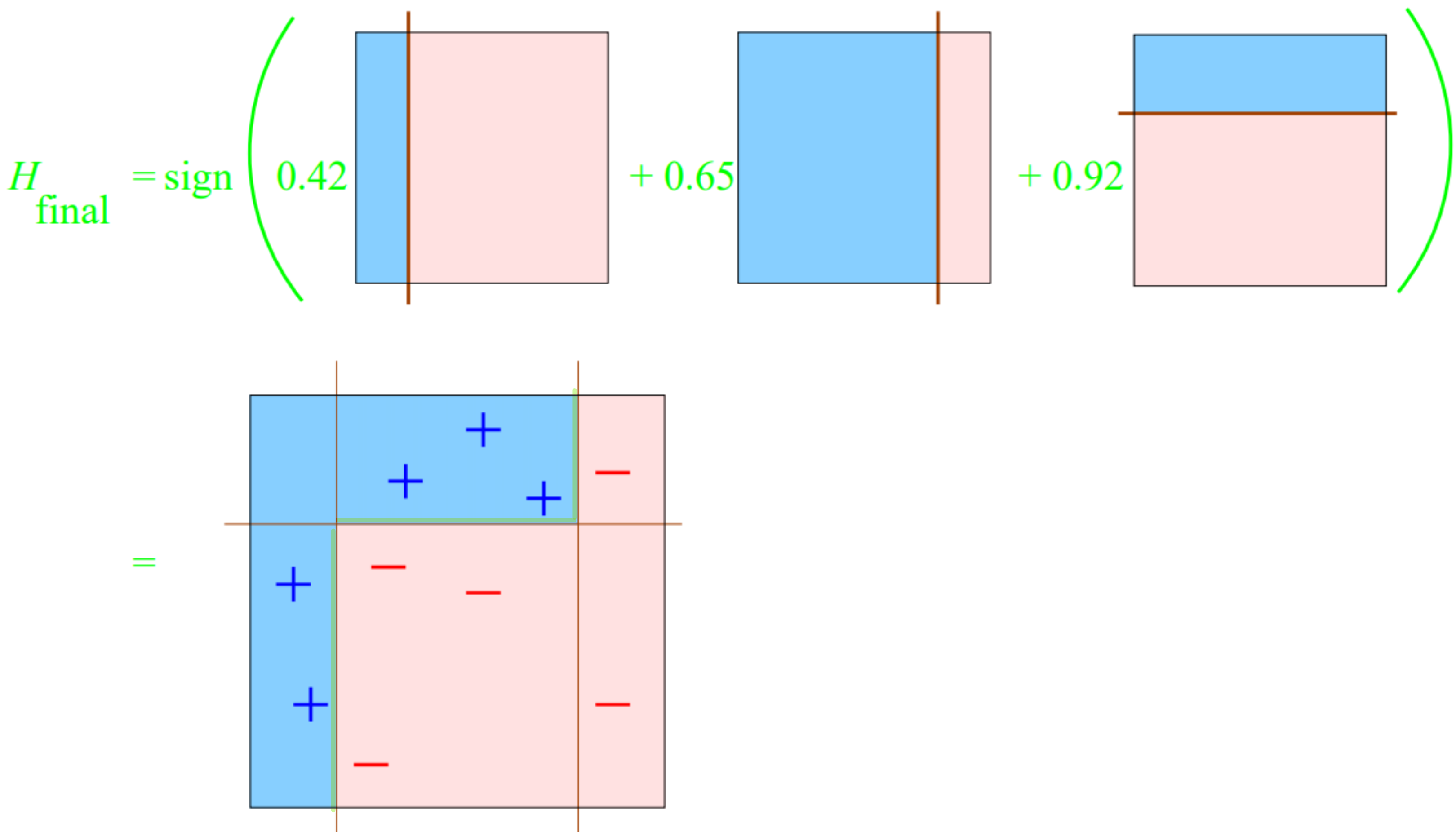
Boosting: AdaBoost

- Illustrative example 1
  - ✓ Round 3



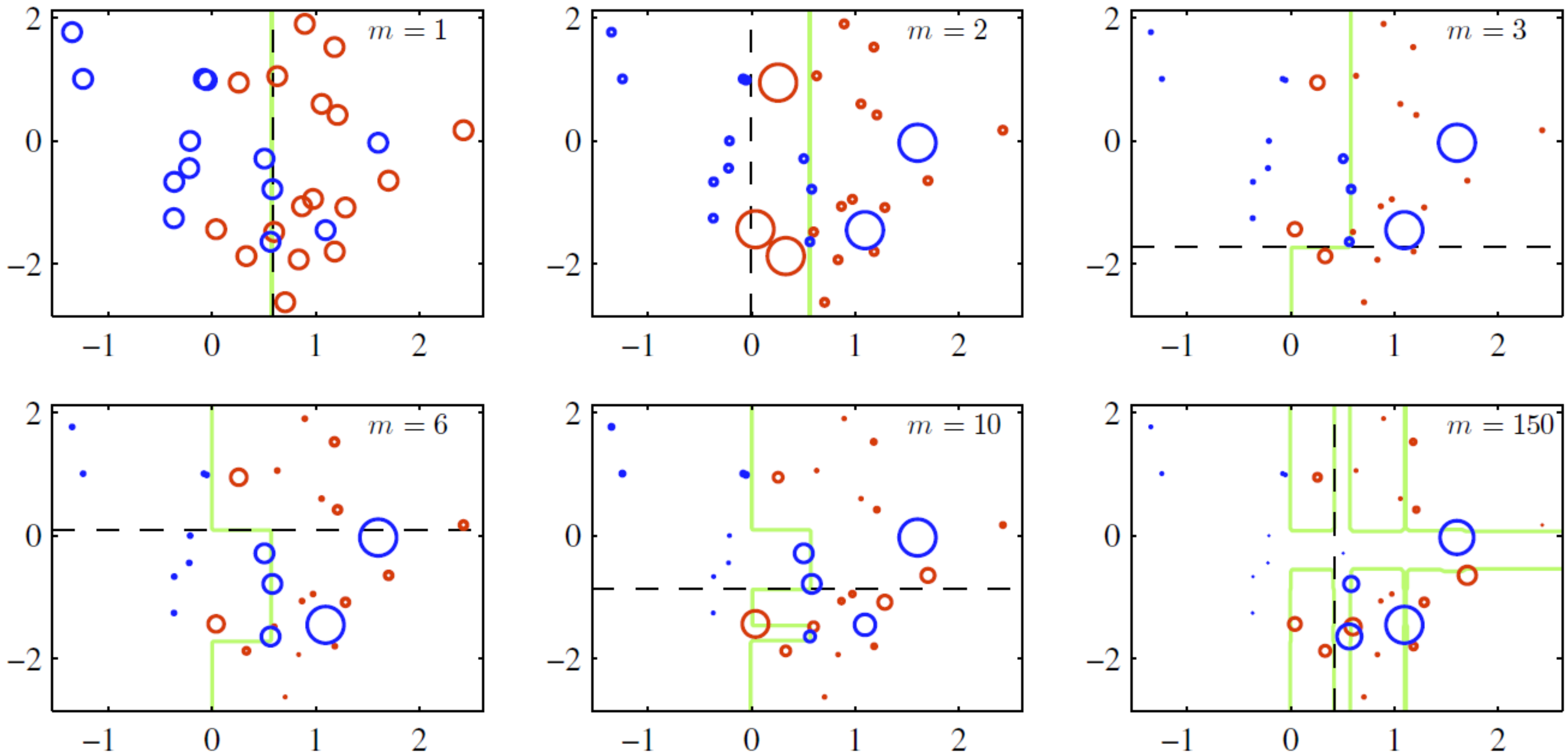
Boosting: AdaBoost

- Illustrative example 1
  - ✓ Final classifier



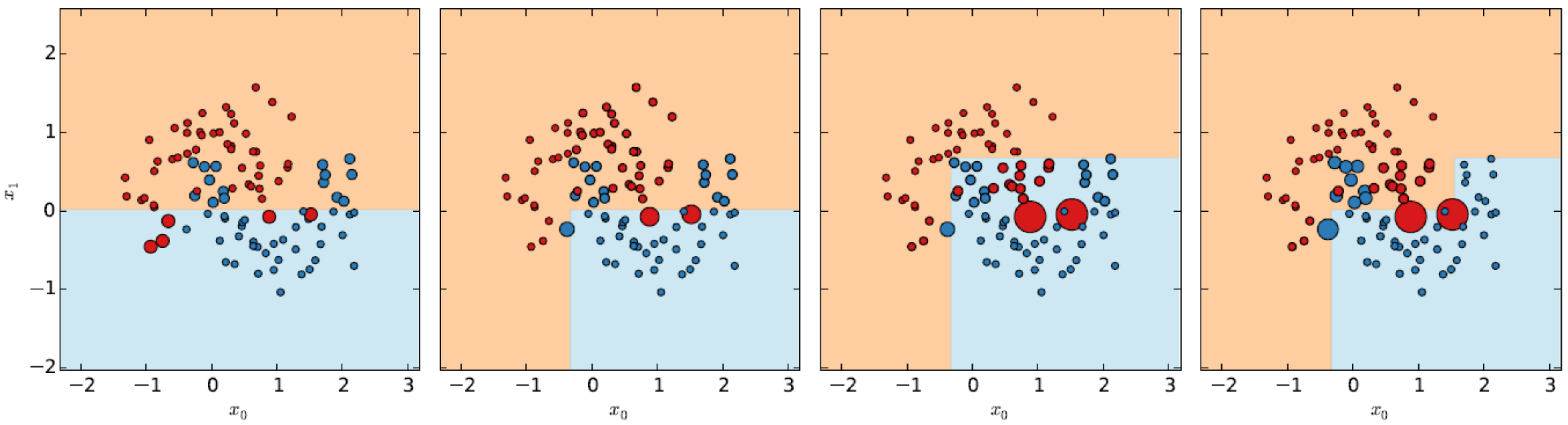
Boosting: AdaBoost

- Illustrative example 2



Boosting: AdaBoost

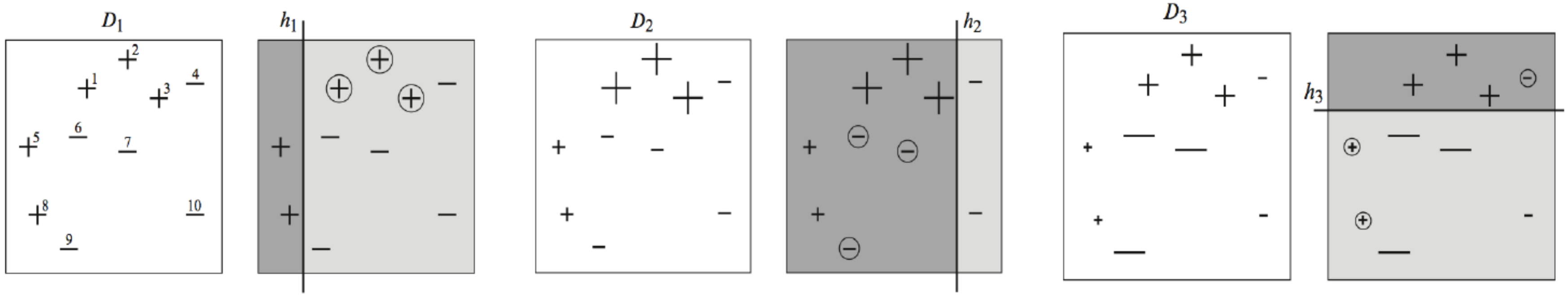
- Illustrative example 3

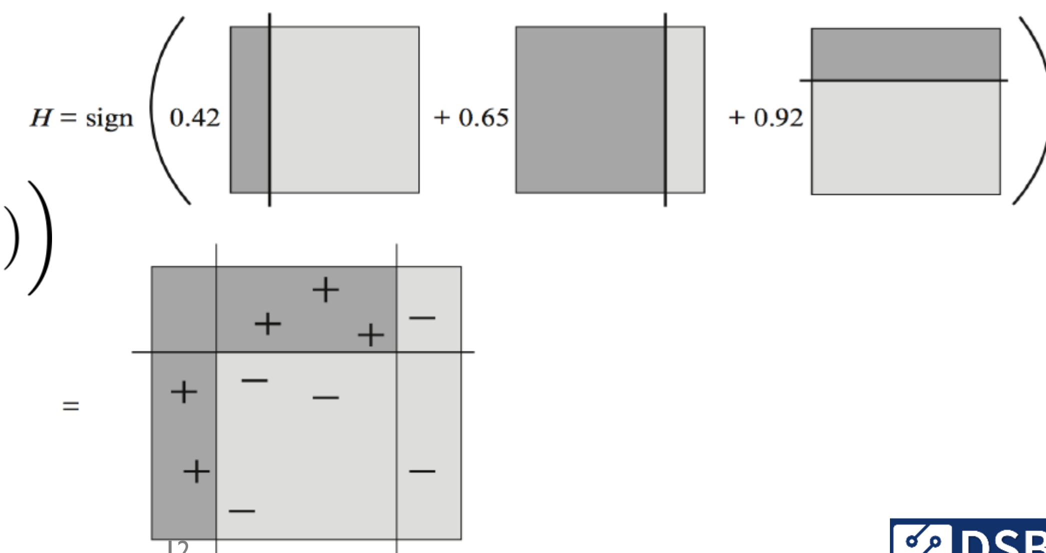


https://www.slideshare.net/DataRobot/gradient-boosted-regression-trees-in-scikitlearn?from_action=save

Boosting: AdaBoost

- Illustrative example 4





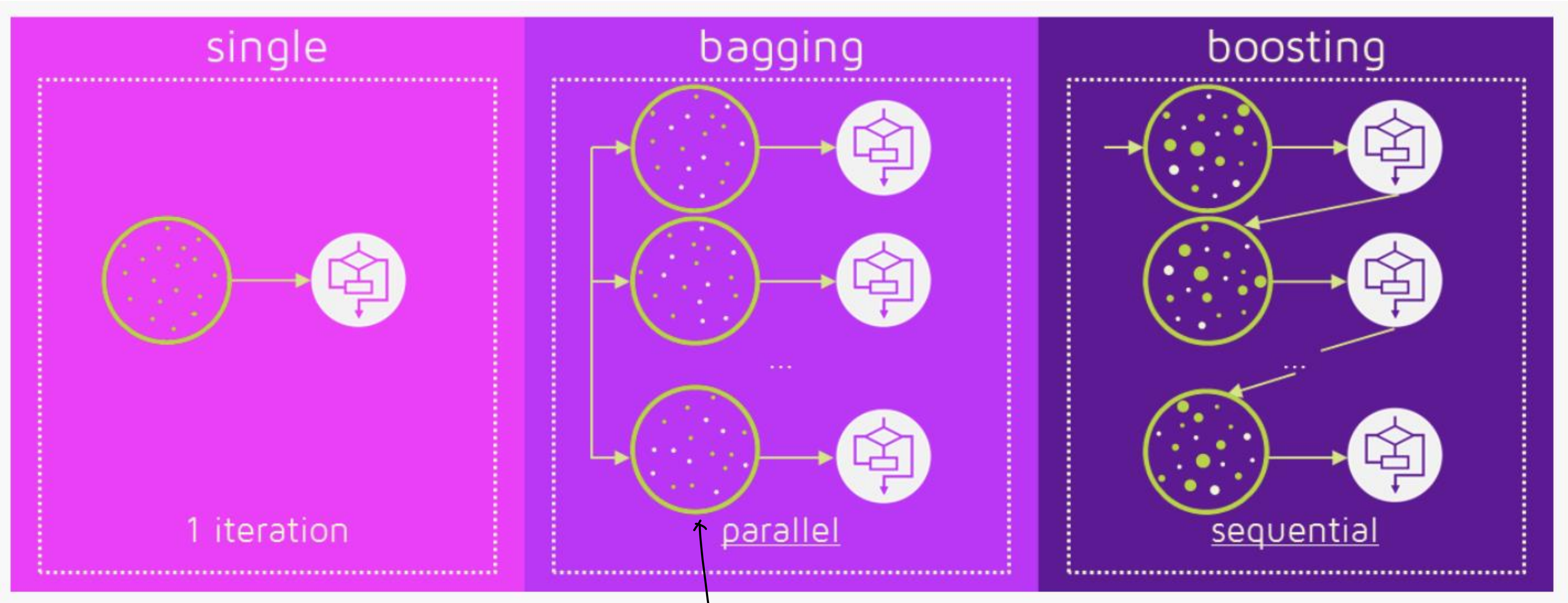
$$H(x') = sign\left(\sum_{t=1}^{T} \alpha_t h_t(x')\right)$$

고려대학교 KOREA UNIVERSITY

DSBA Data Science & Business Analytics

Boosting: AdaBoost

- Single model vs. Bagging vs. Boosting



https://quantdare.com/what-is-the-difference-between-bagging-and-boosting/

Uniform Distribution

복원 추출

Boosting: AdaBoost

- AdaBoost in Action

https://www.youtube.com/watch?v=k4G2VCuOMMg

Boosting: AdaBoost

- Bagging vs. Boosting
  - ✓ Selected instances in each training dataset

A sample of a single classifier on an imaginary set of data.	
	(Original) Training Set
Training-set-1:	1, 2, 3, 4, 5, 6, 7, 8

A sample of Bagging on the same data.	
	(Resampled) Training Set
Training-set-1:	2, 7, 8, 3, 7, 6, 3, 1
Training-set-2:	7, 8, 5, 6, 4, 2, 7, 1
Training-set-3:	3, 6, 2, 7, 5, 6, 2, 2
Training-set-4:	4, 5, 1, 4, 6, 4, 3, 8

→ random

A sample of Boosting on the same data.	
	(Resampled) Training Set
Training-set-1:	2, 7, 8, 3, 7, 6, 3, 1
Training-set-2:	1, 4, 5, 4, 1, 5, 6, 4
Training-set-3:	7, 1, 5, 8, 1, 8, 1, 4
Training-set-4:	1, 1, 6, 1, 1, 3, 1, 5

h_1, h_2, h_3, h_4는 모두
1번 example 정분류를 실패하고 있다

Boosting: AdaBoost

- Face detection with AdaBoost

https://www.youtube.com/watch?v=fCnV4MyI9-M


Q & A