Lecture 7-4: Ensemble Learning AdaBoost

Pilsung Kang

School of Industrial Management Engineering

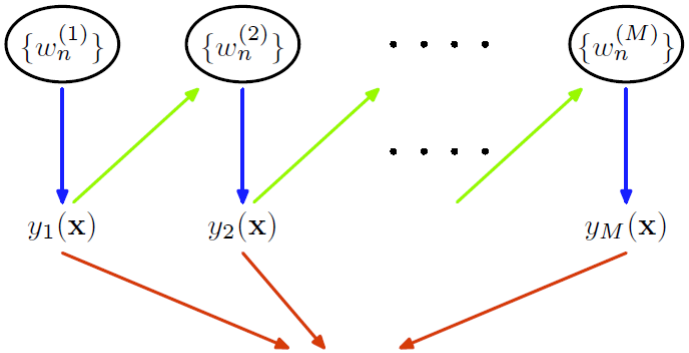
Korea University

Boosting: AdaBoost

- AdaBoosting: Idea
  - ✓ Strong model vs. Weak model
    - 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
    - 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: AdaBoost

- AdaBoosting: Idea
  - ✓ 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



$$Y_M(\mathbf{x}) = \text{sign}\left(\sum_m^M \alpha_m y_m(\mathbf{x})\right)$$

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.

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

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

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

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

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

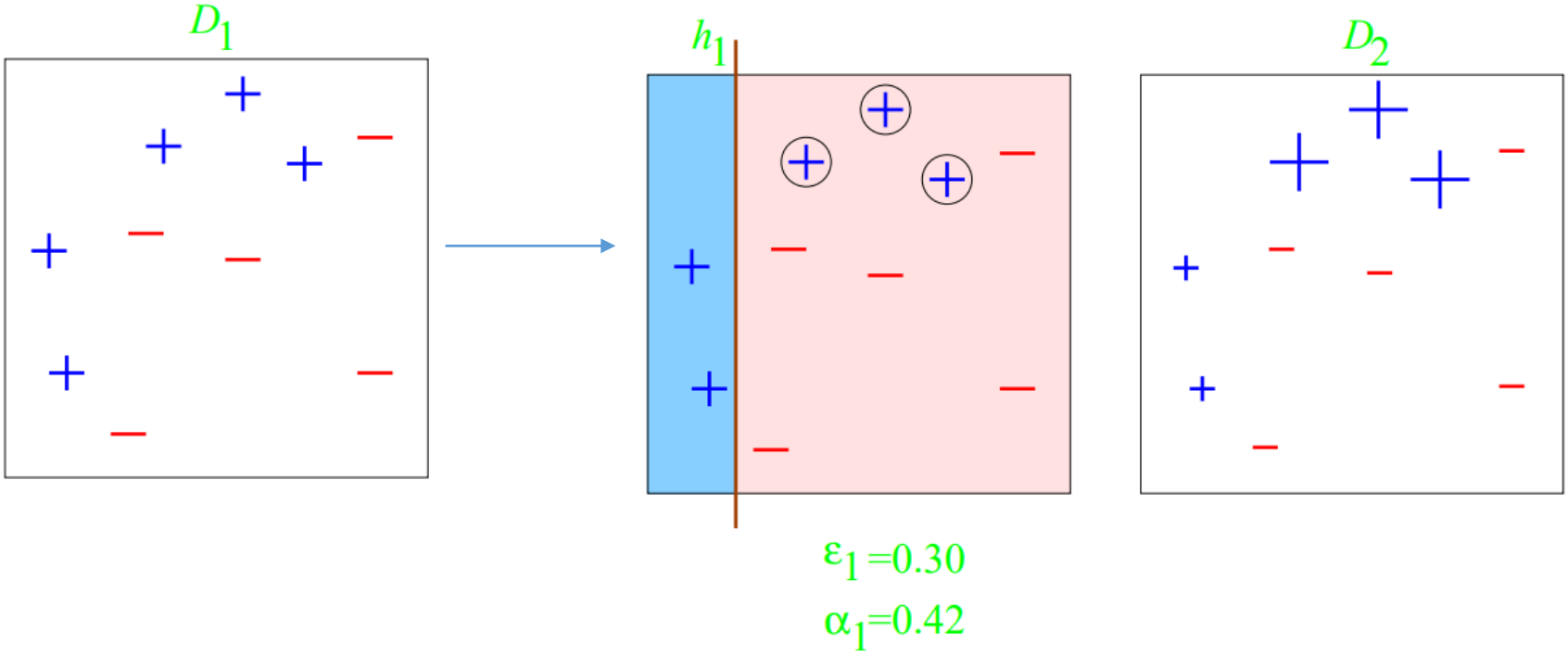
end for

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

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

Boosting: AdaBoost

- Illustrative example 1
  - ✓ Round 1



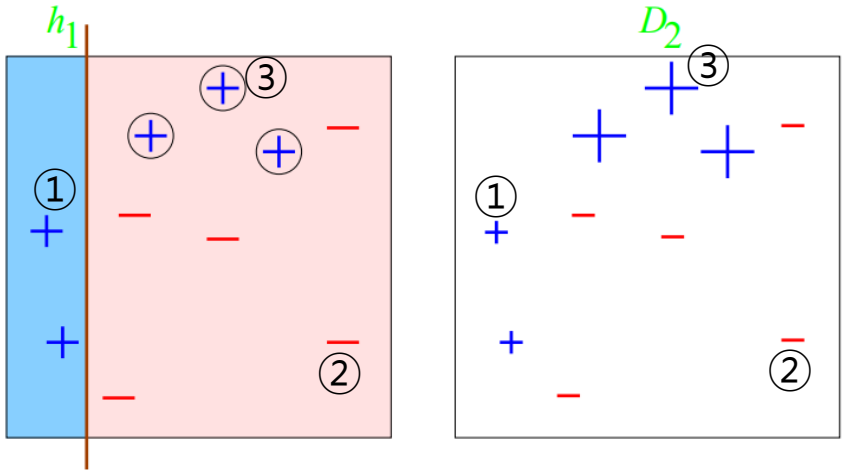
- 3 misclassifications out of 10: $\epsilon_i = 0.30$
- Model confidence: $\alpha_i = \frac{1}{2}\log\left(\frac{1-\epsilon_i}{\epsilon_i}\right) = \frac{1}{2}\log\frac{1-0.3}{0.3} = 0.42$

Boosting: AdaBoost

- AdaBoost Example
  - ✓ The selection probability of x_i for the next training dataset

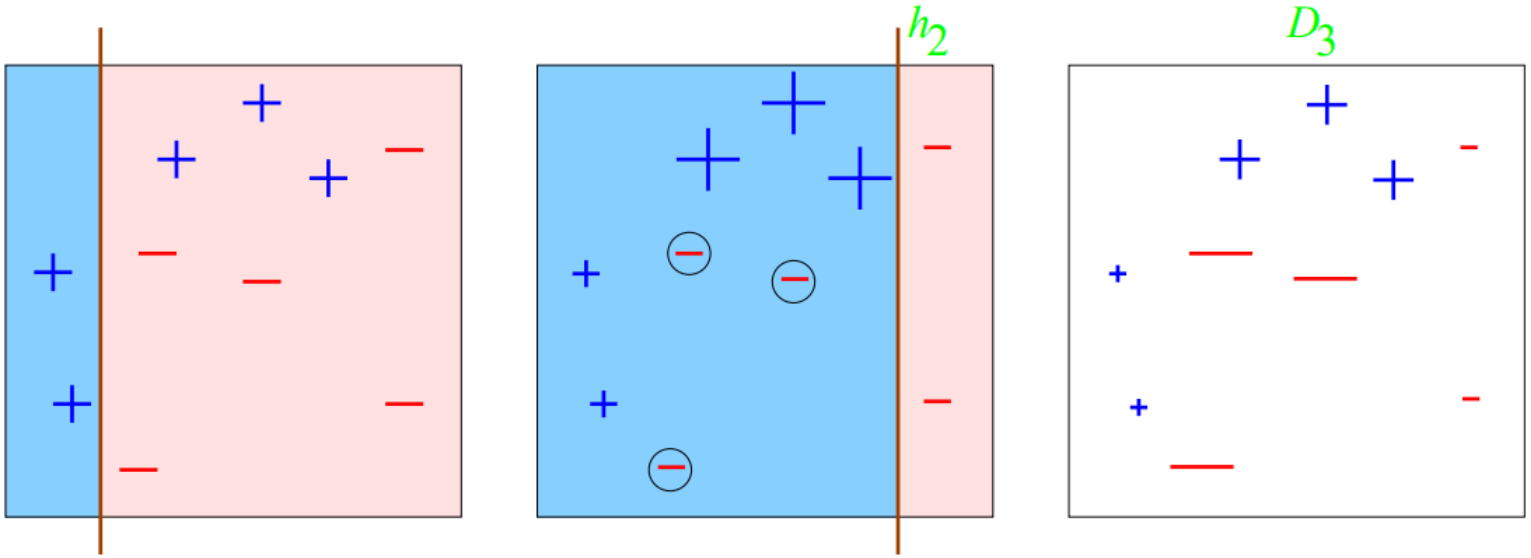
$$D_{t+1}(i) = \frac{D_t(i)\exp(-\alpha_t y_i h_t(x_i))}{Z_t}$$

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



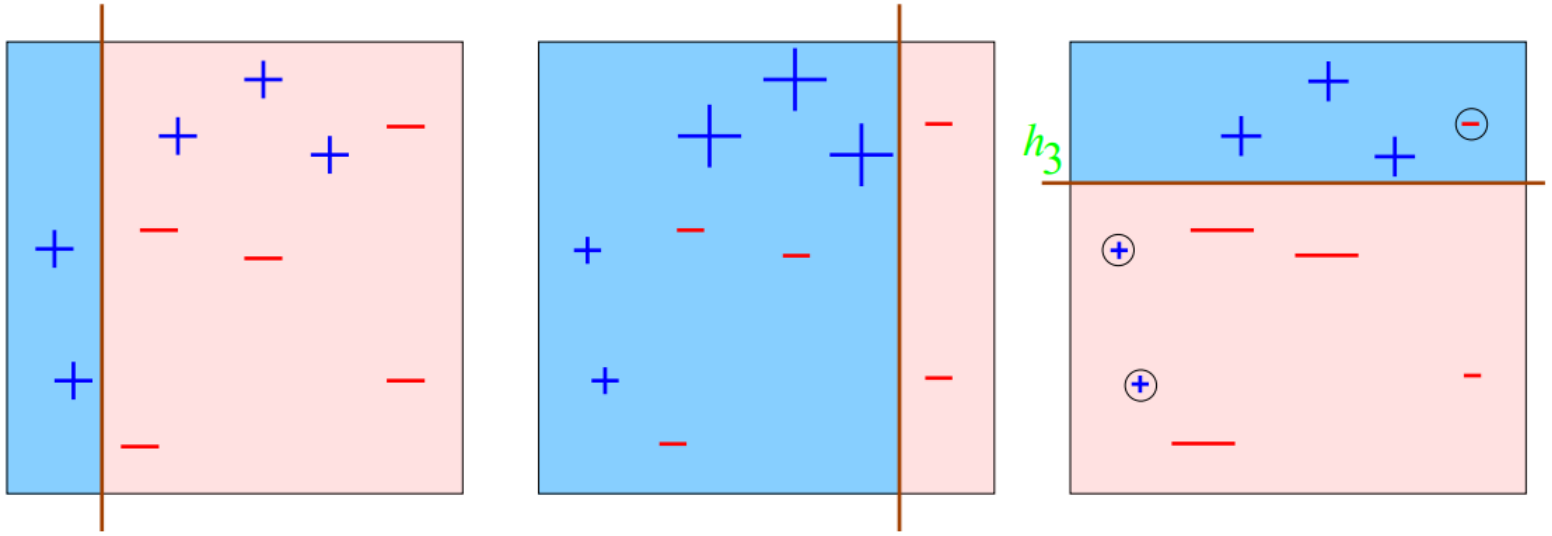
Boosting: AdaBoost

- Illustrative example 1
  - ✓ Round 2



Boosting: AdaBoost

- Illustrative example 1
  - ✓ Round 3

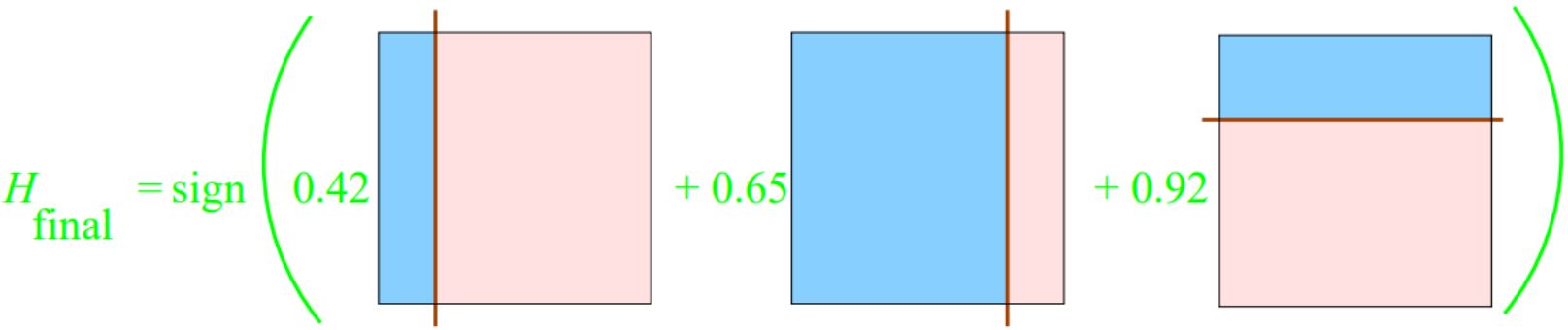


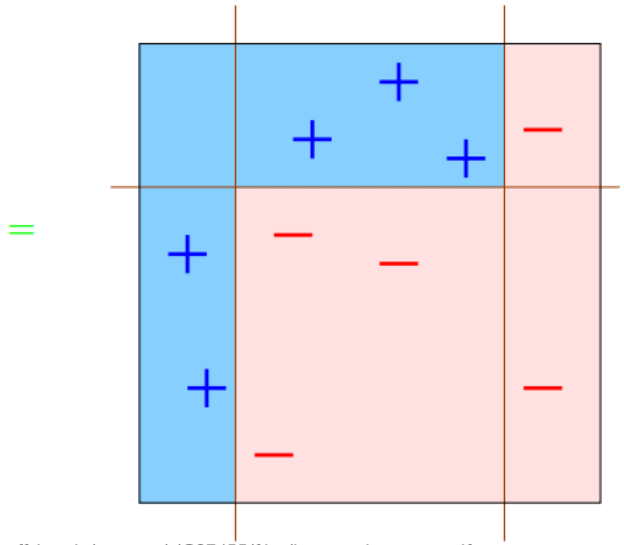
$\varepsilon_3 = 0.14$

$\alpha_3 = 0.92$

Boosting: AdaBoost

- Illustrative example 1
  - ✓ Final classifier

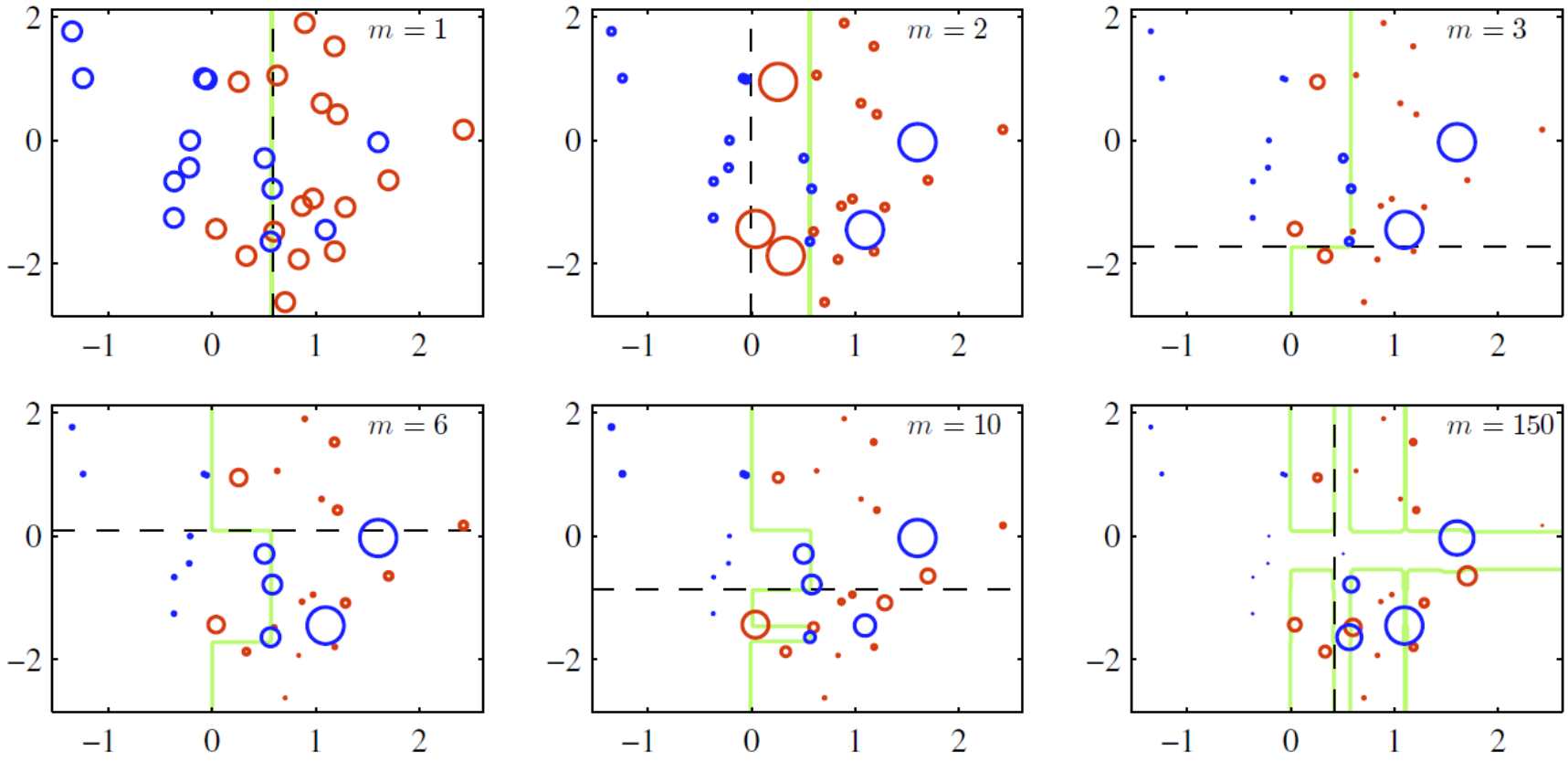




https://www.cse.buffalo.edu/~jcorso/t/CSE455/files/lecture_boosting.pdf

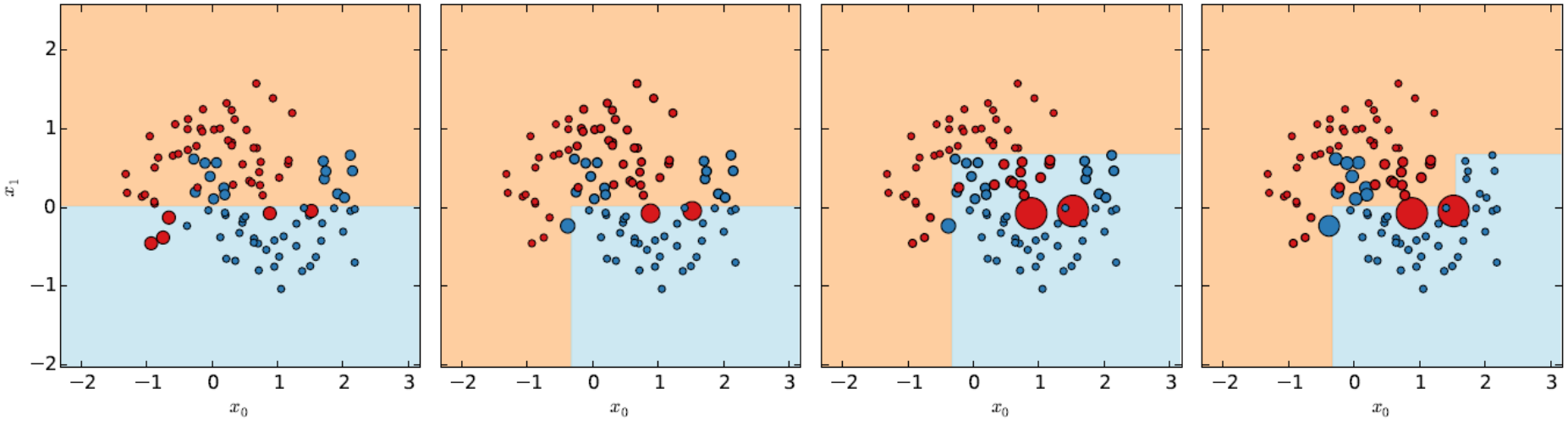
Boosting: AdaBoost

- Illustrative example 2



Boosting: AdaBoost

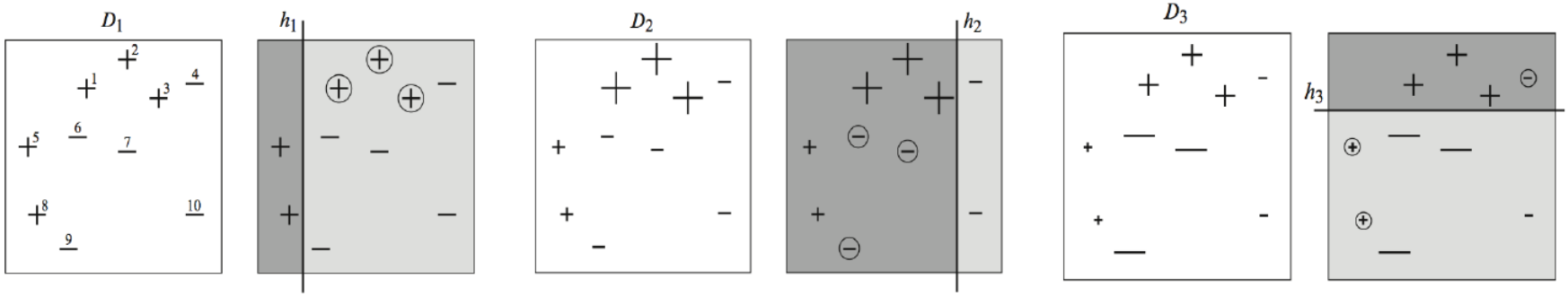
- Illustrative example 3



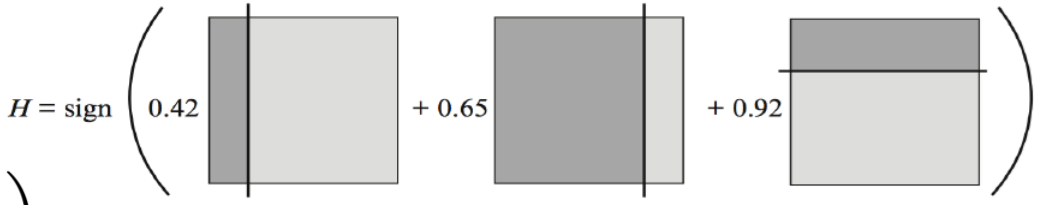
https://www.slideshare.net/DataRobot/gradient-boosted-regression-trees-in-scikitlearn?from_action=save

Boosting: AdaBoost

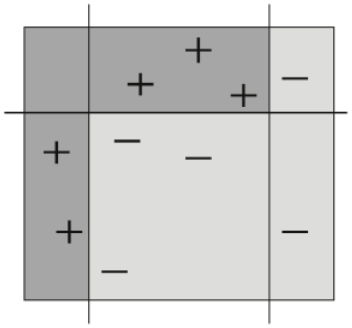
- Illustrative example 4



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

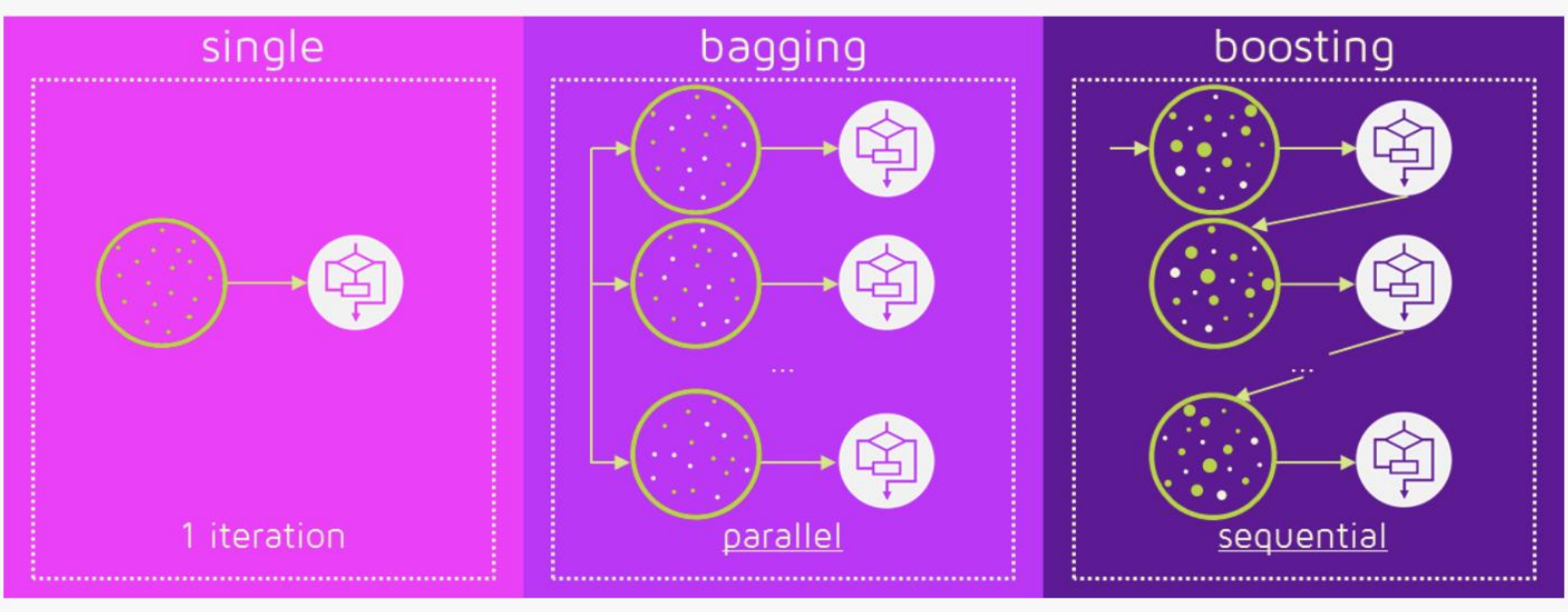


=



Boosting: AdaBoost

- Single model vs. Bagging vs. Boosting



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

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

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

Boosting: AdaBoost

- Face detection with AdaBoost

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


Q & A