# Adaboost

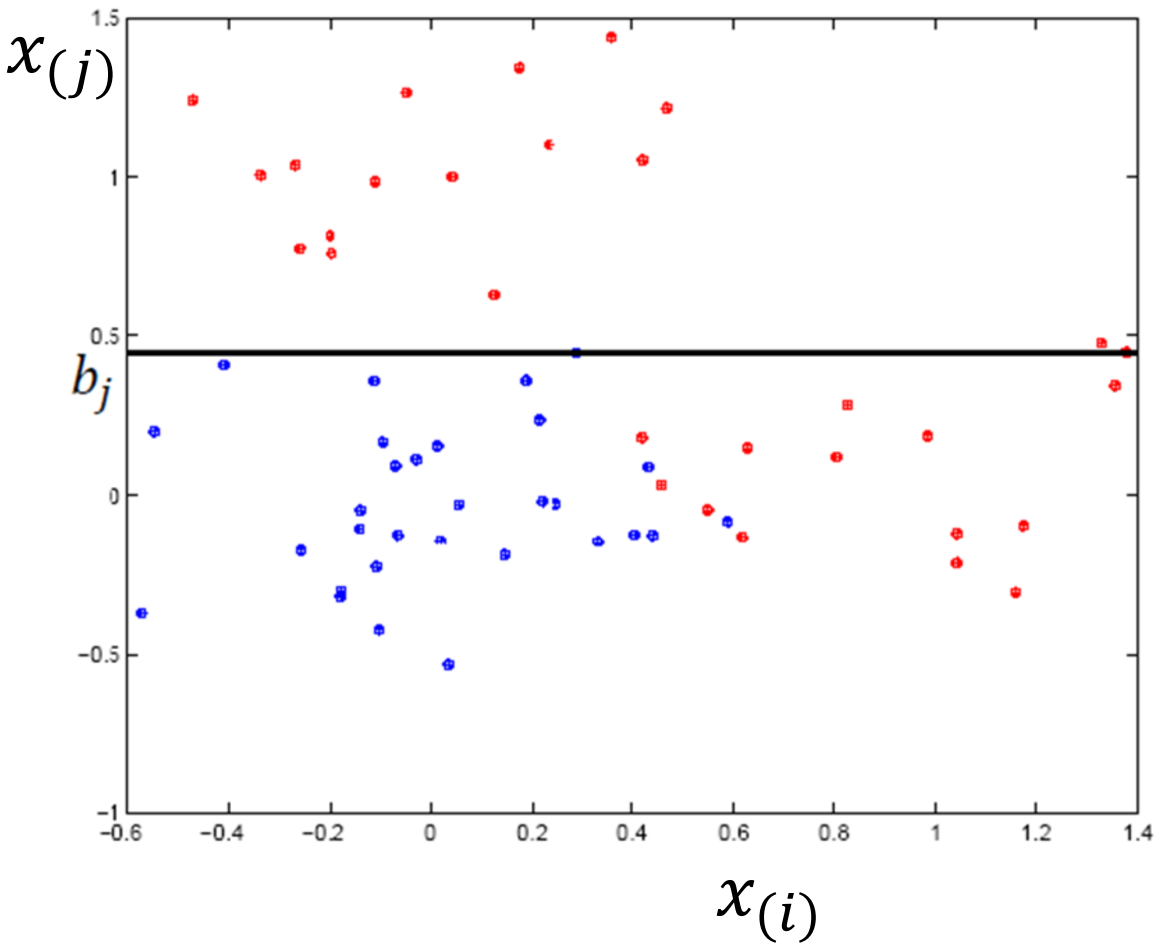
Liang Liang

# Decision stump - a very simple/weak model

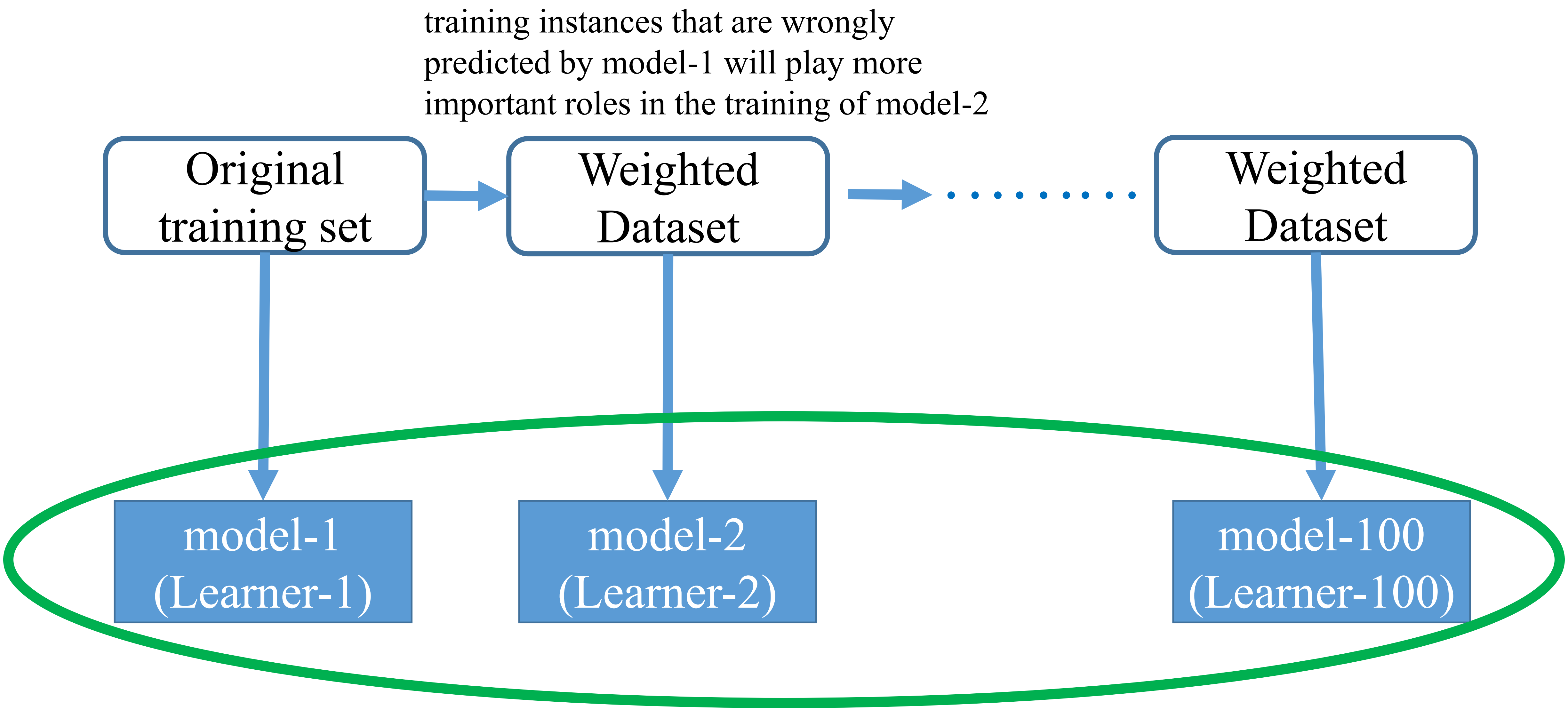
- $y \in \{-1, +1\}$ is class label
- $x_{(j)}$ is a feature in the feature vector $x$
- A decision stump

$$h(x) = sign(w_{(j)} x_{(j)} + b_j)$$

- A decision stump pays attention to only a single dimension/ feature of the input feature vector



# Adaboost flow chart




The final model = Weighted combination

# Adaboost(classification): the learning algorithm

- iteration $t=0$: initialize the weight of each data point $D_t(n)=\frac{1}{N}$, n=1,…,$N$
- Given $D_t(n)$ at iteration $t$
  - find a decision stump $h_t$ to minimize the classification error

$$\epsilon=\frac{1}{N}\sum_{n=1}^{N} D_t(n) I(y_n \neq h_t(x_n))$$

$I(y_n \neq h_t(x_n))=1$ if $y_n \neq h_t(x_n)$ ~ prediction is wrong

$I(y_n \neq h_t(x_n))=0$ if $y_n = h_t(x_n)$ ~ prediction is correct

  - update the weight of every training data point

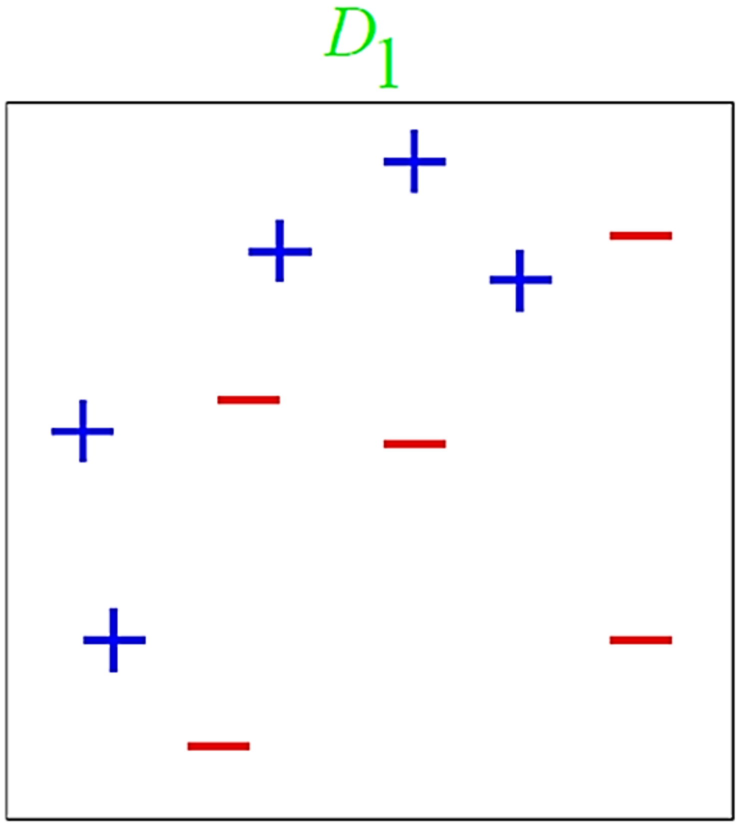
$$D_{t+1}(n)=\frac{D_t(n)}{Z}\times\begin{cases} e^{-\alpha_t}, & if\ y_n = h_t(x_n) \\ e^{\alpha_t}, & if\ y_n \neq h_t(x_n) \end{cases}$$

$\alpha_t=\frac{1}{2}log\left(\frac{1-\epsilon}{\epsilon}\right)>0,\ Z$ is normalization constant

- After many iterations, we get the final classifier: $H_{final}(x)=sign(\sum_t \alpha_t h_t(x))$

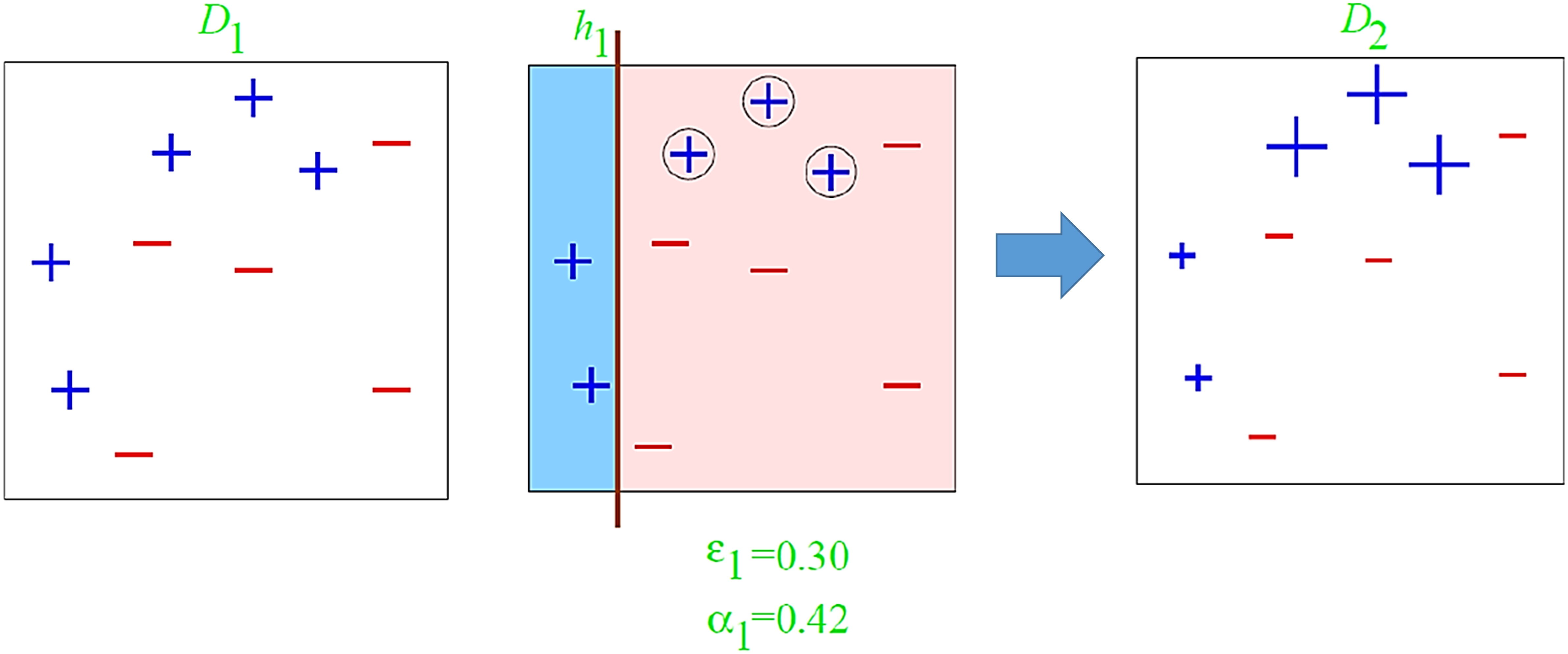
# Toy Example: initialization

- Weak classifiers are decision stumps
- $y \in \{-1, +1\}$ is class label
- Initialize the uniform weights on all data points



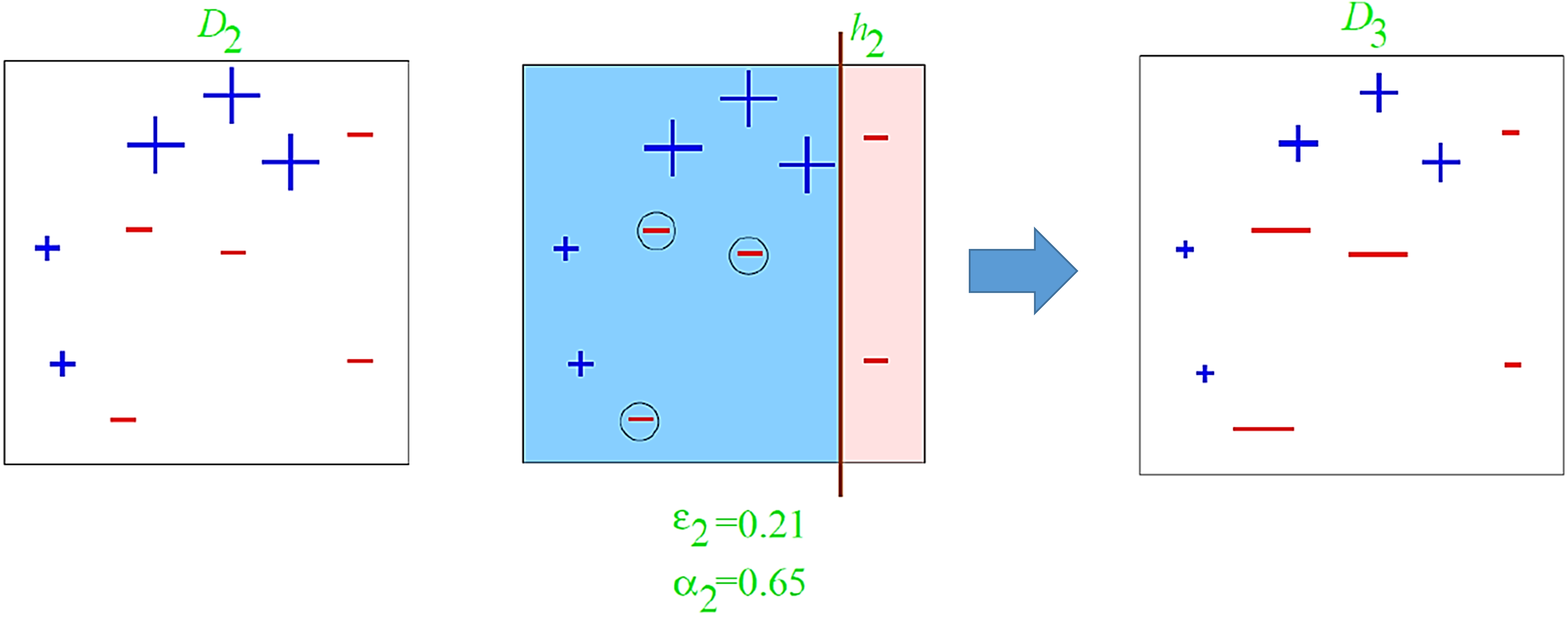
# Toy Example: boosting round 1

- Choose a decision stump (weak classifier)
- Some data points get higher weights because they are misclassified



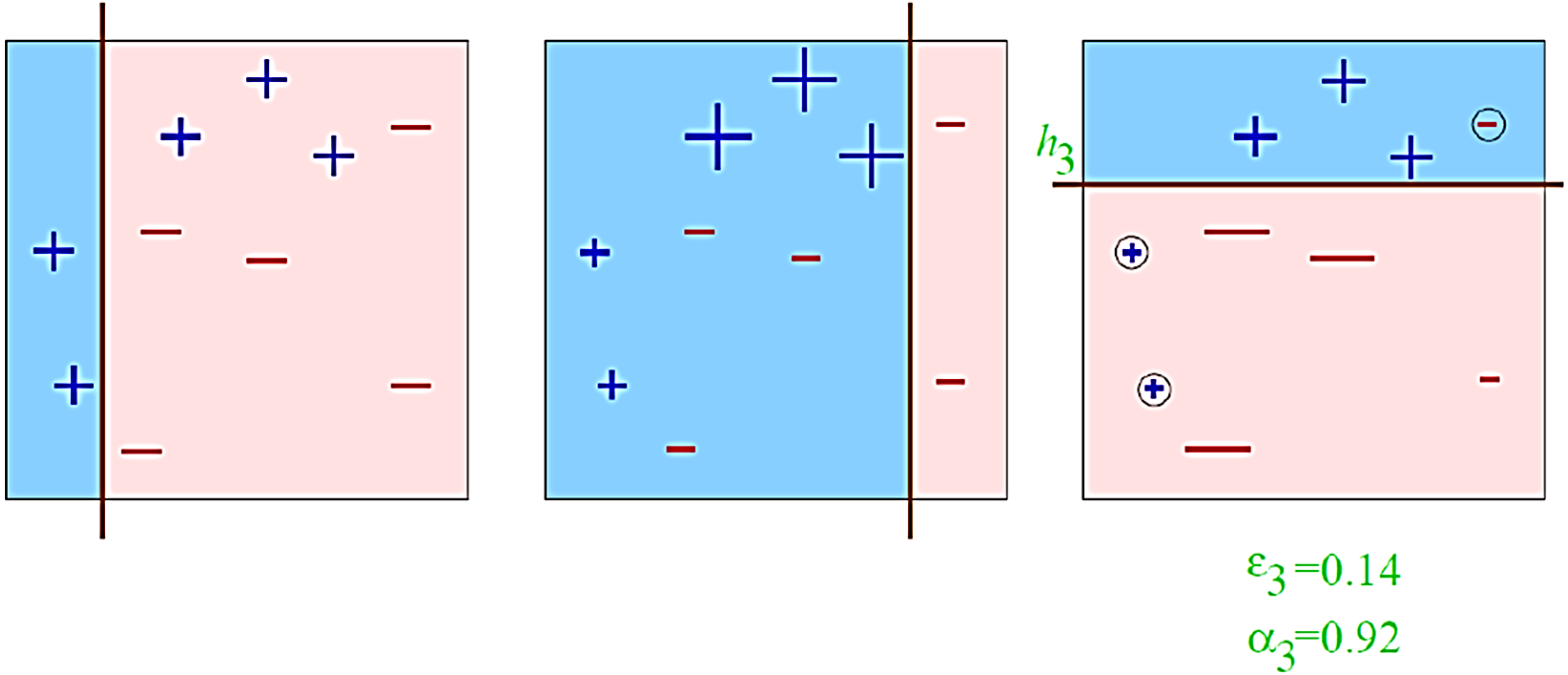
# Toy Example: boosting round 2

- Choose a decision stump (weak classifier)
- Reweight again. increase weights on the misclassified samples



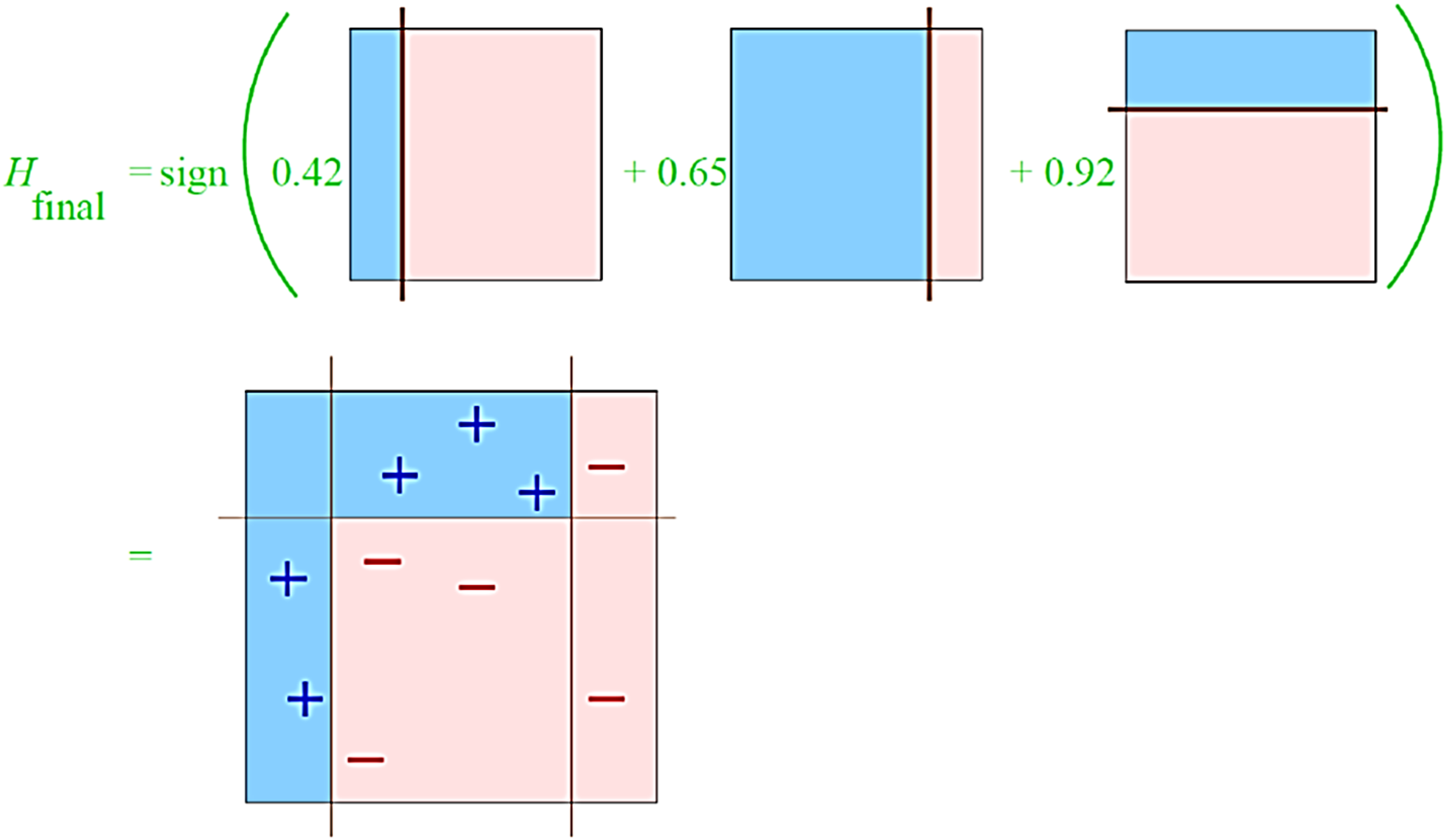
# Toy Example: boosting round 3

- Repeat the same process
- Now we have 3 classifiers



# Toy Example: the final classifier

- Final classifier is the weighted combination of the three weak classifiers



# Adaboost(classification): the loss function

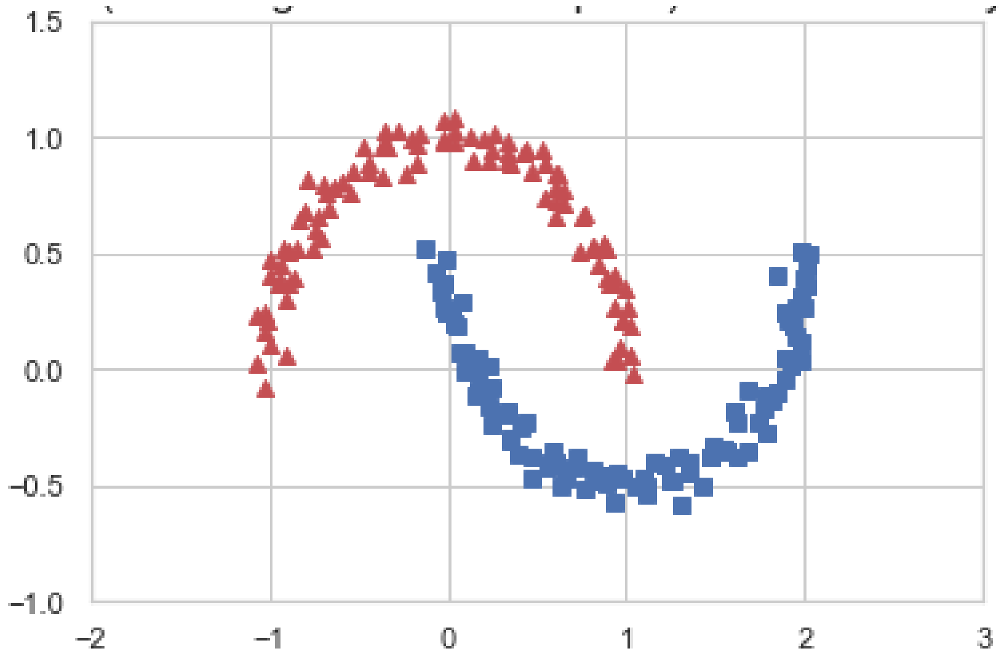
- The combined classifier

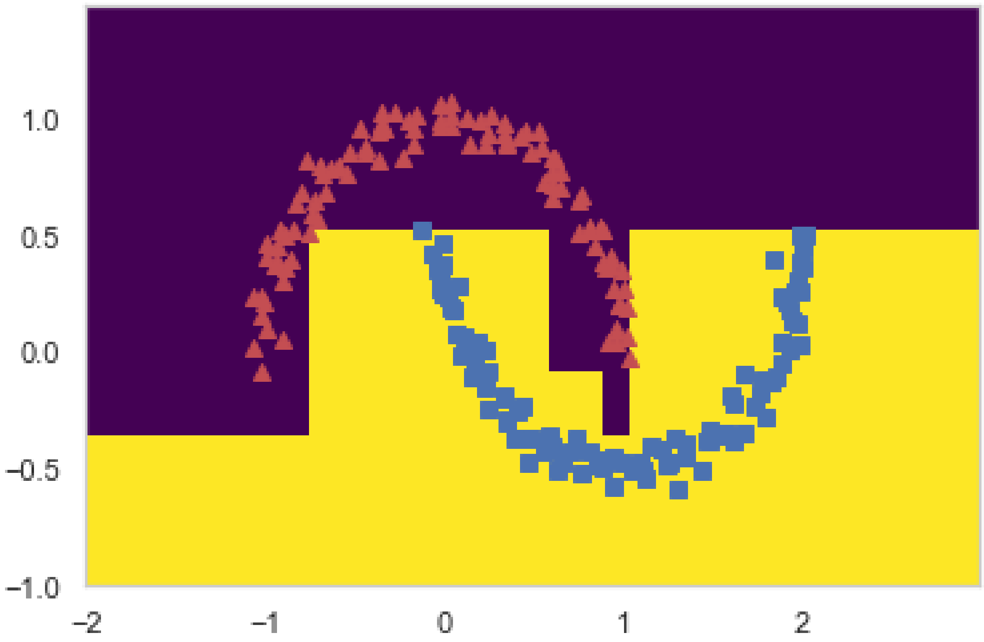
$$f_t(x) = \alpha_1 h_1(x) + \cdots + \alpha_t h_t(x)$$

- Exponential loss

$$L = \frac{1}{N}\sum_{n=1}^{N} exp\left(-y_n f_t(x_n)\right)$$

$$= \frac{1}{N}\sum_{n=1}^{N} \underbrace{exp\left(-y_n f_{t-1}(x_n)\right)}_{D_t(n)} exp\left(-y_n \alpha_t h_t(x)\right)$$






AdaBoost.ipynb