CS 360: Machine Learning

Prof. Sara Mathieson

Fall 2019

HAVERFORD
COLLEGE

Outline for October 24

- Ensemble methods
  - Bagging
  - Random forests
  - Boosting
  - Weighted entropy

In-lab today:
Hand back exam

Outline for October 24

- Ensemble methods
  - Bagging
  - Random forests
  - Boosting
  - Weighted entropy

Bagging

Train

- use bootstrap resampling to generate $X^{(t)}$ for $t = 1 \cdots T$
- train classifier on $X^{(t)}$ to get $h^{(t)}$

Test

$h^{(t)}(\vec{x}) \in \{0,1\}$

for x in test data:

voting

$$h(x) = \underset{y \in \{0,1\}}{\operatorname{argmax}} \sum_{t=1}^{T} \mathbb{1}(h^{(t)}(\vec{x}) = y)$$

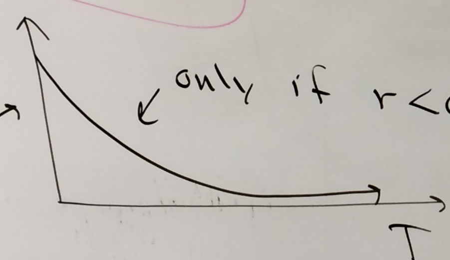
r = prob base classifier is wrong.

R = # of votes for the wrong class.

$$P(R=k) = \binom{T}{k} r^k (1-r)^{T-k}$$

$$\binom{T}{k} = \frac{T!}{k!\,(T-k)!}$$

$$\underbrace{P\left(R > \frac{T}{2}\right)}_{\text{overall wrong!}} = \sum_{k=\frac{T+1}{2}}^{T} \binom{T}{k} r^k (1-r)^{T-k}$$



Example $T=3$, $r=\frac{1}{4}$ $\vec{X}_{test} = [\,.\,.\,.\,.\,.\,]$

$Y_{test} = 1$ ← true

$h^{(1)}$	$h^{(2)}$	$h^{(3)}$	prob
0	0	0	$\frac{1}{4}\cdot\frac{1}{4}\cdot\frac{1}{4} = \left(\frac{1}{4}\right)^3$
0	0	1	$\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)$
0	1	0	$\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)$
0	1	1	
1	0	0	$\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)$
1	0	1	
1	1	0	
1	1	1	

$$P\left(R > \frac{T}{2}\right) = \left(\frac{1}{4}\right)^3 + 3\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right) \approx \boxed{0.16}$$

lower error!

Outline for October 24

- Ensemble methods
  - Bagging
  - Random forests
  - Boosting
  - Weighted entropy

Random Forests

- Idea: choose a different subset of features for every classifier *t*
- Typically use *decision stumps* (depth 1)

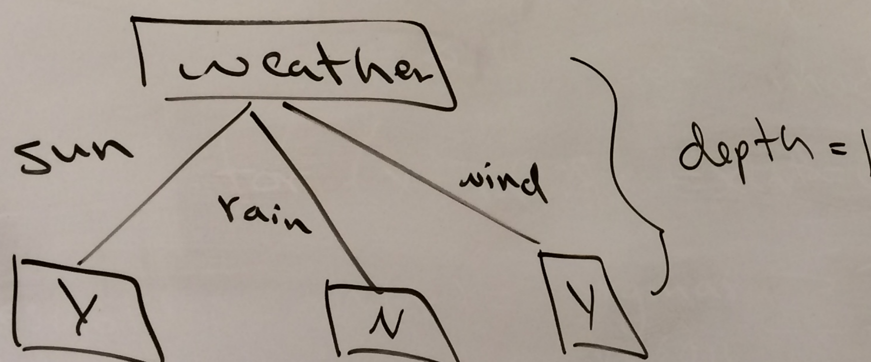
- Goal: decorrelate models

- In practice: choose sqrt(p) features
  - Without replacement for each model
  - Every model: data points and features chosen independently

Random Forests

* idea: need models to be decorrelated!

* choose base classifier ("weak") a **decision stump**



choose subset of features!

* in practice: $\sqrt{p}$ features w/o replacement.

=> decorrelate stumps!

* generate random subset of features independently across models.

Outline for October 24

- Ensemble methods
  - Bagging
  - Random forests
  - Boosting
  - Weighted entropy

Boosting Idea

Train:

- equal weights on all examples to start. ($\rightarrow \frac{1}{n}$)
- for T iterations:
  - learn classifier using weighted examples.
  - change example weights based on train errors.

Test:

- get predictions from T classifiers
- vote based on how well that classifier did during training.

AdaBoost (Adaptive Boosting)

$y \in \{-1, 1\}$

set $w_i^{(1)} = \frac{1}{n}$ for $i = 1 \cdots n$

for $t = 1 \cdots T$: — # of classifiers.

(t: classifier index)

ⓐ fit classifier to weighted training set $\Rightarrow h^{(t)}(\vec{x})$

ⓑ compute weighted classification error.

$$\varepsilon_t = \sum_{i=1}^{n} w_i^{(t)} \mathbb{1}\left(y_i \neq h^{(t)}(\vec{x}_i)\right)$$

ⓒ compute model score.

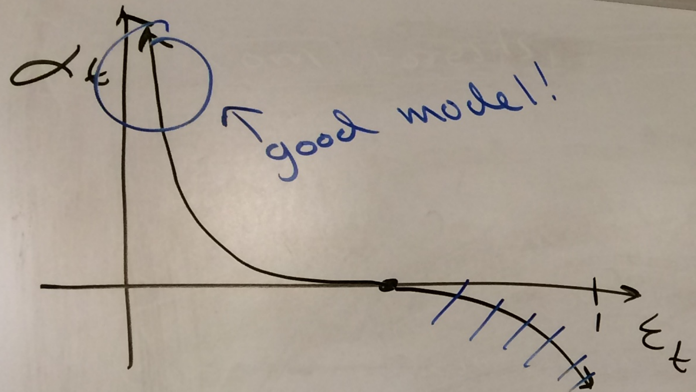
$$\alpha_t = \frac{1}{2} \log\left(\frac{1 - \varepsilon_t}{\varepsilon_t}\right)$$

ⓓ update weights

$$w_i^{(t+1)} = c_t \, w_i^{(t)} \exp\left(-\alpha_t y_i h^{(t)}(\vec{x}_i)\right)$$

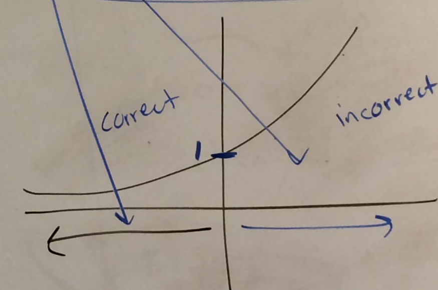
c_t: normalizer. Want weights to sum to 1

$\exp(\ldots)$: less than 1 if correct. else > 1



$$\log\left(\frac{1-\frac{1}{2}}{\frac{1}{2}}\right) = \log(1) = 0$$

$$\begin{cases} e^{-\alpha_t} & \text{if } y_i = h^{(t)}(\vec{x}_i) \\ e^{\alpha_t} & \text{if } y_i \neq h^{(t)}(\vec{x}_i) \end{cases}$$

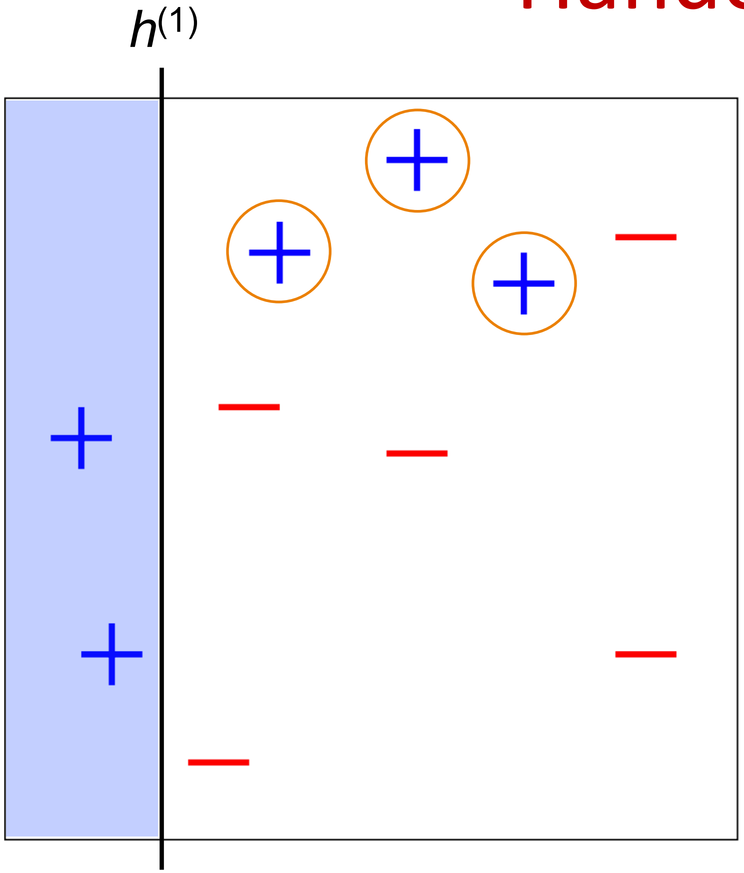


$$\text{Sign}(x) = \begin{cases} +1 & \text{if } x > 0 \\ -1 & \text{if } x \leq 0 \end{cases}$$

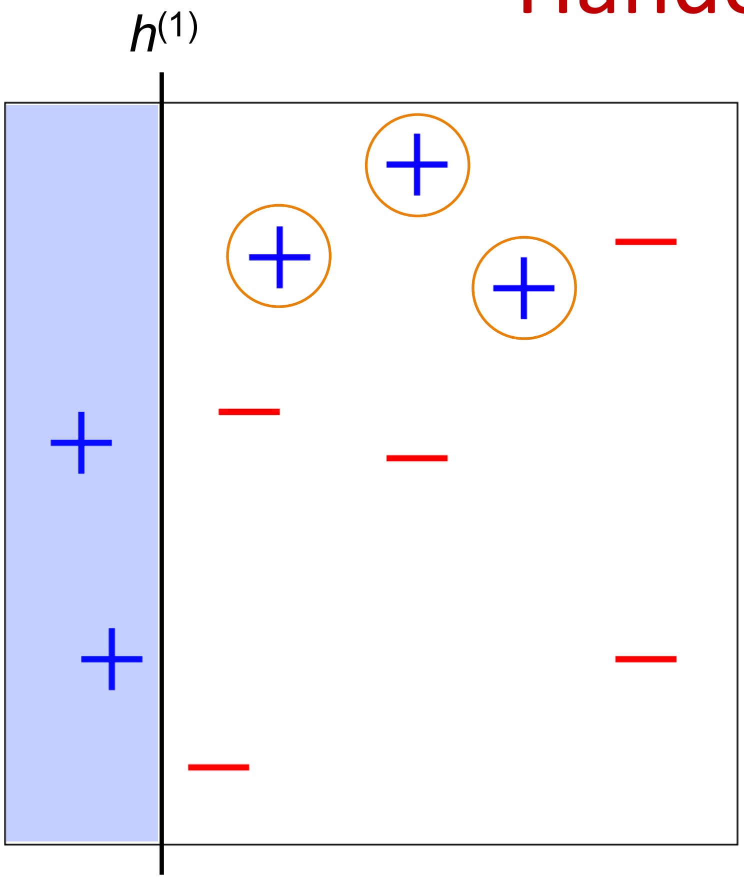
$$\underbrace{h(\vec{x})}_{\text{overall model}} = \text{sign}\left(\sum_{t=1}^{T} \alpha_t h^{(t)}(\vec{x})\right)$$

weighted vote

Handout 14: Round 1



Handout 14: Round 1



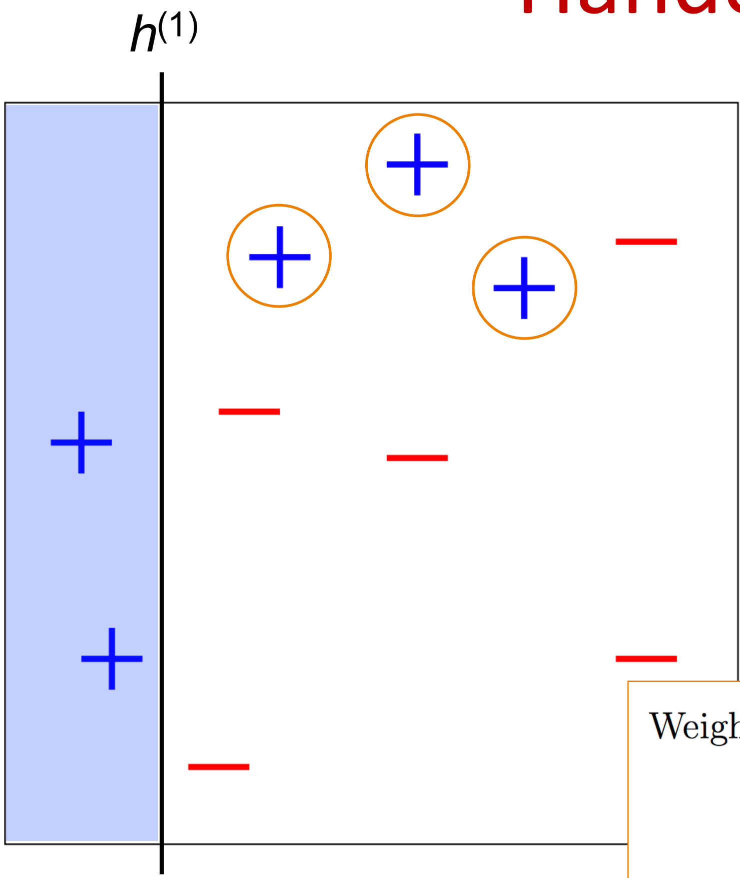
$w_i^{(1)} = \frac{1}{10}$ for all $i = 1, 2, \cdots, 10.$

$\epsilon_1 = \frac{3}{10}$ (three points incorrectly classified, all with weight $\frac{1}{10}$)

$\alpha_1 = \frac{1}{2} \ln\left(\frac{1-\frac{3}{10}}{\frac{3}{10}}\right) = \ln\sqrt{\frac{7}{3}} \approx 0.42$

- correctly classified: $w_i^{(2)} = c_1 \cdot \frac{1}{10} \exp\left(-\ln\sqrt{\frac{7}{3}}\right)$
- incorrectly classified: $w_i^{(2)} = c_1 \cdot \frac{1}{10} \exp\left(\ln\sqrt{\frac{7}{3}}\right)$

Handout 14: Round 1



$w_i^{(1)} = \frac{1}{10}$ for all $i = 1, 2, \cdots, 10$.

$\epsilon_1 = \frac{3}{10}$ (three points incorrectly classified, all with weight $\frac{1}{10}$)

$\alpha_1 = \frac{1}{2} \ln \left(\frac{1 - \frac{3}{10}}{\frac{3}{10}} \right) = \ln \sqrt{\frac{7}{3}} \approx 0.42$

- correctly classified: $w_i^{(2)} = c_1 \cdot \frac{1}{10} \exp \left(- \ln \sqrt{\frac{7}{3}} \right)$
- incorrectly classified: $w_i^{(2)} = c_1 \cdot \frac{1}{10} \exp \left(\ln \sqrt{\frac{7}{3}} \right)$

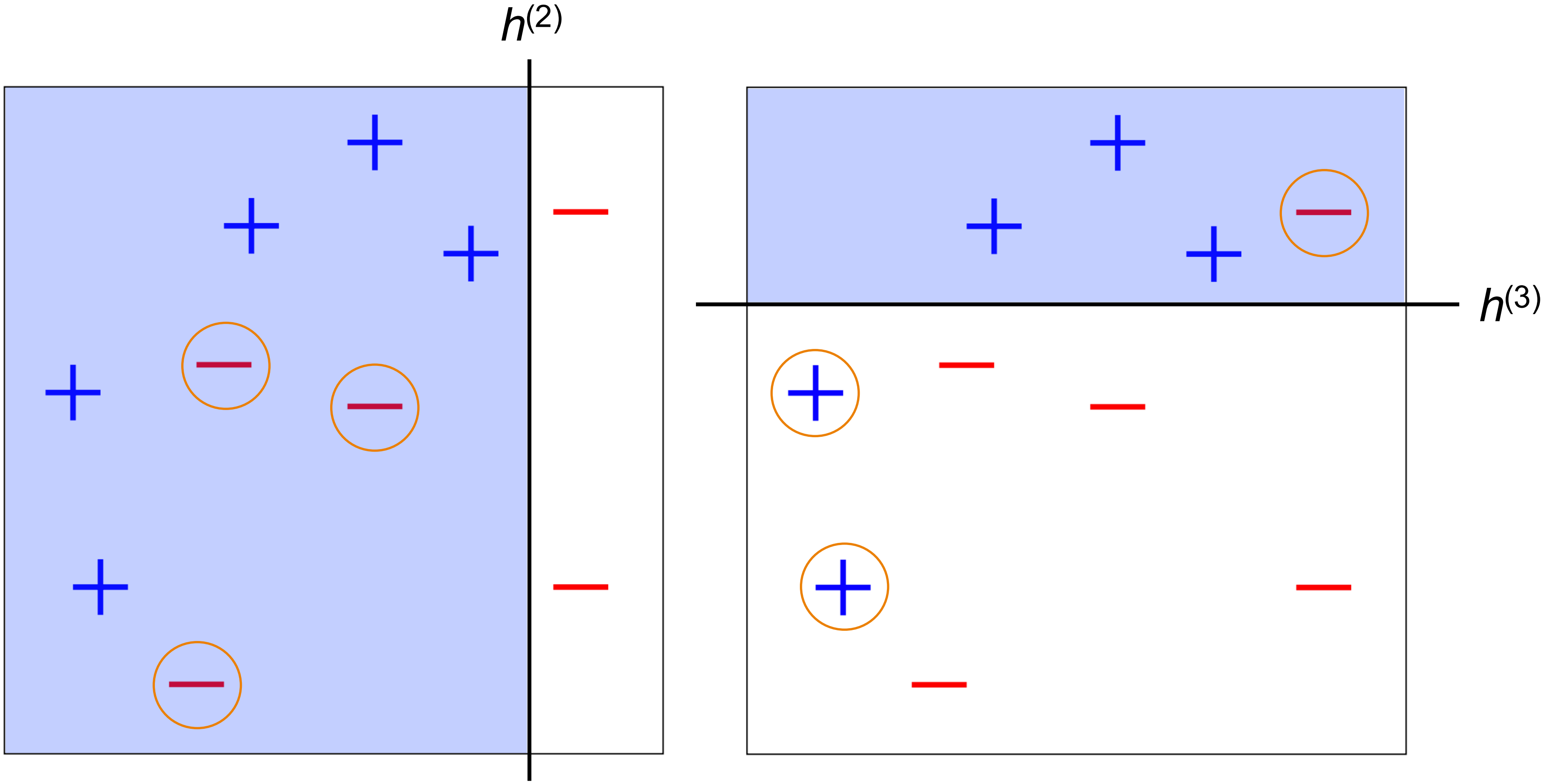
Weights must sum to 1, $\Rightarrow$

$$7 \cdot \frac{c_1}{10} \exp \left(- \ln \sqrt{\frac{7}{3}} \right) + 3 \cdot c_1 \cdot \frac{1}{10} \exp \left(\ln \sqrt{\frac{7}{3}} \right) = 1$$

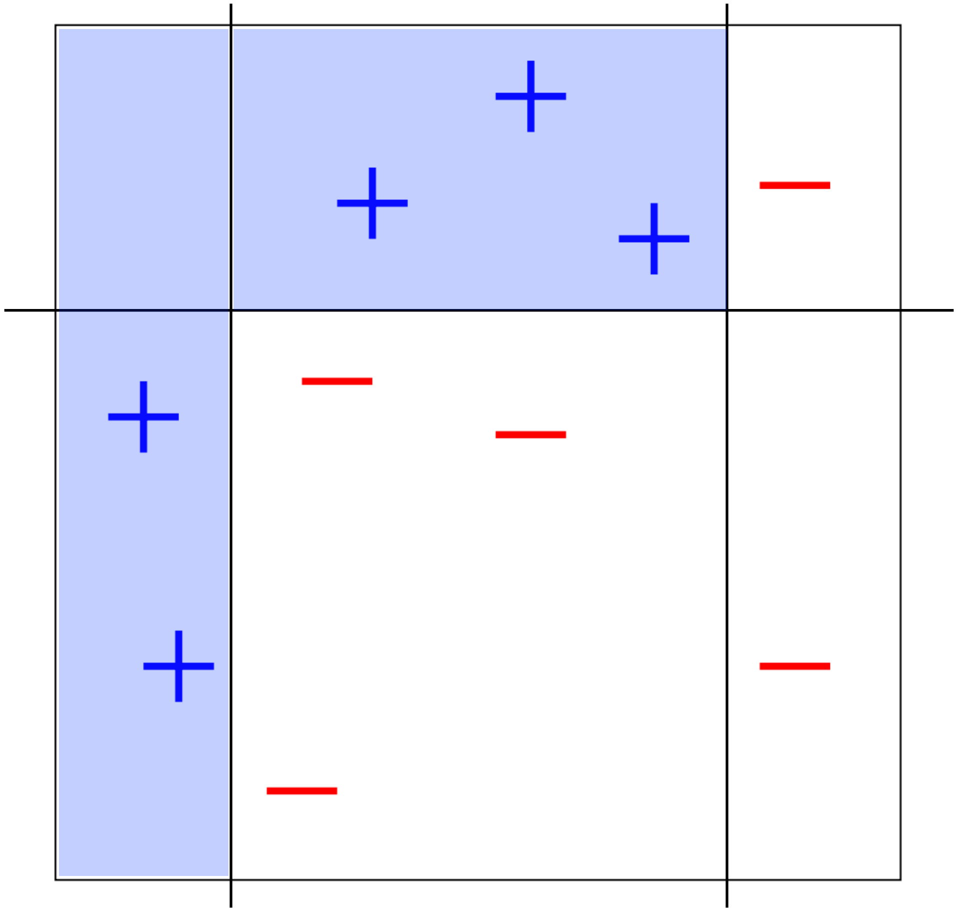
$$\Rightarrow \quad c_1 = \frac{5}{\sqrt{21}}$$

- correctly classified: $w_i^{(2)} = \frac{5}{\sqrt{21}} \cdot \frac{1}{10} \sqrt{\frac{3}{7}} = \frac{1}{14}$ decrease!
- incorrectly classified: $w_i^{(2)} = \frac{5}{\sqrt{21}} \cdot \frac{1}{10} \sqrt{\frac{7}{3}} = \frac{1}{6}$ increase!

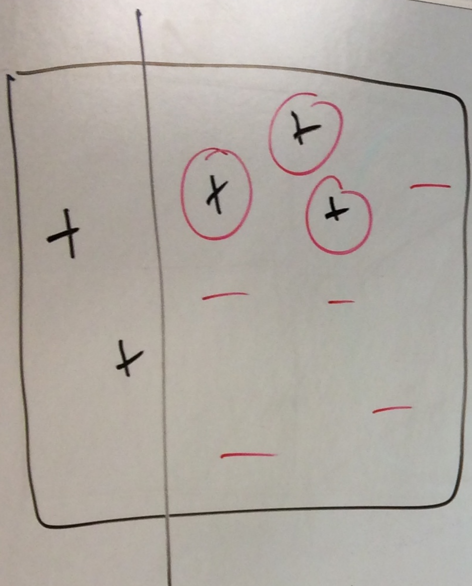
Handout 14: Round 2 & 3 (exercise!)



Handout 14: final classifier



$$h(\boldsymbol{x}) = \text{sign}\left(0.42 \cdot h^{(1)}(\boldsymbol{x}) + 0.65 \cdot h^{(2)}(\boldsymbol{x}) + 0.92 \cdot h^{(3)}(\boldsymbol{x})\right)$$



$$w_i^{(1)} = \frac{1}{10} \quad \forall i$$

(b) weighted error:

$$\varepsilon_1 = \frac{3}{10}$$

(c)

$$\alpha_1 = \frac{1}{2}\log\left(\frac{1-\frac{3}{10}}{\frac{3}{10}}\right) = \frac{1}{2}\log\frac{7}{3}$$

$$= \log\sqrt{\frac{7}{3}}$$

misclassified

$$w_i^{(2)} = c_1 \frac{1}{10} \exp\left(\log\sqrt{\frac{7}{3}}\right)$$

$$= c_1 \frac{1}{10}\sqrt{\frac{7}{3}}$$

correct

$$w_i^{(2)} = c_1 \frac{1}{10}\sqrt{\frac{3}{7}}$$

$$3\left(c_1 \frac{1}{10}\sqrt{\frac{7}{3}}\right) + 7\left(c_1 \frac{1}{10}\sqrt{\frac{3}{7}}\right) = 1$$

<u>incorrect</u>: $\frac{1}{6}$

<u>correct</u>: $\frac{1}{14}$

Outline for October 24

- Ensemble methods
  - Bagging
  - Random forests
  - Boosting
  - Weighted entropy

See back of Handout 13
Note: entropy calculation should have a negative sign in front!