10-701 Recitation

MACHINE LEARNING

BOOSTING

A Formal View of Boosting

- given training set $(x_1, y_1), \ldots, (x_m, y_m)$
- $y_i \in \{-1, +1\}$ correct label of instance $x_i \in X$
- for $t = 1, \ldots, T$:
  - construct distribution D_t on $\{1, \ldots, m\}$
  - find weak hypothesis ("rule of thumb")

    $$h_t : X \to \{-1, +1\}$$

    with small error ϵ_t on D_t:

    $$\epsilon_t = \Pr_{D_t}[h_t(x_i) \neq y_i]$$
- output final hypothesis H_{final}

(Slide stolen from Schapire's Boosting Tutorial)

AdaBoost

[Freund & Schapire]

- constructing D_t:
  - $D_1(i) = 1/m$
  - given D_t and h_t:

$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} \cdot \begin{cases} e^{-\alpha_t} & \text{if } y_i = h_t(x_i) \\ e^{\alpha_t} & \text{if } y_i \neq h_t(x_i) \end{cases}$$

$$= \frac{D_t(i)}{Z_t} \cdot \exp(-\alpha_t \, y_i \, h_t(x_i))$$

where Z_t = normalization constant

$$\alpha_t = \frac{1}{2} \ln\left(\frac{1-\epsilon_t}{\epsilon_t}\right) > 0$$

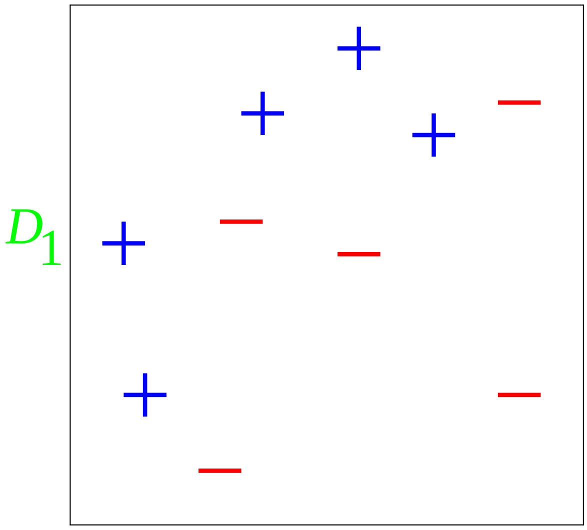
- final hypothesis:
  - $H_{\text{final}}(x) = \text{sign}\left(\sum_t \alpha_t h_t(x)\right)$

$$\epsilon_t = \sum_{i=1}^{m} D_t(i) \cdot \mathbb{1}[y_i \neq h_t(x_i)]$$

$$Z_t = \sum_{i=1}^{m} D_t(i) e^{-\alpha_t y_i h_t(x_i)}$$

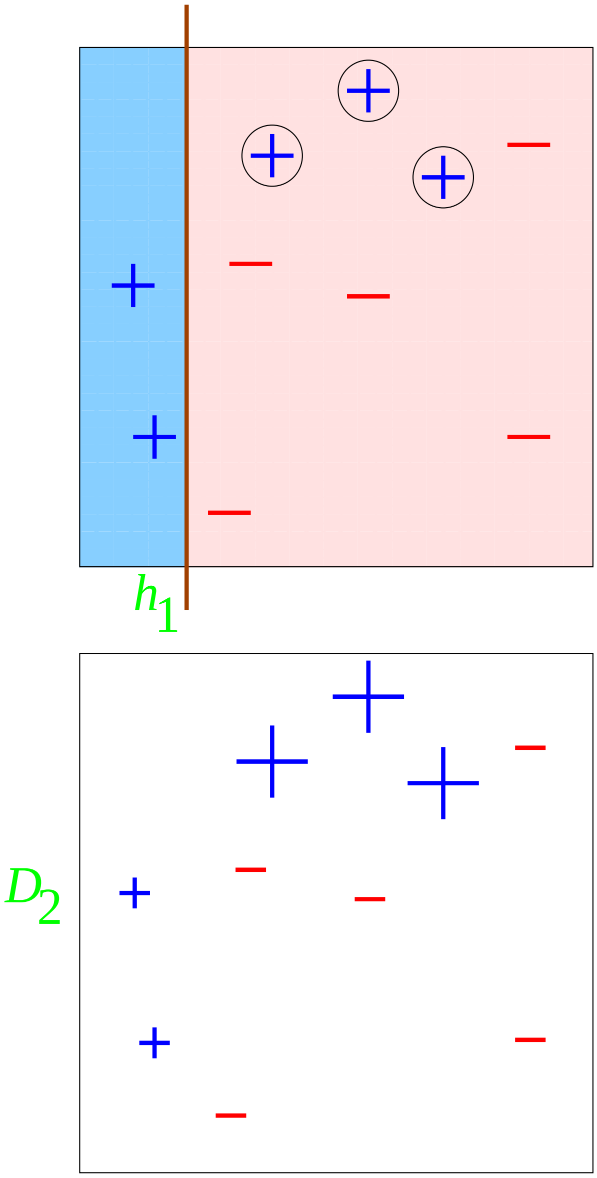
(Slide stolen from Schapire's Boosting Tutorial)

Toy Example



(Slide stolen from Schapire's Boosting Tutorial)

Round 1



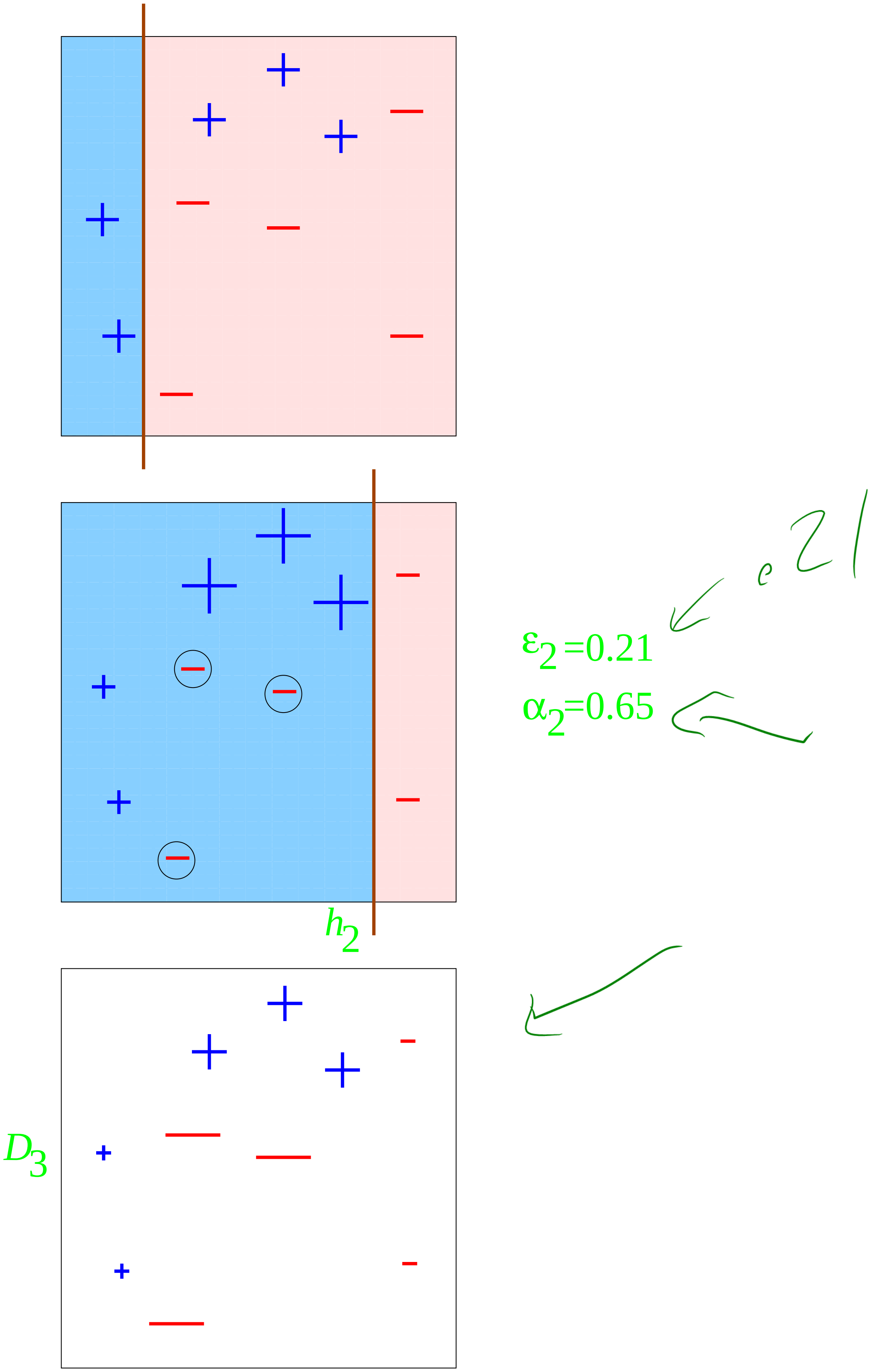
$\varepsilon_1 = 0.30$ $\frac{3}{10}$

$\alpha_1 = 0.42$ $.42$

$\frac{1}{2} \ln\left(\frac{1-\frac{3}{10}}{3/10}\right)$

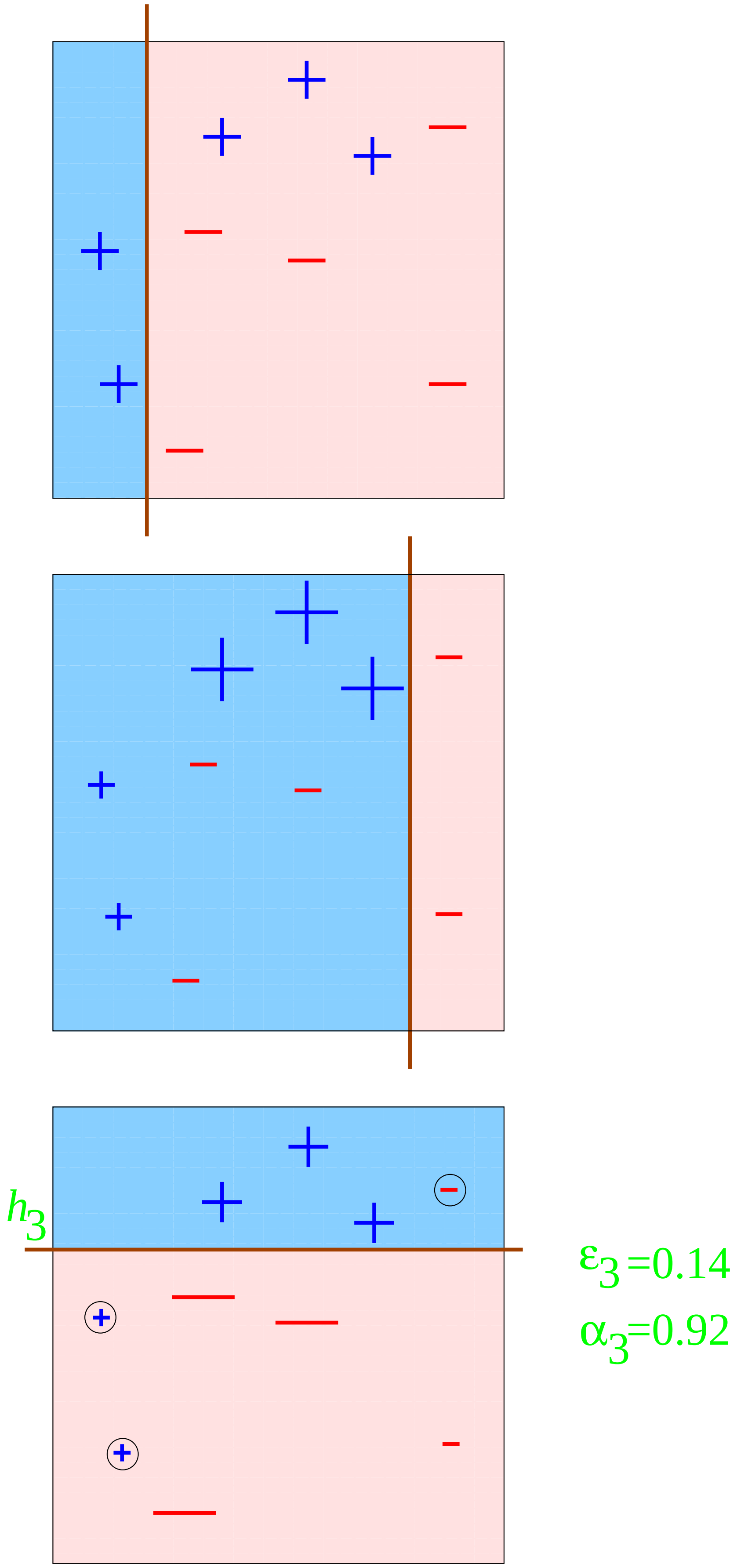
(Slide stolen from Schapire's Boosting Tutorial)

Round 2



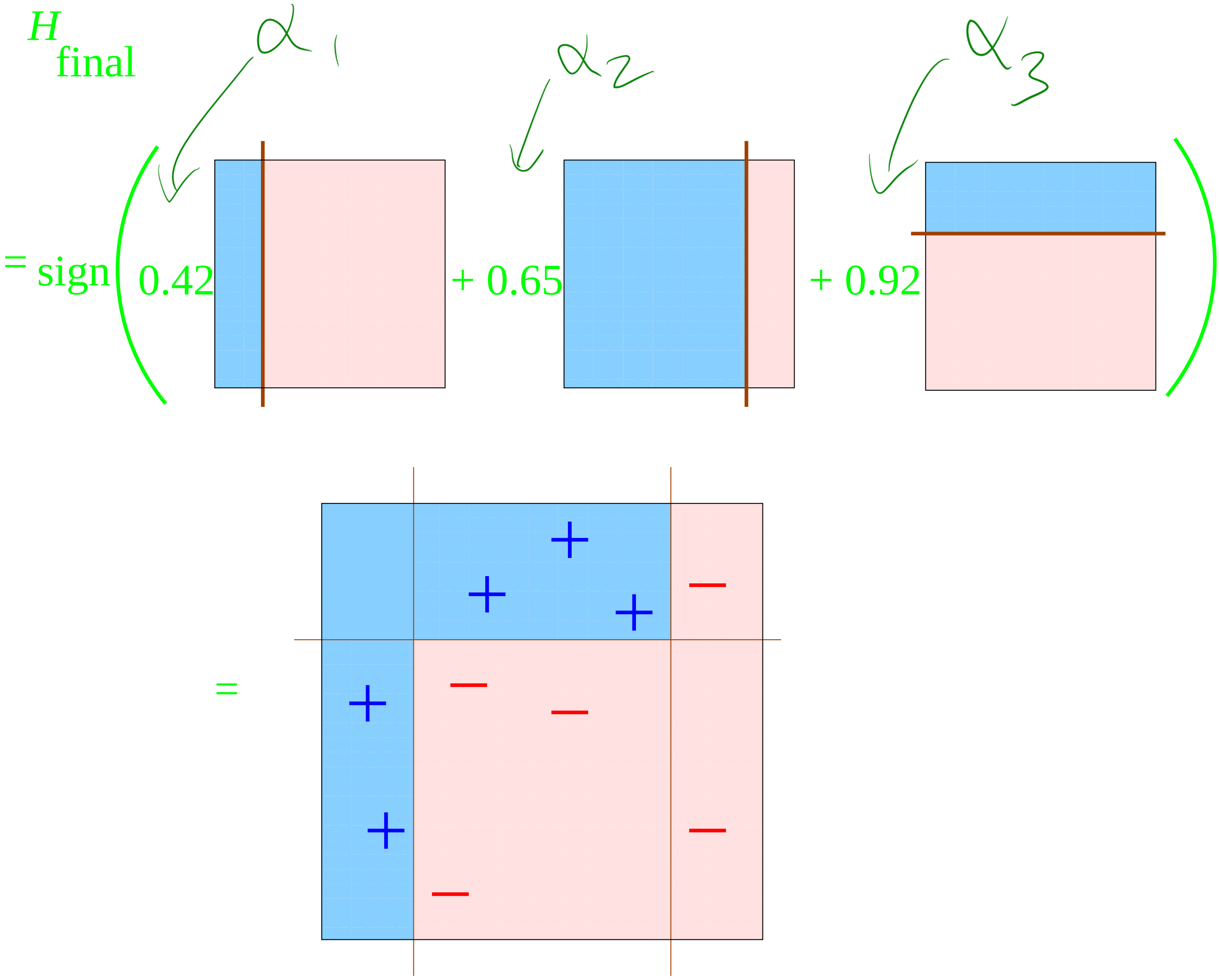
(Slide stolen from Schapire's Boosting Tutorial)

Round 3



(Slide stolen from Schapire's Boosting Tutorial)

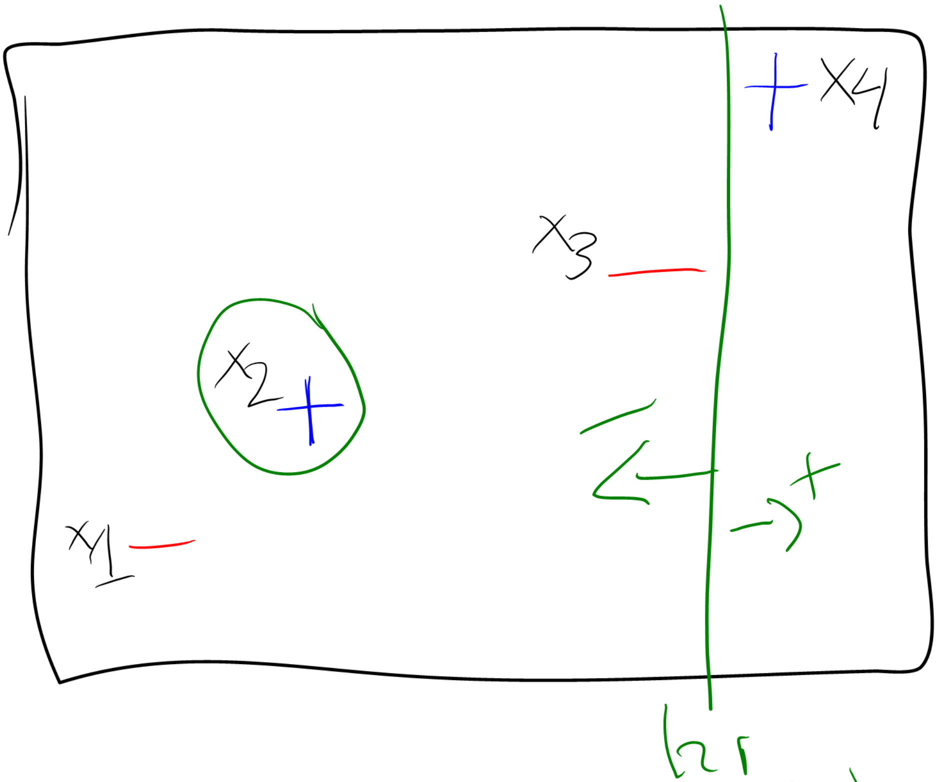
Final Hypothesis



* See demo at
www.research.att.com/˜yoav/adaboost

(Slide stolen from Schapire's Boosting Tutorial)

weak Learner: Pick a decision stump that minimizes ε_t.



$$D_1(1) = D_1(2) = D_1(3) = D_1(4) = \frac{1}{4}$$

$$\varepsilon_1 = \frac{1}{4}, \quad \alpha_1 = \frac{1}{2}\ln\left(\frac{3/4}{1/4}\right) = \frac{1}{2}\ln 3$$

$$e^{\alpha_1} = e^{\frac{1}{2}\ln 3} = \sqrt{3}, \quad e^{-\alpha_1} = \frac{1}{\sqrt{3}}$$

$$D_2(1) = \frac{1}{\sqrt{3}} \cdot \frac{1}{4} \cdot \frac{1}{Z_1}, \quad D_2(2) = \sqrt{3}\,\frac{1}{4}\,\frac{1}{Z_1}$$

$$D_2(3) = \frac{1}{\sqrt{3}}\,\frac{1}{4}\,\frac{1}{Z_1}, \quad D_2(4) = \frac{1}{\sqrt{3}}\,\frac{1}{4}\,\frac{1}{Z_1}$$

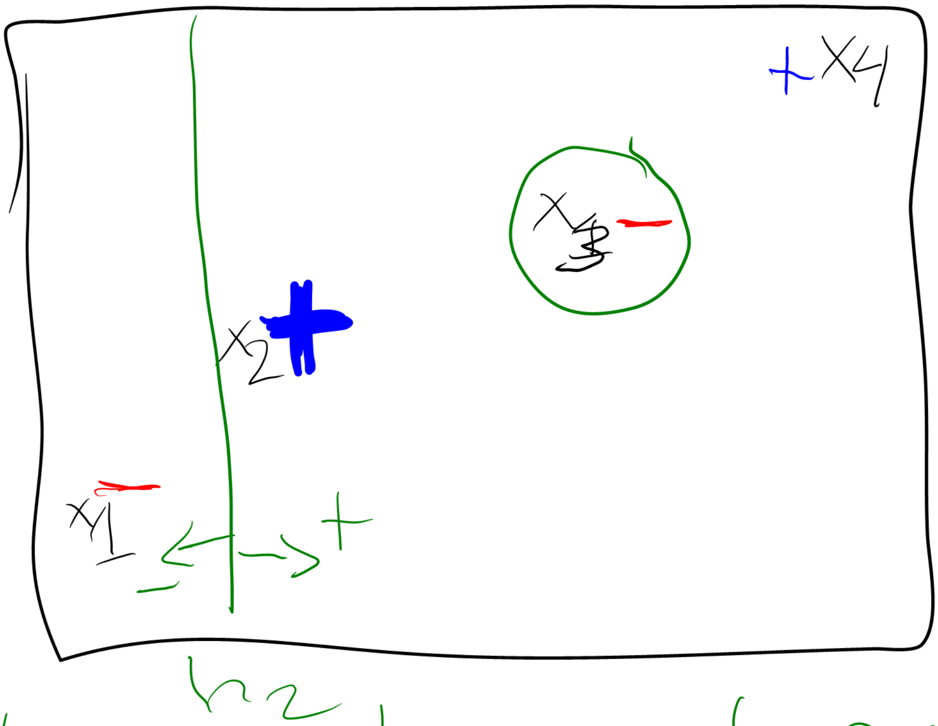
$$Z_1 = \frac{3}{4\sqrt{3}} + \frac{\sqrt{3}}{4} = \frac{\sqrt{3}}{2}$$

$$D_2(1) = \frac{1}{6}, \quad D_2(2) = \frac{1}{2}, \quad D_2(3) = \frac{1}{6}, \quad D_2(4) = \frac{1}{6}$$

$$H(x) = \operatorname{sign}(\alpha_1 h_1(x))$$

$$\boxed{Z_1 = \frac{\sqrt{3}}{2} \approx .87 \geq \frac{1}{4} = \text{error train}}$$

$\alpha_1 = \frac{1}{2}\ln 3$



$D_2(1) = \frac{1}{6}, \; D_2(2) = \frac{1}{2}, \; D_2(3) = \frac{1}{6}, \; D_2(4) = \frac{1}{6}$

$\varepsilon_2 = \frac{1}{6} \qquad \alpha_2 = \frac{1}{2}\ln 5$

$H(x) = \operatorname{sign}(\alpha_1 h_1(x) + \alpha_2 h_2(x))$
$= \operatorname{sign}((\frac{1}{2}\ln 3)h_1(x) + (\frac{1}{2}\ln 5)h_2(x))$

$e^{\alpha_2} = \sqrt{5}, \quad e^{-\alpha_2} = \frac{1}{\sqrt{5}}$

$D_3(1) = \frac{1}{6}\cdot\frac{1}{\sqrt5}\cdot\frac{1}{Z_2}, \quad D_3(2) = \frac{1}{2}\cdot\frac{\sqrt5}{Z_2}$

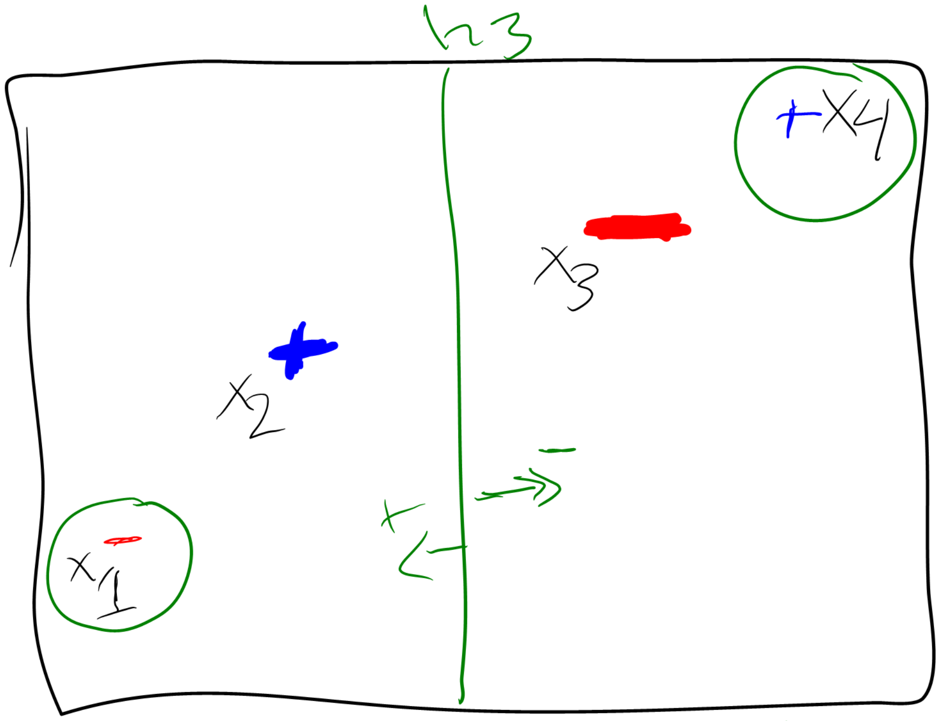
$D_3(3) = \frac{1}{6}\frac{1}{\sqrt5}\frac{1}{Z_2}, \quad D_3(4) = \frac{1}{6}\frac{1}{\sqrt5}\frac{1}{Z_2}$

$Z_2 = 2\frac{1}{6\sqrt5} + \frac{1}{2\sqrt5} + \frac{\sqrt5}{6} \simeq \frac{\sqrt5}{3}$

$D_3(1) = \frac{1}{10}, \quad D_3(2) = \frac{3}{10},$

$D_3(3) = \frac{1}{2}, \quad D_3(4) = \frac{1}{10}$

$Z_1 Z_2 = \frac{\sqrt3}{2}\cdot\frac{\sqrt5}{3} \approx 0.65 \approx 7, \frac{1}{4}$



$D_3(1)=\frac{1}{10},\ D_3(2)=\frac{3}{10},\ D_3(3)=\frac{1}{2},\ D_3(4)=\frac{1}{10}$

$\varepsilon_3=\frac{2}{10}=\frac{1}{5},\quad \alpha_3=\frac{1}{2}\ln\frac{4/5}{1/5}=\frac{1}{2}\ln 4$

$H(x)=\text{sign}(\alpha_1 h_1(x)+\alpha_2 h_2(x)+\alpha_3 h_3(x))$

$=\text{sign}\left((\tfrac{1}{2}\ln 3)h_1(x)+(\tfrac{1}{2}\ln 5)h_2(x)+(\tfrac{1}{2}\ln 4)h_3(x)\right)$

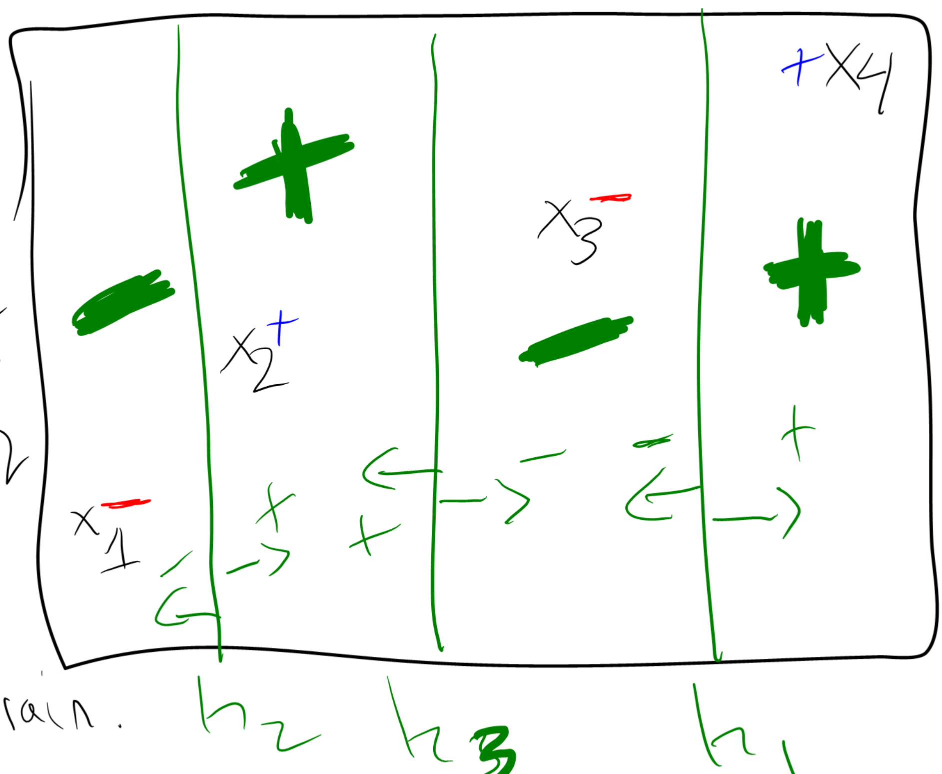
Note:

$Z_1 \cdot Z_2 \cdot Z_3$

$=\frac{\sqrt{3}}{2}\cdot\frac{\sqrt{5}}{3}\cdot\frac{4}{5}$

$=\frac{2}{\sqrt{15}}\approx .52$

$>0=$ errtrain.



see applet too.

guy who discovered boosting!

See also Schapire's video tutorial (linked from the course website).

$\approx 2\frac{1}{2}$ hrs well spent

There was a question about situations where AdaBoost fails. Here's an example