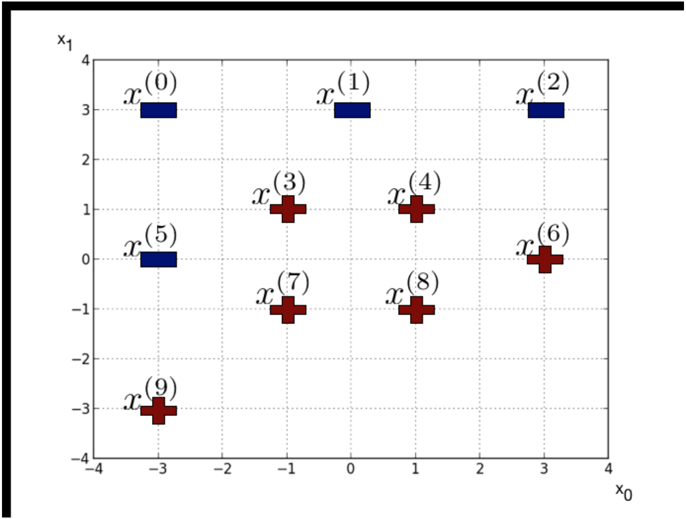
Boosting

Machine Learning: Jordan Boyd-Graber
University of Maryland
SLIDES ADAPTED FROM ROB SCHAPIRE

Content Questions

Administrivia Questions

Boosting Example



Hypothesis 1

- Find the best weak learner weighted by D_1

Hypothesis 1

- Find the best weak learner weighted by D_1
- Return 1.0 if x_1 is less than 2.0, -1.0 otherwise

Iteration 1

- Error: $\epsilon_1 = \sum_{i=1}^{m} D_1(i) \mathbb{1}\left[y^{(i)} \neq h_1(x^{(i)})\right]$

Iteration 1

- Error: $\epsilon_1 = \sum_{i=1}^{m} D_1(i) \mathbb{1}\left[y^{(i)} \neq h_1(x^{(i)})\right]$

$$\epsilon_1 = 0.10_5 = 0.10 \tag{1}$$

Iteration 1

- Error: $\epsilon_1 = \sum_{i=1}^{m} D_1(i) \mathbb{1}\left[y^{(i)} \neq h_1(x^{(i)})\right]$

$$\epsilon_1 = 0.10_5 = 0.10 \tag{1}$$

- $\alpha_1 = \frac{1}{2} \ln\left(\frac{1-\epsilon_1}{\epsilon_1}\right)$

Iteration 1

- Error: $\epsilon_1 = \sum_{i=1}^{m} D_1(i) \mathbb{1}\left[y^{(i)} \neq h_1(x^{(i)})\right]$

$$\epsilon_1 = 0.10_5 = 0.10 \tag{1}$$

- $\alpha_1 = \frac{1}{2} \ln\left(\frac{1-\epsilon_1}{\epsilon_1}\right) = 1.10$
- Update distribution: $D_2(i) \propto D_1(i) \exp(-\alpha_1 y^{(i)} h_1(x^{(i)}))$

Iteration 1

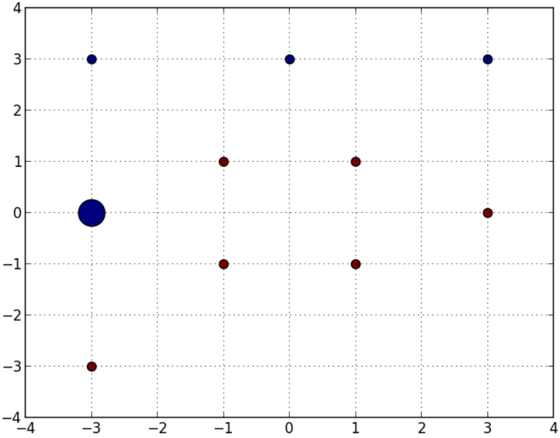
- Error: $\epsilon_1 = \sum_{i=1}^{m} D_1(i) \mathbb{1}\left[y^{(i)} \neq h_1(x^{(i)})\right]$

$$\epsilon_1 = 0.10_5 = 0.10 \quad (1)$$

- $\alpha_1 = \frac{1}{2} \ln\left(\frac{1-\epsilon_1}{\epsilon_1}\right) = 1.10$
- Update distribution: $D_2(i) \propto D_1(i) \exp(-\alpha_1 y^{(i)} h_1(x^{(i)}))$

0	1	2	3	4	5	6	7	8	9
0.06	0.06	0.06	0.06	0.06	0.50	0.06	0.06	0.06	0.06

Distribution 2



Hypothesis 2

- Find the best learner weighted by D_2
- Return 1.0 if x_0 is greater than -2.0, -1.0 otherwise

Iteration 2

- Error: $\epsilon_2 = \sum_{i=1}^{m} D_2(i) \mathbb{1}\left[y^{(i)} \neq h_2(x^{(i)}) \right]$

Iteration 2

- Error: $\epsilon_2 = \sum_{i=1}^{m} D_2(i) \mathbb{1}\left[y^{(i)} \neq h_2(x^{(i)})\right]$

$$\epsilon_2 = 0.06_1 + 0.06_2 + 0.06_9 = 0.17 \tag{2}$$

Iteration 2

- Error: $\epsilon_2 = \sum_{i=1}^{m} D_2(i)\mathbb{1}\left[y^{(i)} \neq h_2(x^{(i)})\right]$

$$\epsilon_2 = 0.06_1 + 0.06_2 + 0.06_9 = 0.17 \tag{2}$$

- $\alpha_2 = \frac{1}{2}\ln\left(\frac{1-\epsilon_2}{\epsilon_2}\right)$

Iteration 2

- Error: $\epsilon_2 = \sum_{i=1}^{m} D_2(i) \mathbb{1}\left[y^{(i)} \neq h_2(x^{(i)})\right]$

$$\epsilon_2 = 0.06_1 + 0.06_2 + 0.06_9 = 0.17 \tag{2}$$

- $\alpha_2 = \frac{1}{2} \ln\left(\frac{1-\epsilon_2}{\epsilon_2}\right) = 0.80$
- Update distribution: $D_3(i) \propto D_2(i) \exp(-\alpha_2 y^{(i)} h_2(x^{(i)}))$

Iteration 2

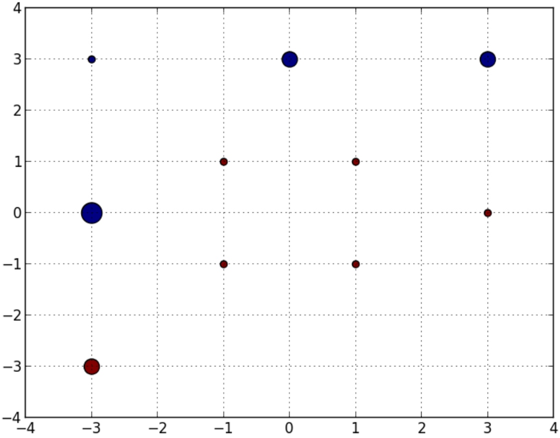
- Error: $\epsilon_2 = \sum_{i=1}^{m} D_2(i) \mathbb{1}\left[y^{(i)} \neq h_2(x^{(i)})\right]$

$$\epsilon_2 = 0.06_1 + 0.06_2 + 0.06_9 = 0.17 \tag{2}$$

- $\alpha_2 = \frac{1}{2} \ln\left(\frac{1-\epsilon_2}{\epsilon_2}\right) = 0.80$
- Update distribution: $D_3(i) \propto D_2(i) \exp(-\alpha_2 y^{(i)} h_2(x^{(i)}))$

0	1	2	3	4	5	6	7	8	9
0.03	0.17	0.17	0.03	0.03	0.30	0.03	0.03	0.03	0.17

Distribution 3



Hypothesis 3

- Find the best learner weighted by D_3
- Return 1.0 if x_1 is less than -0.5, -1.0 otherwise

Iteration 3

- Error: $\epsilon_3 = \sum_{i=1}^{m} D_3(i) \mathbb{1}\left[y^{(i)} \neq h_3(x^{(i)})\right]$

Iteration 3

- Error: $\epsilon_3 = \sum_{i=1}^{m} D_3(i)\mathbb{1}\left[y^{(i)} \neq h_3(x^{(i)})\right]$

$$\epsilon_3 = 0.03_3 + 0.03_4 + 0.03_6 = 0.10 \tag{3}$$

Iteration 3

- Error: $\epsilon_3 = \sum_{i=1}^{m} D_3(i)\mathbb{1}\left[y^{(i)} \neq h_3(x^{(i)})\right]$

$$\epsilon_3 = 0.03_3 + 0.03_4 + 0.03_6 = 0.10 \tag{3}$$

- $\alpha_3 = \frac{1}{2}\ln\left(\frac{1-\epsilon_3}{\epsilon_3}\right)$

Iteration 3

- Error: $\epsilon_3 = \sum_{i=1}^{m} D_3(i) \mathbb{1}\left[y^{(i)} \neq h_3(x^{(i)})\right]$

$$\epsilon_3 = 0.03_3 + 0.03_4 + 0.03_6 = 0.10 \tag{3}$$

- $\alpha_3 = \frac{1}{2} \ln\left(\frac{1-\epsilon_3}{\epsilon_3}\right) = 1.10$
- Update distribution: $D_4(i) \propto D_3(i) \exp(-\alpha_3 y^{(i)} h_3(x^{(i)}))$

Iteration 3

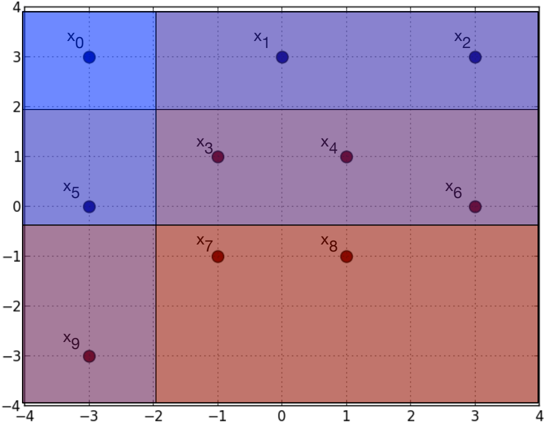
- Error: $\epsilon_3 = \sum_{i=1}^{m} D_3(i) \mathbb{1}\left[y^{(i)} \neq h_3(x^{(i)})\right]$

$$\epsilon_3 = 0.03_3 + 0.03_4 + 0.03_6 = 0.10 \tag{3}$$

- $\alpha_3 = \frac{1}{2} \ln\left(\frac{1-\epsilon_3}{\epsilon_3}\right) = 1.10$
- Update distribution: $D_4(i) \propto D_3(i) \exp(-\alpha_3 y^{(i)} h_3(x^{(i)}))$

0	1	2	3	4	5	6	7	8	9
0.02	0.09	0.09	0.17	0.17	0.17	0.17	0.02	0.02	0.09

Classifier



Final Predictions

$$H(x) = \text{sign}\left(\sum_t \alpha_t h_t(x)\right) \tag{4}$$

- $H(x^{(0)}) =$

Final Predictions

$$H(x) = \text{sign}\left(\sum_t \alpha_t h_t(x)\right) \tag{4}$$

- $H(x^{(0)}) = \text{sign}(-1.10 + -0.80 + -1.10) = \text{sign}(-3.00) = -1.0$
- $H(x^{(1)}) =$

Final Predictions

$$H(x) = \text{sign}\left(\sum_t \alpha_t h_t(x)\right) \tag{4}$$

- $H(x^{(0)}) = \text{sign}(-1.10 + -0.80 + -1.10) = \text{sign}(-3.00) = -1.0$
- $H(x^{(1)}) = \text{sign}(-1.10 + 0.80 + -1.10) = \text{sign}(-1.39) = -1.0$
- $H(x^{(2)}) =$

Final Predictions

$$H(x) = \text{sign}\left(\sum_t \alpha_t h_t(x)\right) \tag{4}$$

- $H(x^{(0)}) = \text{sign}(-1.10 + -0.80 + -1.10) = \text{sign}(-3.00) = -1.0$
- $H(x^{(1)}) = \text{sign}(-1.10 + 0.80 + -1.10) = \text{sign}(-1.39) = -1.0$
- $H(x^{(2)}) = \text{sign}(-1.10 + 0.80 + -1.10) = \text{sign}(-1.39) = -1.0$
- $H(x^{(3)}) =$

Final Predictions

$$H(x) = \text{sign}\left(\sum_t \alpha_t h_t(x)\right) \tag{4}$$

- $H(x^{(0)}) = \text{sign}(-1.10 + -0.80 + -1.10) = \text{sign}(-3.00) = -1.0$
- $H(x^{(1)}) = \text{sign}(-1.10 + 0.80 + -1.10) = \text{sign}(-1.39) = -1.0$
- $H(x^{(2)}) = \text{sign}(-1.10 + 0.80 + -1.10) = \text{sign}(-1.39) = -1.0$
- $H(x^{(3)}) = \text{sign}(1.10 + 0.80 + -1.10) = \text{sign}(0.80) = 1.0$
- $H(x^{(4)}) =$

Final Predictions

$$H(x) = \text{sign}\left(\sum_t \alpha_t h_t(x)\right) \tag{4}$$

- $H(x^{(0)}) = \text{sign}(-1.10 + -0.80 + -1.10) = \text{sign}(-3.00) = -1.0$
- $H(x^{(1)}) = \text{sign}(-1.10 + 0.80 + -1.10) = \text{sign}(-1.39) = -1.0$
- $H(x^{(2)}) = \text{sign}(-1.10 + 0.80 + -1.10) = \text{sign}(-1.39) = -1.0$
- $H(x^{(3)}) = \text{sign}(1.10 + 0.80 + -1.10) = \text{sign}(0.80) = 1.0$
- $H(x^{(4)}) = \text{sign}(1.10 + 0.80 + -1.10) = \text{sign}(0.80) = 1.0$
- $H(x^{(5)}) =$

Final Predictions

$$H(x) = \text{sign}\left(\sum_t \alpha_t h_t(x)\right) \tag{4}$$

- $H(x^{(0)}) = \text{sign}(-1.10 + -0.80 + -1.10) = \text{sign}(-3.00) = -1.0$
- $H(x^{(1)}) = \text{sign}(-1.10 + 0.80 + -1.10) = \text{sign}(-1.39) = -1.0$
- $H(x^{(2)}) = \text{sign}(-1.10 + 0.80 + -1.10) = \text{sign}(-1.39) = -1.0$
- $H(x^{(3)}) = \text{sign}(1.10 + 0.80 + -1.10) = \text{sign}(0.80) = 1.0$
- $H(x^{(4)}) = \text{sign}(1.10 + 0.80 + -1.10) = \text{sign}(0.80) = 1.0$
- $H(x^{(5)}) = \text{sign}(1.10 + -0.80 + -1.10) = \text{sign}(-0.80) = -1.0$
- $H(x^{(6)}) =$

Final Predictions

$$H(x) = \text{sign}\left(\sum_t \alpha_t h_t(x)\right) \tag{4}$$

- $H(x^{(0)}) = \text{sign}(-1.10 + -0.80 + -1.10) = \text{sign}(-3.00) = -1.0$
- $H(x^{(1)}) = \text{sign}(-1.10 + 0.80 + -1.10) = \text{sign}(-1.39) = -1.0$
- $H(x^{(2)}) = \text{sign}(-1.10 + 0.80 + -1.10) = \text{sign}(-1.39) = -1.0$
- $H(x^{(3)}) = \text{sign}(1.10 + 0.80 + -1.10) = \text{sign}(0.80) = 1.0$
- $H(x^{(4)}) = \text{sign}(1.10 + 0.80 + -1.10) = \text{sign}(0.80) = 1.0$
- $H(x^{(5)}) = \text{sign}(1.10 + -0.80 + -1.10) = \text{sign}(-0.80) = -1.0$
- $H(x^{(6)}) = \text{sign}(1.10 + 0.80 + -1.10) = \text{sign}(0.80) = 1.0$
- $H(x^{(7)}) =$

Final Predictions

$$H(x) = \text{sign}\left(\sum_t \alpha_t h_t(x)\right) \tag{4}$$

- $H(x^{(0)}) = \text{sign}(-1.10 + -0.80 + -1.10) = \text{sign}(-3.00) = -1.0$
- $H(x^{(1)}) = \text{sign}(-1.10 + 0.80 + -1.10) = \text{sign}(-1.39) = -1.0$
- $H(x^{(2)}) = \text{sign}(-1.10 + 0.80 + -1.10) = \text{sign}(-1.39) = -1.0$
- $H(x^{(3)}) = \text{sign}(1.10 + 0.80 + -1.10) = \text{sign}(0.80) = 1.0$
- $H(x^{(4)}) = \text{sign}(1.10 + 0.80 + -1.10) = \text{sign}(0.80) = 1.0$
- $H(x^{(5)}) = \text{sign}(1.10 + -0.80 + -1.10) = \text{sign}(-0.80) = -1.0$
- $H(x^{(6)}) = \text{sign}(1.10 + 0.80 + -1.10) = \text{sign}(0.80) = 1.0$
- $H(x^{(7)}) = \text{sign}(1.10 + 0.80 + 1.10) = \text{sign}(3.00) = 1.0$
- $H(x^{(8)}) =$

Final Predictions

$$H(x) = \text{sign}\left(\sum_t \alpha_t h_t(x)\right) \tag{4}$$

- $H(x^{(0)}) = \text{sign}(-1.10 + -0.80 + -1.10) = \text{sign}(-3.00) = -1.0$
- $H(x^{(1)}) = \text{sign}(-1.10 + 0.80 + -1.10) = \text{sign}(-1.39) = -1.0$
- $H(x^{(2)}) = \text{sign}(-1.10 + 0.80 + -1.10) = \text{sign}(-1.39) = -1.0$
- $H(x^{(3)}) = \text{sign}(1.10 + 0.80 + -1.10) = \text{sign}(0.80) = 1.0$
- $H(x^{(4)}) = \text{sign}(1.10 + 0.80 + -1.10) = \text{sign}(0.80) = 1.0$
- $H(x^{(5)}) = \text{sign}(1.10 + -0.80 + -1.10) = \text{sign}(-0.80) = -1.0$
- $H(x^{(6)}) = \text{sign}(1.10 + 0.80 + -1.10) = \text{sign}(0.80) = 1.0$
- $H(x^{(7)}) = \text{sign}(1.10 + 0.80 + 1.10) = \text{sign}(3.00) = 1.0$
- $H(x^{(8)}) = \text{sign}(1.10 + 0.80 + 1.10) = \text{sign}(3.00) = 1.0$
- $H(x^{(9)}) =$

Final Predictions

$$H(x) = \text{sign}\left(\sum_t \alpha_t h_t(x)\right) \tag{4}$$

- $H(x^{(0)}) = \text{sign}(-1.10 + -0.80 + -1.10) = \text{sign}(-3.00) = -1.0$
- $H(x^{(1)}) = \text{sign}(-1.10 + 0.80 + -1.10) = \text{sign}(-1.39) = -1.0$
- $H(x^{(2)}) = \text{sign}(-1.10 + 0.80 + -1.10) = \text{sign}(-1.39) = -1.0$
- $H(x^{(3)}) = \text{sign}(1.10 + 0.80 + -1.10) = \text{sign}(0.80) = 1.0$
- $H(x^{(4)}) = \text{sign}(1.10 + 0.80 + -1.10) = \text{sign}(0.80) = 1.0$
- $H(x^{(5)}) = \text{sign}(1.10 + -0.80 + -1.10) = \text{sign}(-0.80) = -1.0$
- $H(x^{(6)}) = \text{sign}(1.10 + 0.80 + -1.10) = \text{sign}(0.80) = 1.0$
- $H(x^{(7)}) = \text{sign}(1.10 + 0.80 + 1.10) = \text{sign}(3.00) = 1.0$
- $H(x^{(8)}) = \text{sign}(1.10 + 0.80 + 1.10) = \text{sign}(3.00) = 1.0$
- $H(x^{(9)}) = \text{sign}(1.10 + -0.80 + 1.10) = \text{sign}(1.39) = 1.0$