Department of Computer Science
UNIVERSITY OF COLORADO **BOULDER**

Boosting

Jordan Boyd-Graber
University of Colorado Boulder
LECTURE 12

Content Questions

Content Questions

Content Questions

Content Questions

Content Questions

Content Questions

Content Questions

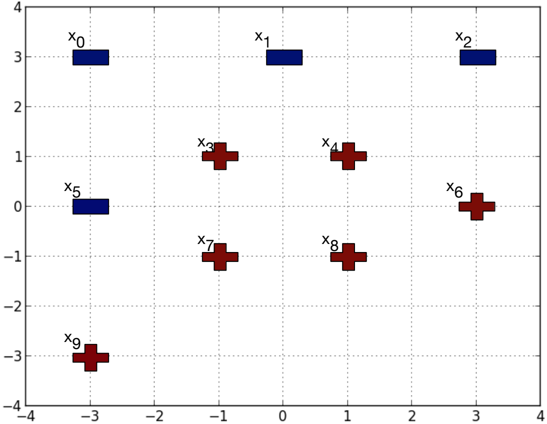
Content Questions

Content Questions

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

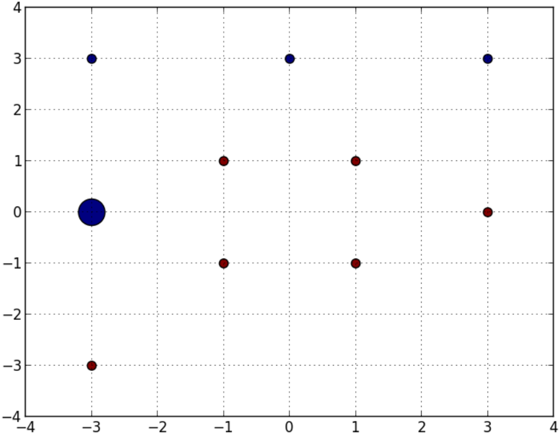
- Error: $\epsilon_1 = \sum_{i=1}^{m} D_1(i) \mathbb{1}\left[y_i \neq h_1(x_i)\right]$

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

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

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

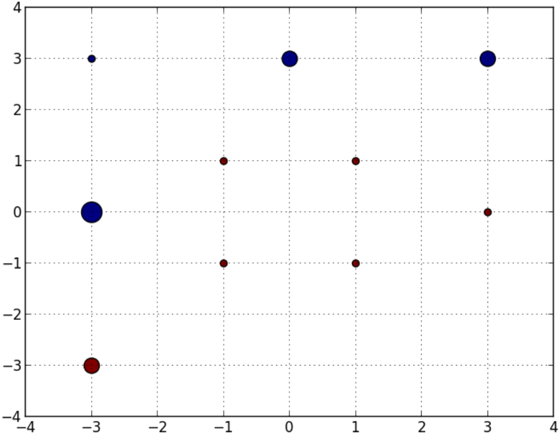
- Error: $\epsilon_2 = \sum_{i=1}^{m} D_2(i) \mathbb{1}\left[y_i \neq h_2(x_i)\right]$

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

- $\alpha_2 = \frac{1}{2} \ln \left(\frac{1-\epsilon_2}{\epsilon_2} \right) = 0.80$
- 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

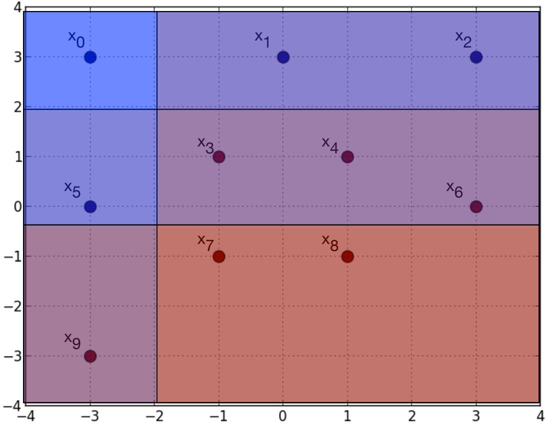
- Error: $\epsilon_3 = \sum_{i=1}^{m} D_3(i) \mathbb{1}\left[y_i \neq h_3(x_i)\right]$

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

- $\alpha_3 = \frac{1}{2} \ln \left(\frac{1-\epsilon_3}{\epsilon_3} \right) = 1.10$
- 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

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

- $H(x_0) = \text{sign}(-1.10 + -0.80 + -1.10) = \text{sign}(-3.00) = -1.0$
- $H(x_1) = \text{sign}(-1.10 + 0.80 + -1.10) = \text{sign}(-1.39) = -1.0$
- $H(x_2) = \text{sign}(-1.10 + 0.80 + -1.10) = \text{sign}(-1.39) = -1.0$
- $H(x_3) = \text{sign}(1.10 + 0.80 + -1.10) = \text{sign}(0.80) = 1.0$
- $H(x_4) = \text{sign}(1.10 + 0.80 + -1.10) = \text{sign}(0.80) = 1.0$
- $H(x_5) = \text{sign}(1.10 + -0.80 + -1.10) = \text{sign}(-0.80) = -1.0$
- $H(x_6) = \text{sign}(1.10 + 0.80 + -1.10) = \text{sign}(0.80) = 1.0$
- $H(x_7) = \text{sign}(1.10 + 0.80 + 1.10) = \text{sign}(3.00) = 1.0$
- $H(x_8) = \text{sign}(1.10 + 0.80 + 1.10) = \text{sign}(3.00) = 1.0$
- $H(x_9) = \text{sign}(1.10 + -0.80 + 1.10) = \text{sign}(1.39) = 1.0$