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

- x_0, x_1, x_2
- x_3, x_4
- x_5, x_6
- x_7, x_8
- x_9

Jordan Boyd-Graber
| Boulder

| Boosting | 4 of 14 |
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: \( \epsilon_1 = \sum_{i=1}^{m} D_1(i) \mathbb{1}[y_i \neq h_1(x_i)] \)
Iteration 1

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

$$\epsilon_1 = 0.105 = 0.10$$ (1)
Iteration 1

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

  $\epsilon_1 = 0.105 = 0.10$ \hfill (1)

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

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

  $$\epsilon_1 = 0.105 = 0.10$$  \hspace{1cm} (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: \( \epsilon_1 = \sum_{i=1}^{m} D_1(i) 1[y_i \neq h_1(x_i)] \)

\[ \epsilon_1 = 0.105 = 0.10 \]  

(1)

- \( \alpha_1 = \frac{1}{2} \ln \left( \frac{1-\epsilon_1}{\epsilon_1} \right) = 1.10 \)

- Update distribution: \( D_2(i) \propto D_1(i) \exp(-\alpha_1 y_i h_1(x_i)) \)

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.50</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
</tr>
</tbody>
</table>
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: $\epsilon_2 = \sum_{i=1}^{m} D_2(i) \mathbb{1}[y_i \neq h_2(x_i)]$
Iteration 2

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

  \[ \epsilon_2 = 0.061 + 0.062 + 0.069 = 0.17 \] (2)
Iteration 2

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

\[
\epsilon_2 = 0.061 + 0.062 + 0.069 = 0.17 \quad (2)
\]

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

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

  \[ \epsilon_2 = 0.06_1 + 0.06_2 + 0.06_9 = 0.17 \] (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**: \( \epsilon_2 = \sum_{i=1}^{m} D_2(i) \cdot 1 [y_i \neq h_2(x_i)] \)

  \[
  \epsilon_2 = 0.06_1 + 0.06_2 + 0.06_9 = 0.17 \quad (2)
  \]

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

- **Update distribution**: \( D_3(i) \propto D_2(i) \exp(-\alpha_2 y_i h_2(x_i)) \)

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.03</td>
<td>0.17</td>
<td>0.17</td>
<td>0.03</td>
<td>0.03</td>
<td>0.30</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
<td>0.17</td>
</tr>
</tbody>
</table>
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: \( \epsilon_3 = \sum_{i=1}^{m} D_3(i) 1[y_i \neq h_3(x_i)] \)
Iteration 3

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

\[ \epsilon_3 = 0.033 + 0.034 + 0.036 = 0.10 \] (3)
Iteration 3

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

\[
\epsilon_3 = 0.03_3 + 0.03_4 + 0.03_6 = 0.10
\]  \hspace{1cm} (3)

- $\alpha_3 = \frac{1}{2} \ln \left( \frac{1-\epsilon_3}{\epsilon_3} \right)$
Iteration 3

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

\[
\epsilon_3 = 0.033 + 0.034 + 0.036 = 0.10
\]

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

- Update distribution: \( D_4(i) \propto D_3(i) \exp(-\alpha_3 y_i h_3(x_i)) \)
Iteration 3

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

\[
\epsilon_3 = 0.033 + 0.034 + 0.036 = 0.10
\]  

(3)

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

- Update distribution: \(D_4(i) \propto D_3(i) \exp(-\alpha_3 y_i h_3(x_i))\)

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.02</td>
<td>0.09</td>
<td>0.09</td>
<td>0.17</td>
<td>0.17</td>
<td>0.17</td>
<td>0.17</td>
<td>0.02</td>
<td>0.02</td>
<td>0.09</td>
</tr>
</tbody>
</table>
Classifier
Final Predictions

\[ H(x) = \text{sign} \left( \sum_t \alpha_t h_t(x) \right) \]  

- \( H(x_0) = \)

\[ H(x) = \text{sign} \left( \sum_t \alpha_t h_t(x) \right) \]  

(4)
Final Predictions

\[ H(x) = \text{sign} \left( \sum_{t} \alpha_t h_t(x) \right) \]  \hspace{1cm} (4)

- \[ H(x_0) = \text{sign}(-1.10 + -0.80 + -1.10) = \text{sign}(-3.00) = -1.0 \]
- \[ H(x_1) = \]
Final Predictions

\[ H(x) = \text{sign} \left( \sum_t \alpha_t h_t(x) \right) \]  \hspace{1cm} (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

\[ H(x) = \text{sign} \left( \sum_{t} \alpha_t h_t(x) \right) \]  \hspace{1cm} (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

\[ H(x) = \text{sign} \left( \sum_t \alpha_t h_t(x) \right) \]  \hspace{1cm} (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

\[ H(x) = \text{sign} \left( \sum_t \alpha_t h_t(x) \right) \]  

\[ H(x) = \text{sign} \left( \sum_t \alpha_t h_t(x) \right) \] (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

\[ H(x) = \text{sign} \left( \sum_t \alpha_t h_t(x) \right) \]  \hspace{1cm} (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

\[ H(x) = \text{sign} \left( \sum_t \alpha_t h_t(x) \right) \]  

- \[ H(x_0) = \text{sign}(-1.10 + -0.80 + -1.10) = \text{sign}(-3.00) = -1.0 \]
- \[ H(x_1) = \text{sign}(-1.10 + 0.80 + -1.10) = \text{sign}(-1.39) = -1.0 \]
- \[ H(x_2) = \text{sign}(-1.10 + 0.80 + -1.10) = \text{sign}(-1.39) = -1.0 \]
- \[ H(x_3) = \text{sign}(1.10 + 0.80 + -1.10) = \text{sign}(0.80) = 1.0 \]
- \[ H(x_4) = \text{sign}(1.10 + 0.80 + -1.10) = \text{sign}(0.80) = 1.0 \]
- \[ H(x_5) = \text{sign}(1.10 + -0.80 + -1.10) = \text{sign}(-0.80) = -1.0 \]
- \[ H(x_6) = \text{sign}(1.10 + 0.80 + -1.10) = \text{sign}(0.80) = 1.0 \]
- \[ H(x_7) = \]
Final Predictions

\[ H(x) = \text{sign} \left( \sum_{t} \alpha_t h_t(x) \right) \] (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

\[ H(x) = \text{sign} \left( \sum_t \alpha_t h_t(x) \right) \]  \hspace{1cm} (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

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

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