Perceptron

CS478 – Machine Learning
Spring 2008

Thorsten Joachims
Cornell University

Reading: Mitchell Chapter 4.4-4.4.2 & Chapter 7.5
Cristianini/Shawe-Taylor Chapter 2.2.1.1

Outline

• Linear classification rules
• Perceptron learning algorithm
• Mistake-bound model
• Perceptron mistake bound

Example: Spam Filtering

<table>
<thead>
<tr>
<th>viagra</th>
<th>learning</th>
<th>the</th>
<th>dating</th>
<th>nigeria</th>
<th>spam?</th>
</tr>
</thead>
<tbody>
<tr>
<td>x_1 = ( 1 0 1 0 0 )</td>
<td>y_1 = 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>x_2 = ( 0 1 1 0 0 )</td>
<td>y_2 = -1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>x_3 = ( 0 0 0 0 1 )</td>
<td>y_3 = 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

• Instance Space \( X \):
  - Feature vector of word occurrences -> binary features
  - \( N \) features (\( N \) typically > 50000)
• Target Concept \( c \):
  - Spam (+1) / Ham (-1)

Linear Classification Rules

• Hypotheses of the form
  - unbiased: \( h_{\text{unbiased}}(x) = \sum w_i x_i + b \)
  - biased: \( h_{\text{biased}}(x) = \sum w_i x_i + b \)

• Hypothesis space \( H \)
  - \( H_{\text{unbiased}} = \{ h_{\text{unbiased}} \} \)
  - \( H_{\text{biased}} = \{ h_{\text{biased}} \} \)

• Notation
  - \( h_{\text{unbiased}}(x) = \cdot \cdot \cdot + w_i x_i + b \) and \( \delta_{\text{unbiased}}(x) = \{ \frac{1}{1+b} \}
  - \( h_{\text{biased}}(x) = \cdot \cdot \cdot + w_i x_i + b \)

(Online) Perceptron Algorithm

• Input: \( S = \{(x_1, y_1), \ldots, (x_n, y_n)\} \), \( x_i \in \mathbb{R}^N \), \( y_i \in \{-1, 1\} \), \( \eta \in \mathbb{R} \)
• Algorithm:
  - \( \bar{w}_0 = 0 \), \( k = 0 \)
  - FOR \( i = 1 \) TO \( n \)
    - IF \( y_i (w \cdot x_i) \leq 0 \) make a mistake
      - \( \bar{w}_{k+1} = \bar{w}_k + \eta y_i x_i \)
      - \( k = k + 1 \)
    - ENDIF
  - ENDFOR
• Output: \( \bar{w}_b \)

Margin of a Linear Classifier

Definition: For a linear classifier \( h_{\alpha} \), the margin \( \delta \) of an example \((x, y)\) is \( \delta = y (\bar{w} \cdot x) \).

Definition: The margin is called geometric margin, if \( ||\bar{w}|| = 1 \). Otherwise, functional margin.

Definition: The (hard) margin of an unbiased linear classifier \( h_{\alpha} \) on a sample \( S \) is \( \delta = \min_{(x, y) \in S} \delta_{\alpha}(x) \).

Definition: The (hard) margin of an unbiased linear classifier \( h_{\alpha} \) on a task \( P \) \((X, Y)\) is \( \delta = \inf_{X \sim P, Y \sim Q_{\text{add}}} \min_{(x, y) \in S} \delta_{\alpha}(x) \).
(Batch) Perceptron Algorithm

Input: \( S = \{(x_1, y_1), \ldots, (x_n, y_n)\} \), \( x_i \in \mathbb{R}^N \), \( y_i \in \{-1, 1\} \), \( \eta \in \mathbb{R} \), \( i \in \{1, 2, \ldots\} \)

Algorithm:
- \( w_0 = \vec{0}, k = 0 \)
- repeat
  - FOR \( i = 1 \) TO \( n \)
    - IF \( y_i (\langle w_k, x_i \rangle) \leq 0 \) ### makes mistake
      - \( w_{k+1} = w_k + \eta y_i x_i \)
      - \( k = k + 1 \)
    - ENDIF
  - ENDFOR
- until / iterations reached

Example: Reuters Text Classification