Online Learning and Perceptron Mistake Bound

CS4780/5780 – Machine Learning
Fall 2013

Thorsten Joachims
Cornell University

Reading: Mitchell Chapter 7.5
Cristianini/Shawe-Taylor Chapter 2-2.1.1
Online Learning Model

• Initialize hypothesis $h \in H$
• FOR i from 1 to infinity
  – Receive $x_i$
  – Make prediction $\hat{y}_i = h(x_i)$
  – Receive true label $y_i$
  – Record if prediction was correct (e.g., $\hat{y}_i = y_i$)
  – Update $h$
(Online) Perceptron Algorithm

- **Input:** \( S = ((\vec{x}_1, y_1), \ldots, (\vec{x}_n, y_n)) \), \( \vec{x}_i \in \mathbb{R}^N \), \( y_i \in \{-1, 1\} \)
- **Algorithm:**
  - \( \vec{w}_0 = \vec{0}, \ k = 0 \)
  - **FOR** \( i=1 \) **TO** \( n \)
    * **IF** \( y_i (\vec{w}_k \cdot \vec{x}_i) \leq 0 \) # makes mistake
      * \( \vec{w}_{k+1} = \vec{w}_k + y_i \vec{x}_i \)
      * \( k = k + 1 \)
    * **ENDIF**
  - **ENDFOR**
- **Output:** \( \vec{w}_k \)
Theorem: For any sequence of training examples $S = ((\tilde{x}_1, y_1), \ldots, (\tilde{x}_n, y_n)$ with

$$R = \max \| \tilde{x}_i \|,$$

if there exists a weight vector $\vec{w}_{\text{opt}}$ with $\| \vec{w}_{\text{opt}} \| = 1$ and

$$y_i (\vec{w}_{\text{opt}} \cdot \tilde{x}_i) \geq \delta$$

for all $1 \leq i \leq n$, then the Perceptron makes at most

$$\frac{R^2}{\delta^2}$$

errors.
Margin of a Linear Classifier

Definition: For a linear classifier $h_w$, the margin $\delta$ of an example $(\bar{x}, y)$ with $\bar{x} \in \mathbb{R}^N$ and $y \in \{-1, +1\}$ is $\delta = y(\bar{w} \cdot \bar{x})$.

Definition: The margin is called geometric margin, if $||\bar{w}|| = 1$. For general $\bar{w}$, the term functional margin is used to indicate that the norm of $\bar{w}$ is not necessarily 1.

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

Definition: The (hard) margin of an unbiased linear classifier $h_{\bar{w}}$ on a task $P(X, Y)$ is

$$\delta = \inf_{S \sim P(X, Y)} \min_{(\bar{x}, y) \in S} y(\bar{w} \cdot \bar{x}).$$