# 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: $S = ((\vec{x}_1, y_1), ..., (\vec{x}_n, y_n))$, $\vec{x}_i \in \Re^N$, $y_i \in \{-1, 1\}$
- 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: $\vec{w}_k$

## Perceptron Mistake Bound

Theorem: For any sequence of training examples $S = ((\vec{x}_1, y_1), \dots, (\vec{x}_n, y_n))$ with

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

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

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

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 $(\vec{x}, y)$ with $\vec{x} \in \Re^N$ and $y \in \{-1, +1\}$ is $\delta = y(\vec{w} \cdot \vec{x})$.*

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

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

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

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