

# Linear Classifiers and Perceptrons

CS4780/5780 – Machine Learning  
Fall 2013

Thorsten Joachims  
Cornell University

Reading: Mitchell Chapter 4.4-4.4.2

# Example: Spam Filtering

	viagra	learning	the	dating	nigeria	<i>spam?</i>	
$\vec{x}_1 = ($	1	0	1	0	0	$)$	$y_1 = -1$
$\vec{x}_2 = ($	0	1	1	0	0	$)$	$y_2 = +1$
$\vec{x}_3 = ($	0	0	0	0	1	$)$	$y_3 = -1$

- 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: 
$$h_{\vec{w}}(\vec{x}) = \begin{cases} +1 & w_1x_1 + \dots + w_Nx_N > 0 \\ -1 & \text{else} \end{cases}$$

- biased: 
$$h_{\vec{w},b}(\vec{x}) = \begin{cases} +1 & w_1x_1 + \dots + w_Nx_N + b > 0 \\ -1 & \text{else} \end{cases}$$

- Parameter vector  $\vec{w}$ , scalar  $b$

- Hypothesis space  $H$

- $H_{unbiased} = \{ h_{\vec{w}}: \vec{w} \in \mathfrak{R}^N \}$

- $H_{biased} = \{ h_{\vec{w},b}: \vec{w} \in \mathfrak{R}^N, b \in \mathfrak{R} \}$

- Notation

- $w_1x_1 + \dots + w_Nx_N = \vec{w} \cdot \vec{x}$     and     $sign(a) = \begin{cases} +1 & a > 0 \\ -1 & \text{else} \end{cases}$

- $h_{\vec{w}}(\vec{x}) = sign(\vec{w} \cdot \vec{x})$

- $h_{\vec{w},b}(\vec{x}) = sign(\vec{w} \cdot \vec{x} + b)$

# (Batch) Perceptron Algorithm

Input:  $S = ((\vec{x}_1, y_1), \dots, (\vec{x}_n, y_n))$ ,  $\vec{x}_i \in \mathbb{R}^N$ ,  $y_i \in \{-1, 1\}$ ,  
 $I \in [1, 2, \dots]$

Algorithm:

- $\vec{w}_0 = \vec{0}$ ,  $k = 0$
  - repeat
    - 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
- until  $I$  iterations reached

Training Data:

	$x_1$	$x_2$	$y$
$\vec{x}_1 = ($	1	2	$y_1 = 1$
$\vec{x}_2 = ($	2	1	$y_2 = 1$
$\vec{x}_3 = ($	-1	-1	$y_3 = -1$
$\vec{x}_4 = ($	-1	1	$y_3 = -1$

# Example: Reuters Text Classification

