Lecture 5

Chiranjib Bhattacharyya

chibha@chalmers.se

February 16, 2012

Agenda

- Perceptron
- Support vector machines

Linear Classifiers

Let us a define a linear classifier w, b on $D = \{(x_i, y_i) | i = 1, \ldots, n\}$ such that

$$y_i = sign(w^\top x_i + b) \;\; i = 1, \ldots, n$$

Linear Classifiers

Let us a define a linear classifier w, b on $D = \{(x_i, y_i) | i = 1, \ldots, n\}$ such that

$$y_i = sign(w^\top x_i + b) \;\; i = 1, \ldots, n$$

equivalent to

$$y_i(w^\top x_i + b) \geq 0 \;\; i = 1, \ldots, n$$

Perceptron Algorithm

Let $\tilde{w} = [w^\top b]^\top \;\; \tilde{x} = [x^\top 1]^\top$.
mistake happens if $y \neq sign(\tilde{w}^\top \tilde{x})$

Perceptron Algorithm

Set $\tilde{w}_1 = 0$, $t = 0$
For each example $x \in D$
check if $\tilde{w}(t)$ makes a mistake on x. then update
$\tilde{w}(t+1) = \tilde{w}(t) + y_t\tilde{x}$, and $t \rightarrow t+1$
Continue till there are no mistakes on the training set

Perceptron Algorithm

Theorem

Let $D = \{(\tilde{x}_i, y_i) | i = 1, \ldots, n\}$ and $\|\tilde{x}_i\| = 1$ $i = 1, \ldots, n$. If there exists $\tilde{w}^*, \|\tilde{w}^*\| = 1$ and γ such that

$$\gamma = min_{x_i \in D} y_i \tilde{x}_i^\top \tilde{w}^*$$

then the perceptron algorithm converges after making atmost $\frac{1}{\gamma^2}$mistakes

Proof Sketch: After every update

$$\tilde{w}_{t+1}^\top \tilde{w}^* \geq \tilde{w}_t^\top \tilde{w}^* + \gamma \quad \|\tilde{w}_{t+1}\|^2 \leq \|\tilde{w}_t\|^2 + 1$$

Perceptron Algorithm

Proof:
if T is the number of mistakes

$$\tilde{w}_{T+1}^\top \tilde{w}^* \geq T\gamma \quad \|\tilde{w}_{T+1}\| \leq \sqrt{T}$$

For any two vectors $u^\top v \leq \|u\| \|v\|$

$$T\gamma \leq \tilde{w}_{T+1}^\top \tilde{w}^* \leq \|\tilde{w}_{T+1}\| \leq \sqrt{T}$$

and the claim follows

Large margin

One can think of an alternate classifier

$$max_{w,b,\gamma}\gamma$$

$$\frac{y_i(w^\top x_i + b)}{\|w\|} \geq \gamma$$

Large margin

One can think of an alternate classifier

$$max_{w,b,\gamma}\gamma$$

$$\frac{y_i(w^\top x_i + b)}{\|w\|} \geq \gamma$$

This is equivalent to

Support Vector Machines

$$min_{w,b}\frac{1}{2}\|w\|^2$$

$$y_i(w^\top x_i + b) \geq 1$$

Some Interesting facts

$$min_x f(x) \ \ s.t. g_i(x) \geq 0 i = 1, \ldots, n \ \ \text{(P)}$$

$$\text{Lagrangian} L(x, \alpha) = f(x) - \sum_{i=1}^{n} \alpha_i g_i(x)$$

$\alpha_i \geq 0$
K.K.T conditons

$$\frac{\partial L}{\partial x} = 0 \ g_i(x) \geq 0$$

$$\alpha_i g_i(x) = 0, \alpha_i \geq 0$$

SVMs

At optimality $\alpha_i \geq 0$

$$w = \sum_{i=1}^{n} \alpha_i y_i x_i$$

$$b = y_i - w^\top x_i \;\; \alpha_i > 0$$

SVM Dual

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^\top x_j$$

$$\alpha_i \geq 0 \;\; \sum_{i=1}^{n} \alpha_i y_i = 0$$