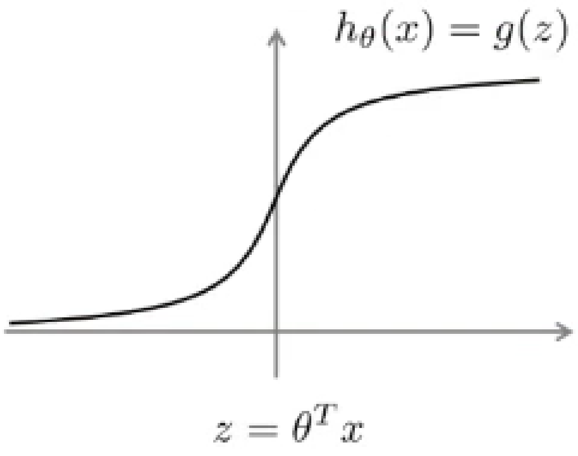
Machine Learning

Support Vector Machines

Large Margin Intuition

Alternative view of logistic regression

$$h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$$



If y=1 we want $h_\theta(x) \approx 1, \theta^T x >> 0$

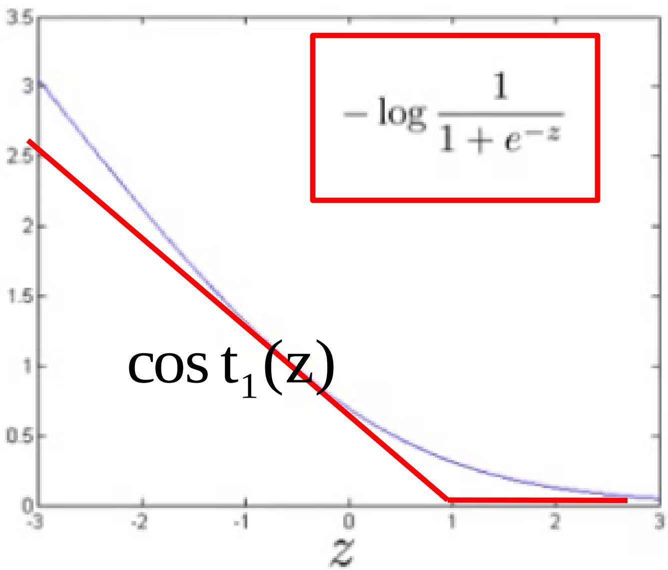
If y=0 we want $h_\theta(x) \approx 0, \theta^T x << 0$

Alternative view of logistic regression

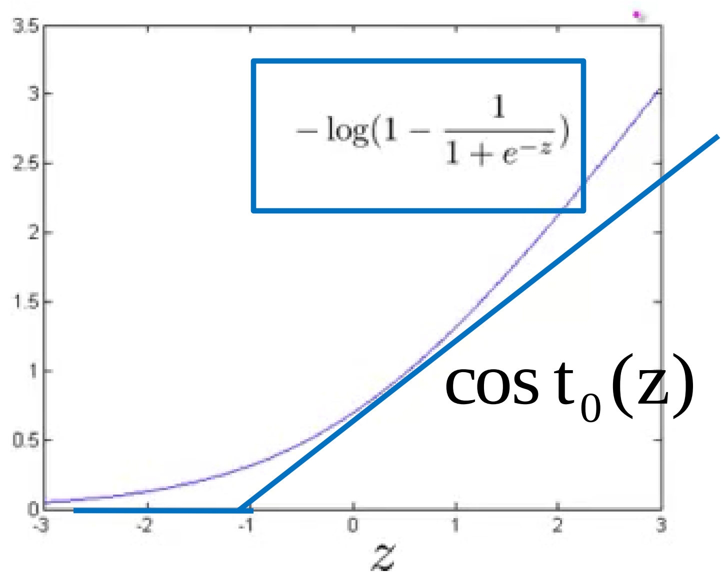
Cost of example: $-(y\log h_\theta(x)+(1-y)\log(1-h_\theta(x)))$

$$=-y\boxed{\log\frac{1}{1+e^{-\theta^T x}}}-(1-y)\boxed{\log(1-\frac{1}{1+e^{-\theta^T x}})}$$

If y=1 (want $\theta^T x >> 0$):

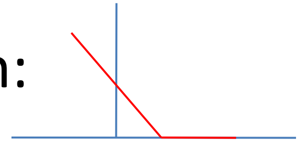


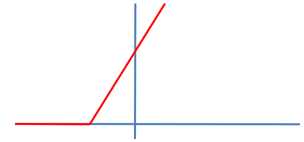
If y=0 (want $\theta^T x << 0$):



Support vector machine

Logistic regression:





$$\min_\theta \frac{1}{m}\left[\sum_{i=1}^{m} y^{(i)}\underbrace{(-\log h_\theta(x^{(i)})}_{\cos t_1(\theta^T x^{(n)})} + (1-y^{(i)})\underbrace{\log(1-h_\theta(x^{(i)}))}_{\cos t_0(\theta^T x^{(n)})})\right] + \underbrace{\frac{\lambda}{2m}\sum_{j=1}^{n}\theta_j^{\,2}}$$

Support vector machine

A B

$$\min_\theta \frac{1}{\cancel{m}}\left[\sum_{i=1}^{m} y^{(i)}\cos t_1(\theta^T x^{(n)}) + (1-y^{(i)})\cos t_0(\theta^T x^{(n)})\right] + \frac{\lambda}{2\cancel{m}}\sum_{j=1}^{n}\theta_j^{\,2}$$

A + λ B

$$C = \frac{1}{\lambda}$$

CA + B

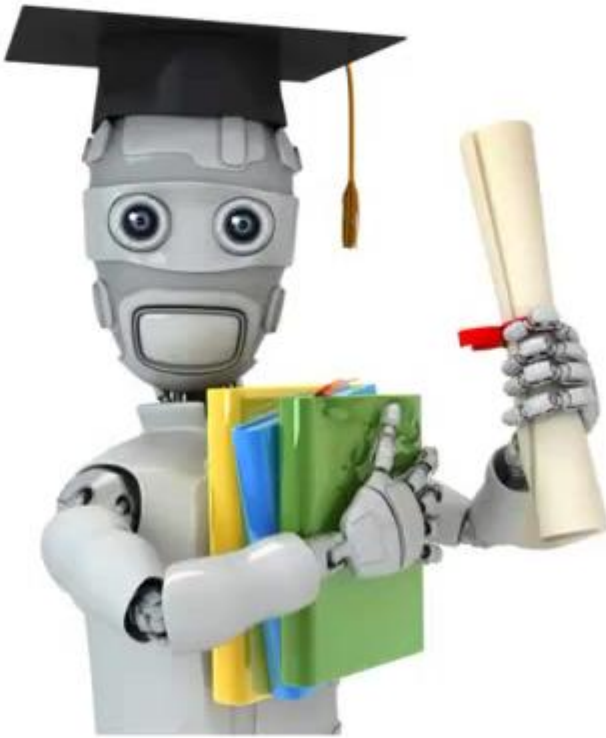
$$\min_\theta C\left[\sum_{i=1}^{m} y^{(i)}\cos t_1(\theta^T x^{(n)}) + (1-y^{(i)})\cos t_0(\theta^T x^{(n)})\right] + \frac{1}{2}\sum_{j=1}^{n}\theta_j^{\,2}$$

SVM hypothesis

$$\min_{\theta} C[\sum_{i=1}^{m} y^{(i)} \cos t_1(\theta^T x^{(n)}) + (1 - y^{(i)}) \cos t_0(\theta^T x^{(n)})] + \frac{1}{2}\sum_{j=1}^{n} \theta_j{}^2$$

Hypothesis:

$$h_{\theta}(x) = \begin{Bmatrix} 1\text{.......if } (\theta^T x \geq 0) \\ 0\text{......otherwise} \end{Bmatrix}$$



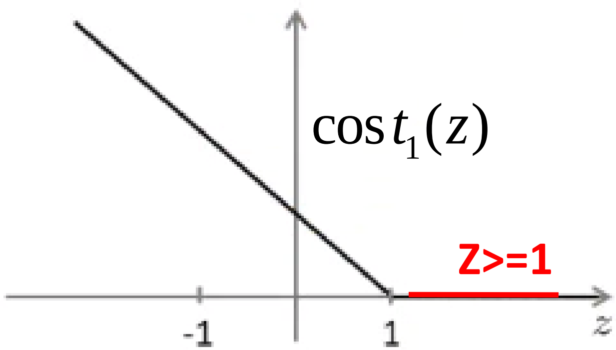
Machine Learning

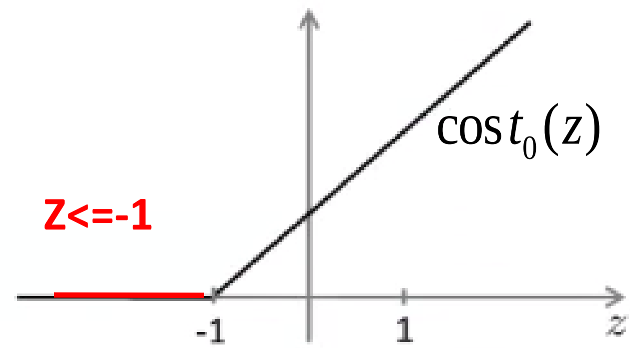
Support Vector Machines

Optimization objective

Support Vector Machine

$$\min C\sum_{i=1}^{m}[y^{(i)}\cos t_1(\theta^T x^{(i)})+(1-y^{(i)})\cos t_0(\theta^T x^{(i)})]+\frac{1}{2}\sum_{j=1}^{n}\theta_j^{\,2}$$





If y=1 ,we want $\theta^T x \geq 1$ (not just>=0) $\theta^T x \geq \cancel{0}$ 1

If y=0 ,we want $\theta^T x \leq -1$ (not just<0) $\theta^T x \leq \cancel{0}$ -1

SVM Decision Boundary

$$\min C\sum_{i=1}^{m}[y^{(i)}\cos t_1(\theta^T x^{(i)})+(1-y^{(i)})\cos t_0(\theta^T x^{(i)})]+\frac{1}{2}\sum_{j=1}^{n}\theta_j^{\,2}$$

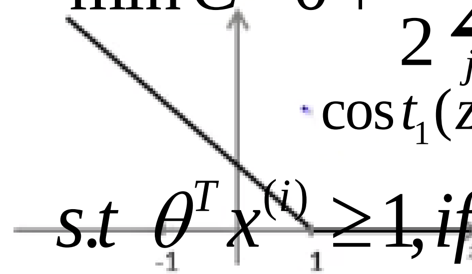
C is very large

=0

$$\min C*0+\frac{1}{2}\sum_{j=1}^{n}\theta_j^{\,2}$$

Whenever $y^{(i)}=1$

$$\theta^T x\geq 1$$

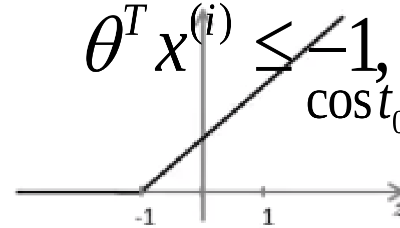


$\cos t_1(z)$

$$s.t\ \ \theta^T x^{(i)}\geq 1, if\,(y^{(i)}=1)$$

$$\theta^T x^{(i)}\leq -1, if\,(y^{(i)}=0)$$

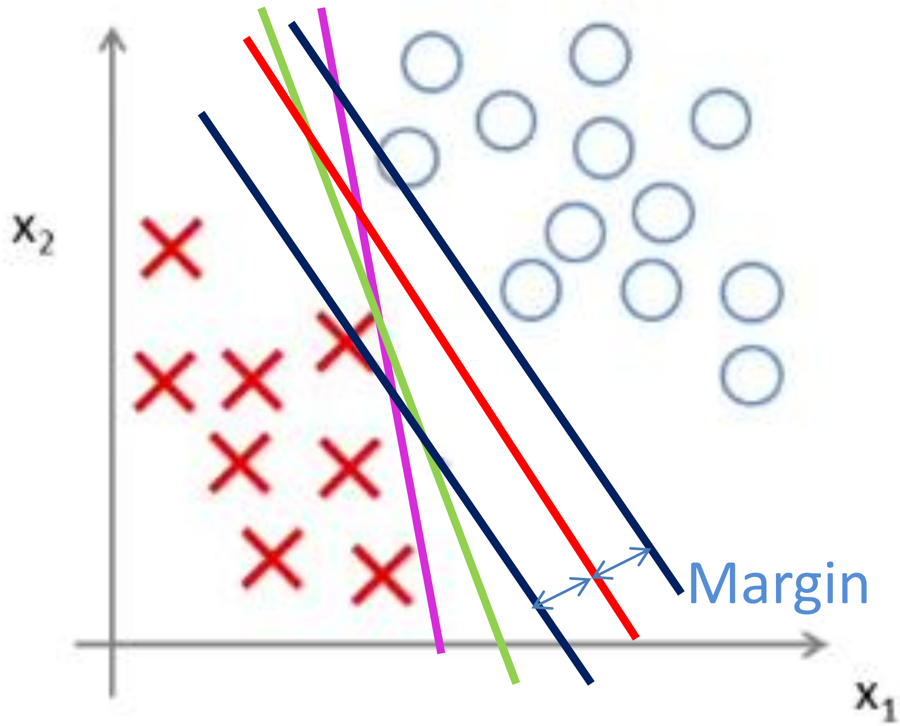
Whenever $y^{(i)}=0$



$\cos t_0(z)$

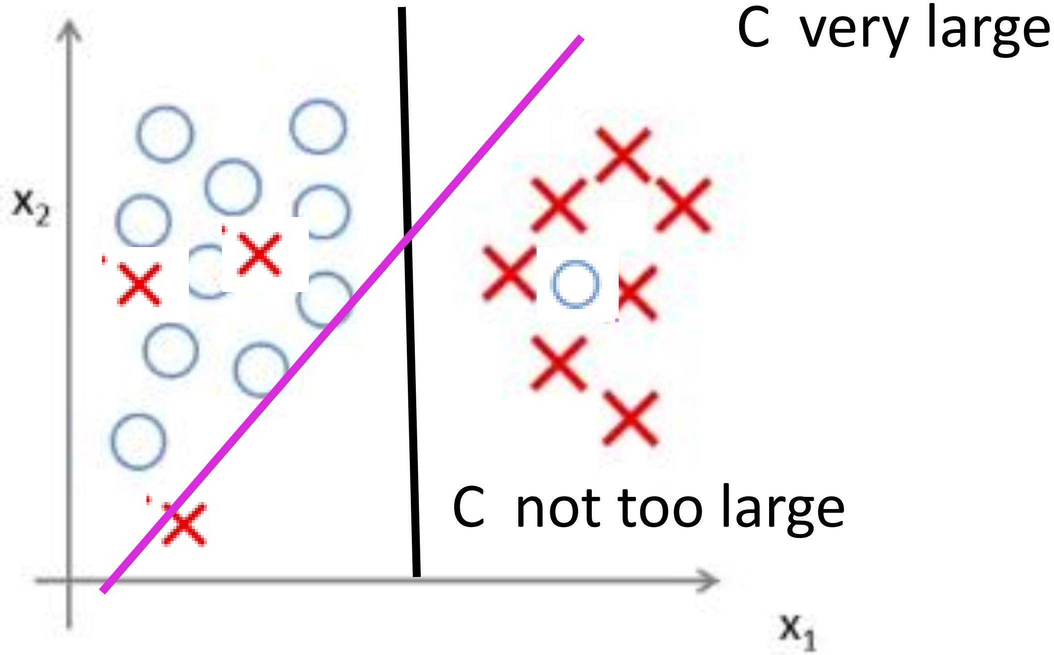
$$\theta^T x\leq -1$$

SVM Decision Boundary:Linearly separable case



Large margin classifier

Large margin classifier in presence of outliers



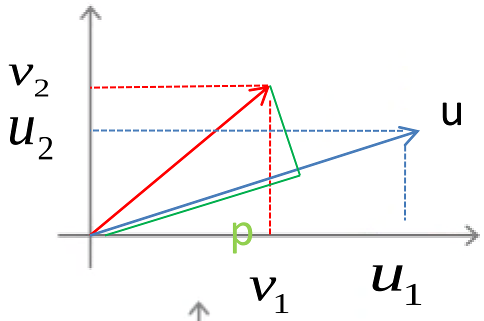
Machine Learning

Support Vector Machines

The mathematics behind large margin classification (optional)

Vector Inner Product

$$u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \qquad v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

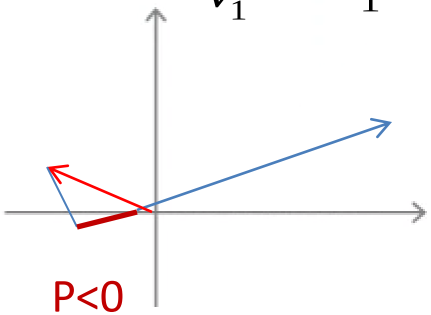


$$u^T v = ?$$

P =length of projection of v onto u

$$u^T v = p.\|u\| \qquad = v^T u$$

$$= u_1 v_1 + u_2 v_2$$



SVM Decision Boundary

$$\min \frac{1}{2}\sum_{j=1}^{n}\theta_j^{\,2} = \frac{1}{2}(\theta_1^{\,2}+\theta_2^{\,2}) = \frac{1}{2}(\sqrt{\theta_1^{\,2}+\theta_2^{\,2}})^2 = \frac{1}{2}\|\theta\|^2$$

$$s.t\ \ \theta^T x^{(i)} \ge 1 \quad if, y^{(i)} = 1$$

$$\theta^T x^{(i)} \le -1 \quad if, y^{(i)} = 0$$

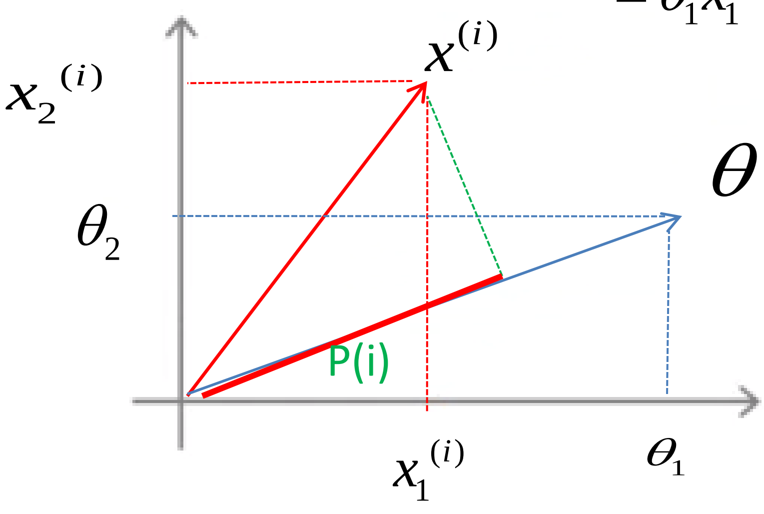
$$simplification: \theta_0 = 0, n = 2$$

$$\theta^T x_2^{(i)} = P^{(i)}\|\theta\|$$

$$= \theta_1 x_1^{(i)} + \theta_2 x_2^{(i)}$$

$$\theta^T x^{(i)} = ?$$

$$u^T v$$



SVM Decision Boundary

$$\min \frac{1}{2}\sum_{j=1}^{n}\theta_j^{\,2} = \frac{1}{2}\|\theta\|^2$$

$$s.t\; P^{(i)}\|\theta\| \geq 1 \; if, y^{(i)} = 1$$

$$P^{(i)}\|\theta\| \leq -1 \; if, y^{(i)} = 0$$

Where $p^{(i)}$ is the projection of $x^{(i)}$ onto the vector θ

$$simplification: \theta_0 = 0$$

$P^{(1)}\|\theta\| \geq 0$