Support vector machines

True nondeterminism: you make a prediction, but the final outcome is yet to be determined.

Why might predictions be wrong: ① Suppose $P(\text{heads}) = 0.5$, yet due to non-determinism, you observe 0001000 ...
which is of partial observability

② Noise in the observation, ③ representational bias, ...

Rephrase the parameters: f is changed to h_θ, w is changed to θ
decision boundary $= w^T x \Rightarrow \theta^T x = 0$, if orthogonal

$$\langle \theta, x \rangle = \theta^T x = \theta_1 x_1 + \cdots + \theta_D x_D = \sum_i \theta_i x_i$$

Decision function: $h(x) = \text{sign}(\theta^T x)$

Maximize the margin to allow for more (measurement & otherwise) error in the representation of samples.

Alternative view on Logistic regression: $\sigma(z) = \frac{1}{1+e^{-z}} = \frac{1}{1+e^{-w^T x}} \Leftrightarrow \frac{1}{1+e^{-\theta^T x}}$

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

Cost: $J(\theta) = -\sum_i^N [y_i \underbrace{\log h_\theta(x_i)}_{\text{cost1}} + (1-y_i) \underbrace{\log(1-h_\theta(x_i))}_{\text{cost0}}]$

Loss: $L(\theta) = -y_i \log \sigma(z) - (1-y_i)\log(1-\sigma(z))$

$$= -(y_i \log h_\theta(x_i) + (1-y_i)\log(1-h_\theta(x_i)))$$

Logistic Regression:

$$\text{Loss} = -t\log y - (1-t)\log(1-y)$$

Notation change $= -y_i \log h_\theta(x_i) - (1-y_i)\log(1-h_\theta(x_i))$

$$= -[(y_i \log h_\theta(x_i) + (1-y_i)\log(1-h_\theta(x_i))]$$

$J_{reg}(\theta)$ for logistic regression =

$\min\limits_{\theta}$ means "minimize this expression's value by changing θ"

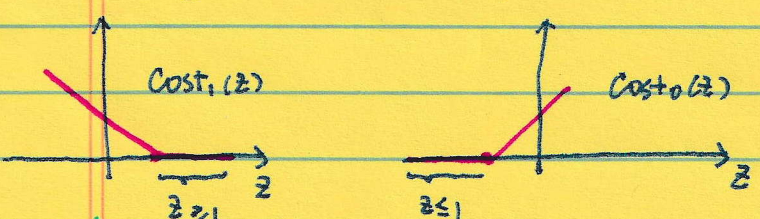
$$J_{reg}(\theta)_{LR} = \boxed{-}\sum_{i}^{N}(y_i \log h_\theta(x_i) + (1-y_i)\log(1-h_\theta(x_i))) + \frac{1}{2}\cdot\boxed{\lambda}\cdot\sum_{j}^{d}\theta_j^2$$

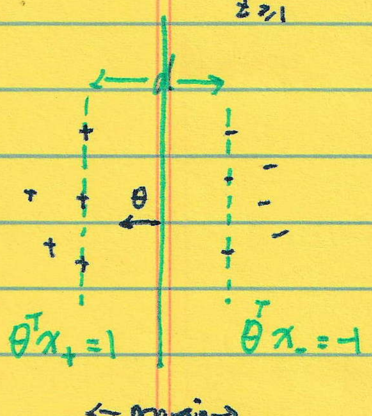
Remark: we didn't divide this J by $\frac{1}{N}$, because minimizing $J_{reg}(\theta)_{LR}$ and $\frac{1}{N}J_{reg}(\theta)_{LR}$ are equivalent. (in terms of what θ yields the minimized vals for both expressions)

$$J_{reg}(\theta)_{SVM} = \boxed{C}\sum_{i}^{N}[y_i\,\text{cost}_1(\theta^T x_i) + (1-y_i)\,\text{cost}_0(\theta^T x_i)] + \frac{1}{2}\sum_{j}^{d}\theta_j^2$$

Remark: In logistic regression, the negative sign is used because we are maximizing likelihood but using a minimization approach. In SVM, C is positive because we are directly minimizing classification error.

$$\ell_{hinge} = \max(0,\ 1-y\cdot h(x))$$





direction · magnitude → Δx

direction represented by unit vector, $\hat{n}$

$\hat{n} = \dfrac{\theta}{\|\theta\|_2}$, $[\theta^T x_+ = 1] - [\theta^T x_- = -1]$

$\Rightarrow \theta^T \cdot \Delta x = 2$

$$\hat{n}\cdot\Delta x = \frac{\theta}{\|\theta\|_2}\cdot\Delta x = \frac{2}{\|\theta\|_2}$$

Cost Function $J_{reg}(\theta)_{svm} = C \cdot \sum_i^N y_i \, Cost_1(\theta^T x) + (1-y_i) Cost_0(\theta^T x) + \frac{1}{2} \sum_j^d \theta_j^2$

C very large: prioritize minimizing classification error
C very small: prioritize maximizing the margin

$u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$ $v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ $\|u\|_2 = \sqrt{u_1^2 + u_2^2}$ = length of u

$\|u\|_2^2 = u_1^2 + u_2^2$

$u^T v = v^T u = \|u\|_2 \cdot \|v\|_2 \cdot \cos\theta$, in which θ is the angle (not the parameter)

$= \|u\|_2 \cdot p$, in which p is the projection $= \|v\|_2 \cdot \cos\theta$

p is small: $\|\theta\|_2$ must be large ; p is large: $\|\theta\|_2$ can be smaller

Suppose we want to only maximize the margin: $\frac{1}{2} \sum_j^d \theta_j^2$. s.t. $y_i(\theta^T x_i + b) \geq 1$

$\frac{1}{2} \sum_{j=1}^d \theta_j^2 - \sum_{i=1}^n a_i (y_i(\theta^T x_i + b) - 1)$. goal: $\theta\downarrow$, margin$\uparrow$, $a_0 \dots a_{N-1} \uparrow$

$= \frac{1}{2} \sum_{j=1}^d \theta_j^2 - (\sum_{i=1}^n a_i (y_i(\theta^T x_i + b) - \sum_{i=1}^n a_i$

$= \frac{1}{2} \sum_{j=1}^d \theta_j^2 - (\sum_i^n a_i y_i \theta^T x_i + b \cdot \sum_{i=1}^n a_i y_i - \sum_{i=1}^n a_i)$

$\frac{\partial J}{\partial \theta} = \theta - \sum_i^n a_i y_i x_i = 0$ (critical point) $\Rightarrow \theta = \sum_i^n a_i y_i x_i$

$\frac{\partial J}{\partial b} = 0 - (0 + \sum_{i=1}^n a_i y_i - 0) = 0$ (critical point) $\Rightarrow -\sum_{i=1}^n a_i y_i = 0$

$\theta = \sum_{i}^{N} a_i y_i x_i \; ; \; \sum_{i=1}^{N} a_i y_i = 0$

$\frac{1}{2} \sum_{j}^{d} \theta_j^2 - \sum_{i=1}^{n} a_i (y_i (\theta^T x_i + b) - 1)$

$= \frac{1}{2} \sum_{j}^{d} \theta_j^2 - \left(\sum_{i=1}^{n} a_i y_i (\theta^T x_i + b) - \sum_{i=1}^{n} a_i \right)$

$= \frac{1}{2} \sum_{j=1}^{d} \theta_j^2 - \left(\sum_{i=1}^{n} a_i y_i \theta^T x_i + b \cdot \overset{0}{\cancel{\sum_{i=1}^{n} a_i}} y_i - \sum_{i=1}^{n} a_i \right)$

$= \frac{1}{2} \sum_{j=1} \sum_{i=1} a_i a_j \, y_i \cdot y_j \cdot x_i \cdot x_j - \left(\sum_{j=1} \sum_{i=1} a_i y_i \cdot a_j \cdot y_j \cdot x_i x_j - \sum_{i=1} a_i \right)$

$= -\frac{1}{2} \sum_{j=1} \sum_{i=1} a_i a_j \, y_i \cdot y_j \cdot x_i \cdot x_j + \sum a_i$

$h(x) = y = \theta^T x + b \Rightarrow h(x) = \mathrm{sign}(\theta^T x + b)$

$= \mathrm{sign}\left(\sum_{i}^{N} a_i y_i \cdot x_i \cdot x + b \right) \; ; \quad y_i = \sum_{j}^{N} a_j y_j x_i \cdot x_j + b$

$b = \frac{1}{N} \sum_{i}^{N} y_i - \sum_{j \in N} a_j y_j x_i \cdot x_j$

$\Leftrightarrow \frac{1}{|SV|} \cdot \sum_{i \in SV} \left(y_i - \sum_{j \in SV} a_j y_j x_i x_j \right)$

Let $x_i = [x_{i1}, x_{i2}]$, $x_j = [x_{j1}, x_{j2}]$; $k(x_i, x_j) = \langle x_i, x_j \rangle^2$

$k(x_i, x_j) = (x_{i1} \cdot x_{j1} + x_{i2} \cdot x_{j2})^2$

$= (x_{i1}^2 \cdot x_{j1}^2 + x_{i2}^2 \cdot x_{j2}^2 + 2 x_{i1} x_{j1} x_{i2} \cdot x_{j2}) = \langle \Phi(x_i), \Phi(x_j) \rangle$

$\Phi(x_i) = [x_{i1}^2, x_{i2}^2, \sqrt{2} x_{i1} \cdot x_{i2}]$

$\Phi(x_j) = [x_{j1}^2, x_{j2}^2, \sqrt{2} x_{j1} \cdot x_{j2}]$

$$J(a) = \sum_{i}^{n} a_i - \frac{1}{2} \sum_{i} \sum_{j} a_i a_j y_i y_j \langle x_i \cdot x_j \rangle$$

$$\Downarrow$$

$$\tilde{J}(a) = \sum_{i} a_i - \frac{1}{2} \sum_{i} \sum_{j} a_i a_j y_i y_j \, k(x_i, x_j)$$

$$k(x_i, x_j) = \exp\left(- \frac{\| x_i - x_j \|_2^2}{2\sigma^2}\right)$$

$\theta_0 + \theta_1(1) + \theta_2(0) + \theta_3(0) = 0.5 > 0$; when at page 47

$\theta_0 + \theta_1(0) + \theta_2(0) + \theta_3(1) = -0.5 < 0$; when at page 41