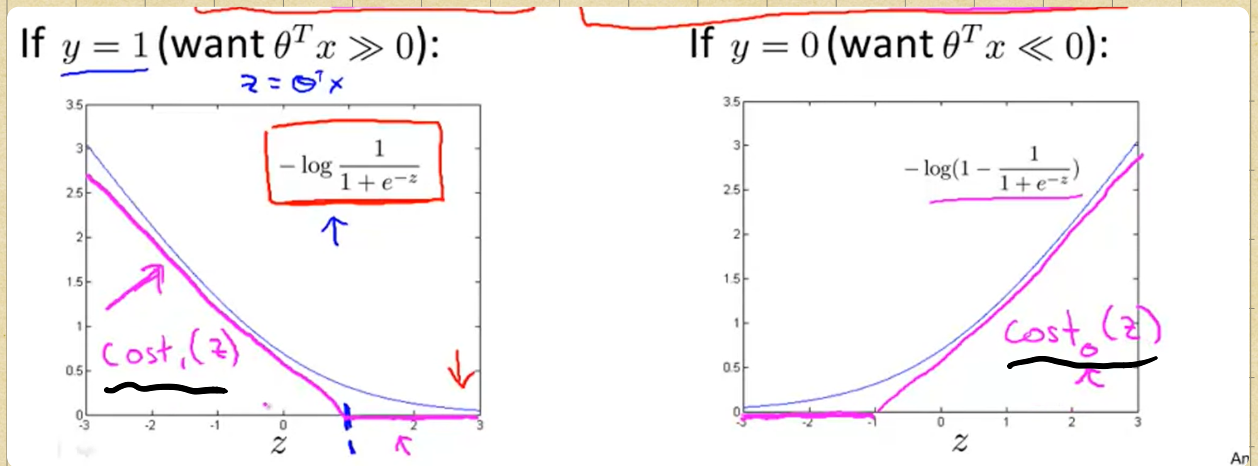
Logistic Regression

$$J = -y\log\frac{1}{1+e^{-\theta^T x}} - (1-y)\log\left(1-\frac{1}{1+e^{-\theta^T x}}\right).$$



SVM:

$$A + \lambda B$$

$$\min_\theta \frac{1}{m}\sum_{i=1}^{m} y^{(i)} cost_1(\theta^T x^{(i)}) + (1-y^{(i)}) cost_0(\theta^T x^{(i)}) + \frac{\lambda}{2m}\sum_{j=1}^{n}\theta_j^2$$

$$\Downarrow$$

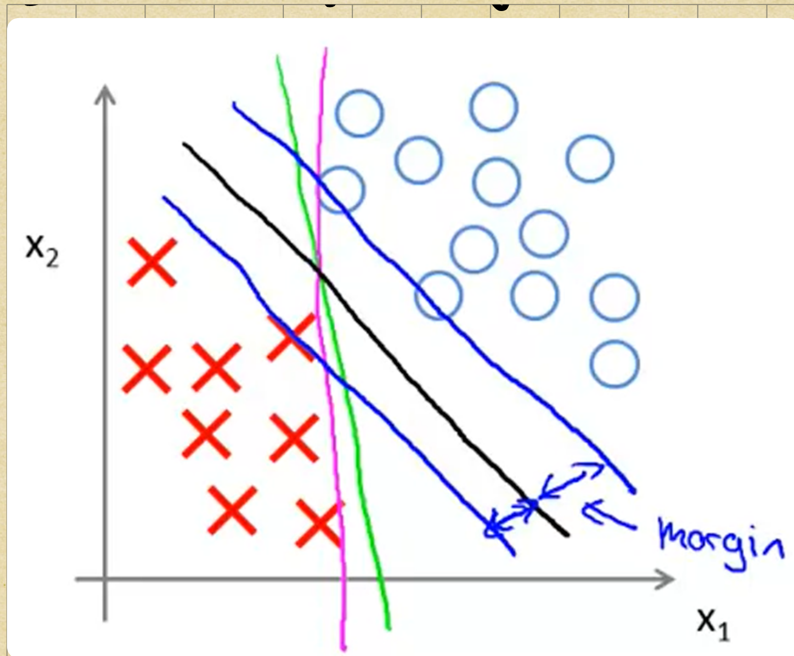
$$\min_\theta C\sum_{i=1}^{m}\left[y^{(i)} cost_1(\theta^T x^{(i)}) + (1-y^{(i)}) cost_0(\theta^T x^{(i)})\right] + \frac{1}{2}\sum_{j=1}^{n}\theta_j^2$$

$$C = \frac{1}{\lambda} \qquad CA + B$$

if $y=1$ $\quad \theta^T x \geq 1$

if $y=0$ $\quad \theta^T x \leq -1$

SVM $\rightarrow$ Large margin classifier



$$h(x) = \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 \cdots$$

$$\Downarrow$$

$$h(x) = \theta_1 f_1 + \theta_2 f_2 + \theta_3 f_3 \cdots$$

distance

$$f_i = \text{similarity}(x, l^{(i)}) = \exp\left(-\frac{\|x - l^{(i)}\|^2}{2\sigma^2}\right)$$

$$\shortparallel$$

kernel (Gaussian kernel) $k(x, l^{(i)})$

For training example $(x^{(i)}, y^{(i)})$

$$x^{(i)} \longrightarrow \begin{array}{l} f_1^{(i)} = \text{sim}(x^{(i)}, l^{(1)}) \\ f_2^{(i)} = \text{sim}(x^{(i)}, l^{(2)}) \\ \vdots \\ f_m^{(i)} = \text{sim}(x^{(i)}, l^{(m)}) \end{array}$$

Training

$$\min C \sum_{i=1}^{m} y^{(i)} cost_1(\theta^T f^{(i)}) + (1-y^{(i)}) cost_0(\theta^T f^{(i)}) + \frac{1}{2}\sum_{j=1}^{m} \theta_j^2$$

Large C bias ↓ variance ↑

Small C bias ↑ variance ↓

Large σ^2 f_i smoothly bias ↑ variance ↓

Small σ^2 f_i less smoothly bias ↓ variance ↑

Polynomial kernel : $(x^T l + \text{const})^{\text{degree}}$