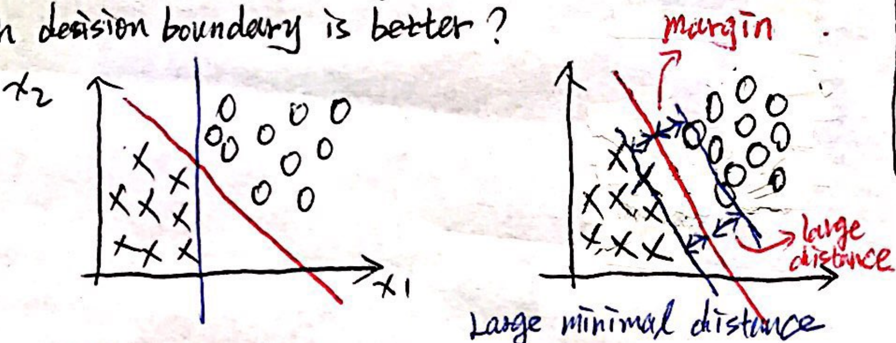
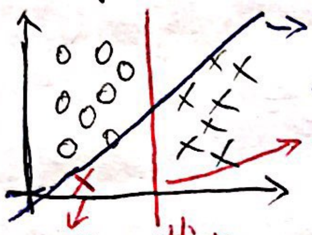
# Large Margin Classification.

Which decision boundary is better?




Large minimal distance

Large Margin classifier is robust to outliers.



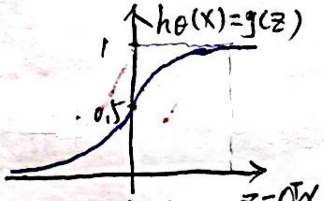

→ not good boundary sensitive to outliers.

→ good decision boundary robust to outliers.

---

# Logistic Regression.

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




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

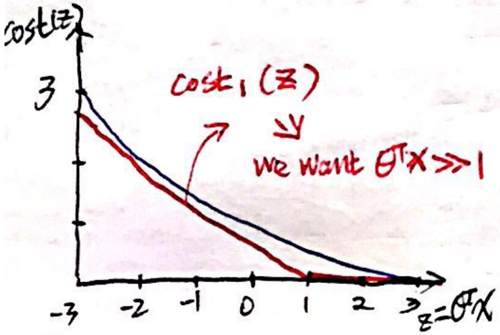
if $y=0$, we want $h_\theta(x) \approx 0 \Rightarrow \theta^T x << 0$

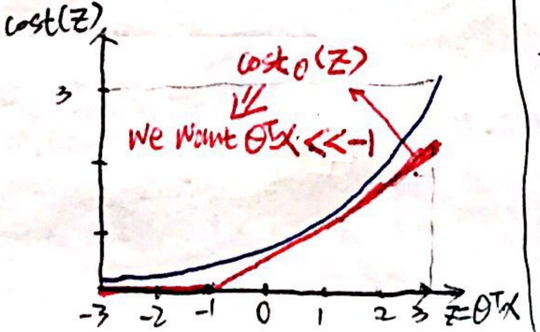
## Cost Function (Logistic Regression)

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log\left(h_\theta(x^{(i)})\right)+(1-y^{(i)})\log\left(1-h_\theta(x^{(i)})\right)\right]$$

$\boxed{\text{For one data point} \quad -y\log\frac{1}{1+e^{-\theta^T x}} - (1-y)\log\left(1-\frac{1}{1+e^{-\theta^T x}}\right)}$

if $y=1 \Rightarrow \theta^T x >> 0$ if $y=0 \Rightarrow \theta^T x << 0$




---

# Support Vector Machine

Cost Function $J(\theta) = \min_\theta \frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\text{cost}_1(\theta^T x^{(i)})+(1-y^{(i)})\text{cost}_0(\theta^T x^{(i)})\right] + \frac{\lambda}{2m}\sum_{j=0}^{n}\theta_j^2$

balance Variance & Bias ⇓ $A+\lambda B$ $CA+B$ $C \approx \frac{1}{\lambda}$

$$\min_\theta \boxed{C}\sum_{i=1}^{m}\left[y^{(i)}\text{cost}_1(\theta^T x^{(i)})+(1-y^{(i)})\text{cost}_0(\theta^T x^{(i)})\right]+\frac{1}{2}\sum_{j=1}^{n}\theta_j^2$$

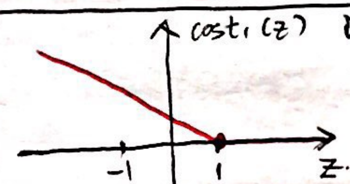
Question: ① $\min_\theta \frac{1}{m}f(\theta)+\frac{\lambda}{2m}\sum_{j=1}^{n}\theta_j^2$ ⇒ if ① ≃ ② the $C = \boxed{\frac{1}{\lambda}}$ $\boxed{\lambda}$ $\boxed{\frac{2}{\lambda}}$ $\boxed{-\lambda}$

② $\min_\theta C f(\theta)+\frac{1}{2}\sum_{j=1}^{n}\theta_j^2$

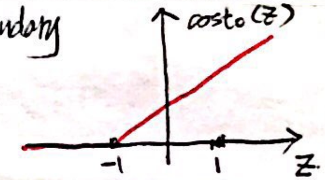
SVM Hypothesis: $h_\theta(x) = \begin{cases}1 & \text{if } \theta^T x \geq 0\\ 0 & \text{otherwise.}\end{cases}$

when testing, same as logistic Regression. when training we want $\theta^T x >> 1$

---




Decision Boundary




if $y=1$, we want $\theta^T x \geq 1$, not just $\geq 0$ } → large margin

if $y=0$, we want $\theta^T x \leq -1$, not just $< 0$

Question: if we want get robust to outlier, should use large or small C?

whenever $y^{(i)}=1 \Rightarrow \theta^T x^{(i)} \geq 1$

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

$\Rightarrow \min \frac{1}{2}\sum_{j=1}^{n}\theta_j^2$

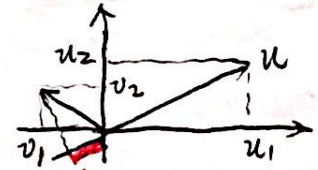
s.t. $\begin{cases}\theta^T x^{(i)} \geq 1 & \text{if } y^{(i)}=1\\ \theta^T x^{(i)} \leq 1 & \text{if } y^{(i)}=0\end{cases}$

---

## Vector Inner Product

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


$\|u\|$ = length $= \sqrt{u_1^2+u_2^2} \in \mathbb{R}^{(1)}$




P = length of the projection of v onto u $\boxed{P>0}$

P = lenght of the projection of v onto u $\boxed{P<0}$

$u^T v = p \cdot \|u\| = u_1 v_1 + u_2 v_2$

$v^T u = p' \cdot \|v\| = u_1 v_1 + u_2 v_2$

Oct 30, 2018