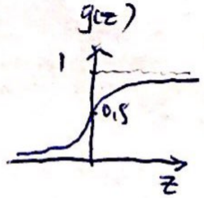
Read : "Bishop" Chapter 4

Logistic Regression.

Hypothesis : $h_\theta(x) = g(\theta^T x)$

sigmoid function : $g(z) = \frac{1}{1+e^{-z}}$ ↗



$\Downarrow$

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

• $h_\theta(x)$ estimated the probability that "$y=1$" given "x"

or we can write $h_\theta(x) = P(y=1 | x; \theta)$

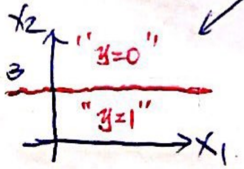
• $P(y=1|x;\theta) + P(y=0|x;\theta) = 1$

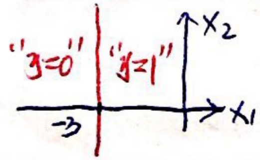
$$\begin{cases} \text{if } \theta^T x \geq 0 \text{, then } h_\theta(x) \geq 0.5 \text{, we predict } y=1 \\ \text{if } \theta^T x < 0 \text{. then } h_\theta(x) < 0.5 \text{, we predict } y=0 \end{cases}$$

Example: Plot the decision boundary of $h_\theta(x)$,

suppose $\theta_0 = 3$. $\theta_1 = 0$, $\theta_2 = -1$

$\theta_0 = 3$, $\theta_1 = +1$, $\theta_2 = 0$.





Given training set $\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \cdots, (x^{(m)}, y^{(m)})\}$

To learn θ, we can define Cost Function :

$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \underline{\frac{1}{2} \left(h_\theta(x^{(i)}) - y^{(i)} \right)^2}.$$ the same as linear regression

$$Cost(h_\theta(x^{(i)}), y^{(i)}) = \frac{1}{2}\left(h_\theta(x^{(i)}) - y^{(i)} \right)^2$$

can be different other then square Errors.

$\Downarrow$

$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} Cost(h_\theta(x^{(i)}), y^{(i)})$$

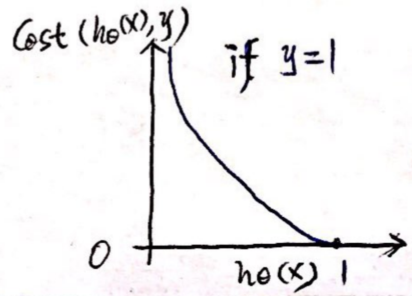
non-convex , inconvenient to find θ that can minize $J(\theta)$

Logistic Regression Cost Function.

groundtruth.

$$Cost(h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)) & \text{if } y=1 \\ -\log(1-h_\theta(x)) & \text{if } y=0 \end{cases}$$

• convex , we can find θ that minimize $J(\theta)$.



If $h_\theta(x) \to 1$, $P(y=1|x;\theta) \to 1$

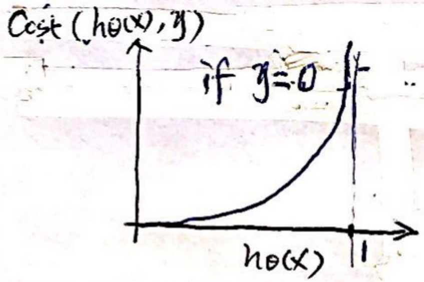
the model predicts "$y=1$", label $y=1$

$\Rightarrow$ Cost $\to 0$, correct prediction.

If $h_\theta(x) \to 0$, $P(y=1|x;\theta) \to 0$

the model predicts "$y=0$", but label $y=1$

$\Rightarrow$ Cost $\to \infty$, incorrect prediction.



If $h_\theta(x) \to 0$, $P(y=1|x;\theta) \to 0$

the model predicts "$y=0$", label $y=0$

$\Rightarrow$ cost $\to 0$, correct prediction

If $h_\theta(x) \to 1$, $\cdots$, $\Rightarrow$ Cost $\to \infty$

Reformat Cost Function.

$$Cost(h_\theta(x), y) = -y \log(h_\theta(x)) - (1-y)\log(1-h_\theta(x))$$

• if true $y=1$, $Cost = -\log(h_\theta(x))$

• if true $y=0$, $Cost = -\log(1-h_\theta(x))$

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

$\Downarrow$

Training: $\theta = \underset{\theta}{\text{minimize}}\ J(\theta)$ Testing : $h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$? 0.5

$\theta^T x$? 0

Gradient Descent:

→ Linear Regression $\theta^T x$

→ Logistic Regression

Repeat {

$$\theta_j := \theta_j - \alpha \sum_{i=1}^{m} \left(\boxed{h_\theta(x^{(i)})} - y^{(i)} \right) x_j^{(i)}$$

}

Vectorized Implementation

Oct 16, 2018.