Logistic Regression

Classification

1. No linear regression for classification

2. Decision Boundary is drawn from hypothesis function after fitting the θ parameters

3. Hypothesis Function:

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

4. Cost Function: convex is important->that is why we use log()

$$Cost(h_\theta(x), y) = \frac{1}{m} \sum_{i}^{m} (-y_i log(h_\theta(x_i)) - (1 - y_i)log(1 - h_\theta(x_i))$$

if $y = 0$:

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

so, if $h_\theta(x) = 0$, the cost function will be close to zero; if $h_\theta(x) = 1$, the cost function will be go infinitiy.

if $y = 1$:

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

so if $h_\theta(x) \sim 0$, the cost function would be infinity while it will be zero if $h_\theta(x) \sim 1$

if we plot the plot for y=1/0, its convex and intuitive.

5. Advanced Optimization instead of Gradient Descent

   - Gadient descent
   - Conjugate gradient
   - BFGS
   - L-BFGS

   Advantages:

- No need to manually pick alpha
- faster than GD(gradient descent)

Disadvantage:

- More complex

6. Multiple-Classification: One vs all method, make one against the remainder max over i for $\max\limits_{i} h_\theta^i(x)$