Coursera Machine Learning Week3

----`

`[Coursera Machine Learning](https://www.coursera.org/learn/machine-learning/home/welcome) Lunar's note`

MachineLearning Coursera`

分类问题 Classfication

1. Linear regression 不适用分类问题

2. Logistic Regression

- Model:Logistic function(Sigmoid function)
  $h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$
- Hypothesis Representation
  $h_\theta(x) = P(Y = 1|x;\theta)$
- Decision boundary决策边界
  $\theta^T x$ 就是决策边界，在边界的不同side做不同决策，比如说$\theta_0 + \theta_1 x_1 + \theta_2 x_2$在图像中就是一条直线，直线上方和下方是不同class
  非线性决策边界 一次多项式不适用时可以使用高阶多项式
- 代价函数cost function 线性回归中的代价函数用在这里会变成非凸函数（non-convex）所以要使用不同的代价函数(极大似然估计 maximum likelihood estimation)
  $Cost(h_\theta(x), y) = \{^{-log(h_\theta(x))(y=1)}_{-log(1-h_\theta(x))(y=0)}$
- 简化版本 $Cost(h_\theta(x), y) = -ylog(h_\theta(x)) - (1-y)log(1-h_\theta(x))$
  $y = 0 or 1$
- $J(\theta) = \frac{1}{m}\sum_{i=1}^{m} Cost(h_\theta(x^{(i)}), y^{(i)})$
- 拟合参数 fit parameters 梯度下降 Gradient Descent 方法和线性回归一样

3. 高级优化 Advanced Optimization

- Conjugate gradient
- BFGS
- L-BFGS
  advantages：
- 不需要选择学习速率
- 比梯度下降更快
  disadvantages：
- 更复杂

4. 多元分类 Multiclass calssification

One-vs-all
分成n次二元分类问题,选取$max(h_\theta^{(i)}(x))$

过度拟合问题 Overfitting

1. 欠拟合（underfit 或 High bias）

预测偏差较大

2. 过拟合（overfitting）

（或 High variance）对于训练集拟合得很好，但是对new example表现不佳。

3. 解决过度拟合

- 减少特征变量
- 正则化（Regularization）

4. 正则化

减小某些特征值的参数值(θ^i)，通常我们能获得一个更“简单”的假设("Simpler"hypothesis)，更不易于过度拟合。

- How
  使用新的cost function
  $J(\theta)=\frac{1}{2m}\left[\sum_{i=1}^{m}(h_\theta(x^{(i)})-y^{(i)})^2+\lambda\sum_{j=1}^{n}\theta_j^2\right]$
  那么最小化$J(\theta)$的过程中（使用梯度下降或者正规方程都可以），θ_j也会逐渐变小。但是过大的λ比如10^9会导致欠拟合。
- - Logistic regression中的正则化，注意cost function中最后一项

$$J(\theta) = -[\frac{1}{m}\sum_{i=1}^{m} -ylog(h_\theta(x)) + (1-y)log(1-h_\theta(x))] + \frac{\lambda}{2m}\sum_{j=1}^{n}\theta_j^2$$