

# Coursera Machine Learning Week2

Coursera Machine Learning Lunar's note

MachineLearning Coursera

## Coursera Machine Learning Week2

### 多变量线性回归(Multiple Features Linear Regression)

1. 基础
2. 梯度下降解决多变量线性回归
3. Normal Equation正规方程
4. 向量化方法

## 多变量线性回归(Multiple Features Linear Regression)

### 1. 基础

- 多个变量的情况下，对于第*i*个训练样本的第*j*个特征变量用 $x_j^{(i)}$ 表示。
- 表示形式:  $h_{\theta}(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \dots + \theta_n x_n$  把所有变量和参数都看成向量，则有 $h_{\theta}(x) = \theta^T x$

### 2. 梯度下降解决多变量线性回归

- cost function :  $J(\theta) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$
- 重复  $\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \dots + \theta_n x_n)$ 
  1. 梯度下降实用技术
- Feature scaling 特征缩放 特征的单位大小如果相差过于悬殊，比如相差数个数量级，代价函数的图会很陡峭，会需要更多时间收敛到局部最小，所以将特征值通过缩放放到同一个范围中可以解决这一问题。比如对于房屋估价的问题，房间面积以英寸为单位和房间个数以个为单位，二者相差较多，可以将面积以十平方米为单位。 X
- Mean normalization 均值归一化

用  $\frac{x_i - \mu_i}{S_i}$  来替换  $x_i$ , 使特征均值为0,  $\mu$  为均值  $S_i$  为极差。

- Learning rate 学习率  $\alpha$

如果梯度下降正常工作, 那么数次迭代后代价函数都应该逐渐decrease。选择阈值决定是否收敛。代价函数若increase或者不断起伏, 那么可以尝试减少  $\alpha$ 。

- 选择特征

如果提供的特征不理想, 可以通过特定角度通过特定的特征的意义将特征运算组合形成新特征, 比如, 特征长和特征宽可以组成面积。

- 多项式回归

可以将幂  $\geq 2$  的项换为多变量线性回归中的  $X_i$

### 3. Normal Equation正规方程

- 相比梯度下降的逐步迭代, 可以一次性解出  $\theta$  的值

- 将特征值增加一全为1的列, 组成  $m \times (n+1)$  矩阵  $X$ , 输出值  $y$  组成矩阵  $Y$ , 那么

$\theta = (X^T X)^{-1} X^T Y$  时会使  $J(\theta)$  最小化

- 不需要归一化变量

Gradient Descent	Normal Equation
需要选择学习速率 $\alpha$	不需要
需要迭代	不需要
特征变量 $n$ 很多时有效	特征变量很多时计算会很慢 $O(n^3)$ 复杂度

### 4. 向量化方法

- Case :  $h_{\theta}(x) = \sum_{j=0}^n \theta_j x_j = \theta^T x$