

斯坦福机器学习整理2——第二周

多元线性回归

多元是指用来预测多个特征量或变量。

$$h_{\theta}(x) = \theta_0 x_0 + \theta_1 x_1 + \dots + \theta_n x_n = \theta^T x, x_0 = 1$$

其采用梯度下降法求代价函数最小值：重复直到收敛

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$$

梯度下降运算中的技巧

feature scaling（数据归一化方法）

如果一个机器学习问题有多个特征，这些特征取值都在一个相近的范围（归一化），这样梯度下降法能更快的收敛。

feature scaling：将特征取值约束到-1到+1之间。

- mean normalization

$$x = \frac{x - \text{mean}}{\text{range}}$$

Normal Equation

与梯度下降法一样，都是用来求代价函数最小值。

$$\theta = (X^T X)^{-1} X^T y$$

梯度下降法和Normal Equation的适用情况

梯度下降法缺点：

- 需要选择学习速率 α

需要选择合适的

- 需要迭代。

Normal Equation缺点：

- 特征较多时，计算量大，耗时多。（如果特征上万维，计算速度会变慢，此时可以考虑梯度下降法）