

斯坦福机器学习整理1——第一周

监督学习（Supervised Learning）

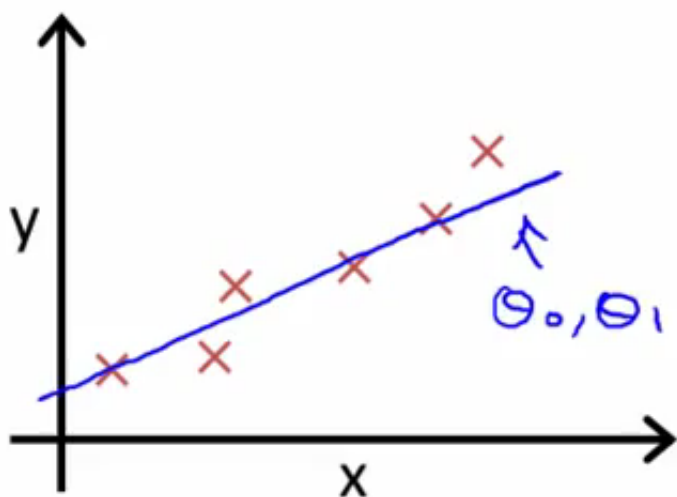
监督学习可以分为分类算法和回归算法。回归：根据之前的数据预测出一个准确的数值。分类：预测离散的输出值。

线性回归

用一条直线来拟合需要的数据

$$h(x) = \theta_0 + \theta_1 x$$

（假设函数）。



- 怎样求假设函数中 θ_0, θ_1 的值：就是让代价函数(cost function)

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^i) - y^i)^2$$

（表示预测值和实际值之间的误差）取得最小值。代价函数有时也被称为平方误差函数。

- 求代价函数最小值的方法：梯度下降法（Gradient descent）。

梯度下降（Gradient descent）

梯度下降法是怎样工作的： 1. 从 θ_0, θ_1 开始，将它们初始化，通常都设置为0；。 2. 不断改变 θ_0, θ_1 的值得到

$J(\theta_0, \theta_1)$, 直到得到期望的最小值。

梯度下降算法 (Gradient descent algorithm)

重复直到收敛

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) \quad (\text{for } j = 0 \text{ and } j = 1)$$

公式中:=为赋值符号 (=为相等符号) ; α 被称为学习速率, 它的值越大, 梯度下降的速度越快, 但如果太大, 梯度下降法可能越过最低点, 甚至无法收敛, 它的值越小, 梯度下降速度越慢。

正确的更新顺序:

$$temp0 := \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$$

$$temp1 := \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$$

$$\theta_0 := temp0$$

$$\theta_1 := temp1$$

代价函数为凸函数 (convex function) 时, 只有全局最优解, 总是收敛到全局最优值, 没有其他的局部最优值。