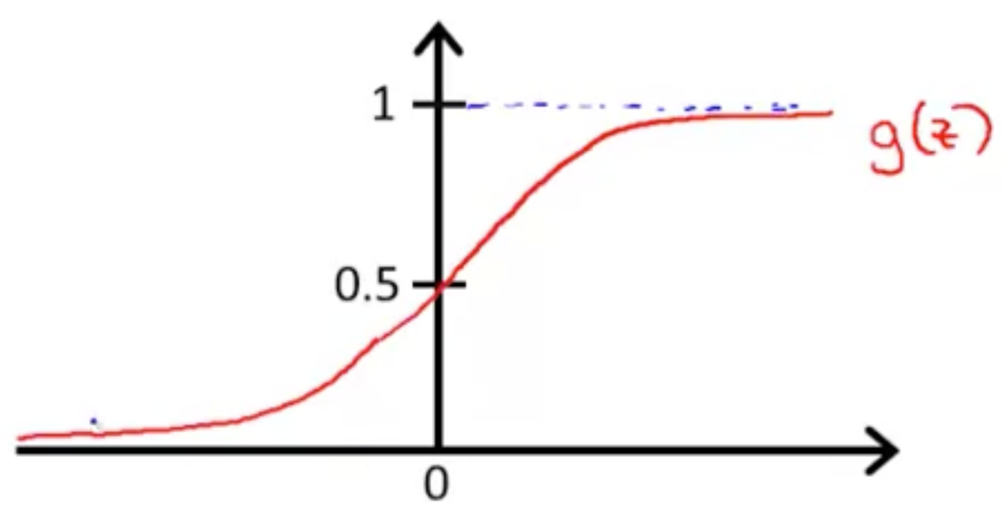
斯坦福机器学习整理3——第三周

逻辑回归（Logistic Regression）

一种分类算法，其输出值一直在0到1之间。

假设函数: $h_\theta(x) = g(\theta^T x)$

$g(z) = \dfrac{1}{1+e^{-z}}$ 称为Sigmoid Function或Logistic Function。如图



假设函数:

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

逻辑回归具体过程

1. 构造假设函数

构造假设函数，确定decision boundary（划分种类的边界），根据当$\theta^T x \geq 0$时，$y = 1$。确定了θ就可以确定decision boundary。

2. 构造cost函数

拟合参数θ。

逻辑回归代价函数:

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

$$Cost(h_\theta(x), y) = \{ \begin{matrix} -log(h_\theta(x)) & if y = 1 \\ -log(1 - h_\theta(x)) & if y = 0 \end{matrix}$$

可以将$Cost(h_\theta(x), y)$的两个等式合为一个：

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

所以逻辑回归代价函数可以写成：

$$J(\theta) = -\frac{1}{m}[\sum_{i=1}^{m} y^i \log h_\theta(x^(i)) + (1 - y^{(i)})\log(1 - h_\theta(x^{(i)}))]$$

3. 梯度下降法求使$J(\theta)$最小的值

根据梯度下降法可以得到θ的更新过程：

$$\theta_j := \theta_j - \alpha\frac{\partial}{\partial\theta_j J(\theta)} = \theta_j - \alpha\frac{1}{m}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)}$$

求代价函数最小的方法

1. Gradient descent（梯度下降法）
2. Conjugated gradient（共轭梯度法）
3. BFGS
4. L-BFGs

后面三种方法的优点：1.不需要手动选择学习率α；2.收敛速度比梯度下降法要快。缺点是算法比梯度下降法要复杂。

正则化（Regularization）

改善或减少过拟合问题（overfitting）也称为高方差（High variance）。过拟合通常发生在当变量较多时，进行分类的代价函数最小值趋近与0。解决过拟合问题有两个方法：1.减少变量的数目（使用重要的变量，舍弃部分变量）2.正则化（使用所有的特征变量）

正则化：

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

其中

$$\lambda \sum_{j=1}^{n} \theta_j^2$$

为正则化项，$\boldsymbol{\lambda}$为正则化参数。