Logistic Regression

逻辑回归算法核心思想

先用线性拟合，然后量化线性拟合的预测结果，即将连续值量化为离散值。(使用**线性回归+阈值**解决分类问题)。逻辑回归将连续值映射到$(0,1)$，采用的是$Sigmoid$函数(也称为$Logistic$函数、S函数)。

Sigmoid函数与决策边界

$Sigmoid$函数：$S(x)=\frac{1}{1+e^{-x}}$

决策边界：分类器对样本进行区分的边界，包括**线性决策边界**和**非线性决策边界**。

线性决策边界

`e.g.` 用函数g表示$Sigmoid$函数，逻辑回归的输出结果由假设函数$h_\theta(x)=g(\theta_0+\theta_1x_1+\theta_2x_2)$得到，这里$\theta_0+\theta_1x_1+\theta_2x_2$就是线性边界。

非线性决策边界

`e.g.` 用函数g表示$Sigmoid$函数，逻辑回归的输出结果由假设函数$h_\theta(x)=g(\theta_0+\theta_1x_1+\theta_2x_2+\theta_3x_1^2+\theta_4x_2^2)$得到，这里$\theta_0+\theta_1x_1+\theta_2x_2+\theta_3x_1^2+\theta_4x_2^2$就是非线性边界。(无法用直线/超平面把不同类别的样本分开)

梯度下降与优化

损失函数和成本本成本函数

1. 损失函数/目标函数/代价函数：量化模型好坏的函数，损失函数(Loss function)是在单个训练样本上的表现。在逻辑回归模型中，损失函数采用**对数损失函数(二元交叉熵损失)**：
   $Cost(h_\theta(x),y)=-[y\cdot\log(h_\theta(x))+(1-y)\cdot\log(1-h_\theta(x))]=\begin{cases}-\log(1-h_\theta(x)) & y=0\\ -\log(h_\theta(x)) & y=1\end{cases}$
2. **成本函数(Cost function)**是在全体训练样本上的表现，也就是对损失函数(Cost function)的求和。公式：
   $J(\theta)=-\frac{1}{m}\sum_{i=1}^{m}[y^{(i)}\log(h_\theta(x^{(i)}))+(1-y^{(i)})\log(1-h_\theta(x^{(i)}))]$

梯度下降(Gradient Descent)

利用梯度下降，要求出最优参数θ，则需要最小化$J(\theta)$同时更新参数$\theta_j := \theta_j - \alpha \frac{\partial J(\theta)}{\partial \theta_j}$。在公式中$\alpha$代表学习率，直观意义是在函数向极小值方向前进时的步长。(太大会导致错过极小值，太小会导致迭代次数过多)

对$Sigmoid$函数和$J(\theta)$函数求导可得更新公式：$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)}$

正则化与防止过拟合

添加正则化项可以防止过拟合，故在成本函数(Cost function)中添加正则化项后为：$J(\theta) = -\frac{1}{m} \sum_{i=1}^{m} [y^{(i)} \log(h_\theta(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_\theta(x^{(i)}))] + \frac{\lambda}{2m} \sum_{j=1}^{n} \theta_j^2$，$\lambda$表示正则化系数，即为惩罚程度：$\lambda$越大，为使$J(\theta)$的值小，参数$\theta$的绝对值就越小。