逻辑回归相关推导

February 27, 2019

1 基本概念

1、逻辑回归基本形式

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

其中

$$g(z) = \frac{1}{1 + e^{-z}}$$

对 $g(z)$ 求导可得：

$$g^{'}(z) = g(z)(1 - g(z))$$

2、目标函数

假设

$$P(y = 1|x;\theta) = h_\theta(x)$$

$$P(y = 0|x;\theta) = 1 - h_\theta(x)$$

即

$$p(y|x;\theta) = (h_\theta(x))^y(1 - h_\theta(x))^{1-y}$$

其中 $y = 0, 1$ 可利用极大似然法估计参数 θ。

$$\begin{aligned} L(\theta) &= p(y|X;\theta) \\ &= \prod_{t=1}^{m} p(y^t|x^t;\theta) \\ &= \prod_{t=1}^{m} (h_\theta(x^t))^{y^t}(1 - h_\theta(x^t))^{1-y^t} \end{aligned} \tag{1}$$

现在需要最大化对数似然：

$$\begin{aligned} l(\theta) &= logL(\theta) \\ &= \sum_{t=1}^{m} y^t logh(x^t) + (1 - y^t)log(1 - h(x^t)) \end{aligned} \tag{2}$$

对该对数似然函数求偏导数得到 (其中上标表示第几个实例，下标表示第几个特征)：

$$\begin{aligned} \frac{\partial}{\partial\theta_j} l(\theta) &= \sum_{t=1}^{m} (y^t \frac{1}{g(\theta^T x^t)} - (1 - y^t \frac{1}{1 - g(\theta^T x^t)}) \frac{\partial}{\partial\theta_j} g(\theta^T x^t) \\ &= \sum_{t=1}^{m} (y^t \frac{1}{g(\theta^T x^t)} - (1 - y^t \frac{1}{1 - g(\theta^T x^t)}) g(\theta^T x^t)(1 - g(\theta^T x^t)) \frac{\partial}{\partial\theta_j} \theta^T x^t \\ &= \sum_{t=1}^{m} (y^t(1 - g(\theta^T x^t)) - (1 - y^t) g(\theta^T x^t)) x_j \\ &= \sum_{t=1}^{m} (y^t - h_\theta(x^t)) x_j^t \end{aligned} \tag{3}$$

写成矩阵形式就是：

$$\frac{\partial}{\partial\theta}l(\theta)=\sum_{t=1}^{m}(y^t-h_\theta(x^t))x^t$$

详细推导见:http://cs229.stanford.edu/notes/cs229-notes1.pdf

3、牛顿法求解

牛顿法用于求解 $f(\theta)=0$ 的根，其更新规则为：

$$\theta:=\theta-\frac{f(\theta)}{f^{'}(\theta)}$$

而我们需要求解 $l(\theta)$ 的最大值，即需要求取 $l^{'}(\theta)=0$ 的根，于是这里的 $f(\theta)=l^{'}(\theta)$，相应地 θ 的更新规则变为

$$\theta:=\theta-\frac{l^{'}(\theta)}{l^{''}(\theta)}$$

写成矩阵形式是：

$$\theta:=\theta-H^{-1}\nabla_\theta l(\theta)$$

其中 H 称为 Hessian 矩阵，其元素由下式决定:

$$H_{ij}=\frac{\partial^2 l(\theta)}{\partial\theta_i\partial\theta_j}$$

下面给出 H_{ij} 的推导：

$$\begin{aligned}H_{ij}&=\frac{\partial^2 l(\theta)}{\partial\theta_i\partial\theta_j}\\&=\frac{\partial}{\theta_j}\sum_{t=1}^{m}(y^t-h_\theta(x^t))x_i^t\\&=\sum_{t=1}^{m}-x_i^t\frac{\partial}{\partial\theta_j}h_\theta(x^t)\\&=\sum_{t=1}^{m}-x_i^t h_\theta(x^t)(1-h_\theta(x^t))\frac{\partial}{\partial\theta_j}(\theta^T x^t)\\&=\sum_{t=1}^{m}h_\theta(x^t)(h_\theta(x^t)-1)x_i^t x_j^t\end{aligned}\tag{4}$$

写成矩阵形式是：

$$H=\frac{\partial^2 l(\theta)}{\partial\theta\partial\theta^T}=\sum_{t=1}^{m}x^t(x^t)^T h_\theta(x^t)(h_\theta(x^t)-1)$$

其中 H 为 $n\times n$ 矩阵，x^t 维度为 $n\times 1$,n 为特征维度总数。

二阶偏导及牛顿法见:https://blog.csdn.net/Fishmemory/article/details/51603836

牛顿法可参考：https://zhuanlan.zhihu.com/p/22862479