逻辑回归

Mia Feng

2018 年 3 月 4 日

1 概述

逻辑回归是广义线性模型的一种，适用的数据为可分的。主要用于解决二分类问题，也可以扩展至多分类问题求解目标：分类超平面，以二维空间为例，求解一条直线

$$y = b + \theta_1 x_1 + \theta_2 x_2 = \theta^{\mathbf{T}} x + b \tag{1}$$

求解思路：求解 θ 的控制方程选择概率描述。使样本点被分类至真实标签的概率最大。推导证明 logistic 分布的鲁棒性最好，（见 LogisticRegression-MaxEnt.pdf）简单来说，因为 $p\left(y|x\right) \sim Bernoulli$，根据最大熵原则求解分布（也就是说这个分布应该在满足假设的前提下越均匀越好），推导得到线性模型 $f\left(\theta x\right)$，结果为 sigmoid 函数，即 Bernoulli 的指数家族形式

求解方法：最大似然估计

迭代方法：gradient ascent

1.1 推导

Sigmoid 函数

$$h_\theta\left(x\right) = \frac{1}{1 + e^{-\theta^{\mathbf{T}} x}} \tag{2}$$

基于二分类的假定，每个样本点的概率为

$$p\left(y|x, \theta\right) = h_\theta\left(x\right)^y \left(1 - h_\theta\left(x\right)\right)^{1-y} \tag{3}$$

这里只是一个技巧性的推导，代入标签值计算一下即可得到。

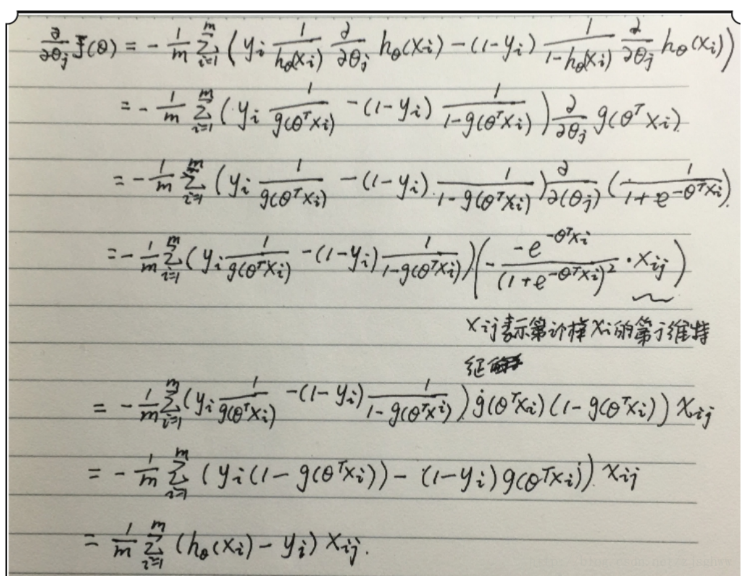
又基于 i.i.d. 假设（独立同分布），m 个样本点的后验概率为：

$$L(\theta) = \prod_{i=1}^{m} p(y_i|x_i,\theta) = \prod_{i=1}^{m} h_\theta(x_i)_i^y \big(1-h_\theta(x_i)\big)^{1-y_i} \tag{4}$$

为便于计算，取对数

$$L(\theta) = \ln L(\theta) = \sum_{i=1}^{m} y_i \ln h_\theta(x_i) + \big(1-y_i\big)\ln\big(1-h_\theta(x_i)\big) \tag{5}$$

接下来计算梯度，设置梯度更新，迭代求解 θ 参考的 blog，打字太多不写了，把他 [1] 的推导粘过来了，推倒还是很明了的。梯度：



$$\frac{\partial}{\partial\theta_j}J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left(y_i\frac{1}{h_\theta(x_i)}\frac{\partial}{\partial\theta_j}h_\theta(x_i) - (1-y_i)\frac{1}{1-h_\theta(x_i)}\frac{\partial}{\partial\theta_j}h_\theta(x_i)\right)$$

$$= -\frac{1}{m}\sum_{i=1}^{m}\left(y_i\frac{1}{g(\theta^T x_i)} - (1-y_i)\frac{1}{1-g(\theta^T x_i)}\right)\frac{\partial}{\partial\theta_j}g(\theta^T x_i)$$

$$= -\frac{1}{m}\sum_{i=1}^{m}\left(y_i\frac{1}{g(\theta^T x_i)} - (1-y_i)\frac{1}{1-g(\theta^T x_i)}\right)\frac{\partial}{\partial(\theta_j)}\left(\frac{1}{1+e^{-\theta^T x_i}}\right)$$

$$= -\frac{1}{m}\sum_{i=1}^{m}\left(y_i\frac{1}{g(\theta^T x_i)} - (1-y_i)\frac{1}{1-g(\theta^T x_i)}\right)\left(-\frac{-e^{-\theta^T x_i}}{(1+e^{-\theta^T x_i})^2}\cdot x_{ij}\right)$$

x_{ij}表示第i个样本x_i的第j维特征

$$= -\frac{1}{m}\sum_{i=1}^{m}\left(y_i\frac{1}{g(\theta^T x_i)} - (1-y_i)\frac{1}{1-g(\theta^T x_i)}\right)g(\theta^T x_i)\big(1-g(\theta^T x_i)\big)x_{ij}$$

$$= -\frac{1}{m}\sum_{i=1}^{m}\Big(y_i\big(1-g(\theta^T x_i)\big) - (1-y_i)g(\theta^T x_i)\Big)x_{ij}$$

$$= \frac{1}{m}\sum_{i=1}^{m}\big(h_\theta(x_i) - y_i\big)x_{ij}.$$

图 1: Partial gradient derivation

梯度更新公式：

$$\theta_j^{t+1} = \theta_j^t - \alpha\frac{\partial}{\partial\theta_j}L(\theta_j) = \theta_j^t - \alpha\frac{1}{m}\sum_{i=1}^{m}\big(h_\theta(x_i) - y_i\big)x_{ij} \tag{6}$$

2 Pesudo Code

1. 初始化回归系数 θ
2. 重复下面步骤直至收敛

   {

   计算整个数据集的梯度

   使用公式 (6) 更新梯度

   }
3. 返回回归系数 θ

参考文献

[1] Maggie 张张. Logistic regression 原理及推导 python 实现. `http://blog.csdn.net/zjsghww/article/details/55211530`, 2017. CSDN-Blog, 2017.