

Linear Classification 04

Chen Gong

1 November 2019

在前面的两小节中我们, 我们讨论了有关于线性分类问题中的硬分类问题, 也就是感知机和 Fisher 线性判别分析。那么, 我们接下来的部分需要讲讲软分类问题。软分类问题, 可以大体上分为概率判别模型和概率生成模型, 概率生成模型也就是高斯判别分析 (Gaussian Discriminate Analysis), 朴素贝叶斯 (Naive Bayes)。而线性判别模型也就是本章需要讲述的重点, Logistic Regression。

1 从线性回归到线性分类

线性回归的问题, 我们可以看成这样一个形式, 也就是 $W^T X$ 。而线性分类的问题可以看成是 $\{0, 1\}$ 或者 $[0, 1]$ 的问题。其实, 从线性回归到线性分类之间通过一个映射, 也就是 Activate Function 来实现的, 通过这个映射我们可以实现 $W^T X \mapsto \{0, 1\}$ 。

而在 Logistic Regression 中, 我们将激活函数定义为:

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

那么很显然会有如下的性质:

$$1. \lim_{z \rightarrow +\infty} \sigma(z) = 1$$

$$2. \lim_{z \rightarrow 0} \sigma(z) = \frac{1}{2}$$

$$3. \lim_{z \rightarrow -\infty} \sigma(z) = 0$$

那么, 通过这样一个激活函数 σ , 我们就可以将实现 $\mathbb{R} \rightarrow (0, 1)$ 。那么我们会得到以下的表达式:

$$p(y|x) = \begin{cases} p_1 = p(y=1|x) = \sigma(w^T x) = \frac{1}{1 + \exp\{-w^T x\}} & y = 1 \\ p_2 = p(y=0|x) = 1 - p(y=1|x) = \frac{\exp\{-w^T x\}}{1 + \exp\{-w^T x\}} & y = 0 \end{cases} \quad (2)$$

而且, 我们可以想一个办法来将两个表达式合二为一, 那么有:

$$p(y|x) = p_1^y \cdot p_0^{1-y} \quad (3)$$

2 最大后验估计

$$\begin{aligned}MLE = \hat{w} &= \operatorname{argmax}_w \log p(y|x) \\&= \operatorname{argmax}_w \log p(y_i|x_i) \\&= \operatorname{argmax}_w \sum_{i=1}^N \log p(y_i|x_i) \\&= \operatorname{argmax}_w \sum_{i=1}^N y \log p_1 + (1-y) \log p_2\end{aligned}\tag{4}$$

我们令,

$$\frac{1}{1 + \exp\{-w^t x\}} = \varphi(x, w) \quad \frac{\exp\{-w^t x\}}{1 + \exp\{-w^t x\}} = 1 - \varphi(x, w)\tag{5}$$

那么,

$$MLE = \operatorname{argmax}_w \sum_{i=1}^N y \log \varphi(x, w) + (1-y) \log(1 - \varphi(x, w))\tag{6}$$

实际上 $y \log \varphi(x, w) + (1-y) \log(1 - \varphi(x, w))$ 就是一个交叉熵 (Cross Entropy)。那么, 我们成功的找到了我们的优化目标函数, 可以表述为 MLE (max) \rightarrow Loss function (Min Cross Entropy)。所以, 这个优化问题就转换成了一个 Cross Entropy 的优化问题, 这样的方法就很多了。Cross Entropy 属于信息论的知识, 同学们可以自行补充。