# logistic 回归

DTDT

2018 年 11 月 19 日

假设存在数据集 $(x_i, y_i)$，$x_i \in R^n$，$y_i \in \{0, 1\}$

令

$$P(Y=1|x) = \frac{exp(wx)}{1+exp(wx)}$$

$$P(Y=0|x) = \frac{1}{1+exp(wx)}$$

如果将 $p$ 看作发生概率，则 $1-p$ 就是不发生概率，$\frac{p}{1-p}$ 被称为几率，对于 logistic 回归有：

$$log\left(\frac{P(Y=1|X)}{P(Y=0|X)}\right) = wx$$

logitic 回归就是要求得 $w$ 矩阵，下面用极大似然估计求 $w$

模型的似然函数为：

$$\prod_{i=1}^{n} \pi(x_i)^{y_i}(1-\pi(x_i))^{1-y_i}$$

其中

$$\pi(x) = P(Y=1|x) = \frac{exp(wx)}{1+exp(wx)}$$

对似然函数取 $log$ 有:

$$\begin{aligned}
L(w) &= \sum_i^N y_i log \pi(x_i) + (1-y_i) log(1-\pi(x_i)) \\
&= \sum_i^N (y_i log \frac{\pi(x_i)}{1-\pi(x_i)} + log(1-\pi(x_i))) \\
&= \sum_i^N (y_i (w \cdot x_i) - log(1+exp(w \cdot x_i))) \qquad (1)
\end{aligned}$$

对 $L(w)$ 求极大值，下面证明 $L(w)$ 是凸的

(1) 式第一部分是凸的，第二部分假设

$$\begin{aligned}
f(x) &= -log(1+exp(w \cdot x_i)) \\
f'(x) &= -\frac{exp(x)}{1+exp(x)} \\
f''(x) &= \frac{exp(x)}{(1+exp(x))^2} \qquad (2)
\end{aligned}$$

(2) 式恒大于 0, 所以 $f(x)$ 是个凸函数。且凸函数的非负线性组合还是凸函数，凸函数和仿射函数的复合是凸函数，因此 $L(w)$ 是凸函数。

对凸函数求解最优值问题有梯度下降法和牛顿法。

1. 牛顿法

函数的极值处偏导等于 0, 且凸函数最值等于极值, 因此最优化等价于求 $f'(x)$ 的零点。

$$w^{t+1} = w^t - \left(\frac{\partial^2 L(w)}{\partial w \partial w^T}\right)^{-1} \frac{\partial L(w)}{\partial w} \tag{3}$$

$$\frac{\partial L(w)}{\partial w} = \sum_i^N y_i x_i - \frac{x_i exp(wx_i)}{1 + exp(wx_i)}$$

$$\frac{\partial^2 L(w)}{\partial w \partial w^T} = \sum_i^N x_i x_i^T \frac{exp(wx_i)}{(1 + exp(wx_i))^2}$$

$$= x_i x_i^T P(Y = 1|x_i) P(Y = 0|x_i)$$

2. 梯度下降法

由于是凸函数，最后收敛的一定是最值。

$$w^{t+1} = w^t - \alpha \frac{\partial L(w)}{\partial w}$$

比较两种方法，1 不需要设置超参数，而且具有良好的收敛速度，2 需要设置合适的 $\alpha$，收敛速度依赖于具体的 $\alpha$。看起来用方法 1 效果不错，但实际中由于矩阵求逆时间复杂度太高，因此大多用第二种方法。