逻辑回归总结

2017-9-28

对于线性回归模型 $y = \omega^T + b$ 输出的预测值 y 都是实值，考虑二分类任务，$y_i \in \{0,1\}$。为了将实值 y 对应到离散值, 使用了 sigmoid 函数。$y = \frac{1}{1+e^{-z}} = \frac{1}{1+e^{-(\omega^T+b)}}$, 对 ω 和 b 作以下变换，令 $\beta = (\omega; b)$,$\hat{x} = (x; 1)$, 则 $y = \omega^T + b = \beta^T \hat{x}$, 带入 sigmoid 函数。

$$y = \frac{1}{1+e^{-\beta^T \hat{x}}} \tag{1}$$

对 sigmoid 函数，$z = \ln \frac{y}{1-y}$, 则式(1)可以变为

$$\ln \frac{y}{1-y} = \beta^T \hat{x} \tag{2}$$

若将 y 视为类后验概率估计 $P(y=1|x)$，即样本属于第一类的概率。则式(2)可以写为：

$$\ln \frac{P(y=1|x)}{1-P(y=1|x)} = \ln \frac{P(y=1|x)}{P(y=0|x)} = \beta^T \hat{x}$$

对于样本集 $\{x_i, y_i\}_{i=1}^m$，样本的概率可以写为：

$$P(y_i|\beta; \hat{x_i}) = P(y_i = 1|\beta; \hat{x_i})^{y_i} \left(1 - P(y_i = 1|\beta; \hat{x_i})\right)^{1-y_i}$$

其中：

$$P(y=1|x) = \frac{e^{\beta^T x}}{1+e^{\beta^T x}} \tag{3}$$

$$P(y=0|x) = \frac{1}{1+e^{\beta^T x}} \tag{4}$$

n 个独立的样本出现的似然函数为 (因为每个样本都是独立的，所以 n 个样本出现的概率就是他们各自出现的概率相乘):

$$L(\theta) = \prod P(y_i|\beta; \hat{x_i}) \tag{5}$$

对上式取对数，

$$
\begin{aligned}
L(\theta) &= \log\prod P(y_i|\beta;\hat{x_i}) \\
&= \log\prod\left(P(y_i=1|\beta;\hat{x_i})^{y_i}\,(1-P(y_i=1|\beta;\hat{x_i}))^{1-y_i}\right) \\
&= \sum\log\left(P(y_i=1|\beta;\hat{x_i})^{y_i}\,(1-P(y_i=1|\beta;\hat{x_i}))^{1-y_i}\right) \\
&= \sum\left[y_i\log P(y_i=1|\beta;\hat{x_i})+(1-y_i)\log(1-P(y_i=1|\beta;\hat{x_i}))\right] \\
&= \sum\left(y_i\log P(y_i=1)-y_i\log(P(y_i=0))+\log(P(y_i=0))\right) \\
&= \sum\left(y_i\log\frac{P(y_i=1)}{P(y_i=0)}+log(\frac{1}{1+e^{\beta^T\hat{x_i}}})\right) \\
&= \sum\left(y_i\beta^T\hat{x_i}-log(1+e^{\beta^T\hat{x_i}})\right)
\end{aligned}
\tag{6}
$$

为了优化求解，对目标函数取反，以进行最小值优化。

式(6) 取反后，对 β 求导数：

$$
\begin{aligned}
\mathrm{d}L(\theta) &= \sum\left(\frac{e^{\beta^T\hat{x_i}}\mathrm{d}\beta^T\hat{x_i}}{1+e^{\beta^T\hat{x_i}}}-y_i\mathrm{d}\beta^T\hat{x_i}\right) \\
&= \mathrm{tr}\sum\left(\frac{\hat{x_i}e^{\beta^T\hat{x_i}}\mathrm{d}\beta^T}{1+e^{\beta^T\hat{x_i}}}-\hat{x_i}y_i\mathrm{d}\beta^T\right) \\
&= \sum\hat{x_i}(P(y_i=1|\beta;\hat{x_i})-y_i)\mathrm{d}\beta^T
\end{aligned}
\tag{7}
$$

$$
\frac{\partial\mathrm{d}L(\theta)}{\partial\beta^T}=\sum\hat{x_i}^T(P(y_i=1|\beta;\hat{x_i})-y_i)
$$

$$
\frac{\partial\mathrm{d}L(\theta)}{\partial\beta}=\sum\hat{x_i}(P(y_i=1|\beta;\hat{x_i})-y_i)
\tag{8}
$$

运用梯度下降法进行优化求解：

$$
\beta_{n+1}=\beta_n-\alpha\frac{\partial\mathrm{d}L(\theta)}{\partial\beta}
$$