对数几率回归

① Generalized Linear Model
1. 广义线性模型 { 1. 指数族分布 / 2. 广义线性模型的三条假设
2. 对数几率回归

指数族分布:

指数族 (Exponential family) 分布是一类分布的总称,该类分布的分布律(或者概率密度函数)的一般形式如下:

$$p(y;\eta)=b(y)\exp(\eta^T T(y)-a(\eta))$$

其中,η 称为该分布的自然参数;$T(y)$ 为充分统计量,视具体的分布而定,通常是等于随机变量 y 本身;$a(\eta)$ 为配分函数;$b(y)$ 为关于随机变量 y 的函数,常见的伯努利分布和正态分布均属于指数族分布。

证明伯努利分布属于指数族分布:

已知伯努利分布的分布律为: $p(y)=\phi^y(1-\phi)^{1-y} \Longrightarrow \begin{cases} y=0,\ p(y)=\phi^0(1-\phi)^1=1-\phi \\ y=1,\ p(y)=\phi^1(1-\phi)^0=\phi \end{cases}$

恒等变形 $\Downarrow$

$$=\exp(\ln(\phi^y(1-\phi)^{1-y}))$$

$$=\exp(\ln\phi^y+\ln(1-\phi)^{1-y})$$

$$=\exp(y\ln\phi+(1-y)\ln(1-\phi))$$

$$=\exp(y\ln\phi+\ln(1-\phi)-y\ln(1-\phi))$$

$$=\exp(y(\ln\phi-\ln(1-\phi))+\ln(1-\phi))$$

$$=\exp\left(y\ln\frac{\phi}{1-\phi}+\ln(1-\phi)\right)$$

对比指数族分布一般形式: $p(y;\eta)=b(y)\exp(\eta^T T(y)-a(\eta))$ 可知:

$b(y)=1$ $\quad \eta=\ln\left(\frac{\phi}{1-\phi}\right)$ $\quad T(y)=y$ $\quad a(\eta)=-\ln(1-\phi)=\ln(1+e^{\eta})$

②

广义线性模型的三条假设:

1.在给定 x 的条件下,假设随机变量 y 服从某个指数族分布;

2. 在给定 x 的条件下,我们的目标是得到一个模型 $h(x)$ 能预测出 $T(y)$ 的期望值;

3. 假设该指数分布中的自然参数 η 和 x 呈线性关系,即 $\eta = w^T x$

对数几率回归 Logistic Regression

1.对数几率回归的广义线性模型推导

2.极大似然估计法

3. 对数几率回归的参数估计

对数几率回归是在对一个二分类问题进行建模,并且假设被建模的随机变量 y 取值为0或1,因此我们可以很自然得假设 y 服从伯努力分布。

此时,如果我们想要构建一个线性模型来预测在给定 x 的条件下 y 的取值的话,可以考虑使用广义线性模型来进行建模。

已知 y 是服从伯努力分布,而伯努力分布属于指数族分布,所以满足广义线性模型的第一条假设,接着根据广义线性模型的第二条假设我们可以推得模型 $h(x)$ 的表达式为: $h(x) = E[T(y|x)]$

由于伯努利分布的 $T(y|x) = y|x$ 所以 $h(x) = E[y|x] \rightarrow E[y]$

$$又 \because E[y|x] = 1 \times p(y=1|x) + 0 \times p(y=0|x) = p(y=1|x) = \phi,$$

$$所以\ h(x) = \phi.$$

$$由前面可知\ \eta = \ln\left(\frac{\phi}{1-\phi}\right) \Rightarrow e^{\eta} = \frac{\phi}{1-\phi} \Rightarrow e^{-\eta} = \frac{1-\phi}{\phi} = \frac{1}{\phi} - 1$$

$$\Rightarrow 1 + e^{-\eta} = \frac{1}{\phi}$$

$$\phi = \frac{1}{1+e^{-\eta}}\ 代入\ h(x) \Rightarrow$$

$$h(x) = \phi = \frac{1}{1+e^{-\eta}}$$

③

根据广义模型的第三条假设：$\eta = w^T x$，$h(x)$ 最终可化为：

$$h(x) = \Phi = \frac{1}{1+e^{-w^T x}} = p(y=1|x)$$ 此即为式3.23

为对数几率回归模型

已知随机变量 y 取 1 和 0 的概率分别为：

$$p(y=1|x) = \frac{e^{w^T x + b}}{1+e^{w^T x + b}}$$

$$p(y=0|x) = \frac{1}{1+e^{w^T x + b}}$$

令 $\beta = (w; b)$，$\hat{x} = (x; 1)$，则 $w^T x + b$ 可简写为 $\beta^T \hat{x}$，于是上式可化简为

$$p(y=1|x) = \frac{e^{\beta^T \hat{x}}}{1+e^{\beta^T \hat{x}}} \triangleq p_1(\hat{x}; \beta)$$

$$p(y=0|x) = \frac{1}{1+e^{\beta^T \hat{x}}} \triangleq p_0(\hat{x}; \beta)$$

$$p(y|x; w, b) = y \cdot p_1(\hat{x}; \beta) + (1-y) \cdot p_0(\hat{x}; \beta)$$ 此即为式3.26

或者：$p(y|x; w, b) = [p_1(\hat{x}; \beta)]^{y} \cdot [p_0(\hat{x}; \beta)]^{1-y}$

根据对数似然函数的定义可知：

$$\ln L(w) = \sum_{i=1}^{m} \ln f(y_i, w_1, w_2 \cdots w_k)$$

由于此时的 y 为离散型，所以将对数似然函数中的概率密度函数换成分布律即：

$$l(w, b) = \ln L(w, b) = \sum_{i=1}^{m} \ln p(y_i | x_i; w, b)$$ 此即为式3.25

将 $p(y|x; w, b) = y \cdot p_1(\hat{x}; \beta) + (1-y) \cdot p_0(\hat{x}; \beta)$ 代入对数似然函数可得：

$$l(\beta) = \sum_{i=1}^{m} \ln\left(y_i p_1(\hat{x}_i; \beta) + (1-y_i) p_0(\hat{x}_i; \beta)\right)$$

由于 $p_1(\hat{x}_i; \beta) = \frac{e^{\beta^T \hat{x}_i}}{1+e^{\beta^T \hat{x}_i}}$，$p_0(\hat{x}_i; \beta) = \frac{1}{1+e^{\beta^T \hat{x}_i}}$ 所以上式可化为

$$l(\beta) = \sum_{i=1}^{m} \ln\left(\frac{y_i e^{\beta^T \hat{x}_i}}{1+e^{\beta^T \hat{x}_i}} + \frac{1-y_i}{1+e^{\beta^T \hat{x}_i}}\right)$$

$$= \sum_{i=1}^{m} \ln\left(\frac{y_i e^{\beta^T \hat{x}_i} + 1 - y_i}{1+e^{\beta^T \hat{x}_i}}\right)$$

$$= \sum_{i=1}^{m} \left(\ln(y_i e^{\beta^T \hat{x}_i} + 1 - y_i) - \ln(1+e^{\beta^T \hat{x}_i})\right)$$

由于 $y_i \in \{0,1\}$，所以

当 $y_i=0$ 时，$l(\beta) = \sum_{i=1}^{m}\left(\ln(0\cdot e^{\beta^T \hat{x}_i} + 1 - 0) - \ln(1+e^{\beta^T \hat{x}_i})\right) = \sum_{i=1}^{m}(-\ln(1+e^{\beta^T \hat{x}_i}))$

当 $y_i=1$ 时，$l(\beta) = \sum_{i=1}^{m}\left(\ln(1\cdot e^{\beta^T \hat{x}_i} + 1 - 1) - \ln(1+e^{\beta^T \hat{x}_i})\right) = \sum_{i=1}^{m}(\ln e^{\beta^T \hat{x}_i} - \ln(1+e^{\beta^T \hat{x}_i}))$

$$= \sum_{i=1}^{m}(\beta^T \hat{x}_i - \ln(1+e^{\beta^T \hat{x}_i}))$$

综合可得 $l(\beta) = \sum_{i=1}^{m}\left(y_i \beta^T \hat{x}_i - \ln(1+e^{\beta^T \hat{x}_i})\right)$　加个负号即为式3.27

若 $p(y|x;w,b) = [p_1(\hat{x};\beta)]^y [p_0(\hat{x};\beta)]^{1-y}$，将其代入对数似然函数可得

$$l(\beta) = \sum_{i=1}^{m} \ln\left([p_1(\hat{x}_i;\beta)]^{y_i} [p_0(\hat{x}_i;\beta)]^{1-y_i}\right)$$

$$= \sum_{i=1}^{m}\left[\ln([p_1(\hat{x}_i;\beta)]^{y_i}) + \ln([p_0(\hat{x}_i;\beta)]^{1-y_i})\right]$$

$$= \sum_{i=1}^{m}\left[y_i \ln(p_1(\hat{x}_i;\beta)) + (1-y_i)\ln(p_0(\hat{x}_i;\beta))\right]$$

$$= \sum_{i=1}^{m}\left\{y_i\left[\ln(p_1(\hat{x}_i;\beta)) - \ln(p_0(\hat{x}_i;\beta))\right] + \ln(p_0(\hat{x}_i;\beta))\right\}$$

$$l(\beta) = \sum_{i=1}^{m}\left[y_i \ln\left(\frac{p_1(\hat{x}_i;\beta)}{p_0(\hat{x}_i;\beta)}\right) + \ln(p_0(\hat{x}_i;\beta))\right]$$

由于 $p_1(\hat{x}_i;\beta) = \dfrac{e^{\beta^T \hat{x}_i}}{1+e^{\beta^T \hat{x}_i}}$，$p_0(\hat{x}_i;\beta) = \dfrac{1}{1+e^{\beta^T \hat{x}_i}}$ 所以上式可化为：

$$l(\beta) = \sum_{i=1}^{m}\left[y_i \ln(e^{\beta^T \hat{x}_i}) + \ln\left(\frac{1}{1+e^{\beta^T \hat{x}_i}}\right)\right]$$

$$= \sum_{i=1}^{m}\left(y_i \beta^T \hat{x}_i - \ln(1+e^{\beta^T \hat{x}_i})\right)$$　加个负号即为式3.27