深度之眼
deepshare.net

法律声明

deepshare.net
深度之眼

西瓜书公式推导

导师：Sm1les（Datawhale南瓜书项目负责人）

对数几率回归公式推导

Derivation of Logistic Regression

本节大纲

Outline

先修内容：西瓜书3.3

1.广义线性模型

2.对数几率回归

广义线性模型

Generalized Linear Model

deepshare.net
深度之眼

1.指数族分布

2.广义线性模型的三条假设

广义线性模型

Generalized Linear Model

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：

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

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。

广义线性模型

Generalized Linear Model

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：

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

其中 $y \in \{0,1\}$ ，ϕ 为 $y=1$ 的概率，即 $p(y=1)=\phi$.对上式恒等变形可得：

$$\begin{aligned} p(y) &= \phi^y(1-\phi)^{1-y} \\ &= \exp\left(\ln\left(\phi^y(1-\phi)^{1-y}\right)\right) \\ &= \exp\left(\ln\phi^y+\ln(1-\phi)^{1-y}\right) \end{aligned}$$

广义线性模型

Generalized Linear Model

$$
\begin{aligned}
p(y) &= \exp\left(y \ln \phi + (1-y)\ln(1-\phi)\right) \\
&= \exp\left(y \ln \phi + \ln(1-\phi) - y\ln(1-\phi)\right) \\
&= \exp\left(y\left(\ln \phi - \ln(1-\phi)\right) + \ln(1-\phi)\right) \\
&= \exp\left(y \ln\left(\frac{\phi}{1-\phi}\right) + \ln(1-\phi)\right)
\end{aligned}
$$

对比指数族分布的一般形式 $p(y;\eta) = b(y)\exp\left(\eta^T T(y) - a(\eta)\right)$ 可知:

广义线性模型

Generalized Linear Model

伯努利分布的指数族分布对应参数为:

$$b(y) = 1$$

$$\eta = \ln\left(\frac{\phi}{1-\phi}\right)$$

$$T(y) = y$$

$$a(\eta) = -\ln(1-\phi) = \ln(1+e^{\eta})$$

广义线性模型

Generalized Linear Model

广义线性模型的三条假设:

1.在给定 $\boldsymbol{x}$ 的条件下，假设随机变量 y 服从某个指数族分布;

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

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

参考文献: Andrew Ng. cs229-notes1

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

Logistic Regression

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服从伯努利分布。此时，如果我们想要构建一个线性模型来预测在给定 $\boldsymbol{x}$ 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。

对数几率回归

Logistic Regression

已知y是服从伯努利分布，而伯努利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 $h(\boldsymbol{x})$ 的表达式为：

$$h(\boldsymbol{x}) = E[T(y|\boldsymbol{x})]$$

由于伯努利分布的 $T(y|\boldsymbol{x}) = y|\boldsymbol{x}$ 所以：

$$h(\boldsymbol{x}) = E[y|\boldsymbol{x}]$$

又因为 $E[y|\boldsymbol{x}] = 1 \times p(y=1|\boldsymbol{x}) + 0 \times p(y=0|\boldsymbol{x}) = p(y=1|\boldsymbol{x}) = \phi$ 所以：

$$h(\boldsymbol{x}) = \phi$$

对数几率回归

Logistic Regression

在前面证明伯努利分布属于指数族分布时我们知道：

$$\eta = \ln\left(\frac{\phi}{1-\phi}\right)$$

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

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

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

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

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

对数几率回归

Logistic Regression

将 $\phi = \dfrac{1}{1+e^{-\eta}}$ 代入 $h(\boldsymbol{x})$ 的表达式可得：

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

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

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

此即为对数几率回归模型。

对数几率回归

Logistic Regression

极大似然估计法：

设总体的概率密度函数（或分布律）为 $f(y, w_1, w_2, ..., w_k)$ ，$y_1, y_2, \ldots, y_m$ 为从该总体中抽出的样本。因为 $y_1, y_2, \ldots, y_m$ 相互独立且同分布，于是，它们的联合概率密度函数（或联合概率）为

$$L(y_1, y_2, ..., y_m; w_1, w_2, ..., w_k) = \prod_{i=1}^{m} f(y_i, w_1, w_2, ..., w_k)$$

其中，$w_1, w_2, \ldots, w_k$ 被看作固定但是未知的参数。当我们已经观测到一组样本观测值 $y_1, y_2, \ldots, y_m$ 时，要去估计未知参数，一种直观的想法就是，哪一组参数值使得现在的样本观测值出现的概率最大，哪一组参数可能就是真正的参数，我们就用它作为参数的估计值，这就是所谓的极大似然估计。

对数几率回归

Logistic Regression

极大似然估计的具体方法：

通常记 $L(y_1, y_2, ..., y_m; w_1, w_2, ..., w_k) = L(\boldsymbol{w})$ ，并称其为似然函数。于是求 $\boldsymbol{w}$ 的极大似然估计就归结为求 $L(\boldsymbol{w})$ 的最大值点。由于对数函数是单调递增函数，所以

$$\ln L(\boldsymbol{w}) = \ln\left(\prod_{i=1}^{m} f(y_i, w_1, w_2, ..., w_k)\right) = \sum_{i=1}^{m} \ln f(y_i, w_1, w_2, ..., w_k)$$

与 $L(\boldsymbol{w})$ 有相同的最大值点，而在许多情况下，求 $\ln L(\boldsymbol{w})$ 的最大值点比较简单，于是，我们就将求 $L(\boldsymbol{w})$ 的最大值点转化为了求 $\ln L(\boldsymbol{w})$ 的最大值点，通常称 $\ln L(\boldsymbol{w})$ 为对数似然函数。

对数几率回归

Logistic Regression

对数几率回归的极大似然估计:

已知随机变量y取1和0的概率分别为

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

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

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

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

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

对数几率回归

Logistic Regression

记

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

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

于是，使用一个小技巧即可得到随机变量y的分布律表达式

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

此即为式3.26

或者

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

对数几率回归

Logistic Regression

根据对数似然函数的定义可知

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

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可

$$\ell(\boldsymbol{w}, b) := \ln L(\boldsymbol{w}, b) = \sum_{i=1}^{m} \ln p(y_i | \boldsymbol{x}_i; \boldsymbol{w}, b)$$ 此即为式3.25

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

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

对数几率回归

Logistic Regression

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

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

对数几率回归

Logistic Regression

deepshare.net 深度之眼

由于 $y_i \in \{0,1\}$ ，所以

当 $y_i = 0$ 时：

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{m} \left(\ln(0 \cdot e^{\boldsymbol{\beta}^T \hat{\boldsymbol{x}}_i} + 1 - 0) - \ln(1 + e^{\boldsymbol{\beta}^T \hat{\boldsymbol{x}}_i}) \right) = \sum_{i=1}^{m} \left(\ln 1 - \ln(1 + e^{\boldsymbol{\beta}^T \hat{\boldsymbol{x}}_i}) \right) = \sum_{i=1}^{m} \left(-\ln(1 + e^{\boldsymbol{\beta}^T \hat{\boldsymbol{x}}_i}) \right)$$

当 $y_i = 1$ 时：

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

综合可得

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

加个负号即为式3.27

对数几率回归

Logistic Regression

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

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

对数几率回归

Logistic Regression

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

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

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

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

加个负号即为式3.27

结 语

在这次课程中，我们学习了西瓜书

对数几率回归的公式推导

那么在下次课程中，我们将会学习西瓜书

决策树的公式推导