General Linear Model

Exponential Family

广义线性模型的核心是Exponential Family，它包含了机器学习中最重要的分布形式（伯努利分布，高斯分布，均匀分布，gamma分布，t分布）。Exponential Family是指一种分布的概率密度函数满足以下条件，

$$P(x|\eta) = h(x)\exp[\eta^T T(x) - A(\eta)]$$

在广义线性模型中，假设

- 变量y的分布满足ExpFamily，即$P(y|\eta) = h(y)\exp[\eta^T T(y) - A(\eta)]$，在二分类问题中y满足伯努利分布
- 对于给定的输入x，目标是输出$h(x) = E(T(y)|x)$。
- 令$\eta = \Theta^T X$，代入目标函数

对数几率回归

基于ExpFamily，广义线性模型可以处理二分类问题，假设分类的输出$y = \{0, 1\}, \phi(y=1) = \phi, \phi(y=0) = 1-\phi$。显然输出y满足伯努利分布，

$$\begin{aligned} P(y) = \phi^y(1-\phi)^{1-y} &= \exp[y\ln\phi + (1-y)\ln(1-\phi)] \\ &= \exp[y\ln\frac{\phi}{1-\phi} + \ln(1-\phi)] \end{aligned}$$

对比ExpFamily形式，$\eta = \ln\frac{\phi}{1-\phi}, \phi(y) = y, a(y) = \ln(1-\phi)$，可以得到sigmoid函数，

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

我们的目标函数是

$$h(y|x) = E(y|x) = \prod_{i=1}^{m} y^i P(y^i)$$

得到他的对数似然函数，并用极大似然法求解系数，

$$\begin{aligned} \ell(\Theta) = \sum_{i=1}^{m}[y^i\Theta^T X^i - \ln(1+e^{\Theta^T X})] \\ \Theta^* = arg_\Theta \max \ell(\Theta) \end{aligned}$$

对对数似然函数求梯度

$$\frac{\partial \ell(\Theta)}{\partial \Theta} = \sum_k \frac{\partial \ell(\Theta)}{\partial \Theta_k} = \sum_{i=1}^{m}[-y^i X^i + \frac{X^i e^{\Theta^T X^i}}{1+e^{\Theta^T X^i}}]$$

Softmax(多分类问题)

Softmax是用于求解多分类问题（类间无交集）的模型，输出y满足多项分布$y^i = \{1, 2, \ldots, k\}; p(y^i = t) = \phi_t$最后一项满足$p(y^i = k) = 1 - \sum_{t=1}^{k-1}\phi_t$

我们定义ExpFamily中的T(y)为这样的形式，

$$T(1)=\begin{pmatrix}1\\0\\0\\\vdots\end{pmatrix}\quad T(k-1)=\begin{pmatrix}0\\0\\\vdots\\1\end{pmatrix}\quad T(k)=\begin{pmatrix}0\\0\\0\\\vdots\end{pmatrix}$$

即等价于对这样的结果分类，各项的元素可以表示成$T(y^i)_t=\mathbb{I}(y^i=t), t=\{1,\cdots,k-1\}$。显然y满足多项分布，

$$P(y)=\prod_{i=1}^{k}\phi_i^{\mathbb{I}(y^i=i)}=\exp[\eta^TT(y)-A(\eta)]$$

其中

$$\eta=\begin{pmatrix}\ln(\frac{\phi_1}{\phi_k})\\\ln(\frac{\phi_2}{\phi_k})\\\vdots\\\ln(\frac{\phi_{k-1}}{\phi_k})\end{pmatrix}\quad A(\eta)=-\ln(\phi_k)$$

得到各分类的概率函数

$$\phi_i=\frac{e^{\eta_i}}{1+\sum_{j=1}^{k-1}e^{\eta_j}}, i=\{1,2,\ldots,k-1\}\quad \phi_k=\frac{1}{1+\sum_{j=1}^{k-1}e^{\eta_j}}$$

令$\eta_k=0$,所以,这里$X^i=[x_1^i,x_2^i,\cdots,x_d^i,1]$,

$$\phi_t=\frac{e^{\eta_t}}{\sum_{j=1}^{k}e^{\eta_j}}=\frac{e^{\Theta_t^TX^i}}{\sum_{j=1}^{k}e^{\Theta_j^TX^i}}, i=\{1,2,\ldots,k\}$$

同理目标函数是

$$h(y|x)=E(y|x)=\prod_{i=1}^{m}P^{\mathbb{I}(y^i=k)}(y^i=k)$$

对数似然函数，并用极大似然法求解系数

$$\ell(\Theta)=\sum_{i=1}^{m}\mathbb{I}(y^i=k)\ln\phi_k=\sum_{i=1}^{m}\mathbb{I}(y^i=k)\ln\frac{e^{\Theta_k^TX^i}}{\sum_{j=1}^{k}e^{\Theta_j^TX^i}}$$

求梯度

$$\begin{aligned}\frac{\partial\ell(\Theta)}{\partial\Theta_t}=\sum_l\frac{\partial\ell(\Theta)}{\partial\Theta_{tl}}&=\sum_l\sum_{i=1}^{m}\mathbb{I}(y^i=k)\frac{1}{\phi_k}\frac{\mathbb{I}(k=t)X_le^{\Theta_t^TX^i}\sum_{j=1}^{k}e^{\Theta_j^TX^i}-e^{\Theta_k^TX^i}e^{\Theta_t^TX^i}X_l^i}{(\sum_{j=1}^{k}e^{\Theta_j^TX^i})^2}\\&=\sum_{i=1}^{m}\left[\mathbb{I}(y^i=k)\mathbb{I}(k=t)X^i-\mathbb{I}(y^i=k)\phi_tX^i\right]\end{aligned}$$

其他方法

对于类间有交集的模型，可以使用One VS One或One VS Rest，他们都可以直接从二分类法拓展出来。