天津大学
TIANJIN UNIVERSITY

300072 TIANJIN, CHINA

2.4.0

参数为η的变量x的指数族分布

$$p(x|\eta) = h(x)g(\eta)\exp\{\eta^T u(x)\}.$$

其中 $g(\eta)\int h(x)\exp\{\eta^T u(x)\}\,dx = 1$ 保证分布归一化

例1：伯努利分布 $p(x|\mu) = \mu^x(1-\mu)^{1-x}$

$\therefore p(x|\mu) = e^{x\ln\mu + (1-x)\ln(1-\mu)} = (1-\mu)e^{x\ln\frac{\mu}{1-\mu}}$

$\therefore \eta = \ln\frac{\mu}{1-\mu}$ 即 $\mu = \frac{1}{1-e^{-\eta}}$ 称为 logistic sigmoid 函数.

例2：多项式分布. $p(x|\mu) = \prod_{k=1}^{M}\mu_k^{x_k} = \exp\{\sum_{k=1}^{M}x_k\ln\mu_k\}$

$\because \sum_{k=1}^{M}x_k\ln\mu_k = [\ln\mu_1, \ln\mu_2\cdots\ln\mu_M][x_1\ x_2\cdots x_M]^T.$

$\therefore \eta = [\ln\mu_1, \ln\mu_2\cdots\ln\mu_M]^T.$

且 $u(x) = x$ $\quad h(x) = 1$ $\quad g(\eta) = 1$

$\because \sum_{k=1}^{M}\mu_k = 1$ 由 $\mu_k = 1 - \sum_{k=1}^{M-1}\mu_k$. 得

$$\exp\{\sum_{k=1}^{M}x_k\ln\mu_k\} = \exp\{\sum_{k=1}^{M-1}x_k\ln\mu_k + (1-\sum_{k=1}^{M-1}x_k)\ln(1-\sum_{k=1}^{M-1}\mu_k)\}.$$

$$= \exp\{\sum_{k=1}^{M-1}x_k\ln\frac{\mu_k}{1-\sum_{j=1}^{M-1}\mu_j} + \ln(1-\sum_{k=1}^{M-1}\mu_k)\}.$$

令 $\eta_k = \ln\frac{\mu_k}{1-\sum_{j=1}^{M-1}\mu_j} \Rightarrow \mu_k = \frac{e^{\eta_k}}{1+\sum_{j=1}^{M-1}e^{\eta_j}}$ 称为 softmax 函数

或归一化指数.

2.4.1

对于 $p(x|\mu)=h(x)g(\eta)\exp\{\eta^T u(x)\}$，求 η 的最大似然估计.

对式子两侧边求梯度:

$$\nabla g(\eta)\int h(x)\exp\{\eta^T u(x)\}dx + g(\eta)\int h(x)\exp\{\eta^T u(x)\}u(x)dx = 0.$$

$$即\quad -\frac{1}{g(\eta)}\nabla g(\eta) = \int h(x)\exp\{\eta^T u(x)\}u(x)\,dx$$

$$= E[u(x)]$$

$$\therefore -\nabla \ln g(\eta) = E[u(x)]$$

$\therefore$ 对于 $X=\{x_1 \cdots x_N\}$ 独立同分布的数据集.

$$-\nabla \ln g(\eta_{ML}) = \frac{1}{N}\sum_{n=1}^{N}u(x_n)$$

可以看出 η_{ML} 只与 $\sum_{n=1}^{N}u(x_n)$ 有关 $\therefore$ $\sum_{n=1}^{N}u(x_n)$ 称为分布的充分统计量.

我们不需要存储整个数据集本身，只需要存储充分统计量的值即可

例如 对于伯努利分布 $u(x)=x$ $\therefore$ 只需存储 $\{x_n\}$ 的和即可

对于高斯分布 $u(x)=(x,x^2)^T$ $\therefore$ 只需存储 $\{x_n\}$ 和 $\{x_n^2\}$ 的和即可

2.4.3.

无信息先验：在对分布应该具有的形式几乎完全不知道时，可寻找一种形式的先验分布.（令先验分布为均匀分布）.

困难一：对于由参数控制的分布 $p(x|\lambda)$ 若 λ 无界 则先验分布无法归一化

因为对 λ 的积分是发散的，这样的先验分布是反常的.

困难二：若概率发生非线性变量的概率密度的变换

例：令 $p_\lambda(\lambda)=C$ $\therefore$ 由 $\lambda=\eta^2$

$$p_\eta(\eta)=p_\lambda(\lambda)\left|\frac{d\lambda}{d\eta}\right| = p_\lambda(\eta^2)2\eta \quad 不再为常数.$$

对①：若对应的后验分布是正常的，即可被归一化，则可以使用反常先验

对②：注意对于参数要使用一个合适的表达形式