朴素贝叶斯分类器

贝叶斯分类器是将数据$X^i=\{x_1^i,x_2^2,\cdots,x_d^i\}$分类成$Y=\{c_1,c_2,\cdots,c_k\}$的决策方法。不同于线性回归模型，贝叶斯分类器是一种生成式模型。机器学习主要有两种策略，

- 判别式模型：给定X，可通过直接建模$P(c|x)$来预测c
- 生成式模型：也可对联合概率分布$P(x,c)$建模，由此获得$P(c|x)$

高斯判别法

高斯判别法是一种生成式模型，假设$X\in R^d$，X是连续属性的变量。

考虑二分类问题，y满足伯努利分布

$$P(y)=\phi^y(1-\phi)^{1-y},\quad y=\{0,1\}$$

利用极大似然法对y建模，可以得到参数ϕ的取值

$$=\sum_{i=1}^m\frac{y^i-\phi}{\phi(1-\phi)}=0,\quad \phi=\frac{\sum_{i=1}^m\mathbb{I}(y^i=1)}{m}$$

假设条件概率在给定y的情况下x的分布为混合高斯分布，则

$$P(X^i|y^i=0)=P_{\mathcal{M}}(X^i;\vec{\mu}_0,\vec{\epsilon}),\quad P(X^i|y^i=1)=P_{\mathcal{M}}(X^i;\vec{\mu}_1,\vec{\epsilon})$$

混合高斯分布的概率密度

$$P_{\mathcal{M}}(x;\vec{\mu},\vec{\epsilon})=\frac{1}{(2\pi)^{d/2}|\epsilon|^{1/2}}\exp(-\frac{1}{2}(x-\vec{\mu})\epsilon^{-1}(x-\vec{\mu})^T)$$

对数似然函数

$$L(\vec{\mu}_0,\vec{\mu}_1,\vec{\epsilon})=\prod_{i=1}^m P_{\mathcal{M}}^{1-y^i}(X^i;\vec{\mu}_0,\vec{\epsilon})P_{\mathcal{M}}^{y^i}(X^i;\vec{\mu}_1,\vec{\epsilon})$$

同理，对条件概率做极大似然估计，求出系数μ_0,μ_1,ϵ,

$$\frac{\partial\ell(\vec{\mu}_0,\vec{\mu}_1,\vec{\epsilon})}{\partial\vec{\mu}_0}=\sum_{i=1}^m\mathbb{I}(y^i=0)\epsilon^{-1}(X^i-\vec{\mu}_0)=0,\quad \vec{\mu}_0=\frac{\sum_{i=1}^m\mathbb{I}(y^i=0)X^i}{\sum_{i=1}^m\mathbb{I}(y^i=0)}$$

同理，可以求出其他参数，

$$\vec{\mu}_0=\frac{\sum_{i=1}^m\mathbb{I}(y^i=0)X^i}{\sum_{i=1}^m\mathbb{I}(y^i=0)},\quad \vec{\epsilon}=\frac{1}{m}\sum_{i=1}^m(X^i-\mu_{y^i})(X^i-\mu_{y^i})^T$$

基于先验概率和似然概率，可以得到我们的目标后验概率，

$$y^*=\arg_y\max P(y|x)=\arg_y\max P(x|y)P(y)$$

朴素贝叶斯

朴素贝叶斯可用于处理离散属性的数据，或者将连续属性离散化。同理，朴素贝叶斯也是一种生成式学习器，其后验概率同样可以由先验概率和似然概率得到，

$$y^*=\arg_y\max P(y|x)=\arg_y\max P(x|y)P(y)$$

考虑y满足多项分布，假设$P(y^i=k)=\phi_k$，

$$P(y)=\prod_{k=1}^{K}\phi_k^{\mathbb{I}(y=k)},\quad \phi_K=1-\sum_{k=1}^{K-1}\phi_k$$

利用极大似然法求到参数，

$$L(\phi_1,\phi_2,\ldots,\phi_K)=\prod_{i=1}^{m}P(y^i),\quad \ell(\phi_1,\phi_2,\ldots,\phi_K)=\sum_{i,k=1}^{m,K}\mathbb{I}(y^i=k)\log\phi_k$$

求梯度

$$\frac{\partial\ell(\phi_1,\phi_2,\ldots,\phi_K)}{\partial\phi_k}=\sum_{i=1}^{m}(\frac{\mathbb{I}(y^i=k)}{\phi_k}-\frac{\mathbb{I}(y^i=K)}{\phi_K})=0$$

$$\frac{\phi_k}{\phi_K}=\frac{\sum_{i=1}^{m}\mathbb{I}(y^i=k)}{\sum_{i=1}^{m}\mathbb{I}(y^i=K)},\quad k=\{1,2,\ldots,K-1\}$$

所以

$$\phi_k=\frac{\sum_{i=1}^{m}\mathbb{I}(y^i=k)}{m},\quad k=\{1,2,\ldots,K\}$$

考虑条件概率$P(x|y)$，假设X的属性条件独立

$$P(X|y)=\prod_{i=1}^{d}P(x_i|y)$$

假设x_i是离散属性，可以取$\{1,2,..,N_i\}$，对每个x_i，$P(x_i|y=k)$满足多项分布，同理可以求得参数

$$\psi_{nk}^{i}=P(x_i=n|y=k)=\frac{\sum_{j=1}^{m}\mathbb{I}(y^j=k)\mathbb{I}(x_i^j=n)}{\sum_{j=1}^{m}\mathbb{I}(y^j=k)},\quad k=\{1,2,\ldots,K\},n=\{1,2,\ldots,N_i\}$$

拉普拉斯平滑

为避免概率变成0，将参数做如下调整

$$\phi_k=\frac{\sum_{i=1}^{m}\mathbb{I}(y^i=k)+1}{m+K},\quad \psi_{nk}^{i}==\frac{\sum_{j=1}^{m}\mathbb{I}(y^j=k)\mathbb{I}(x_i^j=n)+1}{\sum_{j=1}^{m}\mathbb{I}(y^j=k)+N_i}$$