4.1 用极大似然估计法推导朴素贝叶斯法中的先验概率估计公式和条件概率估计公式

$$P(Y=c_k)=\frac{\sum_{i=1}^{N}I(y_i=c_k)}{N}$$

$$P\left(x^{(j)}=a_{jl}|Y=c_k\right)=\frac{\sum_{i=1}^{N}I(x_i^{(j)}=a_{jl},y_i=c_k)}{\sum_{i=1}^{N}I(Y=c_k)}$$

证明：设$p=P(Y=c_k)$

相当于从样本中独立同分布的随机抽取 N 个样本，结果为y_i。

似然概率$P(y_1,y_2,\dots,y_n)=p^{\sum_{i=1}^{N}I(y_i=c_k)}(1-p)^{\sum_{i=1}^{N}I(y_i\neq c_k)}$

$$\frac{dP(y_1,y_2,\dots,y_n)}{dp}=p^{\sum_{i=1}^{N}I(y_i=c_k)-1}(1-p)^{\sum_{i=1}^{N}I(y_i\neq c_k)-1}\left((1-p)\sum_{i=1}^{N}I(y_i=c_k)-p\sum_{i=1}^{N}I(y_i\neq c_k)\right)=0$$

http://blog.csdn.net/xiaoxiao_wen

$$(1-p)\sum_{i=1}^{N}I(y_i=c_k)-p\sum_{i=1}^{N}I(y_i\neq c_k)=0$$

$$\sum_{i=1}^{N}I(y_i=c_k)=p\left(\sum_{i=1}^{N}I(y_i=c_k)+\sum_{i=1}^{N}I(y_i\neq c_k)\right)=pN$$

所以极大似然估计为

$$P(Y=c_k)=\frac{\sum_{i=1}^{N}I(y_i=c_k)}{N}$$

同理容易得到

$$P\left(Y=c_k,x^{(j)}=a_{jl}\right)=\frac{\sum_{i=1}^{N}I(y_i=c_k,x_i^{(j)}=a_{jl})}{N}$$

$$\begin{aligned}P\left(x^{(j)}=a_{jl}|Y=c_k\right)&=\frac{P\left(Y=c_k,x^{(j)}=a_{jl}\right)}{P(Y=c_k)}\\&=\frac{\sum_{i=1}^{N}I(y_i=c_k,x_i^{(j)}=a_{jl})}{N}/\frac{\sum_{i=1}^{N}I(y_i=c_k)}{N}\\&=\frac{\sum_{i=1}^{N}I(y_i=c_k,x_i^{(j)}=a_{jl})}{\sum_{i=1}^{N}I(y_i=c_k)}\end{aligned}$$

http://blog.csdn.net/xiaoxiao_wen

4.2 用贝叶斯估计法推导以下公式

$$P(Y=c_k)=\frac{\sum_{i=1}^{N}I(Y=c_k)+\lambda}{N+K\lambda}$$

$$P\left(x^{(j)}=a_{jl}|Y=c_k\right)=\frac{\sum_{i=1}^{N}I(x^{(j)}=a_{jl},Y=c_k)+\lambda}{\sum_{i=1}^{N}I(Y=c_k)+S_j\lambda}$$

其中λ是参数，K 是 Y 可取的值数，S_j是$x^{(j)}$可取的值数。

解：加入先验概率，在没有任何信息的情况下可以假设先验概率为均匀概率。

即有方程$p=\frac{1}{K}\Leftrightarrow pK-1=0 \quad (1)$

另外上一题中我们已经得到了极大似然下的条件概率约束

$pN-\sum_{i=1}^{N}I(y_i=c_k)=0 \quad (2)$

$$(1)*\lambda+(2)=0$$

http://blog.csdn.net/xiaoxiao_wen

所以有

$$\lambda(pK-1)+pN-\sum_{i=1}^{N}I(y_i=c_k)=0$$

有$P(Y=c_k)=\frac{\lambda+\sum_{i=1}^{N}I(y_i=c_k)}{\lambda K+N}$

同理容易得到

$$P\left(Y=c_k, x^{(j)}=a_{jl}\right)=\frac{\sum_{i=1}^{N} I\left(y_i=c_k, x_i^{(j)}=a_{jl}\right)+\lambda}{N+KS_j\lambda}$$

$$P\left(x^{(j)}=a_{jl} \mid Y=c_k\right)=\frac{P\left(Y=c_k, x^{(j)}=a_{jl}\right)}{P(Y=c_k)}$$

$$=\frac{\sum_{i=1}^{N} I\left(y_i=c_k, x_i^{(j)}=a_{jl}\right)+\lambda}{N+KS_j\lambda} \Big/ \frac{\sum_{i=1}^{N} I(y_i=c_k)+\lambda}{N+K\lambda}$$

http://blog.csdn.net/xiaoxiao_wen

$=(\lambda$可随便取，所以右边取$\lambda=S_j\lambda)$

$$=\frac{\sum_{i=1}^{N} I\left(y_i=c_k, x_i^{(j)}=a_{jl}\right)+\lambda}{N+KS_j\lambda} \Big/ \frac{\sum_{i=1}^{N} I(y_i=c_k)+S_j\lambda}{N+KS_j\lambda}$$

$$=\frac{\sum_{i=1}^{N} I\left(y_i=c_k, x_i^{(j)}=a_{jl}\right)+\lambda}{\sum_{i=1}^{N} I(y_i=c_k)+S_j\lambda}$$