朴素贝叶斯法

用于分类模型 生成模型

输入空间 $X \in R^n$，输出空间 $Y \in \{C_1, C_2, \cdots C_k\}$

原理：

贝叶斯公式：$P(a|b) = \frac{P(ab)}{P(b)} = \frac{P(b|a)P(a)}{P(b)}$

输入X，将后验概率最大的C_j作为输出

$$P(C_j|x) = \frac{P(x|C_k)P(C_k)}{P(x)}$$

$P(x)$ ∵ 输入X给定，$P(x)$固定

$$P(C_j|x) \propto P(x|C_k)P(C_k)$$

条件独立性假设：

设输入X有n维，

则 $P(X=x^{(1)}, \cdots x^{(n)} | C_k)$

$= \prod_{i=1}^{n} P(x^{(i)}|C_k)$

则 输出 $Y = \arg\max_{C_k} \boxed{P(C_k)} \boxed{\prod_{i=1}^{n} P(x^{(i)}|C_k)}$

后验概率最大化

后验概率最大化等价于期望风险最小化（对于朴素贝叶斯来说）

损失函数 $L(Y, f(x)) = \begin{cases} 1 & Y \neq f(x) \\ 0 & Y = f(x) \end{cases}$

期望风险 $R_{exp}(f) = E[L(Y, f(x))]$

$= \sum_x \sum_y L(Y, f(x)) P(x, y)$

$= \sum_x (\sum_y L(Y, f(x)) P(x, y))$ ∵ $Y \in \{C_1, C_2, \cdots C_n\}$

$= \sum_x (\sum_{C_k} L(C_k, f(x)) P(C_k, x))$

$= \sum_x (\sum_{C_k} L(C_k, f(x)) P(C_k|x) P(x))$

$= \sum_x (\sum_{C_k} L(C_k, f(x)) P(C_k|x)) P(x)$

mix $R_{exp}(f)$ 等价于对 $X=x$ 逐个极小化，也就是说

$$\min R_{exp}(f) = \min \sum_x \left(\boxed{\sum_{C_k} L(C_k, f(x)) P(C_k|x))}\right) p(x)$$

(=) 等价于 极小化 这项

$$\propto \min \sum_{C_k} L(C_k, f(x)) P(C_k|x))$$

$$= \min \sum_{C_k} I(f(x) \neq C_k) P(C_k|x))$$

$$= \min \sum_{C_k} (1 - I(f(x) = C_k)) P(C_k|x))$$

$$\propto \max \sum_{C_k} I(f(x) = C_k) P(C_k|x)$$

也就是说要最大化 $\sum_{C_k} I(f(x)=C_k) P(C_k|x)$

因为 $f(x)$ 只能等于一个类 C_i

example: 输入 x，共有3类，C_1, C_2, C_3

$P(C_1|x)=0.2$ $P(C_2|x)=0.6$ $P(C_3|x)=0.2$

那么，若 $f(x)=C_1$ 那么 $\sum_{C_k} I(f(x)=C_k) P(C_k|x) = 1\times0.2+0\times0.6+0\times0.2$
$= 0.2$

若 $f(x)=C_2$，那么 $\sum_{C_k} I(f(x)=C_k) P(C_k|x) = 0\times0.2+1\times0.6+0\times0.2$
$= 0.6$

若 $f(x)=C_3$，则 $\sum_{C_k} I(f(x)=C_k) P(C_k|x) = 0\times0.2+0\times0.6+1\times0.2$
$= 0.2$

因此，取 $P(C_k|x)$ 最大的作为 $f(x)$，可以最大化 $\sum_{C_k} I(f(x)=C_k) P(C_k|x)$

$$\Rightarrow f(x) = \arg\max_{C_k} P(C_k|x)$$

朴素贝叶斯算法

根据上述，我们知道，对于一个输入 x，输出 $y = \arg\max_{C_k} P(C_k) \prod_{i=1}^{n} P(x^{(i)}|C_k)$

那么现在的问题是 $\boxed{P(C_k)}$ 和 $\boxed{P(x^{(i)}|C_k)}$ 怎么求？

极大似然估计

输出空间 $Y=\{C_1, C_2, \dots C_k\}$

对应的概率分布 $C_1\ C_2 \dots C_k$

$\boxed{\theta_1\ \theta_2 \dots \theta_k}$ 习参数θ 要估计的 也就是要估计 $P(k)$

则 $P(Y=y|\theta)=\theta_1^{I(y=C_1)}\theta_2^{I(y=C_2)}\dots\theta_k^{I(y=C_k)}$

有 $y_1, y_2, \dots y_n$

则 $P(y_1, y_2, \dots y_n|\theta)=\prod_{i=1}^{n}\theta_1^{I(y_i=C_1)}\theta_2^{I(y_i=C_2)}\dots\dots\theta_k^{I(y_i=C_k)}$

$=\theta_1^{m_1}\theta_2^{m_2}\dots\theta_k^{m_k}$

m_1 习 $y_1 \dots y_n$ 中有 m_1 个 y 等于 C_1

m_2 习 $y_1 \dots y_n$ 中有 m_2 个 y 等于 C_2

$m_1+m_2+\dots+m_k=n$.

$\max P(y_1, y_2 \dots y_n|\theta) \propto \max L(P(y_1, y_2, \dots y_n|\theta))$

$\propto \max[m_1\log\theta_1+m_2\log\theta_2+\dots+m_k\log\theta_k]$

$\propto \max[m_1\log\theta_1+m_2\log\theta_2+\dots+m_k\log\theta_k+\lambda(\theta_1+\theta_2+\dots+\theta_k-1)]$

对 θ_1 求导, $\frac{m_1}{\theta_1}+\lambda=0$ $\quad \theta_1=\frac{-m_1}{\lambda}$

$\vdots \qquad\qquad \vdots$

对 θ_k 求导 $\frac{m_k}{\theta_k}+\lambda=0$ $\quad \theta_k=\frac{-m_k}{\lambda}$

$\theta_1+\theta_2+\dots+\theta_k=1$ 则有 $\frac{-(m_1+m_2+\dots+m_k)}{\lambda}=1$

$\frac{-n}{\lambda}=1 \quad \lambda=-n$

则有 $\theta_1=\frac{-m_1}{\lambda}=\frac{-m_1}{-n}=\frac{m_1}{n}=\frac{\sum_{i=1}^{n}I(y_i=C_1)}{n}$

$\vdots$

$\theta_k=\frac{\sum_{i=1}^{n}I(y_i=C_k)}{n}$

同理，利用极大似然估计 $P(x^{(i)}|C_k)$

输入空间 $x^{(i)}=\{a_1,\dots a_L\}$

对应的概率分布 $P(x^{(i)}=a_1|C_k)$ $P(x^{(i)}=a_2|C_k)$ $\cdots$ $P(x^{(i)}=a_L|C_k)$

p_1 p_2 p_L.

剩下步骤同理

贝叶斯估计

估计 $P(C_k)$

输出空间 $Y=\{C_1,C_2,\dots C_k\}$

概率分布 C_1 C_2 $\cdots$ C_k

θ_1 θ_2 $\cdots$ θ_k

$$P(\theta_1\cdots\theta_k)=\frac{\Gamma(\alpha_1+\alpha_2+\cdots+\alpha_k)}{\Gamma(\alpha_1)+\Gamma(\alpha_2)+\cdots+\Gamma(\alpha_k)}\theta_1^{\alpha_1-1}\theta_2^{\alpha_2-1}\cdots\theta_k^{\alpha_k-1}$$

$$P(\theta_1\cdots\theta_k|y_1,\cdots y_n)=\frac{P(y_1,y_2\cdots y_n|\theta_1\cdots\theta_k)P(\theta_1\theta_2\cdots\theta_k)}{P(y_1,y_2\cdots y_n)}$$ 与θ无关

$$\propto P(y_1,y_2\cdots y_n|\theta_1\theta_2\cdots\theta_k)P(\theta_1\theta_2\cdots\theta_k)$$

$$\propto(\theta_1^{\alpha_1-1}\theta_2^{\alpha_2-1}\cdots\theta_k^{\alpha_k-1})(\theta_1^{m_1}\theta_2^{m_2}\cdots\theta_k^{m_k})$$

$$\propto\theta_1^{m_1+\alpha_1-1}\theta_2^{m_2+\alpha_2-1}\cdots\theta_k^{m_k+\alpha_k-1}$$

则 $\max P(\theta_1\theta_2\cdots\theta_k|y_1y_2\cdots y_n)\propto\max\log P(\theta_1\theta_2\cdots\theta_k|y_1y_2\cdots y_n)$

$$\propto\max[(m_1+\alpha_1-1)\log\theta_1+(m_2+\alpha_2-1)\log\theta_2+\cdots+(m_k+\alpha_k-1)\log\theta_k]$$

$$\propto\max[(m_1+\alpha_1-1)\log\theta_1+\cdots+(m_k+\alpha_k-1)\log\theta_k+\lambda(\theta_1+\theta_2+\cdots\theta_k-1)]$$

对 θ_1 求导. $\frac{m_1+\alpha_1-1}{\theta_1}+\lambda=0$ $\quad\theta_1=\frac{1-m_1-\alpha_1}{\lambda}$

$\vdots$

$\theta_k=\frac{1-m_k-\alpha_k}{\lambda}$

$$\theta_1+\theta_2+\cdots+\theta_k = \Rightarrow \frac{k-n-(a_1+a_2+\cdots+a_k)}{\lambda} = 1$$

$$\lambda = k-n-(a_1+\cdots+a_k)$$

$$\theta_1 = \frac{1-m_1-a_1}{k-n-(a_1+\cdots+a_k)}$$

一般令 $a_1=a_2=\cdots=a_k=a$

则 $\theta_1 = \dfrac{1-a-m_1}{k-n-ka} = \dfrac{m_1+a-1}{n+ka-k} = \dfrac{m_1+a-1}{n+k(a-1)}$.