朴素贝叶斯（Naive Bayesian Classifier）

限制条件：① $X=\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_M \end{bmatrix}$ 每个维度离散

② X的每个维度独立

应用：垃圾邮件分类

输入：一个文件 d

输出：$d \in C_1$ 或 $d \in C_2$

训练样本：$\{(d_i, c_i) \mid i=1 \sim N\}$

$d=\{w_1, w_2, \dots, w_n\}$（单词）

学习 $P(d|C_1)$ 与 $P(d|C_2)$

$$P(d|C)$$
$$= P(\{w_1, \dots, w_n\}|C)$$
$$= \prod_{i=1}^{n} P(w_i|C) \quad \text{（独立假设）}$$

$$P(w|C_j) = \frac{count(w, C_j)+1}{\sum_{w \in V} count(w, C_j) + |V|} \quad (j=1 \sim 2)$$

（种类）

$P(C_1) = \frac{C_1\text{的个数}}{\text{总个数}}$ $P(C_2) = \frac{C_2\text{的个数}}{\text{总个数}}$

若 $P(C_1) \cdot P(d|C_1) > P(C_2) \cdot P(d|C_1)$ 则 $d \in C_1$

否则 ... $d \in C_2$