步骤一：

设训练集X数量为$N(N_1+N_2=N)$

定义C_1 : 个人收入$> 50K$, 数量为N_1;　C_2 : 个人收入$< 50K$,数量为N_2

步骤二：

假设$\forall x_i \in X$均服从$Gaussian$分布,其中

$x_1, \cdots, x_{N_1} \sim N(\mu_1, \sum)$, 　　$z_1, \cdots, z_{N_2} \sim N(\mu_2, \sum)$

步骤三：

建立极大似然函数

$$L(\mu_1, \mu_2, \sum)(x_1, \cdots, x_{N_1}, z_1, \cdots, z_{N_2}) = f_1(x_1) * \cdots * f_1(x_{N_1}) * f_2(z_1) * \cdots * f_2(z_{N_2})$$

对上式两边取对数再分别对$\mu_1, \mu_2, \sum$求偏导取值为0，可以得到极大似然估计的参数$\mu_1, \mu_2, \sum$，事实上，

$$\frac{\partial lnL}{\partial \mu_1} = 0 \Rightarrow \mu_1 = \frac{1}{N_1}\sum_{i=1}^{N_1} x_i$$

$$\frac{\partial lnL}{\partial \mu_2} = 0 \Rightarrow \mu_2 = \frac{1}{N_2}\sum_{j=1}^{N_2} z_j$$

$$\frac{\partial lnL}{\partial \sum} = 0 \Rightarrow \sum = \frac{N_1}{N}(\frac{1}{N_1}\sum_{i=1}^{N_1}(x_i-\mu_1)(x_i-\mu_1)^T) + \frac{N_2}{N}(\frac{1}{N_2}\sum_{j=1}^{N_2}(z_j-\mu_2)(z_j-\mu_2)^T)$$

步骤四：

得到概率函数

$P(C_1|x) = \frac{1}{1+exp-z}$,其中

$$z = (\mu_1-\mu_2)^T(\sum)^{-1}x - \frac{1}{2}(\mu_1)^T(\sum)^{-1}\mu_1 + \frac{1}{2}(\mu_2)^T(\sum)^{-1}\mu_2 + ln\frac{N_1}{N_2}$$

步骤5：

将被测试样本x的特征数据带入步骤四的概率函数中，计算x属于C_1集合的概率，当$P > 0.5$认为测试样本属于C_1集合，否则认为其属于C_2集合