电子科技大学

电子科技大学（深圳）高等研究院

模式识别

专业	控制科学与工程
班级	**5** 班
学生	陈玉熙 **202122280534**
学生	李育泓 **202122280515**
教师	凡时财
日期	**2021.11.25**

目　　录

1 最小错误率贝叶斯决策

1.1 贝叶斯决策

设实验 E 的样本空间为 S，A 为 E 的事件，$B_1, B_2, \ldots, B_n$为 S 的一个划分，且$P(A) > 0$，$P(B_i) > 0$，(i=1,2,⋯,n)，则贝叶斯公式为：

$$P(B_i \mid A) = \frac{P(B_i A)}{P(A)} = \frac{P(A \mid B_i)P(B_i)}{P(A)}$$
$$= \frac{P(A \mid B_i)P(B_i)}{\sum_{j=1}^{n} P(A \mid B_j)P(B_j)}$$

定义条件概率：

$$P(A)P(B_i \mid A) = P(B_i)P(A \mid B_i)$$
$$p(\vec{x})P(\omega_i \mid \vec{x}) = P(\omega_i)p(\vec{x} \mid \omega_i)$$

其中：

先验概率$P(\omega_i)$表示类ω_i出现的先验概率，简称类ω_i的概率；

后验概率$P(\omega_i \mid \vec{x})$表示$\vec{x}$出现条件下类ω_i出现的概率，称其为类别的后验概率，对于模式识别来讲可理解为$\vec{x}$来自类ω_i的概率；

类概率密度$p(\vec{x} \mid \omega_i)$表示在类ω_i条件下的概率密度，简称为类概率密度。

贝叶斯法则即为最大后验概率准则。

对于两类ω_1, ω_2问题，可以根据后验概率做判决：

若$p(\omega_1 \mid \vec{x}) > p(\omega_2 \mid \vec{x})$ 则$\vec{x} \in \omega_1$

若$p(\omega_1 \mid \vec{x}) < p(\omega_2 \mid \vec{x})$ 则$\vec{x} \in \omega_2$

后验概率$P(\omega_i \mid \vec{x})$可由类ω_i的先验概率$P(\omega_i)$和条件概率密度$p(\vec{x} \mid \omega_i)$来表示，即：

$$p(\omega_i \mid \vec{x}) = \frac{p(\vec{x} \mid \omega_i)P(\omega_i)}{p(\vec{x})} = \frac{p(\vec{x} \mid \omega_i)P(\omega_i)}{\sum_{j=1}^{2} p(\vec{x} \mid \omega_j)P(\omega_j)}$$

1.2 基于最小错误率的贝叶斯决策

定义平均错误率为：

$$P(e)=\int_{-\infty}^{\infty} p(e,x)dx=\int_{-\infty}^{\infty} p(e \mid x)\cdot p(x)dx$$

则在讨论两类问题下，条件错误概率为：

$$P(e \mid x)=\begin{cases} P(w_1 \mid x) & \text{当 } P(w_2 \mid x)>P(w_1 \mid x) \text{ 时} \\ P(w_2 \mid x) & \text{当 } P(w_1 \mid x)>P(w_2 \mid x) \text{ 时} \end{cases}$$

判决门限，即两类的分界面 t，如图 1-1 所示：

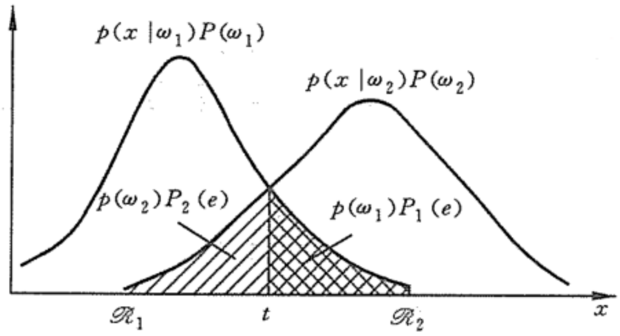

图 1-1 两类分界面图

对于大量样本 x，则总的错误概率是$P(e \mid x)$的数学期望，总错误率为：

$$\begin{aligned} P(e) &= \int_{-\infty}^{\infty} p(e \mid x)p(x)dx \\ &= \int_{-\infty}^{t} P(w_2 \mid x)p(x)dx+\int_{t}^{\infty} P(w_1 \mid x)p(x)dx \\ &= \int_{-\infty}^{t} p(x \mid w_2)P(w_2)dx+\int_{t}^{\infty} p(x \mid w_1)P(w_1)dx \\ &= P(w_2)\int_{-\infty}^{t} p(x \mid w_2)dx+P(w_1)\int_{t}^{\infty} p(x \mid w_1)dx \\ &= P(w_2)p_2(e)+P(w_1)\cdot p_1(e) \end{aligned}$$

最小错误率贝叶斯决策规则，即选择 t 使得对于每个 x，$P(e \mid x)$为最小，则$P(e)$也为最小。

决策规则推广到多类决策：

$$\text{若 } P(w_i \mid x)=\max_{\mathrm{j}=1,2,\dots,c} P\left(w_{\mathrm{j}} \mid x\right) \text{ 则 } x\in w_i$$

1.3 正态分布下最小错误率贝叶斯决策

最小错误率贝叶斯决策的判别函数是：

$$g_i(x)=p(x/w_i)\mathrm{P}(w_i)$$

其中$p(x/w_i)$服从$N(\mu_i, \Sigma_i) i = 1, \dots, c$

$$p(x/w_i) = \frac{1}{(2\pi)^{d/2}|\sum_i|^{1/2}} \exp\left[-\frac{1}{2}(x-u_i)^T \sum_i^{-1} (x-u_i)\right]$$

进行单调的对数变换，则判别函数为：

$$\begin{aligned} g_i(x) &= \ln[p(x/w_i)\mathrm{P}(w_i)] \\ &= -\frac{1}{2}(x-\mu_i)^T \sum_i^{-1} (x-\mu_i) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln\left|\sum_i\right| + \ln \mathrm{P}(w_i) \end{aligned}$$

最小错误率贝叶斯决策的决策面方程是：

$$g_i(x) = g_j(x)$$

即

$$g_i(x) - g_j(x) = 0$$

$$g_i(x) - g_j(x) = -\frac{1}{2}\left[(x-\mu_i)^T \sum_i^{-1} (x-\mu_i) - \left(x-\mu_\mathrm{j}\right)^T \sum_j^{-1} \left(x-\mu_\mathrm{j}\right)\right] - \frac{1}{2}\ln\frac{|\sum_\mathrm{i}|}{|\sum_\mathrm{j}|} + \ln\frac{P(\omega_i)}{P\left(\omega_j\right)} = 0$$

2 最大似然估计（MLE）

2.1 前提条件

（1）待估参数θ是确定的非随机的未知量；

（2）按类别把样本集分为C个子集：$X_1, X_2, \ldots, X_C$，任意一个子集X_i的样本是从总体中独立抽取的，即每一个样本集X_i中样本都是独立同分布的随机变量；

（3）每个类条件概密函数$p(x/w_i)$的形式已知，未知的是参数向量θ_i的值；

（4）不同类别的参数θ_i是独立的，即X_i中的样本不包含$\theta_{\mathrm{j}}(\mathrm{j} \neq \mathrm{i})$的信息，只包含$\theta_i$的信息（$X_i$与$\theta_j$无关）。

2.2 似然函数

设样本子集 $X = \{x_1, x_2, \ldots, x_N\}$，

当样本是独立抽取的，则似然函数为：

$$p(X/\theta) = p(x_1, x_2, \ldots, x_N \mid \theta) = \prod_{K=1}^{N} p(x_K/\theta)$$

当 X 的 N 个样本确定后，$p(X/\theta)$只是θ的函数，记为$l(\theta)$。

2.3 基本思想

（1）若θ已知，当从观测值中抽取样本$x_1, x_2, \ldots, x_N$时，最可能出现的样本是使$l(\theta)$为最大的样本。

（2）若θ未知，X 选定。不同的θ选择，对 N 个样本$x_1, x_2, \ldots, x_N$就有不同的$p(X/\theta)$值，应选择使$x_1, x_2, \ldots, x_N$的似然函数$l(\theta)$为最大的$\hat{\theta}$。

使$p(X/\theta)$达极大值的参数向量$\hat{\theta}$，就是θ的最大似然估计。

显然使$l(\theta)$最大的$\hat{\theta}$是样本$x_1, x_2, \ldots, x_N$的函数，记为$\hat{\theta} = d(x_1, x_2, \ldots, x_N)$。

2.4 正态分布

（1）$\sum$已知，μ未知

似然函数为：

$$H(\mu)=\ln l(\mu)=\ln p(X/\mu)=\ln\prod_{k=1}^{N}p(x_k/\mu)=\sum_{k=1}^{N}\ln p(x_k/\mu)$$

可推导出：

$$\hat{\mu}=\frac{1}{N}\sum_{k=1}^{N}x_k\text{。}$$

（2）μ、$\sum$均未知（考虑一维情况 d=1）

似然函数为：

$$H(\theta)=\ln l(\theta)=\ln p(X/\theta)=\ln\prod_{k=1}^{N}p(x_k/\theta)=\sum_{k=1}^{N}\ln p(x_k/\theta)$$

可推导出：

$$\hat{\mu}=\frac{1}{N}\sum_{k=1}^{N}x_k\text{、}\hat{\sigma}^2=\hat{\theta}_2=\frac{1}{N}\sum_{k=1}^{N}(x_k-\hat{\mu})^2$$

3 贝叶斯估计（BE）

3.1 基本思想

待估参数θ为服从某种先验分布的随机量，其先验概率密度为$p(\theta)$，利用已知的训练样本，使θ的初始密度估计 转化为后验概率密度$p(\theta/X)$。

定义估计风险：$\lambda(\hat{\theta},\theta)$，对应$\lambda(\alpha_i,w_i)$。

总平均风险为：

$$R=\int_{E^d} R(\alpha_i/x)p(x)dx=\int_{E^d}\sum_{j=1}^{C}\lambda\left(\alpha_i,w_j\right)\mathrm{P}\left(w_j/x\right)\cdot p(x)dx$$

$$=\int_{E^d}\sum_{j=1}^{C}\lambda\left(\alpha_i,w_j\right)\mathrm{P}\left(x,w_j\right)dx=\int_{E^d} R(\hat{\theta}/x)p(x)dx$$

定义风险条件：$R(\hat{\theta}/x)=\int_{\Theta}\lambda(\hat{\theta},\theta)p(\theta/x)d\theta$，可得到：

$$R=\int_{E^d} R(\hat{\theta}/x)p(x)dx=\int_{E^d}\int_{\Theta}\lambda(\hat{\theta},\theta)p(\theta/x)\cdot p(x)d\theta dx$$

$$=\int_{E^d}\int_{\Theta}\lambda(\hat{\theta},\theta)p(x,\theta)d\theta dx$$

最终目标是求得的θ的估计值$\hat{\theta}$应使 R 最小，即等价于求使条件风险$R(\hat{\theta}/x)$最小的估计值$\hat{\theta}$。

3.2 正态分布

假设为一维正态分布，已知σ^2，估计μ。

假设概率密度服从正态分布$\mathrm{p(x/}\mu) \sim \mathrm{N}(\mu,\sigma^2),\mathrm{p}(\mu) \sim \mathrm{N}(\mu_0,{\sigma_0}^2)$，已知第 i 类样本集$X=(x_1,x_2,\ldots x_N)^T$，第 i 类联合概率密度$\mathrm{p(X/}\mu,\mathrm{x})=\mathrm{p(X/}\mu)$

μ后验概率为：

$$p(\mu\mid X)=\frac{p(X\mid\mu)\cdot p(\mu)}{\int p(X\mid\mu)p(\mu)d\mu}$$

$p(\mu\mid X)$仍然是一个正态函数，$\mathrm{p}(\mu/\mathrm{X}) \sim \mathrm{N}(\mu_{\mathrm{N}},{\sigma_{\mathrm{N}}}^2)$，可直接写成正态形式：

$$p(\mu \mid X)=\frac{1}{\sqrt{2\pi}\sigma_N}\exp\left[-\frac{1}{2}\left(\frac{\mu-\mu_N}{\sigma_N}\right)^2\right]$$

对μ的估计为：

$$\hat{\mu}=\mu_N=\frac{{\sigma_0}^2}{N_{{\sigma_0}^2}+\sigma^2}\sum_{k=1}^{N}x_k+\frac{\sigma^2}{N_{{\sigma_0}^2+\sigma^2}}\mu_0$$

若令$\mathrm{p}(\mu)=\mathrm{N}(\mu_0,{\sigma_0}^2)=\mathrm{N}(0,1)$，则有：

$$\hat{\mu}_N=\frac{1}{N+1}\sum_{k=1}^{N}x_k$$

贝叶斯估计是样本信息与先验信息的加权平均，反映的是先验知识与样本信息的可靠性度量：

$$\hat{\mu}=\mu_N=\frac{{\sigma_0}^2}{N_{{\sigma_0}^2}+\sigma^2}\sum_{k=1}^{N}x_k+\frac{\sigma^2}{N_{{\sigma_0}^2+\sigma^2}}\mu_0$$

$$=\frac{N{\sigma_0}^2}{N_{{\sigma_0}^2}+\sigma^2}m_N+\frac{\sigma^2}{N_{{\sigma_0}^2}+\sigma^2}\mu_0$$

4 实验

基于提供的数据进行作业的分析。并假定男生、女生的身高、体重、鞋码、50m 成绩、肺活量都服从正态分布。

1.以肺活量为例，画出男女生肺活量的直方图并做对比；

2.采用最大似然估计方法，求男女生肺活量的分布参数；

3.采用贝叶斯估计方法，求男女生肺活量的分布参数（方差已知，注明自己选定的参数情况）；

4.基于身高和体重，采用最小错误率贝叶斯决策，画出类别判定的决策面。并判断某样本的身高体重分别为(165,50)时应该属于男生还是女生？为(175,55)时呢？

原始数据（部分）如图 4-1 所示：

编号	性别 男1女0	籍贯	身高(cm)	体重(kg)	鞋码	50米成绩	肺活量	喜欢颜色	喜欢运动	喜欢文学	喜欢数学	喜欢模式识别
1	1	湖北	163	51	41	7.5	2500	蓝	1	1		
2	1	河南	171	64	41	7.5	3500	蓝	0	0		
3	1	云南	182	68	45	7.8	4900	蓝	1	0		
4	1	广西	172	66	42	8.2	4800	绿	0	1		
5	1	四川	185	80	44	8.5	5100	蓝	0	0		
6	0	河北	164	47	38	9	2500	紫	1	1		
7	0	河南	160	46	38	9	2500	白	1	1		
8	1	重庆	170	46	41	7	3000	蓝	1	1		
9	1	重亲	178	60	41	7	4200	绿	0	0		
10	1	江苏	180	71	43	7.5	3500	紫	0	0		

图 4-1 原始数据（部分）图

4.1 直方图

使用 Python 语言编程，调用 xlrd 库的函数来读取 excel 文件中对应工作表中的数据，将男女生的肺活量数据分别存储在两个数组中。

由于原始数据存在缺失值或异常值等异常数据，所以需要先对数据进行清洗。同时由于原始数据中各数据相对独立，不存在关键数据，所以并未考虑采取插补法、建模法等办法来修复数据，而是直接将异常数据删除。最终编程实现判断肺活量的值是否不为 0，以及值的类型是否为 float，若不满足要求则将数据删除。

最后调用 matplotlib.pyplot 库的函数分别画出女生、男生的肺活量直方图，

如图 4-1 所示：

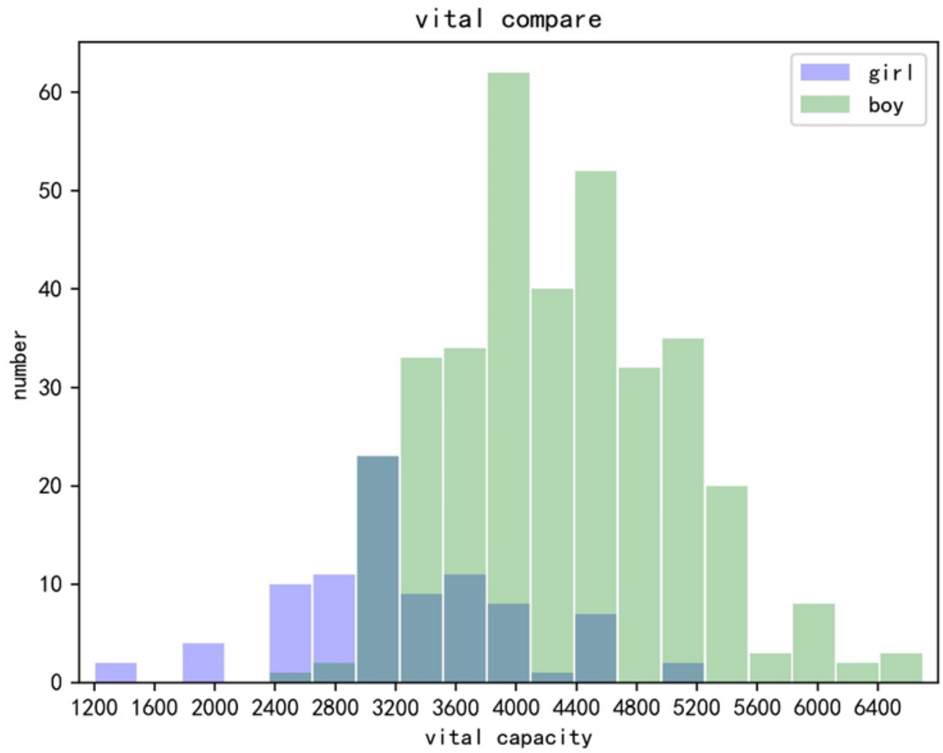

图 4-1 女生、男生肺活量直方图

由图可看出，女生肺活量整体低于男生肺活量，女生肺活量集中在 2000~4000，男生肺活量集中在 3000~5000。同时可以看出，女生、男生肺活量虽然大体上呈现中间人数多，两边人数少，但并未完全服从正态分布，主要原因是因为数据量比较少，特别是女生数据量太少，导致与正态分布模型形式相差较大。

4.2 最大似然估计

我们假设男生女生肺活量分布服从一维正态分布，均值与方差均未知，由第二章分析可得出计算公式，使用 Python 编程可以容易算出参数，具体所求参数如表 4-1 所示：

表 4-1 女生、男生肺活量最大似然估计参数表

	女生肺活量	男生肺活量
均值(μ)	3246.6591	4304.5543
方差(σ)	565218.3383	548884.6471

由表可以看出，女生、男生肺活量的方差都很大，说明数据波动大，也再次说明了由于数据样本少，导致数据的分布较为分散。

4.3 贝叶斯估计

由前述贝叶斯估计的推导，贝叶斯估计是样本信息与先验信息的加权平均：

$$\hat{\mu} = \mu_N = \frac{{\sigma_0}^2}{N_{{\sigma_0}^2} + \sigma^2}\sum_{k=1}^{N} x_k + \frac{\sigma^2}{N_{{\sigma_0}^2+\sigma^2}}\mu_0$$
$$= \frac{N{\sigma_0}^2}{N_{{\sigma_0}^2} + \sigma^2} m_N + \frac{\sigma^2}{N_{{\sigma_0}^2} + \sigma^2}\mu_0$$

由此本题可以对肺活量两种特殊情况下的分布参数进行估计：

（1）当认为样本的先验知识完全可靠，则调和均数的样本信息成分近似为0。方差已知且为 1 的情况下，贝叶斯估计的 $\hat{\mu}$ 即为先验知识 μ_0。

（2）当认为样本的先验知识完全不可靠，则调和均数的先验信息成分近似为 0。方差已知且为 1 的情况下，贝叶斯估计的 $\hat{\mu}$ 即为样本知识 μ。

实验具体结果如表 4-2 所示：

表 4-2 女生、男生肺活量贝叶斯估计参数表

	女生肺活量	男生肺活量
特殊情况 1: （先验知识可靠，样本不起作用） μ_0 = 3240； σ_0 = 0； σ = 1	3240	3240
特殊情况 2: （先验知识十分不确定，完全依靠样本信息） μ_0 = 3240； σ_0 = 10000； σ = 1	3246.6591	4304.5543

4.4 最小错误率贝叶斯决策

本题的最小错误率贝叶斯决策基于男女生的身高体重数据，我们假设男女生身高体重分别服从正态分布。最终画出最小错误率的贝叶斯决策图如图4-2所示：

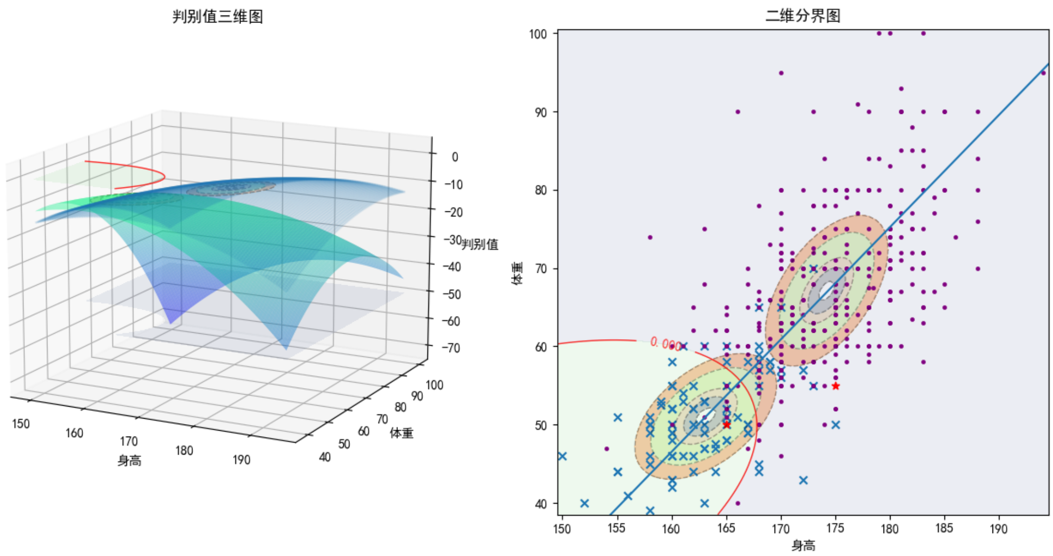

图 4-2 基于最小错误率的贝叶斯决策图

首先由题设，类数$c=2$，特征数$d=2$，女生和男生类分别为w_1、w_2。

由输入数据，计算训练样本的相关参数：

	女生身高	女生体重	男生身高	男生体重
均值(μ)	163.125	51.045	174.204	67.107
协方差矩阵 (Σ)	[23.214 14.356 14.356 35.475]		[34.959 32.611 32.611 102.986]	

然后，我们计算矩阵中各类的后验概率：

$P(\omega_1)$	20.09%
$P(\omega_2)$	79.91%

接着，利用均值矩阵、协方差矩阵和后验概率，由判别函数计算基于样本的w_1和w_2判别值：

$$g_i(x) = -\frac{1}{2}(x-\mu_i)^T \sum_i^{-1} (x-\mu_i) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln\left|\sum_i\right| + \ln \mathrm{P}(w_i), \qquad i=1,2$$

根据二维采样点的判别值，我们可以分别绘制出空间内w_1和w_2判别函数的等密度环以及密度曲面，如图 4-2 所示。

最后，我们计算决策面：

$$g(x)=g_1(x)-g_2(x)=-\frac{1}{2}\left[(x-\mu_1)^T\sum_1^{-1}(x-\mu_2)-(x-\mu_1)^T\sum_2^{-1}(x-\mu_2)\right]$$
$$-\frac{1}{2}\ln\frac{|\sum_1|}{|\sum_2|}+\ln\frac{P(\omega_1)}{P(\omega_2)}=0$$

由决策面方程，我们可以绘制出一条二次曲线，如图 4-2 上红线所示。这条线将二维空间上的决策线，落在小于 0 区域的即属于w_1，反之属于w_2。同时也可以从左侧立体图看到，密度曲面的相交投影即为二维空间上分类决策线。

同时，也易知题目中(165,50)和(175,55)两样本（如图 4-2 上红星所示），分别属于女生和男生的区域。由此通过样本的最小错误率贝叶斯决策计算出身高 165，体重 50 的样本属于女生；身高 175，体重 55 的样本属于男生。

5 分工与感想

5.1 分工

陈玉熙：负责实验第 3、4 题编程实现与报告撰写。

李育泓：负责实验第 1、2 题编程实现与报告撰写；负责报告模板格式整理，负责报告前三章理论知识的梳理与撰写。

5.2 感想

陈玉熙、李育泓：这一次的作业是关于两种参数估计方法——最大似然法跟贝叶斯估计，以及贝叶斯决策的内容，在假定正态分布的情况下，对提供的数据进行分析与处理。在对理论知识的梳理之中，我加深了对知识的理解，而在编程实现中，又加深了对公式的印象。

在拿到数据后，首先自然要对数据进行分析，在读取之后取出有效的部分，画直方图其实也是对数据进行分析，可以直观地感受到数据的实际分布，对数据有一个更清晰的把握。之后对数据的参数进行估计，最大似然估计把待估参数看作确定的量，最佳估计即概率最大，在假定正态分布下，其均值就是算术平均值；而贝叶斯估计认为参数服从某种分布，在假定正态分布下，其均值其实是样本信息跟先验信息的加权平均，需要权衡两者的可靠性，在选择不同参数的情况下会有不同的结果，同时存在两种极端情况。而最后的基于最小错误率的贝叶斯决策，其实就是要求最大后验概率，先验概率是男女生比例可以容易求得，而类概率密度我们假定其满足正态分布，跟课程例子有区别，题目是二维的数据，通过跟队友讨论，类比一维正态分布的交点 t，我们假定二维数据满足二维正态分布，那其类概率密度图像应该是类似于小山坡，男女生对应的两个图会有一个交平面，这个面就是决策面了。

在小组合作中，既要清晰分工，完成自己的负责的内容，但也要参与进其他的部分，不管哪一题是谁实现的，最终都是要一起掌握方法与技巧。同时还要积

极沟通，一个人可能会想得不够全面，或者有所缺漏，而沟通就是一个很好的机会去互相学习，去接触不同的思路，去学习不同的方法，才能得到更大的收获。