《模式识别》 任务报告

Course: Pattern Recognition
Author: Xiping.Yu

Email: Amoiensis@outlook.com

Github: https://github.com/Amoiensis

Date: 2019.12.11

一、实验概述

本次实验通过对人群收集得到的 10 个维度的特征数据，基于这些特征的分析和处理，以预期受调查人群的性别情况。

从一定程度上讲，本问题是一个典型的分类问题。在本次实验中，在实验要求范围内，采用了 Bayes 决策、FLD(LDA)fisher 线性判别、SVM(Linear)支撑向量机、MLP 多层感知机，；此之外，还尝试了 Naïve Bayes 决策、概率密度函数估计结合 Bayes 决策求解、SVM 的自行实现等。

二、实验内容与原理理解

1. 实验内容：

实验基于所搜集得的 10 组特征，进行挑选和组合，以期望更准确地推测被调查人员的性别特征，并根据训练所得的模型，预测一组标签未知的样本的性别特征。其中使用不同的模式识别方法进行分类，学习和运用不同的方法（其中要求的方法有：Bayes\Fisher-LDA\SVM\ MLP），观察并记录不同方法在真实数据集中的直观表现情况（误判率、灵敏度、特异度、AUC 等）。

2. 实验原理与理解

Bayes 决策：

Bayes 方法本质上是通过后验概率判断，寻找最小错误率的分类器。

值得一提的是 Bayes 是在知道样本来自的总体的分布情况下，最小错误率估计，在样本足够大的情况下，具有一致最小错误率的性质。

在给定先验概率 $P(\omega_i)\ (i=1,2,3,...,n)$ 的情况下，通过样本估计所得的不同分组的分布或概率密度情况。

计算待预测样本属于不同种类的后验概率，并认为该待预测样本属于所得后验概率最大的类别。根据实验说明，本次实验数据满足如下条件:

$$P(\omega_i \mid X) \sim N(\mu_i, \Sigma_i)\ \ X \in R^{10}$$
$$P(\omega_i) = 0.5\ (i=1,2,...,n)\ \text{n=2}$$

得到求解的数学表达，如下：

$$P(\omega_i \mid X) = \frac{P(\omega_i \mid X)P(\omega_i)}{\sum_{j=1}^{n} P(\omega_j \mid X)P(\omega_j)}\ \ (i=1,2,...,n)$$
$$X \in \{\omega_k \mid P(\omega_k \mid X) = \max\{P(\omega_i \mid X)\}\}$$

表示为线性分类模型为：

$$w = \Sigma^{-1}(\mu_1 - \mu_2)$$
$$\omega_0 = -\frac{1}{2}(\mu_1 - \mu_2)^T \Sigma^{-1}(\mu_1 - \mu_2) - \ln(\frac{P(\omega_2)}{P(\omega_1)})$$

Fisher 线性判别(LDA)

Fisher(LDA),可以看作是把原数据降维，降到类别差距最大的低维空间。

Fisher 方法其实是一种基于实例的方法，并不需要知道样本来自总体的概率分布情况，通过类内离散度和类间离散度的目标函数便可以完成分类器的求解。

Fisher 线性判别，通过使得总类内离散度 $\tilde{S}_{\mathrm{w}}$ 尽量小，类间离散度 $\tilde{S}_b$ 尽量大，从而构建并求解 $\max J_F(\omega)=\dfrac{\tilde{S}_b}{\tilde{S}_{\mathrm{w}}}$。原优化问题，经变形后通过构建并求解 Rayleigh 商 $\max J_F(w)=\dfrac{w^T S_{\mathrm{b}} w}{w^T \tilde{S}_{\mathrm{w}} w}$ 来求解分类器。在本次实验的二分类问题中，可以推导解析求解器形式。

推导得，如下结果：

$$w^*=S_{\mathrm{w}}^{-1}(m_1-m_2) \quad ,m_i=\frac{1}{N_i}\sum_{x_j\in\chi_i} x_j$$

$$\omega_0=-\frac{1}{2}(\tilde{m}_1+\tilde{m}_2) \ \ ,\tilde{m}_i=\frac{1}{N_i}\sum_{y_j\in\gamma_i} y_j=\frac{1}{N_i}\sum_{x_j\in\chi_i} w^T x_j=w^T m_i$$

SVM 支撑向量机

（个人感受理解*）

SVM 和 LDA 有相似之处，二者都是基于实例的方法，二者在信息使用量和处理方式上都有差异。在使用的信息量上，SVM 的分类面仅仅取决于支持向量，但是 LDA 和所有样本都有关系。在处理的方法上，SVM 寻找的是一个超平面（配合核函数），将数据进行划分。值得说明的是 SVM 方法是最小化结构风险的处理方法。

经典的求解 SVM 线性分类器的方法，是通过求解一个正定系数的二次规划问题，而得到线性分类器的方向 w，但是求解这个二次规划问题往往比较复杂，而且容易出现计算误差等情况（这里将会在后面“自行实现 SVM”中详细介绍），现在常采用的是 SMO 算法作为这个二次规划问题的求解。

从直观上，可以很好的理解 SVM 的求解。对于 SVM 的求解（以二分类问题为例），可以理解为寻找到一个超平面，使得两类中距离该超平面最近的点之间的距离最大。一个自然的想法，是遍历所有的方向并分别计算最近的点之间的距离，但是这样的作法是困难的。因为方向是一个连续的变量，从而根据离散对于连续逼近，设计的求解算法只能求得近似解，而且因为步长的原因还可能遗漏最优解。

但是再深入考虑，会发现用遍历所有方向的而求解最优解的算法可能是低效率的，因为用于训练的样本点的数目是有限的，遍历也不应该是盲目搜索，而 LD-Classifier 的方向也应该由样本决定。

在这里在引入 SVM 的数学解法就是很自然的。这里先不再具体展开，SVM 的求解方案（将在“自行实现 SVM”中详细介绍）。其最重要的过程是，通过拉格朗日松弛将原问题变换为（最优超平面）对偶问题，通过正定二次规划（凸优化）求解，从而找出 Support Vector(支撑向量)以及分类器方向 w，再根据 KKT 条件求解出偏移量 b 。

值得说明的是，这里的二次规划问题求解是凸的，即 SVM 的求解问题是 P 问题，不是 NP 难问题，因为其二次项系数矩阵是正定的，这里将“自行实现 SVM”中给出。

SVM 的核心，在完全线性可分的情况下，待求解的二次规划问题，分量表示如下：

$$\max Q(\alpha)=\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{N}\alpha_i\alpha_j y_i y_j(x_i\bullet x_j)$$

$$s.t.\begin{cases}\sum_{i=1}^{N}y_i\alpha_i=0\\ \alpha_i\geq 0,\ \mathrm{i}=1,\ldots,\mathrm{N}\end{cases}$$

$$\mathrm{w}^*=\sum_{i=1}^{N}\alpha_i^* y_i x_i$$

SVM 在非完全线性可分的情况下，采用引入松弛变量的方法（其中 C 为惩罚系数，为给定值），求解类似的二次规划问题。

$$\max Q(\alpha)=\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{N}\alpha_i\alpha_j y_i y_j(x_i\bullet x_j)$$

$$s.t.\begin{cases}\sum_{i=1}^{N}y_i\alpha_i=0\\ 0\leq\alpha_i\leq C,\ \mathrm{i}=1,\ldots,\mathrm{N}\end{cases}$$

$$\mathrm{w}^*=\sum_{i=1}^{N}\alpha_i^* y_i x_i$$

MLP 多层感知器

对于本体的二分类问题，实际上感知器神经元就是在高维空间内寻找一个超平面将空间划分为两块，但是单个神经元只能处理线性可分的问题，对于线性不可分的问题需要多个神经元共同作用。

其数学表达式，如下：

$$y=\theta\{\sum_{j=1}^{n}\upsilon_j\theta(\sum_{i=1}^{m}\omega_{ij}x_i+\omega_{0j})+\upsilon_0\}$$

$$\theta(\bullet)\quad \text{is } Sigmoid/\mathrm{sgn}/step$$

使用 BP 算法重要引入了 Sigmoid 函数替代原来使用的阈值函数，实现了误差的反向传播从，不断修正连接上的权值，从而实现网络的训练。

三、特征选择

1. 特征分析：

根据“额外探究部分”估计概率密度函数中，所拟合的概率函数大致分布情况可知：

a. 第 1、2、3、4、5 维特征，在男女中表现出比较显著差异（近似不同均值、同方差）
b. 第 6、9、10 维特征，在男女中表现出差异不明显（近似同均值、不同方差）
c. 第 7、8 维特征，在男女中表现出非常不明显的差异（近似同均值、同方差）

2. 特征选择：

因为从所有的 10 维特征中选择 2 维作为特征训练模型，从而需要提取最能反应不同样本之间区别的特征，即采用 PCA 方法进行降维处理。

采用主成分分析的方法，对于训练集的 10 维特征进行主成分分析，可以得到这 10 维特征对应的特征值：

$$\begin{bmatrix}952.58 & 188.03 & 158.28 & 125.94 & 98.91 & 66.73 & 24.36 & 1.47 & 0.04 & 0.02\end{bmatrix}^T$$

在取二维特征的情况下，选取特征值最大的两维特征，即第 1、2 维特征。

四、实验结果与讨论

1. 错误率：

训练样本数	特征数	Bayes	FLD	Linear SVM	MLP
10+10	10	0.112204724	0.145669291	0.104330709	0.122047244
	2	0.094488189	0.125984252	0.098425197	0.104330709
Dataset3	10	0.092519685	0.112204724	0.090551181	0.094488189
	2	0.094488189	0.094488189	0.104330709	0.090551181

（注：2 维特征中，选择第 1、2 维特征）

2. 总结:

根据得到的不同方法处理下，分别对应不同训练样本数和训练特征所得到的错误率如上。

可以看出各种方法，基本上在大样本训练时错误率都要优于小样本训练的到的分类器，但是在特征数目上并没有统一的明显区别。

在样本数不算多的情况下，SVM 都是一种比较稳定的方法，在实际的测试中 MLP 方法在小样本的情况下表现很不稳定，可能是因为样本量太少，随机方法在其中产生的影响比较明显，得到的训练模型波动大，不稳定，相较之下 Bayes 和 fisher 的表现稳定但是在测试集上的分类一般。值得注意的是，这四种方法在不同的情况下建立的模型最终的在测试集上表现一致。一方面，这可以说明在寻优问题中，在不同的情况下各种方法可以找到相同的极值点；另一方面，这也在侧面反映出了不同方法在内在上的相似和互通。

3. 具体分析:（附上 AUC，ROC，等具体信息并进行分析）

   a) 正态分布概率模型下的最小概率 Bayes 决策

   由于假设数据满足正态分布，从而使用多为正态分布进行拟合，又根据“额外探究部分”中各维特征其实正态性良好，从而假设正态分布的 Bayes 决策在理论上，在这个问题中可行，且在样本足够大的时候应该具有一致最小错误率估计的性质。但是本题中训练集数据量并不多，从而和最优分类面容易产生一定的偏差。

   所以在求解错误率中，可以看出 Bayes 方法整体表现不错，但是因为数据量的不充足，而且存在正态假设的近似性，从而没有达到最优分类情况。

   b) Fisher 线性判别分析(LDA)

   由于 Fisher 线性判别的方法，是基于样本的实例方法，从而不需要假设样本来自的总体的概率分布情况。

   在求解的错误率中，Fisher 方法在大样本时表现明显优于小样本，但是在小样本时，少特征优于多特征，这可能是因为因为样本少，少量的特征可以使得分类信息更加集中。

   c) SVM (Linear)支持向量机

   SVM 方法在各种方法中，表现最为稳定，特别是在小样本训练时，分类效果很好，这也体现出了 SVM 作为基于实例的方法的优势。在小样本时，少特征优于多特征，这可能是因为因为样本少，少量的特征可以使得分类信息更加集中。在大样本时，特征更多效果可能更好，因为不像小样本时，特殊值容易影响分类器的选择，从而训练的分类器更加稳定。

   d) MLP 多层感知机

   MLP 方法，在小样本时，表现不好，但是在大样本时表现优秀，这很大程度上是因为 MLP 算法对于样本的依赖度很高，参数很多，在小样本时，很难训练完成一个成熟的 MLP 模型，而在样本量较大的情况下，参数能够得到不错的训练，从而结果表现更好。

五、额外探究部分

1. Naive Bayes 决策

   介绍 Naïve Bayes 的情况和结果情况。

2. 估计概率密度函数，加权的 Naïve Bayes 决策（最小风险）

   对于概率密度函数的估计和分析。

在假设的高斯分布的条件下，对 Train 集的数据的各个维度做高斯正态分布函数的拟合，可以得到二分类问题情况下的两类分别的分布情况。

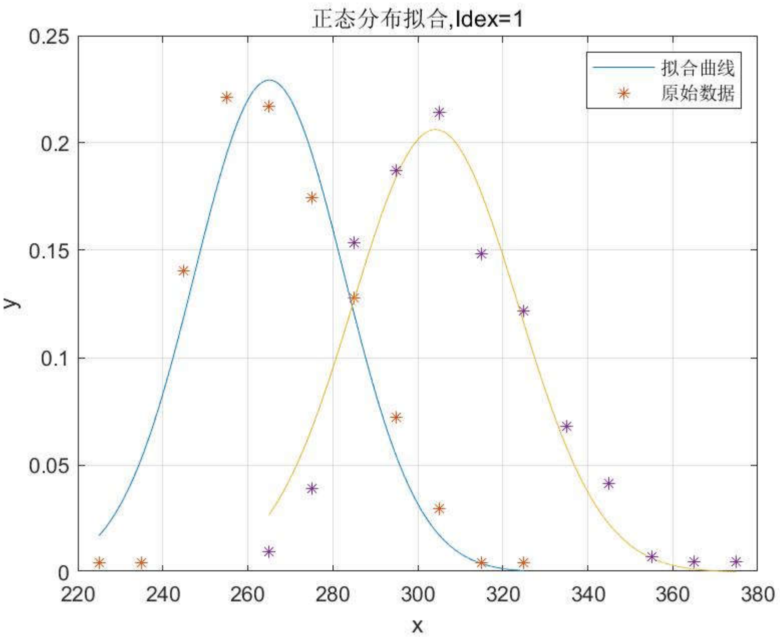


图 1 – 第 1 维特征分布

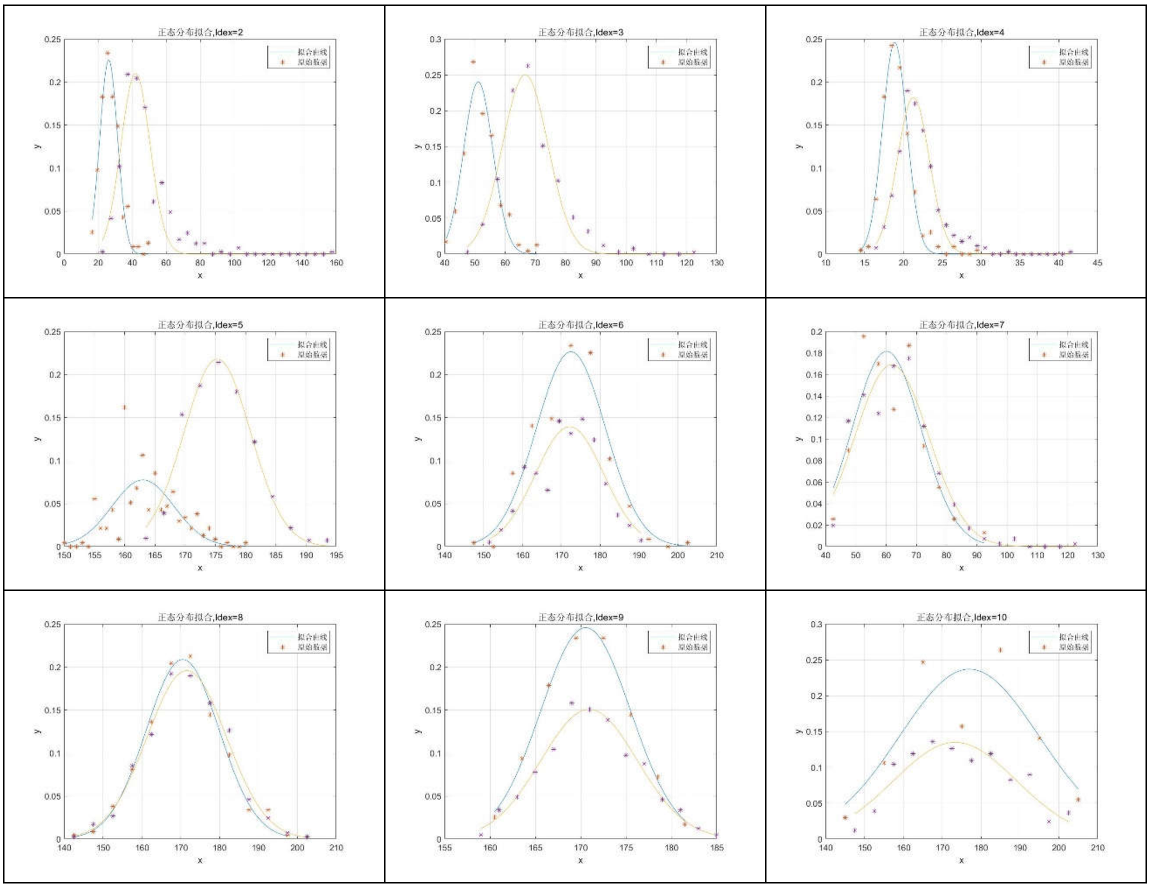


图 2 – 依次 2~10 维特征分布

说明：

根据最小错误率方法，从一定的程度上看，在二分类问题两类先验概率相同的情况下，Bayes 方法其实是

在求解两类情况下的特征概率密度函数的方程的解。

在面对多特征的情况下，三种解决方案：

(1) 朴素贝叶斯方法(Naïve Bayes)，假设各特征之间相互独立。
(2) 估计概率密度函数，加权 Naïve Bayes 决策，对于各类特征不假设相互独立，分别预测，最后加权判断。
(3) 正态分布概率模型下的最小错误率贝叶斯，一般情况。

相对于其他方法，这种方法简单易行，但是其各特征维度带有权重，从而又类似于最小风险的 Bayes 决策。在本次实验中因为二者先验概率相同，从而：

$$\frac{\mathrm{f}_{\mathrm{X1}}^{i}(x)}{\mathrm{f}_{\mathrm{X2}}^{i}(x)}=\frac{P(\omega_1)}{P(\omega_2)}\ (\mathrm{i}=1,2,3...,10)$$

即求解每维特征的概率函数分布函数构成的方程；在二者先验概率相等条件下，二者概率密度函数交点，即为解（阈值）。

观察密度函数分布图，选择合适的方程，并通过解方程，求得阈值后，在不同特征上赋予不同的权值，在由加权和确定样本归于哪一类（类似于不同特征进行带优先级的投票表决）。

3. SVM (Based on QP) 支撑向量机自行实现(基于二次规划)

SVM 的实现/二次规划的矩阵构建/正定的证明、凸优化/

由于收集数据并非为完全线性可分问题，从而传统的完全可分的 SVM 求解需要再引入一组松弛变量，类似于实验原理

$$\max Q(\alpha)=\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{N}\alpha_i\alpha_j y_i y_j(x_i\bullet x_j)$$

$$s.t.\ \begin{cases}\displaystyle\sum_{i=1}^{N}y_i\alpha_i=0\\ \alpha_i\ge 0,\ \mathrm{i}=1,...,\mathrm{N}\end{cases}$$

$$\mathrm{w}^*=\sum_{i=1}^{N}\alpha_i^* y_i x_i$$

将上述分量形式可以改写为矩阵形式：

$$\max Q(\alpha)=\frac{1}{2}\alpha^T H\alpha+f\alpha$$

$$s.t.\ \begin{cases} A\cdot\alpha=b \\ lb\le\alpha\le ub \\ \alpha\in R^{n\times 1} \end{cases}$$

$$\mathrm{w}^*=\sum_{i=1}^{n}\alpha_i^* y_i x_i$$

$$Coeffcient:$$

$$f=[1,1,...,1]\in R^{1\times n}$$

$$G=0.5\times I+diag(0.5,0.5,...,0.5)\in R^{n\times n}$$

$$H=G*(Y\cdot Y^T)*(X\cdot X^T)\quad (Hadamard\ product)$$

$$A=Y^T$$

$$b=0$$

$$lb=[1,1,...,1]^T\in R^{n\times 1}$$

$$ub=C=[c,c,...,c]^T\in R^{n\times 1}$$

$$X\in R^{n\times p}\ is\ the\ features\ of\ the\ samples.$$

$$Y\in R^{n\times 1}\ is\ the\ labels\ of\ the\ samples.$$

有理论证明，当 H 正定时，用椭圆法可在多项式时间内解二次规划问题。当 H 负定时，二次规划问题是NP 困难的(NP-Hard)。即使 H 存在一个负特征值时，二次规划问题就是 NP 困难的。

下面证明 SVM 求解问题不是 NP 难问题，在多项式时间内可求解。

证明，H 矩阵是半正定矩阵：

因为$H=G*(Y\cdot Y^T)*(X\cdot X^T)\quad (Hadamard\ product)$

引理:

H 可以看成三个矩阵的哈达马积(Hadamard product)（即对应位置元素相乘），首先证明两个正定矩阵的哈达马积，结果仍然是正定矩阵，见文献。

即当 A，B 为正定矩阵时，$X=A*B$ 也为正定矩阵。

由 G 的定义，容易看出$G=0.5\times I+diag(0.5,0.5,...,0.5)\in R^{n\times n}$是正定矩阵；

而$(Y\cdot Y^T)$、$(X\cdot X^T)$，即不难证明其是半正定矩阵；

由引理，即证 H 为半正定矩阵。从而 SVM 问题，是 P 问题，可在多项式时间内求解。

但是值得注意的是在$(Y\cdot Y^T)$、$(X\cdot X^T)$是正定的情况下，则 H 矩阵为正定矩阵，有稳定的方法求解；但是当 H 为半正定情况下时数值求解不稳定，容易出现偏差，从而有人提出了 SMO 算法，一方面减少运算时间，另外一方面使得可以稳定求解。

H 矩阵正定得证，从而可以使用求解二次规划方法求解 SVM 分类器。其中 C 为人为选定，反映对错误的容忍情况（C 越小，反映对错分更加容忍，更强调分类的间隔）。

六、程序运行方法说明

（一） Naïve Bayes 决策

文件内容

“***\提交版本\Naive_Bayes”目录下，有如下文件：

1. Bayes_trainClassifier.m
2. Estimate_result.m
3. dataset3.mat
4. dataset3_cut.mat
5. vali_500_with_tag.mat

使用方法：

1. 点击运行“Bayes_trainClassifier.m “，训练模型
2. 点击运行“Estimate_result.m”,输出预测结果和正确情况

（二） Fisher 线性判别分析(LDA)

文件内容

“***\提交版本\Naive_Bayes”目录下，有如下文件：

6. Fisher_trainClassifier.m
7. Estimate_result.m
8. dataset3.mat
9. dataset3_cut.mat
10. vali_500_with_tag.mat

使用方法：

1. 点击运行“Fisher _trainClassifier.m “，训练模型
2. 点击运行“Estimate_result.m”,输出预测结果和正确情况

（三） SVM (Linear)支持向量机

文件内容

“***\提交版本\Naive_Bayes”目录下，有如下文件：

1. SVM_trainClassifier.m
2. Estimate_result.m
3. dataset3.mat
4. dataset3_cut.mat
5. vali_500_with_tag.mat

使用方法：

1. 点击运行“SVM _trainClassifier.m “，训练模型
2. 点击运行“Estimate_result.m”,输出预测结果和正确情况

（四） MLP 多层感知机

文件内容

“***\提交版本\Naive_Bayes”目录下，有如下文件：

1. Caculate.m
2. Cut_data.m
3. Estimate_result.m
4. my20_2_12_NeuralNetworkFunction.m
5. my20_10_NeuralNetworkFunction.m
6. myAll_2_12_NeuralNetworkFunction.m
7. myAll_10_NeuralNetworkFunction.m

8. merge_data.mat
9. vali_500_with_tag.mat
10. dataset3.mat
11. dataset3_cut.mat

使用方法：

在“Caculate.m“中设置训练集大小（20/dtaset3）和特征，训练模型；自动会输出结果和错误率情况(“Wrong_all”).

（五） 基于 QP 的 SVM（自行实现）

文件内容

“***\提交版本\ Quadratic Programming”目录下，有如下文件：

1. final_Formula.m
2. Muti_Formula.m
3. Estimate_result.m
4. vali_500_with_tag.mat
5. dataset3.mat
6. dataset3_cut.mat

使用方法：

二维情况：

在“final_Formula “中设置训练集大小（20/dtaset3）和特征，训练模型；运行“Estimate_result.”输出结果和错误率情况(“Wrong_all”).

三（高）维情况：

在“final_Formula “中设置训练集大小（20/dtaset3）和特征，训练模型，将自动绘制出分类面

七、程序运行结果截图

程序参数设置：

交叉验证：

Fisher、SVM 均采用 5 折交叉验证；

具体结果：

1. Bayes 决策：

```
Featrues = [1 2]
```

Train: dataset3

工作区：

Wrong_all	0.1122
Wrong_female	0.0683
Wrong_male	0.1544

2. Fisher 线性判别分析(LDA)

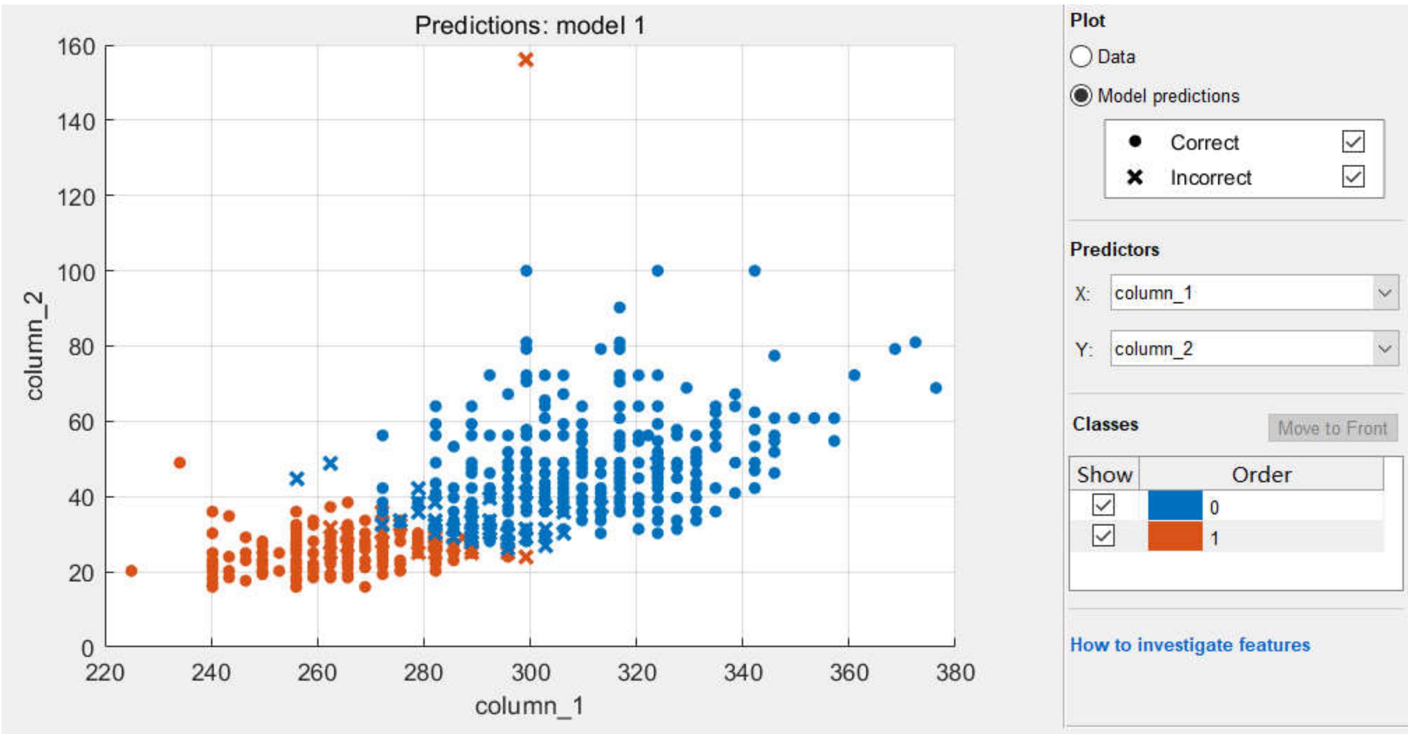
```
Featrues = 1:10
```

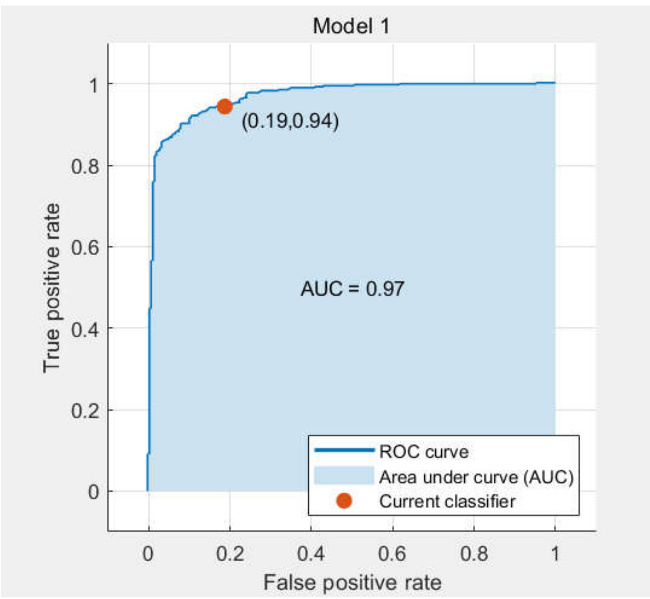
Train: dataset3

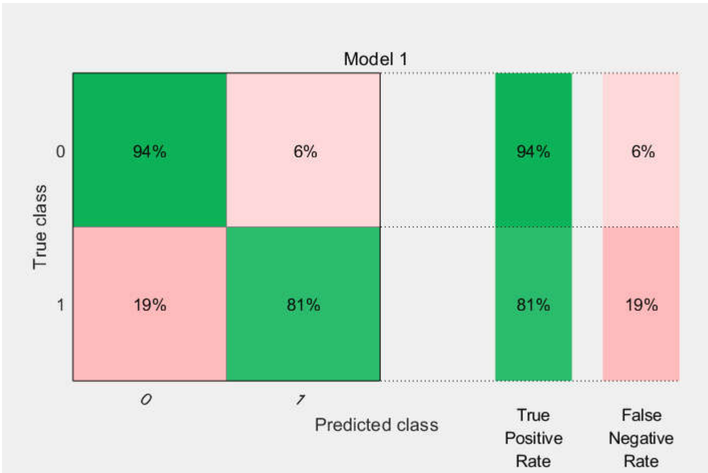
工作区：

Wrong_all	0.1122
Wrong_female	0.2088
Wrong_male	0.0193

工具箱：（`Featrues = [1 2 3 4 5 6 7 8 9 10]` Train: dataset3）







3. SVM (Linear)支持向量机

Featrues = [2 3 6 4 7 10]

Train: dataset3

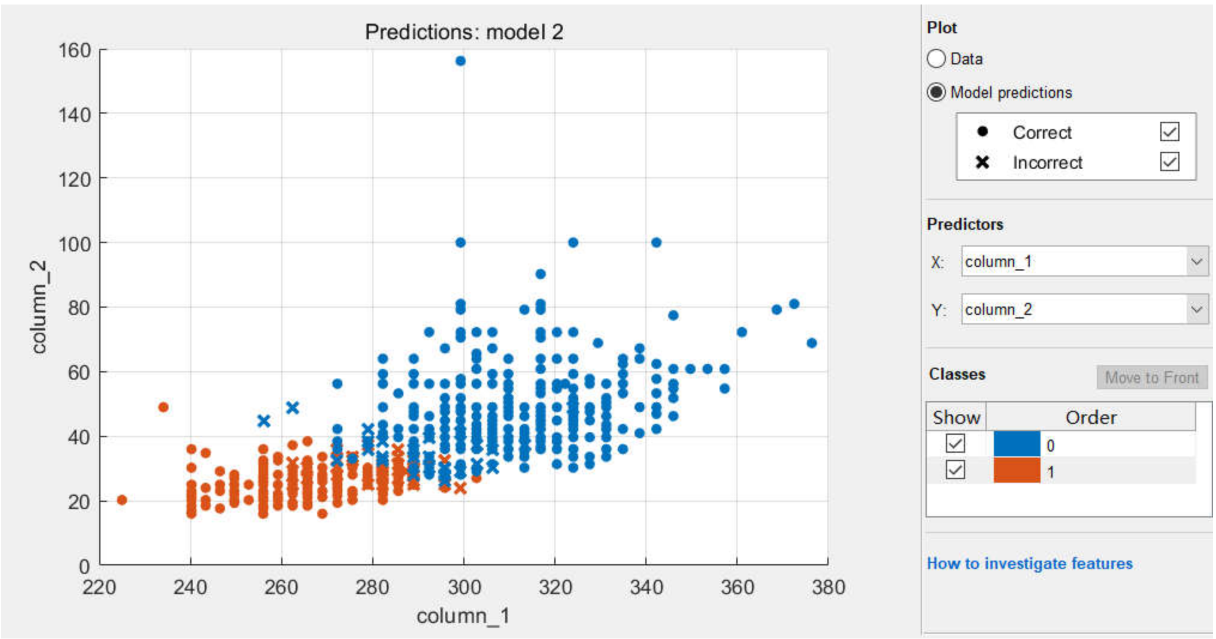
工作区：

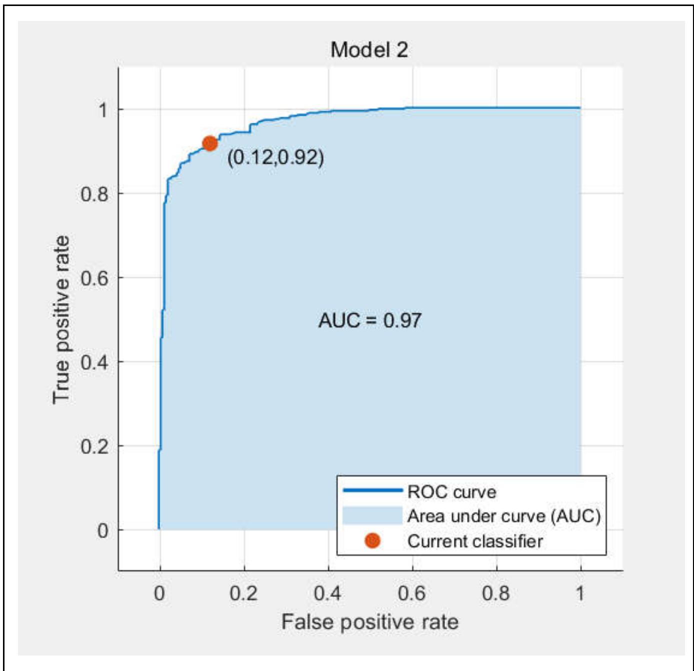
Wrong_all	0.0906
Wrong_female	0.1446
Wrong_male	0.0386

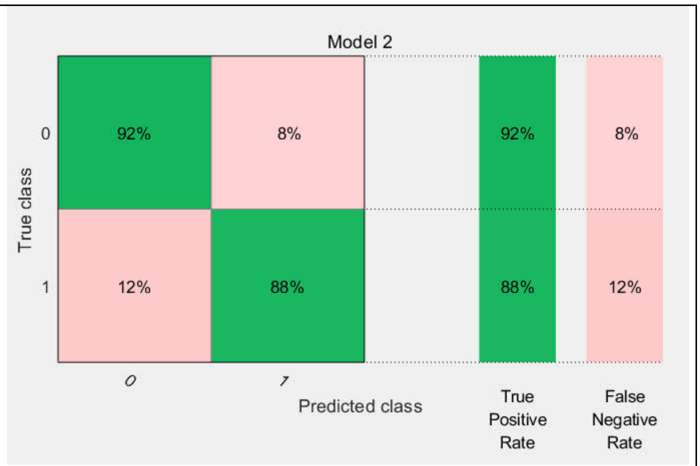
工具箱：

Featrues = [1 2 3 4 5 6 7 8 9 10] Train: dataset3)

a. （线性 SVM/核函数：线性）







a. （线性 SVM/核函数：其他核函数）

核函数: Quadratic SVM/ Cubic SVM/ Fine Gaussian SVM/ Medium Gaussian SVM/ Coarse Gaussian SVM/

在 dtatset3 上训练结果如下：

可以注意到 Medium Gaussian SVM 和 Coarse Gaussian SVM，表现效果最佳，但是二者结果一致，且结构比较复杂，在 dataset3 数据量不大的情况下，容易出现过拟合。结合其他方法的表现情况，同时结合实现的容易程度，采用 Linear SVM 为主体，虽然 Quadratic SVM 效果比 SVM 略好，但是在小样本训练情况下，选择尽量简单的模型。

综上，在自行实现的过程中，仍然选择 Linear SVM 模型，即线性核函数.

1.1 SVM	Last change: Linear SVM	Accuracy: 90.6% 10/10 features
1.2 SVM	Last change: Quadratic SVM	Accuracy: 90.7% 10/10 features
1.3 SVM	Last change: Cubic SVM	Accuracy: 85.8% 10/10 features
1.4 SVM	Last change: Fine Gaussian SVM	Accuracy: 68.0% 10/10 features
1.5 SVM	Last change: Medium Gaussian SVM	Accuracy: **91.2%** 10/10 features
1.6 SVM	Last change: Coarse Gaussian SVM	Accuracy: **91.2%** 10/10 features

4. MLP 多层感知机

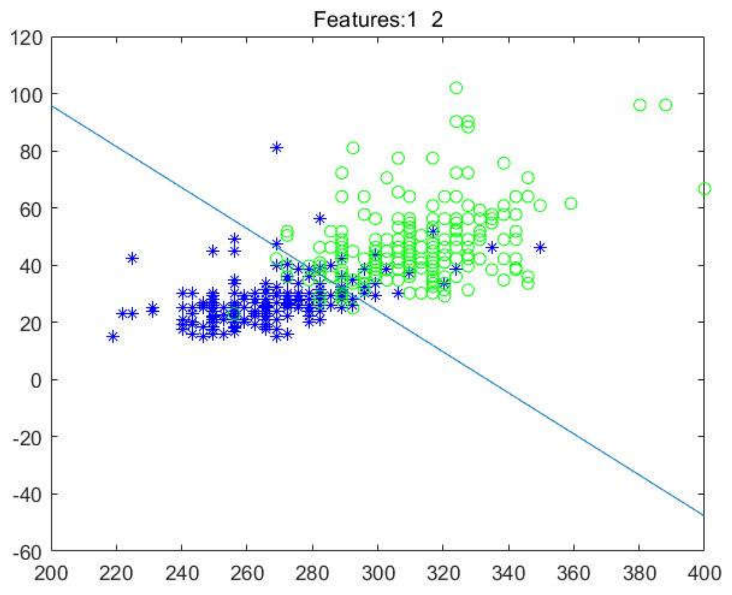
Featrues = 1:10;

Train: dataset3

工作区：

Wrong_all	0.0945
Wrong_female	0.1606
Wrong_male	0.0309

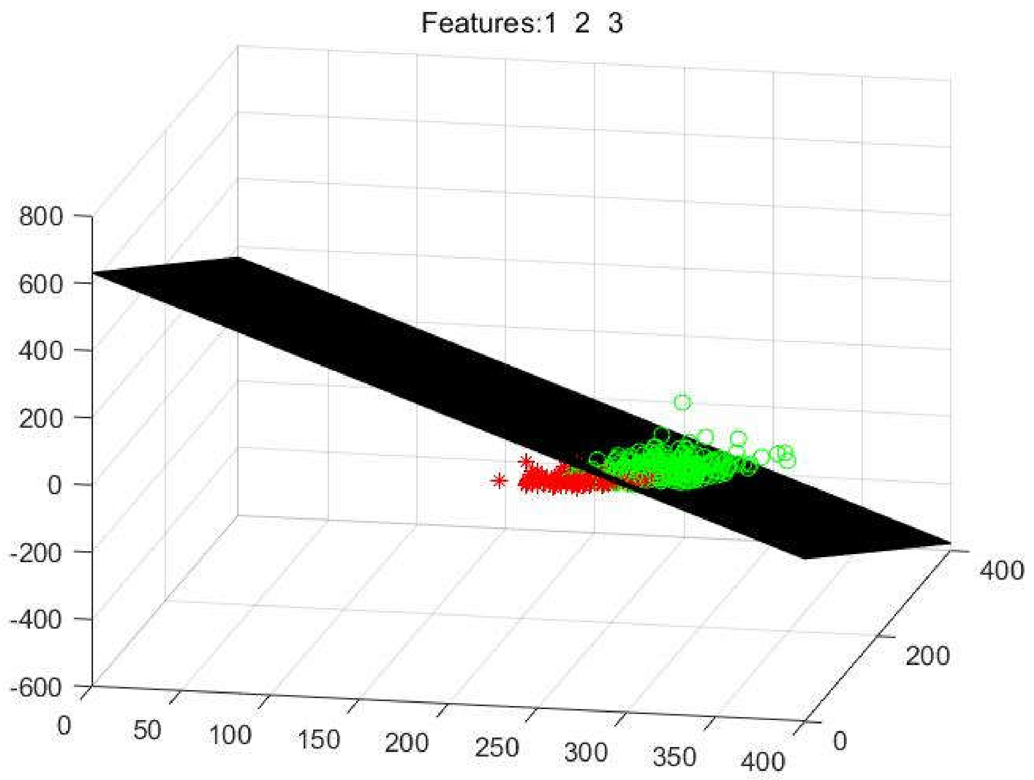
5. 基于 QP 的 SVM（自行实现）



工作区

名称	值
axi_x	[200,400]
b	239.4163
Error	508x1 double
Error_female2male	1
Error_male2female	-1
Features	[1,2]
female	259x1 double
Female	1
i	508
k	-0.7178
male	249x1 double
Male	0
num_Female	249
num_Male	259
Predict_error_female2m...	28
Predict_error_male2fem...	16
Predict_right	464
Predict_value	508x1 double
Pridict_num	508
right	0
Size_test	[508,11]
temp	19.7359
test	508x11 double
vali500withtag	508x11 double
Wrong_all	0.0866
Wrong_female	0.1124
Wrong_male	0.0618
x_p	1x201 double
x_v	508x2 double
y_p	1x201 double
y_v	508x1 double

二维 SVM(QP)



三维-SVM(QP)

6. Naive Bayes 决策

Featrues = [1 3];

Train: dataset3

工作区：

Wrong_all	0.0925
Wrong_female	0.1406
Wrong_male	0.0463

7. 估计概率密度函数，加权的 Naïve Bayes 决策（最小风险）

Featrues = [1 2 3 4 5];

Train: dataset3

工作区：

wrong_all	0.1201
wrong_female_num	11
wrong_female_rate	0.0425
wrong_male_num	50
wrong_male_rate	0.2008

八、实验收获与小结

本次实验，通过探究的方式使用多种不同分类方法对实际问题进行处理。其中一部分的方法通过调用现有工具箱，另外一部分自行实现，在算法的实现过程中（例如 Bayes\SVM）,对方法的原理有了一个更加深刻的认识，更加深刻的理解和认识了分类方法和问题。

在实验的过程中，很多问题需要自己决定和解决，例如，特征的确定和选择，参数的确定等，这些都需要自己在原理理解的基础上通过编程的方法寻找优化的参数。在这个过程中编程能力的得到了锻炼和提高。

在实验结果的分析和认识过程中，能够更加感性和直观的认识到不同分类方法的适用情况和优缺点，同时在对于分类后的结果可以进行相应的解释，对于待分类问题能够得出一个比较好的解释和判断。

特别的是在本次实验中，通过对于 SVM 方法更加深入的理解，理解了 SMO 算法引入求解 SVM 分类器的必要性，明白并能够证明使用二次规划的方法求解 SVM 分类器可能存在的问题和问题的原因。

九、程序来源说明

1. 正态分布概率模型下的最小概率 Bayes 决策
   使自行编写；
2. Fisher 线性判别分析(LDA)
   使用 matlab 自带 Classifier Liner 工具箱中，Linear Discriminant Analysis(Fisher)方法；
3. SVM (Linear)支持向量机
   a) 使用 matlab 自带自带 Classifier Liner 工具箱中，SVM(Linear)方法；
   b) 基于 QP(二次规划)实现 SVM 分类器；
4. MLP 多层感知机
   使用 Matlab 自带 Neural Net Pattern Recognition 工具箱；
5. SVM 自行实现，基于 QP（二次规划）
   在证明本实验问题为正定二次规划问题后，构建 SVM 分类器和求解器。自行实现问题的二次规划模型的构建与表达，其中正定二次规划问题，调用 Matlab 求解工具求解。
6. Naïve Bayes 决策
   使用 Matlab 朴素贝叶斯模型训练器，其中模型评价部分为自行实现。
7. 估计概率密度函数，加权的 Naïve Bayes 决策（最小风险）
   自行实现，概率密度函数估计与阈值求解，并自行构建模型训练与评价部分。

十、 参考文献

[1]王周宏,王能超,钟毅芳.求解一般半正定二次规划的数值稳定方法[J].华中科技大学学报(自然科学版),2002(04):110-112.

[2]贾正华.Hadamard 乘积矩阵的一些性质[J].工科数学,1998(03):154-158.

[3] Wikipedia. 二次规划(Quadratic programming)和 NP-hard 的关系. zh.wikipedia.org/zh/二次规划