第二部分：大作业

作业 3：

1、理论：

(a). 读取图片数据为矩阵 *X*，标签数据为向量 *t*；

$$X=\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \qquad \text{（公式 1）}$$

$$\vec{t}^{\,T}=\begin{bmatrix} t_1 & t_2 & \cdots & t_m \end{bmatrix} \qquad \text{（公式 2）}$$

(b). 特征工程（公式 3）和回归模型（公式 4）；

$$f_1\left(\vec{x}\right)=1,\ f_i\left(\vec{x}\right)=x_{i-1},\ \ i=2,\cdots,n+1 \qquad \text{（公式 3）}$$

$$\hat{y}=\sum_{i=1}^{n+1}\theta_i f_i\left(\vec{x}\right)=\theta_1+\theta_2 x_1+\cdots+\theta_{n+1}x_n \qquad \text{（公式 4）}$$

- 所以运算得出矩阵 *A* 和向量***θ***

$$A=\begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1n} \\ 1 & x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \qquad \text{（公式 5）}$$

$$\vec{\theta}^{\,T}=\begin{bmatrix} \theta_1 & \theta_2 & \cdots & \theta_{n+1} \end{bmatrix} \qquad \text{（公式 6）}$$

(c). 生成向量 *b*；

$$\vec{b}^{\,T}=\begin{bmatrix} g\left(t_1\right) & g\left(t_2\right) & \cdots & g\left(t_m\right) \end{bmatrix} \qquad \text{（公式 7）}$$

$$g\left(x\right)=\begin{cases} 1 & x=0 \\ -1 & x\neq 0 \end{cases} \qquad \text{（公式 8）}$$

(d). 求解方法

- 对矩阵 *A* 进行 QR 分解

$$A=QR \qquad \text{（公式 9）}$$

- 计算向量***θ***

$$\vec{\theta}=R^{-1}Q^T\vec{b} \qquad \text{（公式 10）}$$

2、程序简介

程序运行需要 Src 文件夹和 Data 文件夹放在同一目录下，线性求解请运行“Least_Squares_Classification_Project.m”文件。

运行前准备：

(a). 文件“Least_Squares_Classification_Project.m”中 13、14、25、33、34、35 行注释掉；

(b). 文件“Least_Squares_Classification_Project.m”中 29、30、31 行取消注释；

(c). 文件“Results_output_xlsx.m”中 6~43 行取消注释；

(d). 文件“Results_output_xlsx.m”中 46~82 行注释掉。

(e). 修改文件“Least_Squares_Classification_Project.m”中第 11 行设置要识别的数字；（为满足题目要求，作业中设置为 0）

(f). 是否增加 5000 个随即特征的设置：

- 文件“FeatureExtraction.m”中 11 行注释掉，则取消添加 5000 个随即特征；
- 文件“FeatureExtraction.m”中 11 行取消注释，则添加 5000 个随即特征。

3、运行结果（位于 Src 文件夹内的 xlsx 文档，具体对应见文件 Results_output_xlsx.m 中的设置）

实验结果显示，在添加 5000 个随机特征后，正确率有明显提高。

(a). 矩阵 *A* 列数：494

表 1 训练集结果（1.6%）

输出预	测值：+1	预测值：-1	合计
真值：+1	5158	765	5923
真值：-1	167	53910	54077
合计	5325	54675	60000

表 2 测试集结果（1.6%）

输出预	测值：+1	预测值：-1	合计
真值：+1	864	116	980
真值：-1	42	8978	9020
合计	906	9094	10000

(a). 矩阵 *A* 列数：5494

表 3 训练集结果（0.21%）

输出预	测值：+1	预测值：-1	合计
真值：+1	5818	105	5923
真值：-1	21	54056	54077
合计	5839	54161	60000

表 4 测试集结果（0.26%）

输出预	测值：+1	预测值：-1	合计
真值：+1	963	17	980
真值：-1	9	9011	9020
合计	972	9028	10000

作业 4：

1、理论：

(a). 读取图片数据为矩阵 *X*（公式 1），标签数据为向量 *t*（公式 2）；

(b). 特征工程（公式 3）和回归模型（公式 11）；

$$\tilde{f}\left(x\right)=\sum_{i=1}^{n+1}\beta_i f_i\left(\vec{x}\right)=\beta_1+\beta_2 x_1+\cdots+\beta_{n+1}x_n \qquad \text{（公式 11）}$$

- 所以运算得出矩阵 *A*（公式 5）和向量$\boldsymbol{\beta}$（公式 12）；

$$\vec{\beta}^T=\left[\beta_1 \quad \beta_2 \quad \cdots \quad \beta_{n+1}\right] \qquad \text{（公式 12）}$$

(c). 生成向量 *b*（公式 7）；

(d). 求解方法

- 根据文件“05 线性优化与最小二乘_PPT__least_square_07.pdf”中公式（10.6）、公式（10.7）、公式（10.10）推导出迭代公式如公式 13 所示。公式 13 中λ为正则化参数，λ_k为 LM 法的信任参数；公式 13 中矩阵 J 为 $f\left(A,\vec{\beta},\vec{b}\right)$ 对 $\vec{\beta}$ 的 Jacobian 矩阵；假设矩阵 A 为 $m\times n$ 维，公式 13 中矩阵 B 为一个(n-1)×n 的矩阵，具体表达式如公式 14 所示，B 是由一个 n-1 维的 0 列向量和一个(n-1)×(n-1)的单位阵左右拼接而成，设置 B 的原因为此处的 $\vec{\beta}$ 于原题目中的 $\vec{\beta}$ 不同；公式 13 中函数 $f\left(A,\vec{\beta},\vec{b}\right)$ 的表达式如公式 15 所示，arrayfun(@fun,M)的含义为将 fun 作用于 M 的每一个元素；sigmoid 函数如公式 16 所示；根据复合函数求导公式，Jacobian 矩阵的计算公式如公式 17 所示，其中“.*”意为元素对应相乘，其中 sigmoid_derivative 为 sigmoid 函数的导数。

$$\beta_{k+1}=\beta_k-\left(J^TJ+\lambda B^TB+\lambda_k I_{n维}\right)^{-1}\left(J^Tf\left(\beta_k\right)+\lambda B^TB\beta_k\right) \qquad \text{(公式 13)}$$

$$B=\left[\vec{0} \quad I_{\text{n-1维}}\right]=\begin{bmatrix}0&1&0&\cdots&0\\0&0&1&\cdots&0\\\vdots&\vdots&\vdots&\ddots&\vdots\\0&0&0&\cdots&1\end{bmatrix} \qquad \text{（公式 14）}$$

$$f\left(A,\vec{\beta},\vec{b}\right)=arrayfun\left(@sigmoid,A\vec{\beta}\right)-\vec{b} \qquad \text{（公式 15）}$$

$$sigmoid(u)=\frac{e^u-e^{-u}}{e^u+e^{-u}} \qquad \text{（公式 16）}$$

$$J=arrayfun\left(@\text{sigmoid\_derivative},arrayfun\left(@sigmoid,A\vec{\beta}\right)\right).*A$$

（公式 17）

- 迭代方式为文件“05 线性优化与最小二乘_PPT__least_square_07.pdf”中的“算法 10.2”

2、程序简介

程序运行需要 Src 文件夹和 Data 文件夹放在同一目录下，非线性求解有两个文件，若

果需要单独求解请运行“Least_Squares_Classification_Project.m”文件，如果需要对 LM 最大循环次数和正则化参数进行扫描请运行“Least_Squares_Classification_Project_NonLearScan.m”。

“Least_Squares_Classification_Project.m”运行前准备：

(a). 文件“Least_Squares_Classification_Project.m”中 13、14、25、33、34、35 行注释掉；

(b). 文件“Least_Squares_Classification_Project.m”中 29、30、31 行取消注释；（作业程序中，采用线性求解的向量θ作为非线性求解的初始点，初始的信任参数λ_k设置为 0.1）

(c). 文件“Results_output_xlsx.m”中 6~43 行注释掉；

(d). 文件“Results_output_xlsx.m”中 46~82 行取消注释。

(e). 修改文件“Least_Squares_Classification_Project.m”中第 11 行设置要识别的数字；（为满足题目要求，作业中设置为 0）

(f). 是否增加 5000 个随即特征的设置：

- 文件“FeatureExtraction.m”中 11 行注释掉，则取消添加 5000 个随即特征；
- 文件“FeatureExtraction.m”中 11 行取消注释，则添加 5000 个随即特征。

“Least_Squares_Classification_Project_NonLearScan.m”运行前准备：

(a). 修改文件“Least_Squares_Classification_Project_NonLearScan.m”中第 10 行设置要识别的数字；（为满足题目要求，作业中设置为 0）

(b). 修改文件“Least_Squares_Classification_Project_NonLearScan.m”中第 11、12 行设置正则化参数和最大循环次数的扫描区间及密度；

(c). 是否增加 5000 个随即特征的设置：

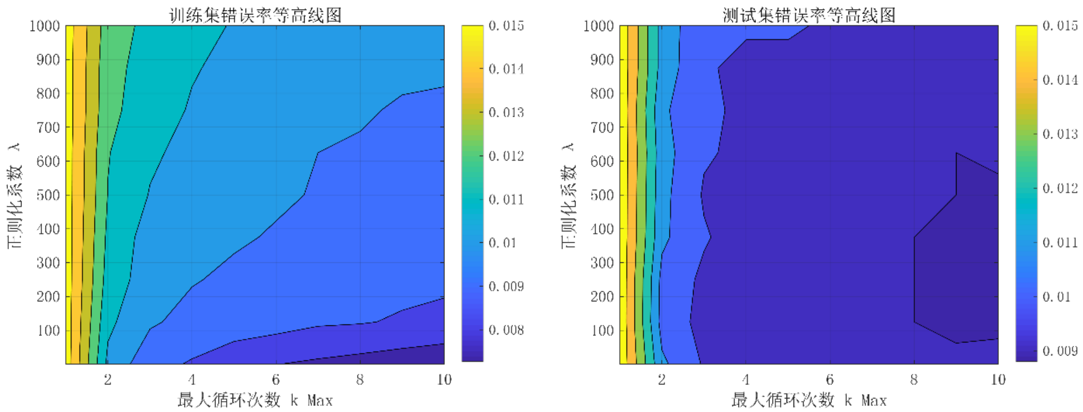
- 文件“FeatureExtraction.m”中 11 行注释掉，则取消添加 5000 个随即特征；
- 文件“FeatureExtraction.m”中 11 行取消注释，则添加 5000 个随即特征。

3、运行结果（位于 Result 文件夹内）

实验结果显示，相较于线性求解的结果，非线性求解后的正确率提升明显，在添加 5000 个随机特征后，正确率再次有明显提高。并且根据图 1 和图 2 可以看出，训练集的测试正确率随着最大循环次数的提高正确率会不断提高，在图 2(a)中可以看出左下角的一众结果正确率为 100%，但测试集正确率不随最大循环次数的增大而无限增大，而是存在一个最优循环次数。根据图 1 和图 2，对于正则化参数λ，在训练集中对于单一的最大循环次数，正则化参数越大正确率越低，且最大循环次数越大该现象越明显，而对于测试集，正则化参数同样存在一个最优值，从图 1(b)看除在正则化参数 100~600，最大循环次数 8~10 处存在一个错误率低的谷地，该现象在图 2(b)中更为明显，该谷地位于正则化参数约 300~1000，最大循环次数约 1.5~3 处。因此若想得到理想的识别结果，需要对正则化参数和最大循环次数进行优化。　（图见下一页）

原始运算结果下载链接（输入浏览器使用，校内免流）：
http://[2001:da8:9000:a191:211:32ff:feb7:86e7]:5000/d/f/523433019713822731

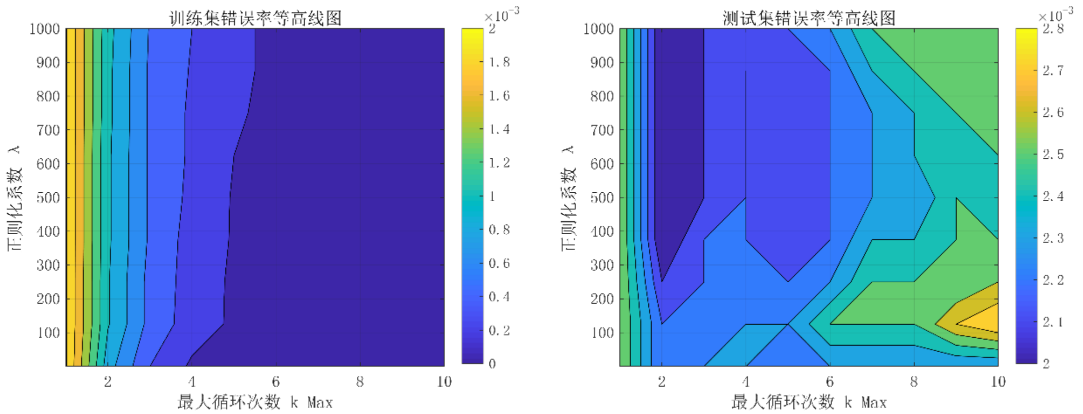
(a). 矩阵 *A* 列数：494



(a) 训练集　　　　　　　　　　(b) 测试集

图 1 矩阵 A 为 494 列时的测试正确率

(a). 矩阵 *A* 列数：5494



(a) 训练集　　　　　　　　　　(b) 测试集

图 2 矩阵 A 为 5494 列时的测试正确率