第7章：特征提取与特征选择

第一部分：简述题

1. 简述PCA的原理、学习模型和算法步骤。
2. 简述LDA的原理和学习模型，给出多类LDA的计算步骤。
3. 作为一类非线性降维方法，简述流形学习的基本思想。
4. 根据特征选择与分类器的结合程度，简述特征选择的主要方法，指出各类方法的特点。

第二部分：编程题

编程实现1：PCA+KNN：即首先PCA进行降维，然后采用最近邻分类器（1近邻分类器）作为分类器进行分类。

编程实现2：LDA +KNN，即首先LDA进行降维，然后采用最近邻分类器（1近邻分类器）作为分类器进行分类。

任务：采用80%作样本作训练集，20%样本做测试集，报告降至不同维数时的分类性能。

1. 所用数据集1：AT&T 40个人脸数据集（即著名的ORL数据集）。样本个数：400，样本维数：644，类别总数：40

提示：降维时可以以5为间隔，比如，10, 15, 20，…

读取数据和类标签信息的Matlab代码如下：

clear all;

load ORLData\_25;

X = ORLData';

X = double(X);

[n, dim] = size(X);

labels = X(:, dim); %获取各样本的类别标签

labels = floor(double(labels));

c = max(labels); % c = 40

X(:, dim) = []; % 获取样本数据

clear ORLData;

1. 所用数据集2：Vehicle数据集。样本个数：846，样本维数：18，类别总数：4

读取数据和类标签信息的Matlab代码如下：

clear all;

load vehicle;

out = UCI\_entropy\_data.train\_data;

X = out';

X = double(X);

[n, dim] = size(X);

labels = X(:, dim);

labels = floor(double(labels)); % 获取各样本的类别标签

c = max(labels); % c = 4

X(:, dim) = []; % 获取样本数据

clear UCI\_entropy\_data;

clear out;