简述题

1、PCA原理：寻找方差较大的方向，将原始数据向该方向进行投影，保留方差较大方向的信息，达到尽可能保留原始数据信息和降维的目的。

PCA学习模型：寻找投影矩阵 $W \in R^{d\times m}$，使得 $Y=[y_1 \cdots y_n]=W^T[x_1 \cdots x_n]=W^TX$

PCA算法：① $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$

② $C=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})^T$

③ 对 C 进行特征值分解，取最大的 m 个特征值 $(\lambda_1 \geq \lambda_2 \cdots \geq \lambda_m)$ 对应的特征向量 $w_1 \cdots w_m$，构成投影矩阵 $W=[w_1 \cdots w_m] \in R^{d\times m}$

④ 对每个数据进行投影：$y_i=W^Tx \in R^m$，$i=1,2\cdots n$

2. LDA原理：寻找投影方向，使样本投影后的类内样本尽可能靠近，类间样本点尽可能相互远离，提升样本表示分类的鉴别能力

LDA学习模型：寻找投影矩阵 $W \in R^{d\times m}$，最大化 $J=\frac{W^TS_bW}{W^TS_wW}$，s.t. $W^TW=I$，其中 S_b 为类间散度矩阵，S_w 为类内散度矩阵

多类LDA：全局散度矩阵：$S_t=S_w+S_b=\sum_{i=1}^{n}(x_i-\mu)(x_i-\mu)^T$，$\mu=\frac{1}{n}\sum_{i=1}^{n}x_i$

类内散度矩阵：$S_w=\sum_{j=1}^{C}S_{wj}$，$S_{wj}=\sum_{x\in X_j}(x-\mu_j)(x-\mu_j)^T$，$\mu_j=\frac{1}{n_j}\sum_{x\in X_j}x$

类间散度矩阵：$S_b=\sum_{j=1}^{C}n_j(\mu_j-\mu)(\mu_j-\mu)^T$

$\max \frac{tr(W^TS_bW)}{tr(W^TS_wW)}$，s.t. $W^TW=1$

或 $\max \frac{|W^TS_bW|}{|W^TS_wW|}$，s.t. $W^TW=1$

3. 流形学习的基本思想：某些高维数据是一种低维的流形结构嵌入在高维空间中。流形学习的目标是将其映射回低维空间同时能够反映高维数据原本的结构特征。

4、(1) 过滤式特征选择：首先定义一个评价函数，用它来度量某个给定特征与类别标签之间的相关度，选取具有最大相关度的m个特征作为结果。

特点：① 特征选择过程与后续学习器无关

② 无法获取最优特征子集

③ 计算量小

(2) 包裹式特征选择：先对数据做特征选择，再训练分类器，特征选择过程与分类算法~~进~~的，特征选择评价判据间接反应分类性能。

特点：① 特征选择与分类性能相结合，特征评价判据为分类器性能，对给定分类方法，选择最有利于提升该性能的特征子集

② 无法保证获取最优的特征子集

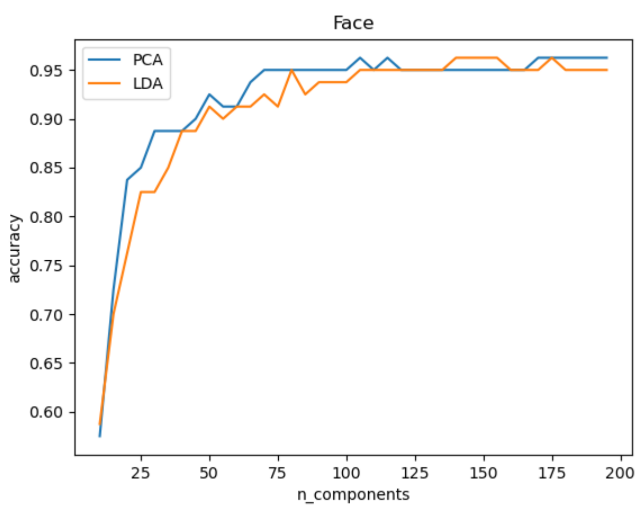
③ 特征选择效果好，但计算量大

(3) 嵌入式特征选择：在学习投影W时，对W进行限制，使其不仅满足训练样本误差最小，同时使W中非零元素尽可能少

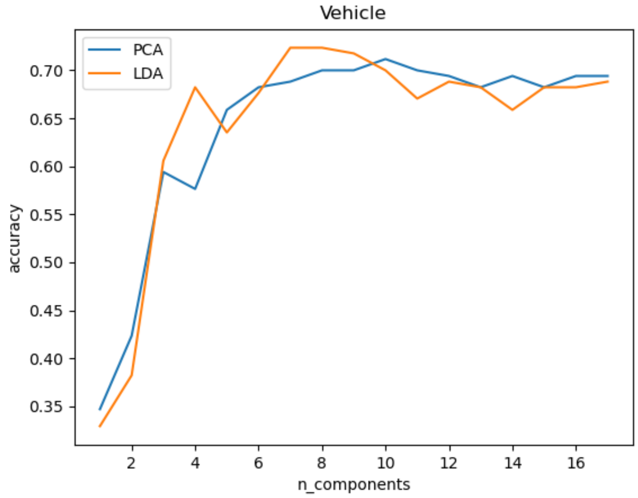
特点：① 特征选择与分类器训练同时进行

编程题

数据集1：



数据集2：



分类器性能大体上随着维度的增加而提升，但维度过小时，信息损失过多导致性能很差，同时对比两个数据集，由于第一个数据集在的特征数较多，第二个较少，从而PCA和LDA在数据集1上性能表现更好。