

# 机器学习练习题 6

## 1. Principal Component Analysis 算法填空并排序 (15 分)

输入：数据集(维度 $>k$ )，降到的维度数  $k$

输出：PCA 降维后的样本集  $D$

- A 计算协方差矩阵  $Cov$
  - B 用特征(▲)矩阵乘原数据得到降维结果
  - C 原始数据进行标准化,均值为(▲),方差为 1
  - D 计算协方差矩阵的(▲)和(▲)
  - E 选取特征值中的前(▲)列 (假设要降为  $K$  维)
- 对 ABCDE 进行排序 (▲)(▲)(▲)(▲)(▲)

## 2. Linear Discriminant Analysis 算法填空并排序 (15 分)

输入：数据集，降到的维度数  $d$

输出：LDA 降维后的样本集  $D$

- A 计算矩阵  $S_w^{-1}S_w$
  - B 计算类(▲)散度矩阵  $S_w$
  - C 得到输出的样本
  - D 计算  $S_w^{-1}S_w$  的最大的(▲)个特征值和对应的(▲),得到投影矩阵  $W$
  - E 计算类(▲)散度矩阵  $S_b$
- 对 ABCDE 进行排序 (▲)(▲)(▲)(▲)(▲)

## 3. Canonical Correlation Analysis 算法填空并排序(15 分)

输入：样本  $X$  和  $Y$  数量都为  $N$ ，其中  $X$  和  $Y$  的维度都  $> 1$

输出： $X, Y$  的相关系数,  $X$  和  $Y$  的线性系数向量  $w$  和  $v$

- A 由特征值计算  $X$  和  $Y$  的线性系数向量  $w$  和  $v$ ,
- B 计算  $X$  的  $S_{xx}$ ,  $Y$  的  $S_{yy}$ ,  $X$  和  $Y$  的(▲),  $Y$  和  $X$  的(▲)
- C 对矩阵  $S_{xx}^{-1}S_{xy}S_{yy}^{-1}S_{yx}$ ,  $S_{yy}^{-1}S_{yx}S_{xx}^{-1}S_{xy}$  进行(▲)
- D 计算矩阵  $S_{xx}^{-1}S_{xy}S_{yy}^{-1}S_{yx}$ ,  $S_{yy}^{-1}S_{yx}S_{xx}^{-1}S_{xy}$
- E 原始数据进行标准化,均值为(▲),方差为(▲)

对 ABCDE 进行排序 (▲)(▲)(▲)(▲)(▲)

编程题(使用 numpy 满分 55 分,直接调 sklearn 库则满分 40 分)

使用 python, Q#,C++,java,C#,R,Go,Pascal ,Ada,Perl,Ruby 任意一种实现上述 3 种算法.(建议用 Python).

注 1:数据和读取数据 python 代码已给出.

注 2:把每个矩阵的 Shape 都要在代码中用注释说明.

注 3:抄袭 0 分