Data Mining, Spring 2018

Problem Set #3: Unsupervised Learning

(Due on XXXX)

**Submission Instructions**

These questions require thought but do not require long answers. Please be as concise as possible. You should submit your answers as a write-up in PDF format to [DataMining\_2018@126.com](mailto:DataMining_2018@126.com). The email title is formatted as “hwk2\_学号\_姓名”.

**Questions**

1. **主成分分析（Principal Component Analysis，PCA）**

请从课程网站或[此链接](https://pan.baidu.com/s/1djZy69OmLWqJDpStjGcQaA)下载Yale人脸数据集进行降维。通过MATLAB命令load('yale\_face.mat')读取数据，包含一个矩阵。此矩阵的每一列是由一张灰度人脸图像所转成的向量。例如，可以使用imshow(reshape(X(:,1),[64 64]),[])命令显示第一张人脸图像。

1. 试使用MATLAB中的svd函数实现PCA算法，并显示均值图像和前五个特征向量所对应的图像；
2. 试对协方差矩阵使用MATLAB中的eig函数计算特征值，显示前五个最大的特征向量所对应的图像，并比较对数据矩阵使用svd函数的所得出的特征向量与运算时间；
3. 试计算当降维后的维数分别是10和100时，保留的方差的比例，并分别利用10维和100维坐标恢复原高维空间中的人脸图像，对前三张人脸图像，对比原图和两张恢复的图像；
4. **关联规则**