**Homework2**





**The PCA and KPCA transformations of data**

**电 话 13312055621**

**专 业 水利水电工程**

**课程名称 机器学习算法与应用**

**班 级 8班**

**学 号 2018205075**

**姓 名 董志明**

**2018年 10 月 28 日**

**Abstract**

It should be an explicit summary of the report that states the problem, the methods used, the major experiments, discussions and conclusions.

**1. Introduction**

**2. PCA and KPCA**

2.1 Theory of PCA and KPCA

In this section, the background, principle and derivation formula about PCA and KPCA should be described soundly.

PCA所解决的问题：给定一个由n维向量组成的，包含N个数据的数据集X，目标是将其压缩成p维，同时使得数据集中的主要特征被保留。

假设数据被投影到的方向向量为v，整个数据集在v上投影为:

投影的方差为：

将平方写成本身乘以其转置的形式，展开整理得：

其中：

假设：

从而可知：

PCA的目标是降维，所以我们的目标是找到使得最大的v。使用拉格朗日乘子的方法，可以得到下式：

从而可以求出C矩阵的特征向量和特征值。

最后，计算主成分的贡献率和累计贡献率。

贡献率：

累积贡献率：

一般取累计贡献率达85—95%的特征值，即为所对应的第一、第二、…、第k（k≤n）个主成分。

Kpca：基于kernel的pca。

2.2 The PCA and KPCA for "data\_PCA.txt"

In this section, an actual project is required to carry out with PCA and KPCA. The steps and results of processing should be given and analyzed in detail. The original date was given as follow, the different data sets are marked as red (1st-128th data in "data\_PCA. txt") and blue (129th-256th data in "data\_PCA.txt") respectively.



**3. Conclution**

**Appendix**

Give the used Codes

Specify your environment and how to execute.