**Project1-Part 2**

****



**The LDA and KLDA 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 9班**

**学 号 2018205075**

**姓 名 董志明**

**2018年 11 月 4 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**

LDA(Linear Discriminant Analysis)，中文名为线性判别分析，是一种经典的降维方法，LDA在模式识别领域（比如人脸识别，舰艇识别等图形图像识别领域）中有非常广泛的应用，因此我们有必要了解下它的算法原理。

LDA是一种监督学习的降维技术，也就是说它的数据集的每个样本是有类别输出的。这点和PCA不同。PCA是不考虑样本类别输出的无监督降维技术。LDA的思想可以用一句话概括，就是“投影后类内方差最小，类间方差最大”。什么意思呢？ 我们要将数据在低维度上进行投影，投影后希望每一种类别数据的投影点尽可能的接近，而不同类别的数据的类别中心之间的距离尽可能的大。

**2. LDA and KLDA**

2.1 Theory of LDA and KLDA

In this section, the background, principle and derivation formula on LDA and KLDA should be described soundly.

LDA的原理是找到一个满足如下特征的向量：

类间均值在该向量上的投影最大，每一个类中的样本投影与相应类均值投影之间的距离最小。

第i类的均值是：

类内的平方和距离可以表示为：

整理得：

所有类中所有样本的方差为：

整理得：

因此，要求的向量可以通过如下等式求得：

通过拉格朗日乘子法即可求解。

基于kernel的LDA:

将原式中的x替换为，经过相同的推导过程，可以得到：

式中，，

这等于求解如下的特征值问题：

最后，数据在新的投影空间内可以表示为：

2.2 The LDA and KLDA for "data\_LDA.txt"

In this section, an actual project about classification of data sets is required to carry out with LDA and KLDA. The steps and results of processing should be given in detail. The original data was given as follows, the data from different classes in which were marked as red (1st-201th data in "data\_LDA.txt") and blue (202th-402th data in "data\_LDA.txt") respectively.



**3. Conclution**

**Appendix**

Give the used Codes

Specify your environment and how to execute.