**数据降维方法**

# 线性降维方法

侧重让不相似的点在低维表示中分开。

## 主成分分析(Principal Components Analysis, PCA)

### 数据预处理









### 算法推导

标记：与均为列向量；

**目标：使得投影后的数据点尽可能地分散，即最大化投影后方差**

投影长度：

目标函数：



约束条件：



拉格朗日函数：



求导：



即是的特征向量，是对应的特征值。







## 线性判别分析（Linear Discriminant Analysis, LDA）

同类样本点之间的协方差矩阵尽可能小，异类样本点的投影中心距离尽可能大。



### 类间散度矩阵

#### 二分类问题



#### 多分类问题



### 类内散度矩阵

#### 二分类问题



#### 多分类问题



### 目标函数





约束条件



拉格朗日函数：



求导：





即是的特征向量，是对应的特征值。

### LDA与PCA

LDA用于降维，和PCA有很多相同，也有很多不同的地方，因此值得好好的比较一下两者的降维异同点。

#### 相同点

1）两者均可以对数据进行降维。

2）两者在降维时均使用了矩阵特征分解的思想。

3）两者都假设数据符合高斯分布。

#### 不同点

1）LDA是有监督的降维方法，而PCA是无监督的降维方法

2）LDA降维最多降到类别数k-1的维数，而PCA没有这个限制。

3）LDA除了可以用于降维，还可以用于分类。

4）LDA选择分类性能最好的投影方向，而PCA选择样本点投影具有最大方差的方向。

### LDA优缺点

LDA算法既可以用来降维，又可以用来分类，但是目前来说，主要还是用于降维。在进行图像识别相关的数据分析时，LDA是一个有力的工具。下面总结下LDA算法的优缺点。

#### 优点

1）在降维过程中可以使用类别的先验知识经验，而像PCA这样的无监督学习则无法使用类别先验知识。

2）LDA在样本分类信息依赖均值而不是方差的时候，比PCA之类的算法较优。

#### 缺点

1）LDA不适合对非高斯分布样本进行降维，PCA也有这个问题。

2）LDA降维最多降到类别数k-1的维数，如果我们降维的维度大于k-1，则不能使用LDA。当然目前有一些LDA的进化版算法可以绕过这个问题。

3）LDA在样本分类信息依赖方差而不是均值的时候，降维效果不好。

4）LDA可能过度拟合数据。

## 多维缩放（Multidimensional Scaling, MDS）

保证所有数据点对在低维空间中的距离等于在高维空间中的距离。

个样本，，。距离矩阵，其中第i行第j列的元素dij表示第i个实例和第j个实例之间的距离。把数据降维到维空间中去，得到所有样本点在中的表示。并且任意两个实例在维空间中的距离等于原始空间中的距离。

由保持距离原则可知



假设低维空间中的实例点是中心化的



左右两边求和







内积矩阵





对矩阵B做特征分解



中，是由B的特征值生成的对角矩阵，是特征向量作为列的矩阵

降到维空间中，那么选择前个最大特征值及对应的特征向量，得到和，则降维后的特征表示为





### 算法流程：

已知个样本，

1. 计算距离矩阵；



1. 计算内积矩阵



1. 对矩阵B做特征分解



1. 取选择前个最大特征值及对应的特征向量，得到和，则降维后的特征表示为



## 奇异值分解 (Singular Value Decomposition，SVD)





矩阵U的列向量称为矩阵A的左奇异向量，是的特征向量；矩阵V的列向量称为矩阵A的右奇异向量，是的特征向量。奇异值是特征值的平方根。



# 非线性降维方法

侧重让相似的近邻点在低维表示中靠近。

## 局部线性嵌入(Locally Linear Embedding, LLE)

## 拉普拉斯特征映射(Laplacian Eigenmaps, LE)

## 局部保留投影(Locality Preserving Projections, LPP)

## 等距映射（Isometric Mapping，ISOMAP)

## 随机近邻嵌入（Stochastic Neighbor Embedding，SNE）