线性判别分析

机器学习笔记 create by siwanghu v1.0

LDA（线性判别分析），又称Fisher线性判别。将高维的样本投影到最佳鉴别空间，以达到抽取分类信息和压缩特征空间维数的效果。用于分类或数据降低维数，LDA与PCA相比最大的特点是LDA是有监督的数据降维，PCA是无监督的数据降维。

在空间中样本，也就是说样本总共有m个，每个样本是n维的，代表每个样本有n个特征。样本总共分为c个类别，每个类别的数目大小用，也就是说。

LDA要做这样一件事情，它将样本投影到空间中，，代表将样本数据特征降维。在投影的空间中，样本类别很好分离。LDA算法的思想是，投影到另一个空间中时让不同类别的样本尽量离的远，同一个类别的样本尽量离的尽。

为了方便叙述算法，定义如下符号：

类间离散度矩阵

类内离散度矩阵

属于第i类样本的个数

代表第i个样本，是一个多维的列向量，存储了样本的每个特征

所有样本的均值

代表第i类样本的均值

LDA算法的关键步骤是计算类间离散度矩阵和类内离散度矩阵

如果只有两类

然后求出矩阵的特征值，取前最大个特征值对应的特征向量组成矩阵

然后取出样本X中的每个样本，经过投影矩阵W，求出在投影特征空间中的对应点即可，计算方式为：

**LDA与PCA的区别**

**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