# Project\_2 LDA

目录

[一、 线性判别分析有哪些假设](#_Toc8819_WPSOffice_Level1) [2](#_Toc8819_WPSOffice_Level1)

[二、 线性判别分析如何学习多类特征](#_Toc27475_WPSOffice_Level1) [2](#_Toc27475_WPSOffice_Level1)

[三、 编程实现](#_Toc25538_WPSOffice_Level1) [4](#_Toc25538_WPSOffice_Level1)

组内同学信息：

叶晓英（111701436，电子信息类）

陈家瑄（021700440，机械设计制造及其自动化）

吴智鸿（011702236，自动化）

曾世缘（031702332，计算机类）

作业心得、踩坑记录、总结思考：https://www.cnblogs.com/FormerAutumn/p/10914753.html

1. 线性判别分析有哪些假设
2. 样本数据服从正态分布。
3. 各个类别的协方差相等。
4. 样本的特征从统计上来说相互独立。
5. 矩阵可逆。

（事实上，即使违背上述假设，LDA仍能正常工作）

1. 线性判别分析如何学习多类特征

假定存在N个类，且第i类,我们先定义“全局散度矩阵”:

--------（1）

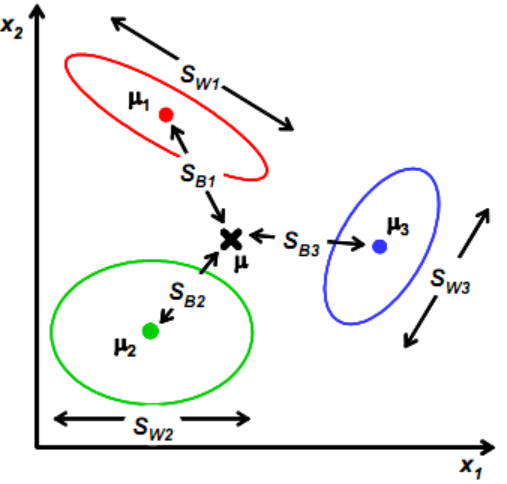
其中是所有示例的均值向量。将类内矩阵重定义为每个类别的散度矩阵之和，即：

--------（2）

其中 --------（3）

由式（1）~（3）可得：

例如：三类问题如下直观图所示：



显然，多分类LDA可以有多种实现方法：使用，三者中任意两个即可。

如下面的一种实现优化： --------（4）

其中tr()表示矩阵的迹（trace），式（4）可通过如下广义特征值问题求解： --------（5）

W的闭式解则是的N-1个最大广义特征值所对应的特征向量组成的矩阵，其中的tr()为矩阵的迹，一个的对角矩阵A的主对角线（从左上方至右下方的对角线）上各个元素的总和被称为矩阵A的迹（或迹数），一般记作tr(A)。这个优化目标实际上等价于求解N-1个w（特征向量）组合成W。若将W视为一个投影矩阵，则多分类LDA将样本投影到N-1维空间，N-1通常远小于数据原有的属性数（维度）。于是，可通过这个投影来减小样本点的维数，且投影过程中使用了类别信息，因此LDA也常被视为一种经典的监督降维技术（可用于特征提取）。

另一种多类推广原理：  
 假设我们的数据集D = {(), ()...()}其中任意样本为n维向量， ，我们定义为第j类样本的个数， (j = 1,2...k)为第j类样本的集合，而为第j类样本的均值向量，定义为第j类样本的协方差矩阵。在二类LDA里面定义的公式可以很容易的类推到多类LDA。  
 由于我们是多类向低维投影，则此时投影到的低维空间就不是一条直线，而是一个超平面了。假设我们投影到的低维空间的维度为d,对应的基向量为)，基向量组成的矩阵为W,它是一个的矩阵。  
 此时我们的优化目标应该可以变成为:  
   
  
 其中，为所有样本均值向量。

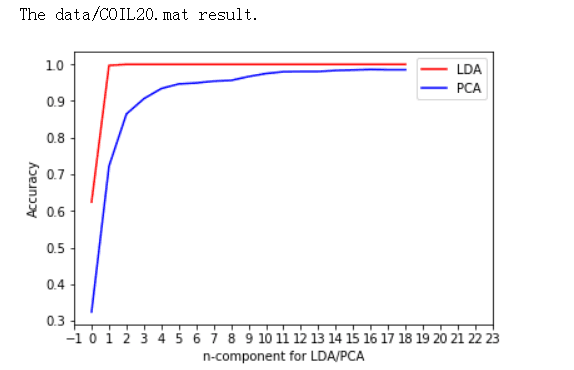
和都是矩阵，不是标量，无法作为一个标量函数来优化，也就是说，我们无法直接用二类LDA的优化方法，一般来说， 我们可以用其他的一些替代优化目标来实现。  
 常见的一个LDA多类优化目标函数定义为:  
 其中为A的主对角线元素的乘积，W为的矩阵。J(W)的优化过程可以转化为:  
 上式最右边是广义瑞利商，最大值是矩阵的最大特征值，最大的d个值的乘积就是矩阵的最大的d个特征值的乘积，此时对应的矩阵W为这最大的d个特征值对应的特征向量张成的矩阵。  
 由于W是一个利用了样本的类别得到的投影矩阵，因此它的降维到的维度d最大值为k- 1。这是因为中每个的秩为1,因此协方差矩阵相加后最大的秩为k(矩的秩小于等于各个相加矩阵的秩的和)，但是由于如果我们知道前k-1个后，最后一个可以由前k-1个线性表示，因此的秩最大为k-1,即特征向量最多有k-1个。

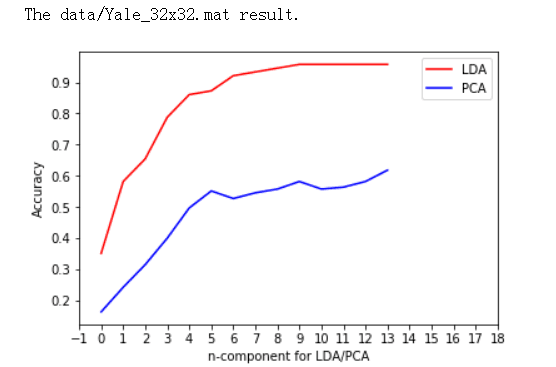
简单来说，当涉及多分类问题时，；而二分类时，。

1. 编程实现

LDA与PCA的对比（Y轴：Accuracy，X轴：枚举参数(n\_component)为横轴进行多分类），其中COIL有20类，Yale\_32x32有15类，下图展示的为多分类结果。

**注：LDA、KNN均为手动编写，运用准确度（Accuracy）进行评估。（即分类正确的样本占所有样本的比例），用LDA实现二分类时的准确度为1（Accuracy=1）。实现KNN时，由于测试出来K值（1-10）的选择差异并不大，所以最后选择K=7，进行的KNN算法。**





结果分析：

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。在不同数据分布下LDA和PCA降维优势不同。二者各有优缺。

PCA是将数据投影到方差最大的几个相互正交的方向上，以期待保留最多的样本信息。样本的方差要是越大表示样本的多样性越好，若是样本很多他们彼此相似或者相同，提供的样本信息将相同，相当于只提供了少量信息。样本信息不足将导致模型性能不够理想。

在物品数据集（COIL20）中，由于不同类别的差异性较大，PCA与LDA都能进行较好的效果。在人脸数据集（Yale\_32x32）中，外在因素如光线，角度和人脸尺寸等带来的图像的差异和人脸本身带来的差异是无法区分的。而LDA提取的特征向量集，强调的是不同人脸的差异而不是人脸表情、照明条件等条件的变化，从而有助于提高识别效果。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。

而至于LDA实现二分类时，即将维度降至类-1（一维），以中间值为阈值，左右即为正类负类，实现过程是在源代码中进行的调参，并未单独另写代码。

