## LDA的材料整理

LDA和PCA之间的区别：

1. LDA和PCA都是数据降维的很好的技术和方法
2. PCA主要是从特征的协方差角度，去找到比较好的投影方式。LDA更多的是考虑了标注，即希望投影后不同类别之间数据点的距离更大，同一类别的数据点更紧凑。同类之间距离短，不同类之间距离大

# python实现LDA多类问题

【机器学习】python实现LDA多类问题

1.读取数据集

2.进行LDA特征提取

2.1 将样本分类

2.2 求类内散度矩阵

2.3求类间散度矩阵

2.4 计算Sw-1\*Sb的特征值和特征矩阵

2.5 特征值排序，提取前k个特征向量

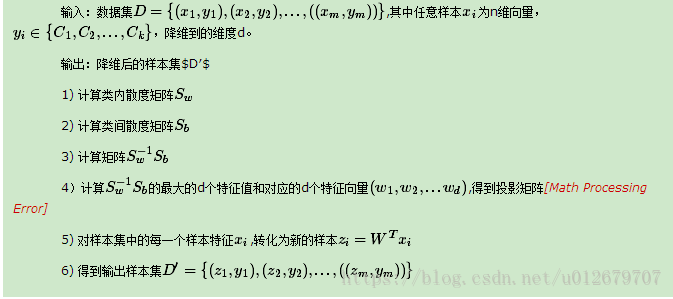
关于LDA的原理已经在之前讲过了，详情可戳：【机器学习】LDA线性判别分析

    本节主要用python实现LDA推广到多类的问题。

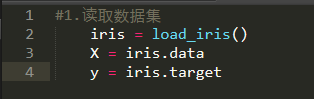
    【注意】本文的运行环境是windows+Pycharm+python3.6。

 【参考】线性判别分析(LDA)和python实现（多分类问题）https://blog.csdn.net/z962013489/article/details/79918758

LDA算法流程:



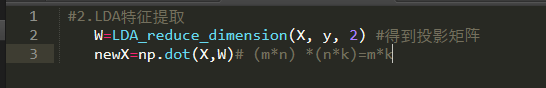
1. ****读取数据集****



注意：读取到的X,y均为矩阵类型

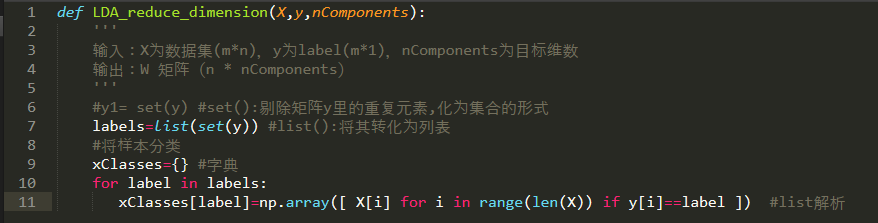


1. ****进行LDA特征提取****

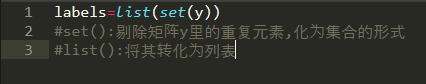


下面,详细介绍ＬＤＡ算法实现过程:

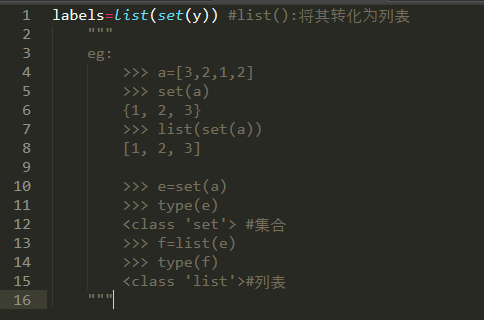
* 1. 将样本分类



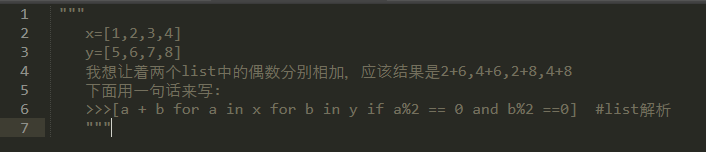
其中， set()是转化为集合形式，这样可剔除类别中的重复元素。list()是将其转化为列表，便于后续操作。



set()和list()的实例如下：

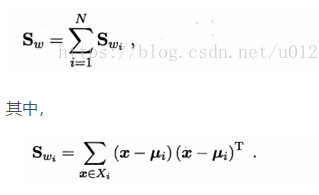


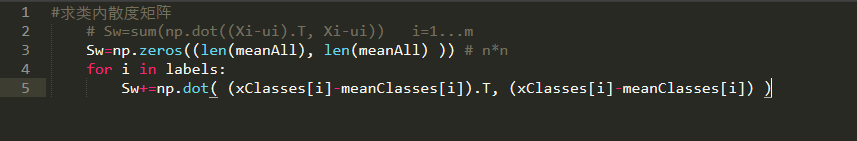
注意，list解析部分，此处非常灵活。可通过下边的例子进行理解。



* 1. 求类内散度矩阵

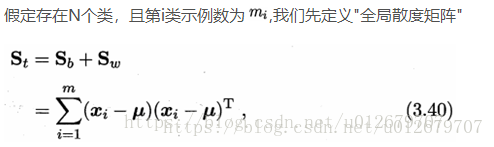
表达式如下：

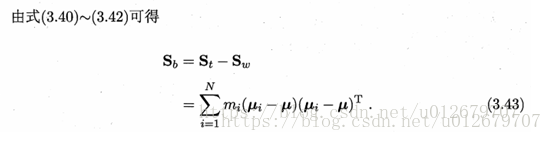


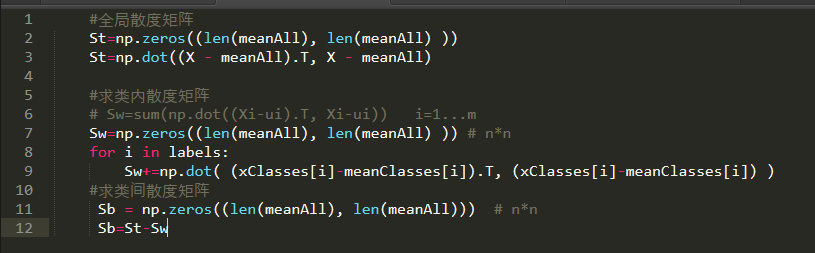


2.3求类间散度矩阵

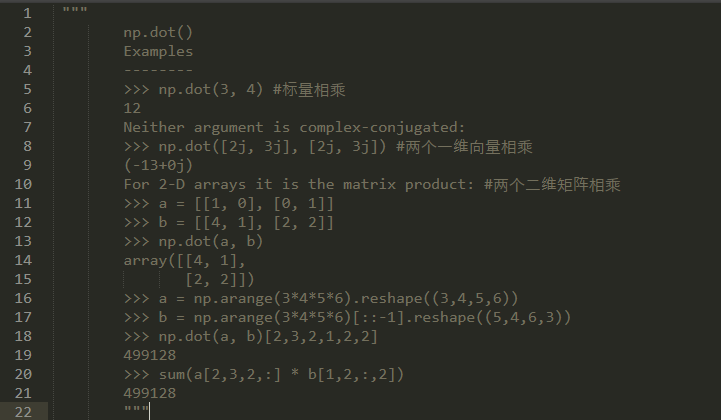
类间散度矩阵Sb 可以通过 全局散度矩阵St 和 类内散度矩阵Sw 来求；



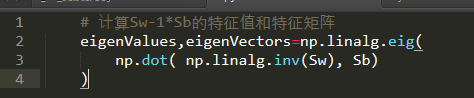




np.dot()矩阵乘法的具体实例如下：



* 1. 计算Sw-1\*Sb的特征值和特征矩阵



* 1. 特征值排序，提取前k个特征向量

