常用的数据降维方法有PCA、LDA和NCA，下面简述它们的基本原理并稍作对比。

**主成分分析PCA**

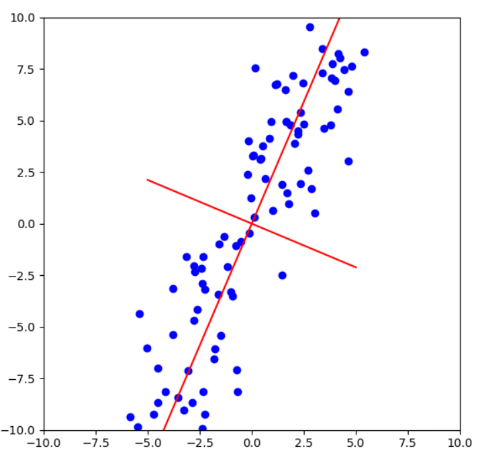
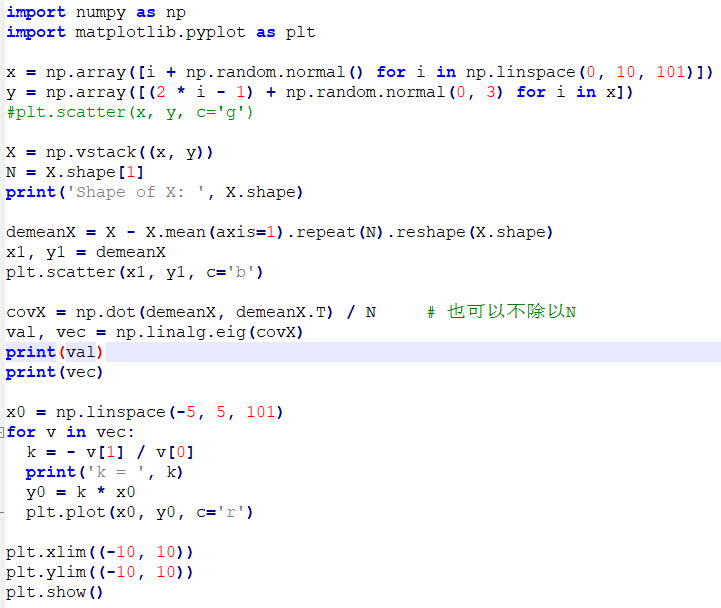
PCA方法的直觉：一个n维数据立方中，任何n-1个点都形成一个超平面，因此它们所对应的数据向量是线性相关的。那么，如果所有数据点都严格分布在同一个n-1维超平面上，那么将所有数据去中心化后就可以直接降一个维度。次之，如果所有数据点都分布在某个n-1维超平面附近不太远的位置上，则这个超平面的法向维度可以视作噪声扰动而被主动忽略，将所有数据点沿法向投影到该超平面上也将使得数据立方降低一个维度。籍此，重复多次去掉噪声维度（方差较小的维度），余下的数据立方空间的正交坐标基即为原数据的主成分。

简单说来：PCA将n维特征映射到k维空间（k<n），新空间的正交坐标基作为特征特征称为主成分；注意k维特征是构造出来的，而非从原空间简单删去维度得到。

基于最大可分性的PCA推导：N个d维既去中心化数据表示为d×N的矩阵X，d维向量w用来标记每个维度的线性权重(亦即待求主成分方向)，作线性投影，现欲使y的方差尽可能大。由于X已中心化、故y亦然，知y的方差，其中为原数据的协方差矩阵。由拉格朗日乘子法求并约束可得，则求w和λ即对S进行特征值分解。

PCA流程：计算原数据矩阵X的协方差矩阵S，对S进行特征值分解、舍去谱尾而保留前k个特征值，将这k个特征值所对应的特征向量转置组成一个k×d的投影矩阵P，可得降维后的低维空间数据立方即为

试着写了个二维的例子：

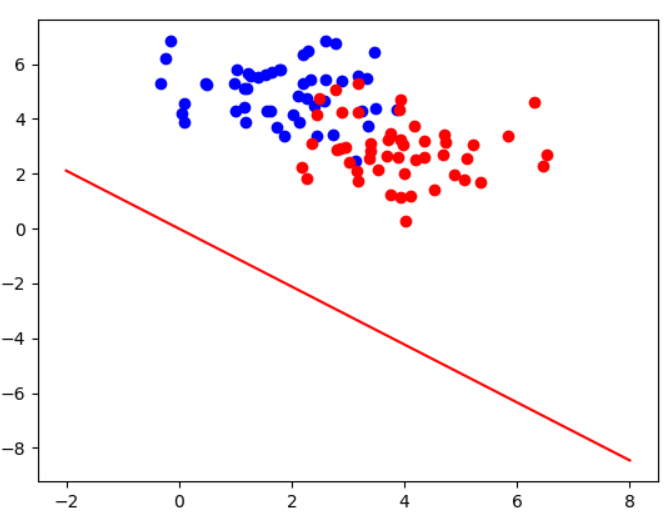
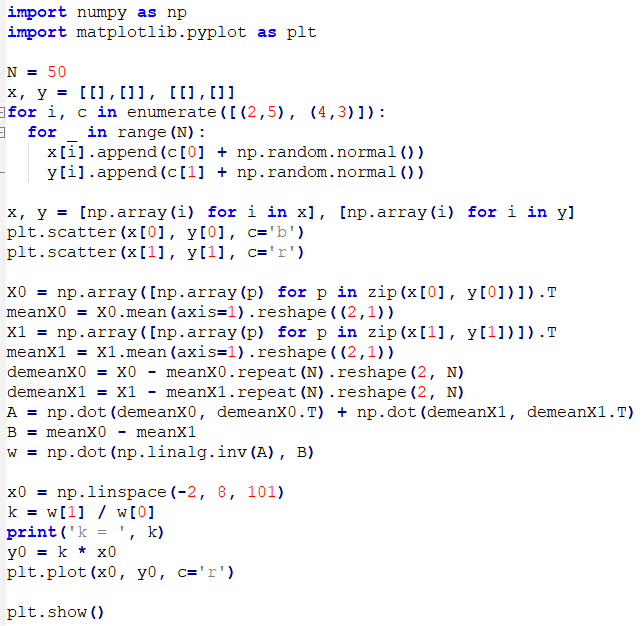
**线性判别分析LDA**

LDA方法的直觉：对于分类任务，考虑到最终只有C个目标分类，我们把数据立方降维到C-1维通常就足以完成分类任务了；考虑同样使用线性投影的方法，我们将d维数据的一些维度合并、坍缩到k维（k∈[C-1,d]），并力图使得投影后的空间中，属于同一个目标分类的数据点尽可能聚集成簇，而不同分类的数据团簇之间尽则可能疏远。

简单说来：LDA借助数据标记作为监督信息，将n维特征映射到k维空间（k<C），使得新空间中不同类数据的团簇内聚外疏。

基于均值最大差异的LDA推导：以二分类问题为例，我们试图把d维数据降到1维，则投影后两个团簇的均值差，其中w依然为线性权重亦即投影方向，而团簇的内聚性由该类中数据点的协方差来衡量，，其中表示在所属分类下去中心化后的向量；则优化目标即为约束条件下的，由乘子法可解得。

依然写了个小例子：

** **

**近邻成分分析NCA**

NCA方法：切近于解决先做流形上的度量学习、然后用kNN处理分类任务。NCA先学习得到一个距离矩阵来合理衡量两个数据点之间的相似度（而非直接使用欧氏距离、曼哈顿距离等人工设定的距离度量），本质是寻找一个对称半正定矩阵W，使得在该变换下近邻分类的效果最优。NCA算法使用随机选择近邻、然后用留一法进行交叉验证以不断优化该矩阵W。

马氏距离的直觉：从基本的欧氏距离入手，任意两个数据点之间的欧氏距离为，我们假设每个分量之间的距离重要程度不一样，因此引入一个线性加权向量，即得马氏距离，其中W为对角阵。

NCA降维原理：如果NCA算法学习所得的距离矩阵W是个降秩矩阵或者有很长的特征谱尾，含义即某些维度上的数据点语义在该维度下距离为0或者很小，则按照PCA的想法同样可降W进行特征谱分解、截断谱尾，从而把原数据投影到一个低维空间中去，并保持变换前后的马氏距离差异不太大；可以说这是某种意义下的保距映射。

这个我就写不动例子了……略过略过……

**方法对比**

1. 三个方法都用到了特征谱分解的思想
2. PCA和LDA是线性方法；NCA则不是
3. PCA和LDA预设数据符合高斯分布；NCA则不必
4. PCA无监督，LDA和NCA有监督
5. LDA选择分类最好的方向投影、NCA以kNN的反馈作为距离度量的参考，因此这两者有过拟合风险；PCA选择方差最大的方向投影、不使用监督信息，因此没有
6. PCA降维的目标空间维度可控性较大，LDA和NCA则有一定限制
7. PCA和NCA对方差敏感，LDA对均值敏感