1. 原始数据图（Figure\_1）
2. Pca数据图（Figure\_2）

基于pca的改进：按照学生所在的专业标签（0、1、2、3、4）用excel分成五张表格。接着利用Matlab制作聚类分析（建模书p274页，R型聚类是对变量进行聚类而非样本聚类，因此反映的是各个变量关于标签的关联程度）图像判断五门科目的分数和专业之间的联系（对应五张聚类分析图。越靠近x轴表明越早开始聚类，也就说明该分数与专业关联度越大）

选择与每种专业对应关系最紧密的两门科目的分数作为原始数据，重新进行PCA。产生的图（Figure\_2-）中已经呈现出了较明显的数据区分。由此可以较为清楚地辨别出专业的分配与不同科目分数之间的联系。

3、T-SNE数据图（Figure\_3）:基于T分布（具有长尾特性，因此会自动舍弃异常数据）的SNE(随机相邻嵌入)无监督学习方法。

4、非负矩阵分解（NMF）（Figure\_4）: 给定矩阵V，寻找非负矩阵W和非负矩阵H，使得VWH。这时用系数矩阵代替原始矩阵，就可以实现对原始矩阵进行降维，得到数据特征的降维矩阵。

5、LDA（线性判别分析模型）(Figure\_5)：也称为 Fisher 线性判别，1936 年由 Ronald Fisher 提出。有监督学习，优化目标是投影后类内方差小类间方差大，也即最大化广义瑞利商。可用来分类或降维。

6、标准差（STD）（Figure\_6）：统计每一列的标准差

Task1论文内容顺序：1——6（标准差展示了各科成绩的离散程度）——5（监督学习方法）——2、3、4（无监督学习方法）

Task 2 四个有监督分类器

7、SVM：在不设定交叉验证集的情况下将整个数据集作为训练集和测试集，得到的准确率为58%（SVMlearning1）。同时作出预测结果的图像（SVMlearning），得出机器学习结果集中在选择programme1和2上。设定原数据集的70%作为训练集，剩下的作为测试集。对训练集采用网格搜索算法（GridSearchCV）进行交叉验证cross-validation，并且设置了防止过拟合的判断。最终得到的最高准确度为54%（交叉验证后的准确度）。作出的预测结果图像为（svmcvresult），可以看出相比于先前的预测结果图像，该图像显示了机器在programme4和0上也做出了一定的选择。另一张图片（svmcv）采用3D网格展示了不同的超参c、g组合对应的准确度。从图中可以看出当c趋近于，g趋近于时，准确度达到一个极大值。这符合代价函数系数c和高斯核函数参数g关于准确度的变化规律。

8、KNN：利用train\_test\_split函数将数据随机分成70%的训练集和30%的测试集并进行交叉验证。最终得到测试样本的给定标签与机器生成的标签之间的准确率在46%～56%之间（KNNaccuracy）

9、随机森林：调用随机森林训练函数，将树的层数控制在10层（防止过拟合）。通过交叉验证（和KNN一致）得出准确率的变化曲线在48%～60%之间。（随机森林）

10、神经网络：使用神经网络对数据进行分类。优化器选择两种进行比较：第一种是Adam优化方案。它是一种逐参数适应学习率方法。经过1000次迭代（每250次后降低一次学习率（自乘0.9），每100次学习输出一次当前的训练集和测试集的准确度以及误差），得到最终的准确度为测试集54%、训练集58%、平均误差1.7。第二种采用SGD+nesterov动量的方案，同样经过1000次迭代后准确度为训练集58%、测试集57%、平均误差1.05。（两种优化方案的原理见CSDN收藏夹）

Task 3 一个无监督分类方案

采用K-Means无监督学习方法对数据进行重新分类。类别划定采用四种类别（合并原先五种类别中的相似项）。数据特征上根据R型聚类得出的结果，选择关联度最大的两个特征作为数据的新特征（原先每种programme对应的两个特征均不相同，因此重新拆分数据原表，经过拼接重组得到新的一张514\*2的表格作为分类的数据集。选取前4个数作为初始聚类中心，通过计算两点之间的欧式距离判断剩余点所属类别，再利用平均值重新计算新的聚类中心。以此类推最终得到分类结果。在进行可视化后采用轮廓系数对聚类效果进行评判。判断距离依然采用欧式距离。最终得到的系数为0.46。由于轮廓系数处在-1至1之间且越大说明聚类效果越好。因此可以看出采用K-Means无监督学习能够取得较好的聚类效果，并且数据点之间重合冲突现象较小。（见图K-Means）

Reference

1. <https://ieeexplore-ieee-org.ez.xjtlu.edu.cn/document/4633689>
2. <https://ieeexplore-ieee-org.ez.xjtlu.edu.cn/document/6002199>
3. <https://distill.pub/2016/misread-tsne/>
4. <https://xueshu.baidu.com/usercenter/paper/show?paperid=a64c6dc5222b0fdda2b1778c6253bbea&site=xueshu_se>
5. <https://xueshu.baidu.com/usercenter/paper/show?paperid=d0d494139782851cfd0e8f1062408ad6&site=xueshu_se>
6. <https://xueshu.baidu.com/usercenter/paper/show?paperid=317eb3940acedafc6b1a4e34a5a8bd9d&site=xueshu_se>
7. <https://xueshu.baidu.com/usercenter/paper/show?paperid=903f59b4e51b5ce5e63c2b0426471bbc&site=xueshu_se>
8. <https://ieeexplore.ieee.org/document/5408784>
9. <https://xueshu.baidu.com/usercenter/paper/show?paperid=2dd18408fda30f6438efd00598c5035e&site=xueshu_se>
10. <https://xueshu.baidu.com/usercenter/paper/show?paperid=2815fe2e7eaf7485735d130eac76d330&site=xueshu_se>