

1. 任务一：找出各个因素对学习满意度的影响情况。

建模思路：是一个特征选择的问题（可以理解为对学习满意度影响因素进行一个打分，看看哪些比较重要，哪些不这么重要）。

建模技术和路线选择：（最后选一到两个就行）

第一类：直接用 python 计算基尼系数，皮尔森相关系数等等（比较易于理解）

第二类：基于 sklearn 的特征选择（递归式特征消除/用 SelectFromModel 选取特征）等等很多，文档链接 <https://www.sklearn.cn/14/>

第二类：可以用一些机器学习的方法（决策树，随机森林，SVM 支持向量机）反过来推，看一看哪些特征对最终影响结果比较大（就很计算机专业化）

2. 任务二：找出各个因素之间的相关性。

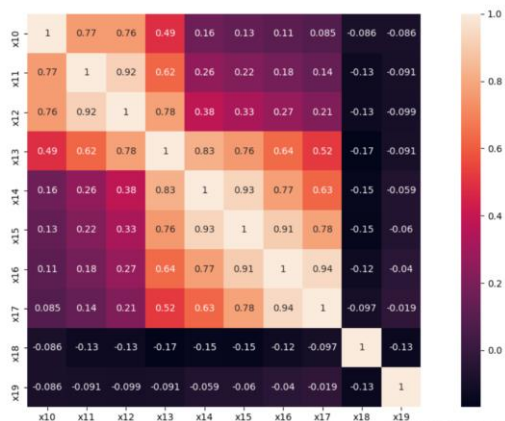
建模技术路线：（这个 python 代码比较简单，就两三行）

1. 可以直接计算特征之间的皮尔森相关系数，协方差等等

2. 根据上面一问的结果，根据打分排名看相关性。

3. 可以做一个预测系统：即根据任意一个同学现在内部因素和外部因素各种影响因素的分数，我们可以预测出他现在的学习满意度怎么样。

Tips：大概建模思路就是上面这些，主要还是得看数据怎么样，做下来看最后结果怎么样，要是大家学习满意度都很高，没有学习满意度差的，这样的数据估计会很难处理，包括样本量到底有多少等等问题，然后做的话可以做的比较丰富，包括各个维度的影响因素都可以用，特征的数量并不太影响，然后结果以表格或者热力图的形式展现等等。



然后老师你这边的数据我还没能太明白每一列的含义，我不知道老师发的这个 excel 是否是最终的数据。