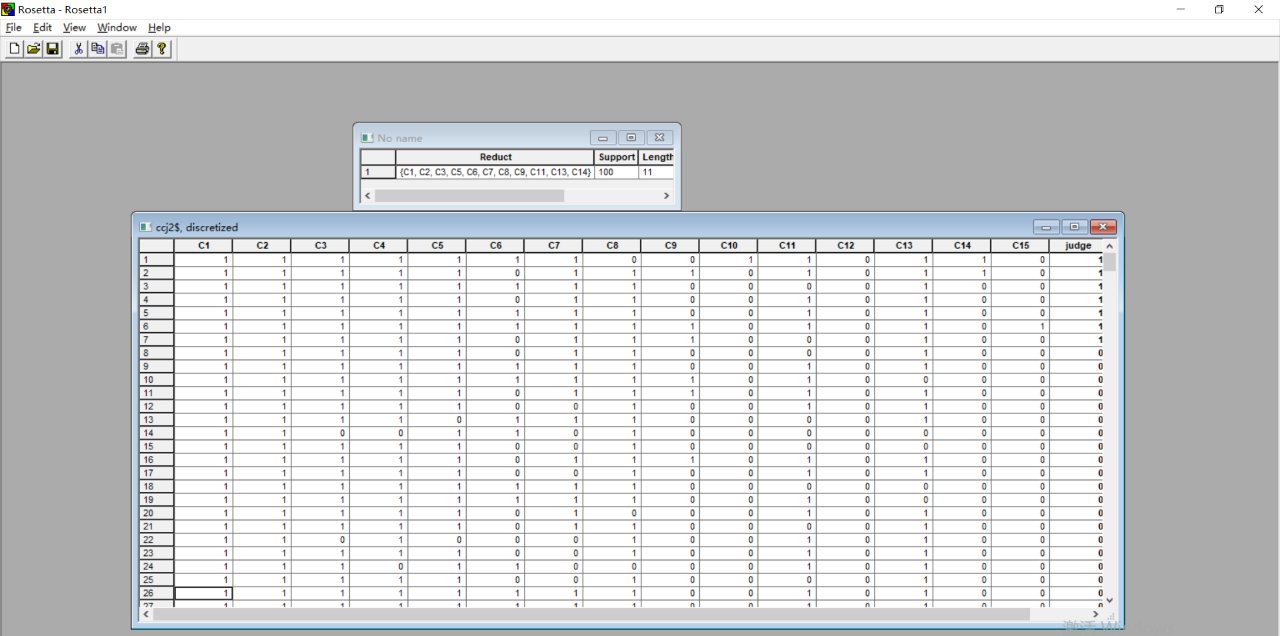
我的数据是电影票房相关数据，共15个条件属性，包括类型、豆瓣评分等，1个决策属性，电影最终票房；共选取将近500部电影。数据都已经量化了，也标准化了。论文实证部分流程是先通过粗糙集属性约简将15个属性约简一下，去掉几个不重要的指标；然后通过模糊聚类分成几类后面分别建立BP神经网络模型；通过k折交叉验证优化参数；通过BP神经网络训练，与普通神经网络对比效果，并拿一部新电影进行预测。我现在的需求是可以将粗糙集属性约简、模糊聚类、k折交叉验证和BP神经网络以最基础不考虑各种复杂优化方式的情况实现一下，但需要解释关键参数为什么这样选择，然后做预测。我大体做了一遍过程，但是包括步骤中用什么算法、什么参数、需要得到哪些结果图表，怎么分析结果还是不清楚。因为没系统学习python和R，所以粗糙集属性约简我用的rosetta，模糊聚类和BP神经网络我用的matlab，k折交叉验证我还没看怎么整，如果不好整我想也可以把这个在论文里整个删掉。用什么软件无所谓，只要结果清楚好解释就行。

下面是我这两天做的一遍流程：

1.粗糙集

用的rosetta软件，下图结果离散化用的信息熵方法。属性约简用的Johnson算法，遗传算法会出现好几种结果，我不清楚啥意思。

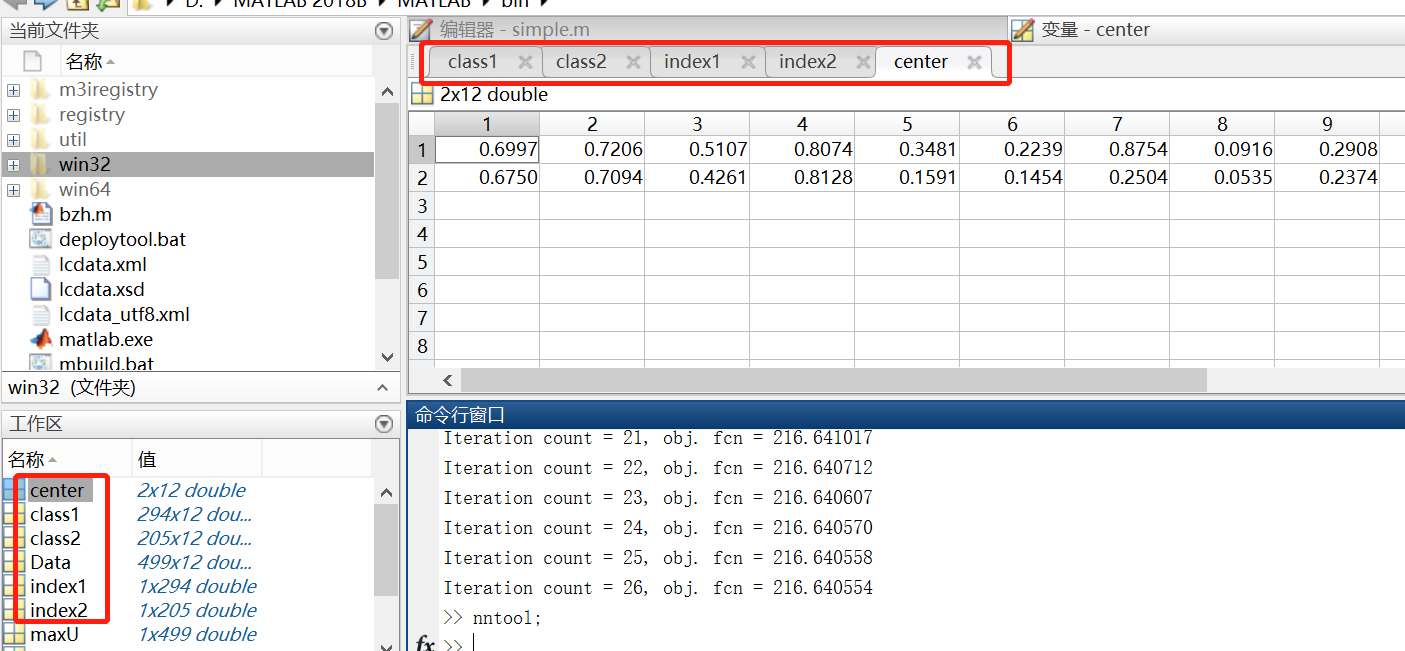
结果：

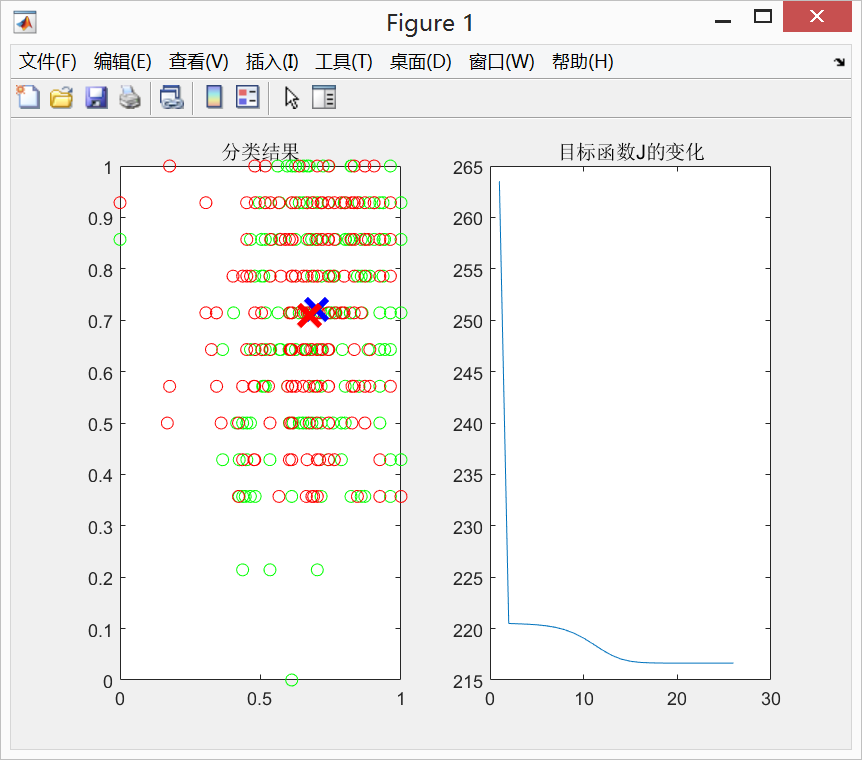


然后我就删掉了C4、C10、C12、C15。

2.模糊聚类

用的matlab，能得到分完类的各类的数据和结果





3.bp神经网络

用的matlab神经网络工具箱，参数几乎没动，就把节点数变成指标数11，因为需要输出一个确切的数而不是分类，应该用回归，输出层激励函数我用的purelin，结果里epoch次数是自动的，没法改，只能设置最大迭代次数，不知道怎么整。下面三个图是结果。

