**实验目的：**

• 编程实现变量：a, b, c, d

• 编程实现Jaccard系数

• 编程实现DB指数

• 编程实现距离公式，p=1,p=2

**实验内容：**

1. 使用iris数据集，将划分好的n=105条训练数据进行聚类。
2. 通过kmeans函数，将数据得到的簇划分为3类，可以直接得到对应的簇标记Idx，自定义Idx\*为：[ones(35,1);2\*ones(35,1);3\*ones(35,1)]
3. 通过二重循环i=1:n和j=(i+1):n组合出不同条件下的SS、SD、DS、DD
4. 计算出a=|SS|、b=|SD|、c=|DS|、d=|DD|
5. 根据求得的a、b、c、d计算Jaccard系数
6. 二重循环计算两两数据之间的闵可夫斯基距离，p=1时为曼哈顿距离，p=2时为欧氏距离。

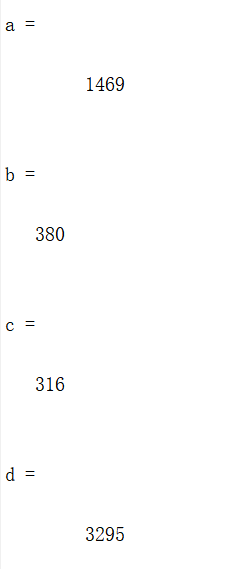
**实验结果，截图并解释：**

1. **簇标记：**

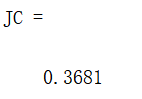
Idx=[2;2;2;2;2;2;2;2;2;2;2;2;2;2;2;2;2;2;2;2;2;2;2;2;2;2;2;2;2;2;2;2;2;2;2;1;1;1;1;1;1;1;1;1;1;1;1;1;1;1;1;1;1;1;1;1;1;1;1;1;1;1;1;1;1;1;1;1;1;1;1;3;1;1;3;3;3;3;1;3;3;1;3;3;1;3;1;1;3;3;1;3;1;1;3;3;3;3;3;3;3;3;3;3;3]

Idx\*=[ones(35,1);2\*ones(35,1);3\*ones(35,1)]

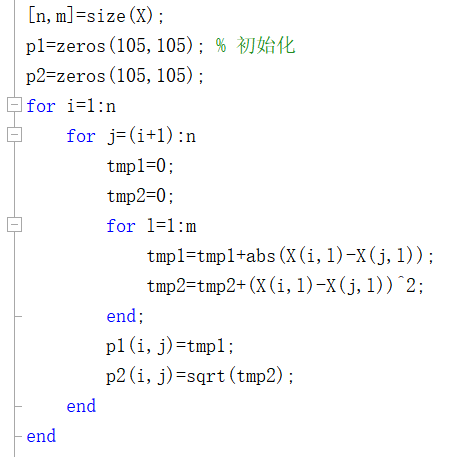
1. **a b c d值**



1. **Jaccard值**

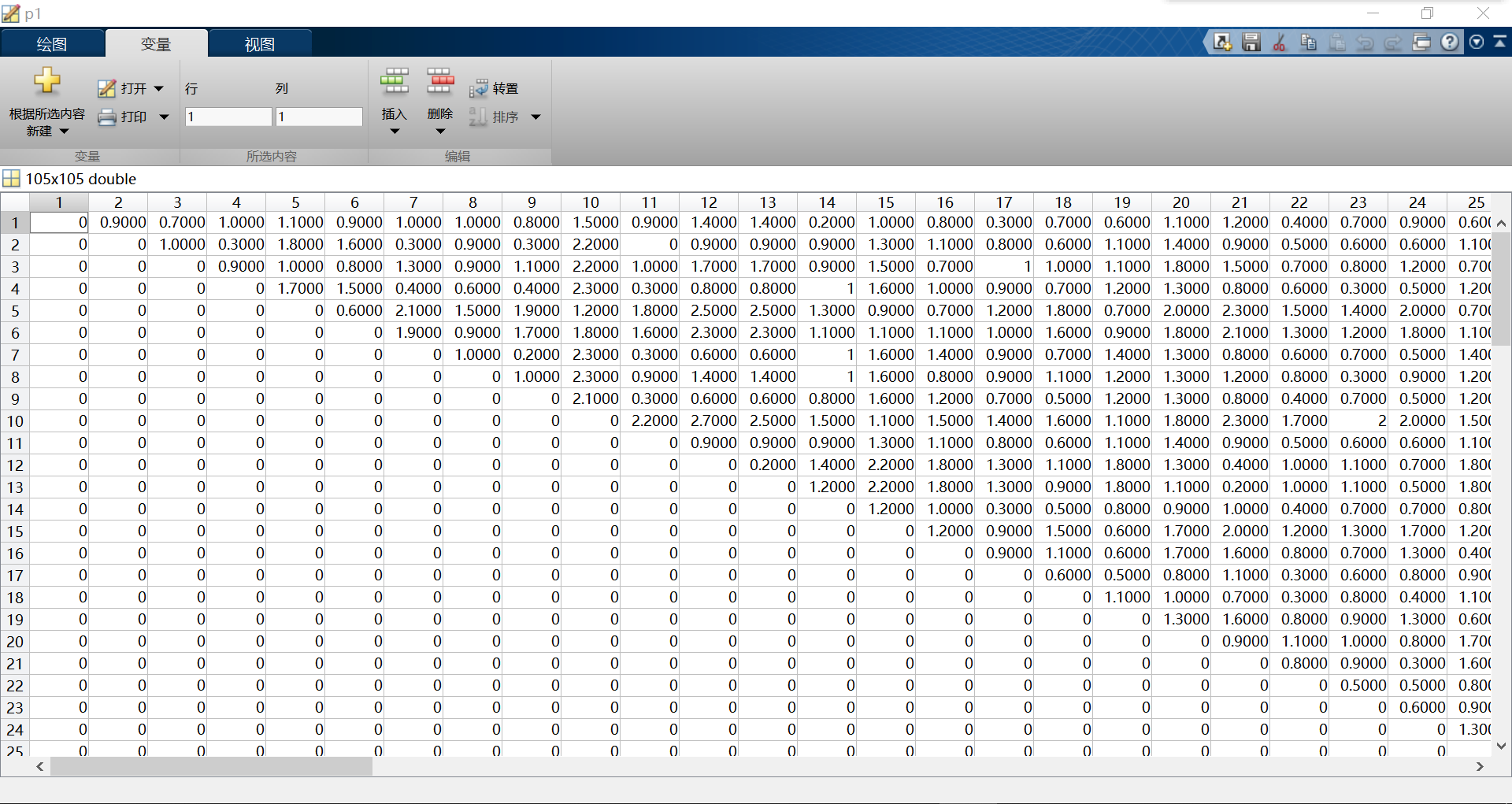


1. **闵可夫斯基距离**

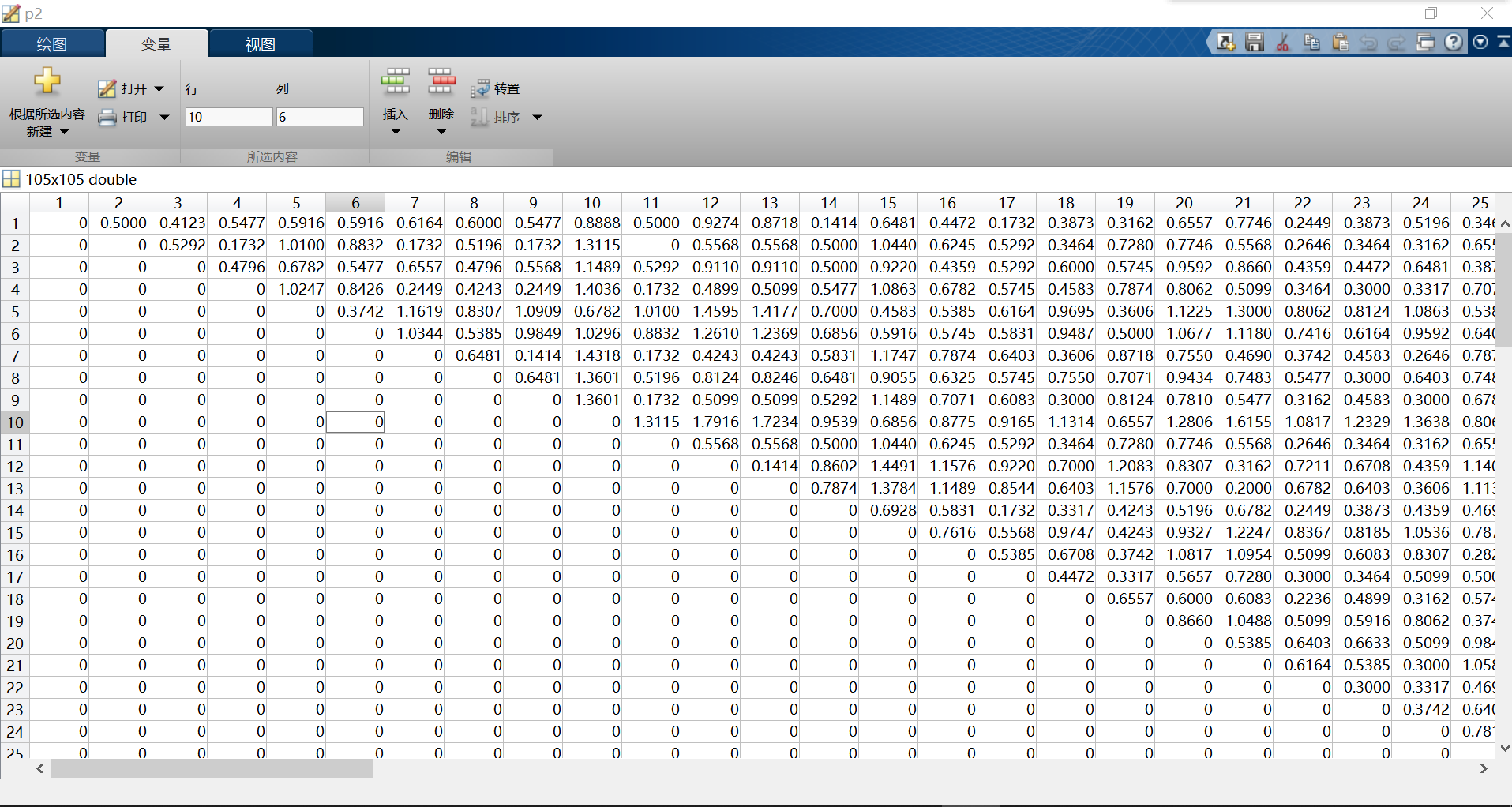
****

鸢尾花数据酸的两两数据之间的距离一共有105\*104/2个，这里只截一小部分结果

P=1时，曼哈顿距离：

****

P=2时，欧氏距离：



举例：

X=

1 1 1

2 2 2

3 3 3

得到：p=1时，dist(x1,x2)=3，dist(x1,x3)=6，dist(x2,x3)=3

P=2时，dist(x1,x2)=1.732，dist(x1,x3)=2.449，dist(x2,x3)=1.732

**实验总结：**

聚类算法有很多种分法，体系也很大，聚类效果的好坏依赖于两个因素：衡量距离的方法（distance measurement）和聚类算法（algorithm）。

外部指标(external index)和内部指标(internal index)能够体现聚类的性能。其中外部指标有Jaccard系数、FM指数、Rand指数，他们的范围在[0,1]区间，值越大性能越好。内部指标有DB指数和Dunn指数，DB指数越小越好，Dunn指数越大越好。