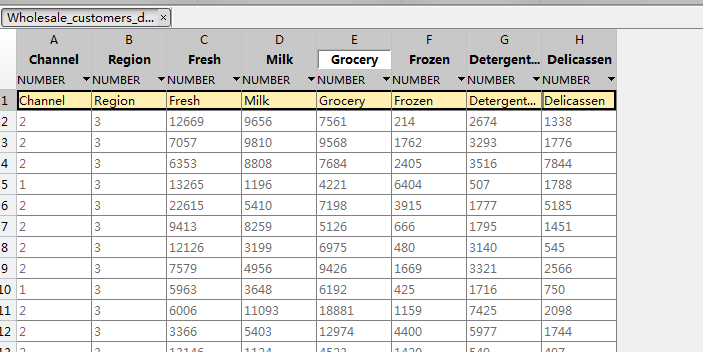
**数据挖掘project**

**大概过程和结果如下：**

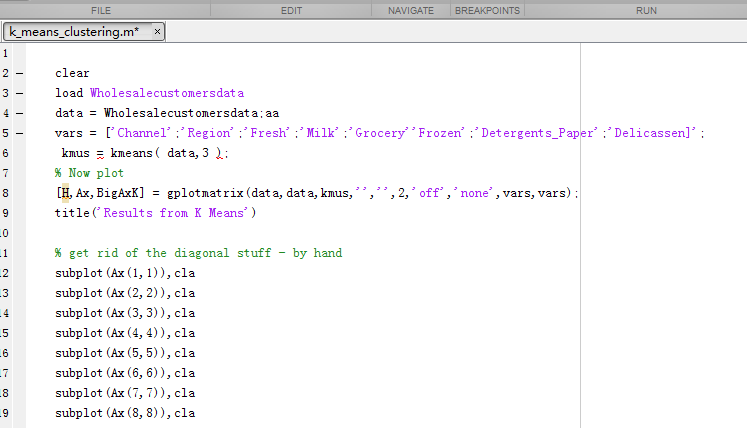
**在很久以前，**

**我们确定了数据集<Wholesalecustomersdata.Svl》，经过MATLAB处理数据**

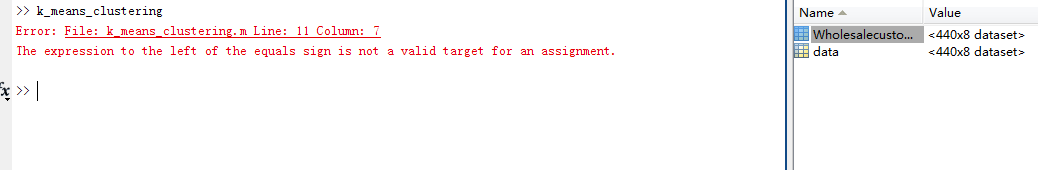


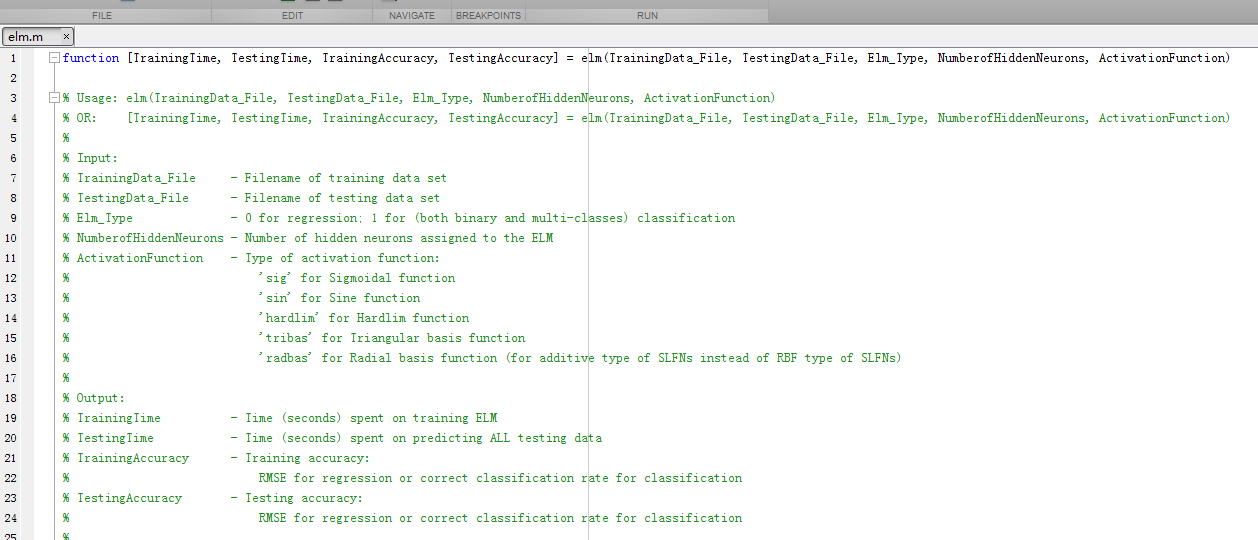
**然后，**

**我们想用K-均值聚类方法和极限学习机（神经网络）分类器处理数据**

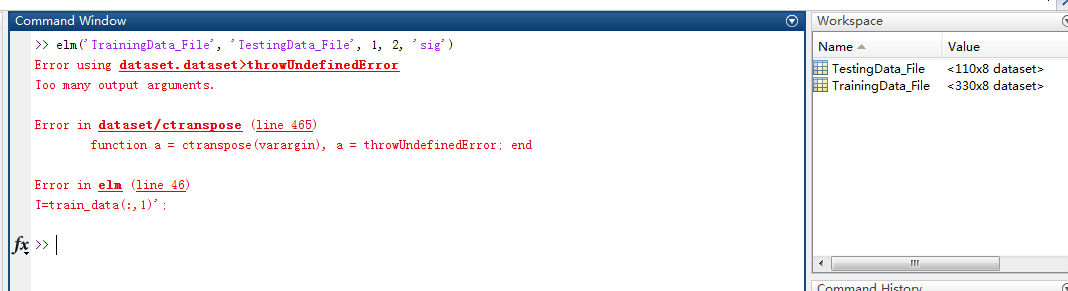


**K均值聚类程序代码在附件里，在运行时，由于属性过多，作图模糊不清，无法分析，随后想换个聚类方法。**

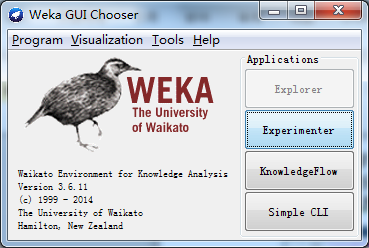




**elm（极限学习机（神经网络）分类器）程序代码在附件里，在运行时，首先写将原始数据分为训练集和测试集，但是由于训练集读取时总是报错，调整多次，无果，随后想换个分类方法。**



**然后，想到老师提到的Weka软件，想用上面的源代码做分类和聚类，随后就在网上下载了该软件，但是却无法运行，原因是需要Java运行环境，随后下载，终于可以运行Weka了**



**于是，我们用Weka上的k-均值聚类源代码处理数据，**

**达到结果如下：**

kMeans

======

Number of iterations: 2

Within cluster sum of squared errors: 85.1169283376534

Missing values globally replaced with mean/mode

Cluster centroids:

Cluster#

Attribute Full Data 0 1

(440) (298) (142)

===================================================

Channel 1.3227 1 2

Region 2.5432 2.5101 2.6127

Fresh 12000.2977 13475.5604 8904.3239

Milk 5796.2659 3451.7248 10716.5

Grocery 7951.2773 3962.1376 16322.8521

Frozen 3071.9318 3748.2517 1652.6127

Detergents\_Paper 2881.4932 790.5604 7269.507

Delicassen 1524.8705 1415.9564 1753.4366

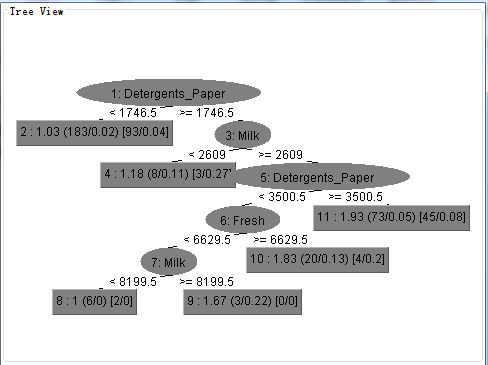
Time taken to build model (full training data) : 0.03 seconds

=== Model and evaluation on training set ===

Clustered Instances

0 298 ( 68%)

1 142 ( 32%)

**然后，我们又用Weka上的决策树源代码处理我们的数据，以销售渠道（零销、批发售给餐饮业）为目标类，其余7个属性做决策树分层，达到结果，但是观察各属性数据结构，Weka源代码自动舍弃了一些属性，比如地域属性。决策树分类器源代码在附件里，运行结果如下**：

**分类结果分析：**

**目标类有两个，**

**当Detergents\_paper<1746.5时，**

**当Detergents\_paper>=1746.5，milk<2609时，**

**当Detergents\_paper>1746.5，milk<8199.5,fresh<3500.5时，**

**目标类为零售渠道；**

**其余情况时，**

**目标类为批发售给餐饮业。**

**分类器参数如下：**

**相关系数0.8849**

**平均绝对误差0.0948**

**均方根误差0.2177**

**相对绝对误差21.6923％**

**根相对误差平方和46.575％**