## 大作业IV：MATLAB部分第二次大作业

使用MATLAB实现数据非监督聚类算法

**1. K-means算法的基本原理**

K-means算法也被称为K均值算法，是最为常见的聚类算法之一。这里的K为一个常数，代表欲聚类的数量，可由用户指定。K-means是一个非监督的聚类过程(即在类别信息的引导下完成)，将未标注的数据进行聚类。在聚类过程中，利用样本间的距离作为指标完成划分操作，这里，可采用基本的欧式距离完成距离测算。

该算法的执行步骤如下：

1. 选取*K*个点做为初始聚集的簇心（也可选择非样本点）；
2. 分别计算每个样本点到*K*个簇核心的距离（这里可采用欧氏距离），找到离该点最近的簇核心，将它归属到对应的簇；
3. 所有点都归属到簇之后，*M*个点就分为了*K*个簇。之后重新计算每个簇的重心（平均距离中心），将其定为新的“簇核心”；
4. 反复迭代2 - 3步骤，直到达到某个中止条件(可选的条件是簇的中心变化小于某个值*ε*)。

**2.本次作业题目：**

请同学们读取“sonar.xls”文件中的数据，这里可将参数*K*设定为2；完成K-means聚类过程（算法过程请自己编写函数实现）。记录迭代过程中分类正确率，可采用图片表示变化曲线。给出最终的中心点、涉及的代码，以及自己对于该问题的心得体会。



(注：最后一列为数据的原始列表，用于最后测算分类正确率时使用)。