# 模式识别实验报告

## 实验一 K-均值聚类

学院：计算机科学与技术学院

姓名：李星晨

学号：16S003084

1. **实验内容**
2. 使用C或Matlab编程实现K-均值聚类算法：要求独立完成算法编程，禁止调用已有函数库或工具箱中的函数；
3. 使用仿真数据测试算法的正确性：将下列19个样本聚成2个聚类：



1. MNIST数据集测试：ClusterSamples中的10000个784维特征手写数字样本聚类为10个类别，根据SampleLabels中的标签统计每个聚类中不同样本的数量。测试不同初始值对聚类结果的影响。
2. **程序代码**

（K-均值算法部分代码）

while 1

pre\_u = u;

quan = zeros(m,k);

for i=1:m

c=[];

for j=1:k

c=[c norm(date(i,:) - u(j,:))];

end

[value index] = min(c);

quan(i,index) = 1;

end

for i=1:k

for j=1:n

u(i,j)=sum(quan(:,i).\*date(:,j))/sum(quan(:,i));

end

end

if norm(u - pre\_u)<0.1

break;

end

end

re =[];

for i=1:m

tmp=[];

for j=1:k

tmp=[tmp norm(date(i,:)-u(j,:))];

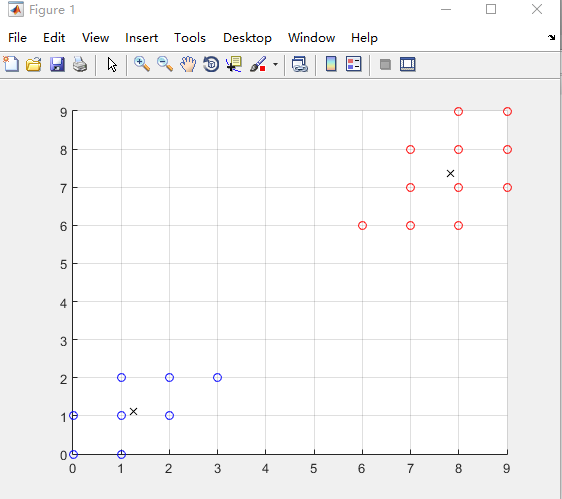
end

[value index] = min(tmp);

re = [re;date(i,:) index];

end

1. **实验结果**
2. 仿真数据实验结果：（可以列出每个聚类中包含的样本，也可以画图显示不同聚类）



1. MNIST数据集实验结果：

**每个聚类中包含不同类别样本数量统计表**

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 聚类0 | 5 | 425 | 54 | 15 | 68 | 179 | 49 | 67 | 101 | 28 |
| 聚类1 | 1 | 1 | 32 | 31 | 492 | 42 | 15 | 290 | 32 | 468 |
| 聚类2 | 22 | 3 | 44 | 604 | 0 | 304 | 7 | 1 | 190 | 12 |
| 聚类3 | 2 | 2 | 726 | 24 | 5 | 3 | 20 | 5 | 13 | 3 |
| 聚类4 | 452 | 0 | 1 | 2 | 2 | 7 | 14 | 1 | 4 | 7 |
| 聚类5 | 0 | 691 | 63 | 56 | 28 | 14 | 33 | 52 | 45 | 34 |
| 聚类6 | 420 | 0 | 17 | 20 | 1 | 35 | 27 | 1 | 4 | 0 |
| 聚类7 | 31 | 1 | 20 | 220 | 5 | 285 | 16 | 1 | 540 | 15 |
| 聚类8 | 25 | 0 | 33 | 11 | 32 | 19 | 798 | 3 | 14 | 1 |
| 聚类9 | 2 | 3 | 4 | 7 | 326 | 53 | 0 | 637 | 36 | 446 |