# 模式识别实验报告

## 实验一 K-均值聚类

学院：计算学部

姓名： 张亚博

学号：1170300521

1. **实验内容**
2. 使用python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）

def K\_means(data, centroids, K):

clusters = [[] for i in range(K)]

for i, d in enumerate(data):

dists = []

for c in centroids:

dists.append(distance(c - d))

dists = np.array(dists)

idx = np.argsort(dists)[0]

clusters[idx].append(i)

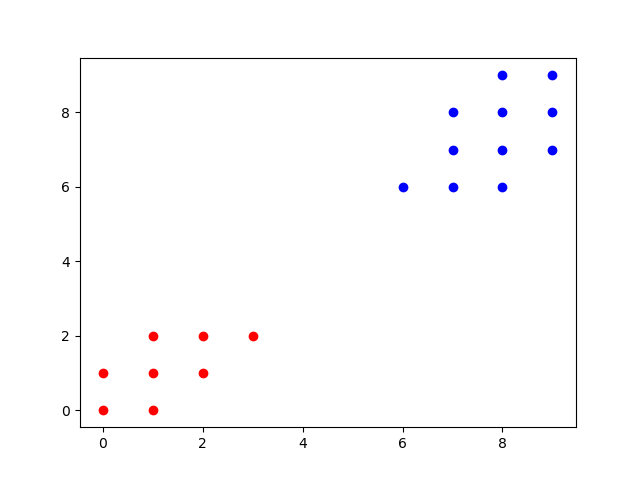
centroids = []

for c in clusters:

centroids.append(data[c].mean(axis=0))

return clusters, np.array(centroids)

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 | 453 | 0 | 2 | 2 | 1 | 6 | 12 | 1 | 4 | 5 |
| 聚类1 | 425 | 0 | 22 | 23 | 1 | 36 | 27 | 0 | 4 | 0 |
| 聚类2 | 19 | 1 | 40 | 202 | 2 | 225 | 17 | 1 | 552 | 12 |
| 聚类3 | 0 | 1 | 23 | 23 | 448 | 41 | 14 | 166 | 33 | 388 |
| 聚类4 | 0 | 471 | 110 | 5 | 4 | 7 | 9 | 26 | 47 | 3 |
| 聚类5 | 21 | 2 | 642 | 18 | 30 | 21 | 58 | 41 | 48 | 24 |
| 聚类6 | 0 | 645 | 60 | 58 | 30 | 21 | 58 | 41 | 48 | 24 |
| 聚类7 | 26 | 3 | 66 | 624 | 0 | 312 | 15 | 1 | 177 | 13 |
| 聚类8 | 13 | 0 | 21 | 24 | 228 | 245 | 8 | 251 | 82 | 187 |
| 聚类9 | 3 | 3 | 8 | 11 | 213 | 20 | 0 | 571 | 14 | 379 |