# 数学建模大作业笔记

## K 均值

设样本坐标集合为 $D=\{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n\}$，对应的权重为该建筑中人口数 $\{w_1,\cdots,w_n\}$，初始 $k$ 个质心位置随机从 $n$ 个点坐标中选取记为 $\{\boldsymbol{\mu}_1,\cdots,\boldsymbol{\mu}_k\}$.

**第一步**：计算每个样本点到最近的质心编号，对于样本点 $\boldsymbol{x}_j$ 其所属的质心为

$$\lambda_j=\mathop{\arg\min}_{1\leqslant i\leqslant k}||\boldsymbol{x}_j-\boldsymbol{\mu}_i||_2,$$

**第二步**：将其划分到该质心的集合中

$$C_{\lambda_j}\leftarrow C_{\lambda_j}\cup\{j\},$$

**第三步**：计算新质心向量

$$\boldsymbol{\mu}_i'=\frac{1}{\sum\limits_{j\in C_i}w_j}\sum_{j\in C_i}w_j\boldsymbol{x}_j.$$

若 $\max\limits_{1\leqslant i\leqslant k}||\boldsymbol{\mu}_i-\boldsymbol{\mu}_i'||_2<10^{-5}$ 则停止算法，否则更新 $\mu_i\leftarrow\mu_i'$, $(1\leqslant i\leqslant k)$，回到第一步继续迭代.

## 打分模型

对于 K 均值得到的质心 $\{\boldsymbol{\mu}_1,\cdots,\boldsymbol{\mu}_k\}$，如下进行评分

$$score=\frac{\dfrac{2}{k(k-1)}\left(\sum\limits_{1\leqslant i<j\leqslant k}||\boldsymbol{\mu}_i-\boldsymbol{\mu}_j||_2\right)}{\sum\limits_{1\leqslant i\leqslant k}\dfrac{1}{||C_i||}\sum\limits_{j\in C_i}\dfrac{||\boldsymbol{x}_i-\boldsymbol{\mu}_i||_2}{w_j}}$$

分子表示质心两两之间的平均距离尽可能大；分母表示每个质心包含的样本距离应尽可能近，并且人口数量越多的所具有的权重更高，即对每个质心所包含的的样本距离进行加权平均，最后再对所有质心进行求和.

当 $k=3$ 时，执行 2000 次 K 均值算法得到稳定的最优解，三个质心分别为：
$(23,22),(44,55),(67,25)$（对坐标进行四舍五入）.

当 $k=6$ 时，执行 2000 次 K 均值算法得到稳定的最优解，六个质心分别为：
$(13,32),(37,8),(38,57),(63,41),(64,11),(75,61)$（对坐标进行四舍五入）.

## 压缩监测方案

设混检比例为 $k:1$，人群中感染新冠的概率为 $p$，则一组中无阳性的概率为 $(1-p)^k$，若存在至少一个阳性的概率为 $1-(1-p)^k$，该组检测出阳性，则其中每个人进行单人单测. 设每个

人检测的次数为 $X$，则

$$P(X=\frac{1}{k})=(1-p)^k, \quad P(1+\frac{1}{k})=1-(1-p)^k,$$

于是 $\mathbb{E}(X)=\frac{1}{k}(1-p)^k+\left(1+\frac{1}{k}\right)(1-(1-p)^k)=1+\frac{1}{k}-(1-p)^k$.

问题转化为给定 $p$，求解 $\mathbb{E}(X)$ 的最小值，该最小值无解析解，通过近似计算可得.

## 第三问计算可转移采样点个数

$$k=\left[\frac{28\times(N_1+N_2)}{12\times 3600-20*60}\right]+1$$