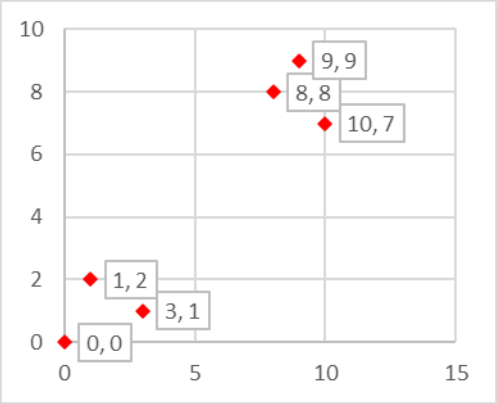
C-均值聚类手工计算作业

- C-均值聚类：
  - 聚类个数：2个
  - 初始聚类中心固定为：P1和P2
  - 迭代：
    - 计算到类中心的距离
    - 聚类结果
    - 更新聚类中心

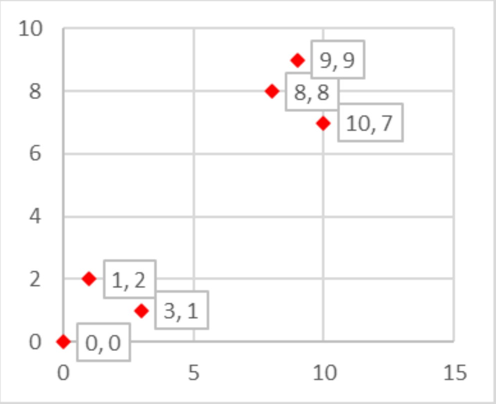
	X	Y
P1	0	0
P2	1	2
P3	3	1
P4	8	8
P5	9	9
P6	10	7



C-均值聚类手工计算作业答案

- C-均值聚类：
  - 聚类个数：2个
  - 初始聚类中心固定为：P1和P2
  - 迭代：
    - 计算到类中心的距离
    - 聚类结果
    - 更新聚类中心

	X	Y
P1	0	0
P2	1	2
P3	3	1
P4	8	8
P5	9	9
P6	10	7



	第一轮		第二轮		第三轮	
初始化聚类中心	C1	C2	C1	C2	C1	C2
	(0,0)	(1,2)	(0,0)	(6.2,5.4)	(1.33,1)	(9,8)
	C1	C2	C1	C2	C1	C2
P1	0.00	2.24	0.00	8.22	1.66	12.04
P2	2.24	0.00	2.24	6.21	1.05	10.00
P3	3.16	2.24	3.16	5.44	1.67	9.22
P4	11.31	9.22	11.31	3.16	9.67	1.00
P5	12.73	10.63	12.73	4.56	11.08	1.00
P6	12.21	10.30	12.21	4.12	10.54	1.41
聚类结果	P1	P2,P3,P4,P5,P6	P1,P2,P3	P4,P5,P6	P1,P2,P3	P4,P5,P6
更新聚类中心	(0,0)	(6.2,5.4)	(1.33,1)	(9,8)	(1.33,1)	(9,8)

描峰法计算

k	x	y
1	0.36	0.85
2	0.65	0.89
3	0.62	0.55
4	0.50	0.75
5	0.35	1.00

根据描峰法，对表格中的点的信息，计算样本点k=1的ρ_k值，其中 σ为1。

$$\rho_k = \rho(x_k) = \frac{1}{N}\sum_{l=1}^{N} \exp(-\frac{\|x_k - x_l\|^2}{2\sigma^2})$$

答案

k	x	y	L2范数的平方	指数函数值
1.000	0.360	0.850		
2.000	0.650	0.890	0.086	0.958
3.000	0.620	0.550	0.158	0.924
4.000	0.500	0.750	0.030	0.985
5.000	0.350	1.000	0.023	0.989
平均				0.964

求出结果：$\rho_{1=0.964}$

论述题

- 论述C均值算法和模糊C均值算法的区别，给出模糊C均值算法的问题描述、求解过程。

- 解：C均值算法为硬划分聚类算法，一个样本只能属于一个类，而模糊C均值算法则允许一个样本以概率的形式属于多类。

- 对于模糊C均值算法，假定样本集为$\{x_1, \ldots, x_N\}$，给定类别数目C和加权隶属度指标m，我们需要找到类别中心$X_1, \ldots, X_C$以及加权隶属度$u_{i,k} \geq 0$使得下述目标函数达到最小

- $J_{\mathrm{FCM}} = \sum_{i=1}^{C} \sum_{k=1}^{N} u_{i,k}^{m} ||x_k - X_i||^2$,

- 其中$\sum_{i=1}^{C} u_{i,k}$=1对于任何指标k都成立。

- 上述优化问题的拉格朗日函数是
- $L(U, X_1, \dots, X_C, \lambda_1, \dots, \lambda_N) = \sum_{i=1}^{C} \sum_{k=1}^{N} u_{ik}^m ||x_k - \underline{X_i}||^2 + \sum_{k=1}^{N} \lambda_k \left(\sum_{i=1}^{C} u_{i,k} - 1\right)$。

- 上述函数对求导可知
- $X_i = \frac{\sum_{k=1}^{N} u_{ik}^m x_k}{\sum_{k=1}^{N} u_{ik}^m}$ 以及 $u_{i,k} = 1/\sum_{l=1}^{C} \left(\frac{||x_k - X_i||^2}{||x_k - X_l||^2}\right)^{\frac{1}{m-1}}$。

- 由此可设计迭代算法如下：
- 输入: 数据集X、初始划分U^0、模糊指数m、迭代次数t、收敛阈值ϵ、聚类个数C
- 输出: 聚类结果
- 迭代过程：
  - （1）令迭代次数t =1
  - （2）用划分矩阵U^{t-1}更新$X^t = \{X_1{}^t, \ldots, X_C{}^t\}$
  - （3）用聚类中心X^t更新划分矩阵U^t
- （4）重复步骤（2）和（3）直到$\|U^t - U^{t-1}\| \leq \epsilon$, 输出聚类结果。