

Kmeans聚类过程实例推导(附必要公式) 【选取K=3】

下面数据是对一批电池产品两种条件测试结果
“黑”良品，“一白一黑”次品，“白”坏品

白		数据重组	
421.95	446.34	代号	Xi Yi
448.78	470.73		
419.51	434.15	C1	421.95 446.34
480.49	503.66	C2	348.78 348.78
431.71	446.34	C3	348.78 351.22
一白一黑		C4	431.71 443.9
431.71	443.9	C5	448.78 470.73
424.39	435.37	C6	424.39 435.37
424.39	437.8	C7	424.39 437.8
431.71	442.68	C8	339.02 339.02
419.51	426.83	C9	419.51 434.15
黑		C10	480.49 503.66
348.78	348.78	C11	431.71 442.68
348.78	351.22	C12	351.22 351.22
339.02	339.02	C13	431.71 446.34
351.22	351.22	C14	356.1 347.56
356.1	347.56	C15	419.51 426.83

第1轮质心	C1	421.95	446.34	C2	348.78	348.78	C3	348.78	351.22
-------	----	--------	--------	----	--------	--------	----	--------	--------

数据代号		距离	$d_i = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2}$	最近点	辅助列
C1	0	121.95	120.01	C1	0
C2	121.95	0	2.439	C2	1
C3	120.01	2.439	0	C3	2
C4	10.056	126.19	124.37	C1	3
C5	36.259	157.71	155.83	C1	4
C6	11.243	114.95	113.13	C1	5
C7	8.8782	116.8	114.95	C1	6
C8	135.62	13.797	15.617	C2	7
C9	12.437	110.86	108.99	C1	8
C10	81.925	203.31	201.46	C1	9
C11	10.42	125.28	123.46	C1	10
C12	118.54	3.4493	2.439	C3	11
C13	9.7561	128.04	126.19	C1	12
C14	118.72	7.418	8.1807	C2	13
C15	19.664	105.33	103.54	C1	14
辅助行	C1	C2	C3		

第1轮
聚类结果

C1	421.95	421.95							
			C2	348.78	348.78				
						C3	348.78	348.78	
C4	431.71	431.71							
C5	448.78	448.78							
C6	424.39	424.39							
C7	424.39	424.39							
			C8	339.02	339.02				
C9	419.51	419.51							
C10	480.49	480.49							
C11	431.71	431.71							
						C12	351.22	351.22	
C13	431.71	431.71							
			C14	356.1	356.1				
C15	419.51	419.51							

分别对X,Y求平均值，得到新的质心坐标(有可能为虚拟坐标)

数据代号	距离			最近点	辅助列
C1	17.278	123.09	120.24	A1	0
C2	119.69	1.1498	1.7247	A2	1
C3	117.98	3.3521	1.7247	A3	2
C4	10.626	127.34	124.47	A1	3
C5	40.357	158.85	155.99	A1	4
C6	9.2329	116.1	113.23	A1	5
C7	10.036	117.95	115.08	A1	6
C8	133.49	12.647	15.522	A2	7
C9	13.922	112.01	109.14	A1	8
C10	84.558	204.45	201.59	A1	9
C11	9.4242	126.43	123.56	A1	10
C12	116.24	4.5991	1.7247	A3	11
C13	13.039	129.19	126.32	A1	12
C14	115.54	8.1402	6.5673	A3	13
C15	15.383	106.48	103.61	A1	14
辅助行	A1	A2	A3		

C1	421.95	446.34
C2	348.78	348.78
C3	348.78	351.22
C4	431.71	443.9
C5	448.78	470.73
C6	424.39	435.37
C7	424.39	437.8
C8	339.02	339.02
C9	419.51	434.15
C10	480.49	503.66
C11	431.71	442.68
C12	351.22	351.22
C13	431.71	446.34
C14	356.1	347.56
C15	419.51	426.83

第2轮 聚类结果

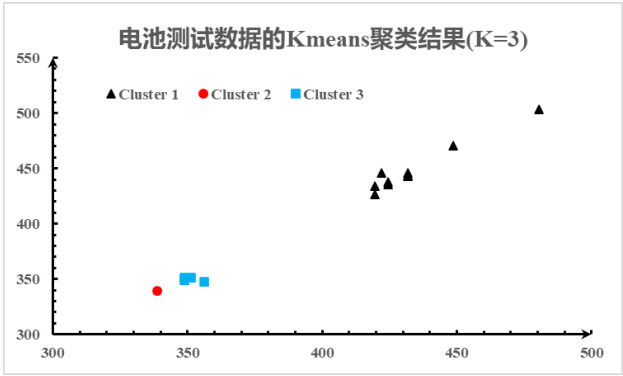
[illegible]

第4轮质心 A1 433.41 433.41 A2 339.02 339.02 A3 351.22 351.22

C1	421.95	446.34
C2	348.78	348.78
C3	348.78	351.22
C4	431.71	443.9
C5	448.78	470.73
C6	424.39	435.37
C7	424.39	437.8
C8	339.02	339.02
C9	419.51	434.15
C10	480.49	503.66
C11	431.71	442.68
C12	351.22	351.22
C13	431.71	446.34
C14	356.1	347.56
C15	419.51	426.83

数据代号	距离			最近点	辅助列
C1	17.278		135.62	118.54	A1 0
C2	119.69		13.797	3.4493	A3 1
C3	117.98		15.617	2.439	A3 2
C4	10.626		139.96	122.75	A1 3
C5	40.357		171.44	154.28	A1 4
C6	9.2329		128.72	111.51	A1 5
C7	10.036		130.56	113.36	A1 6
C8	133.49	0		17.247	A2 7
C9	13.922		124.61	107.43	A1 8
C10	84.558		217.06	199.87	A1 9
C11	9.4242		139.05	121.84	A1 10
C12	116.24		17.247	0	A3 11
C13	13.039		141.8	124.61	A1 12
C14	115.54		19.088	6.0976	A3 13
C15	15.383		119.11	101.89	A1 14
辅助行	A1		A2	A3	

第4轮
聚类结果



C1	421.95	421.95
C2	348.78	348.78
C3	348.78	348.78
C4	431.71	431.71
C5	448.78	448.78
C6	424.39	424.39
C7	424.39	424.39
C8	339.02	339.02
C9	419.51	419.51
C10	480.49	480.49
C11	431.71	431.71
C12	351.22	351.22
C13	431.71	431.71
C14	356.1	356.1
C15	419.51	419.51

第4轮聚类与第3轮聚类结果一致，停止聚类。
可见聚类结果将“次品”和“坏品”聚为一类，良品大多在一类中，少数分到另一类。
但这仍可以实现允许“过杀”的条件下，进行良品的筛选要求。