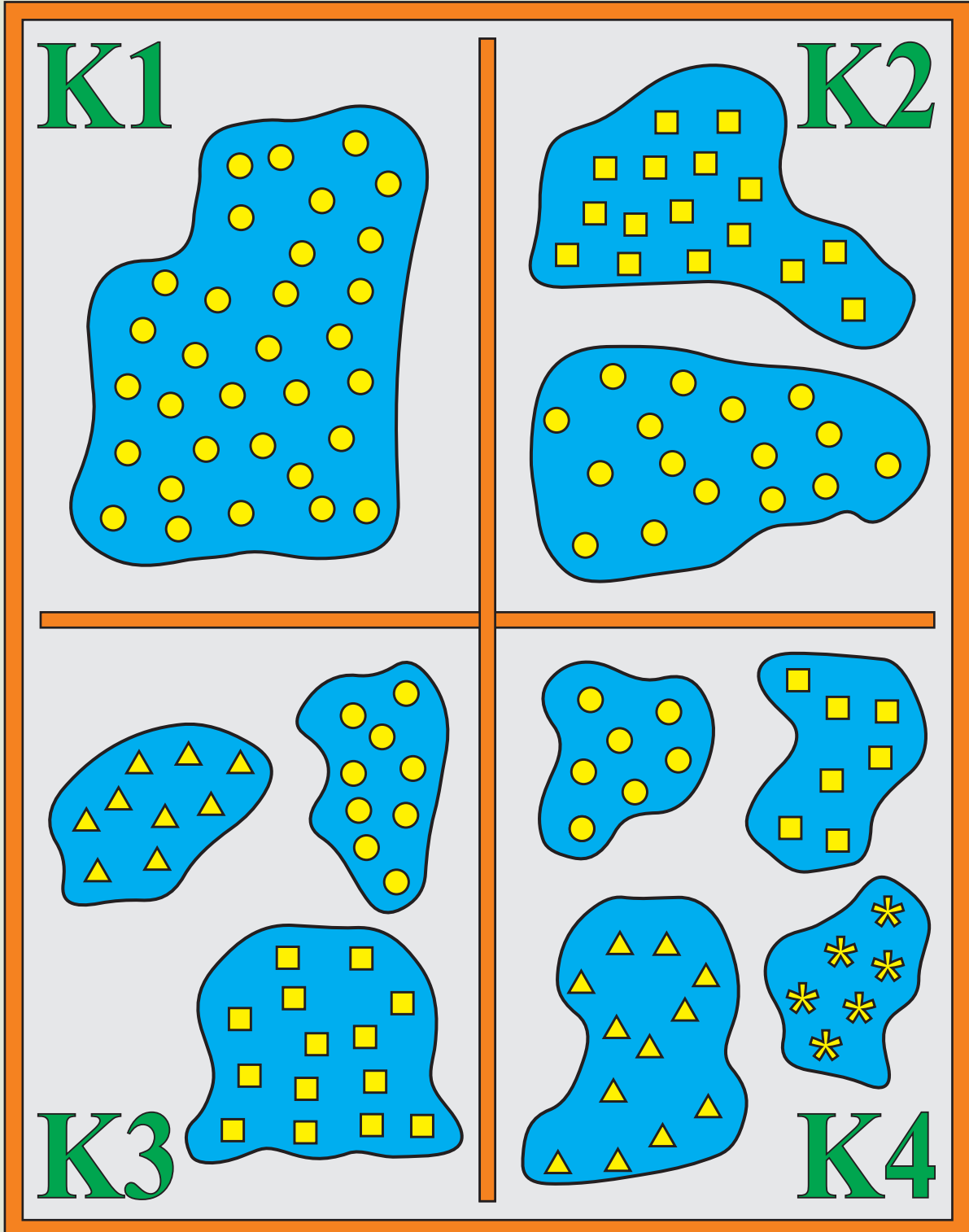




Kmeans de MacQueen



Methods for Classification and Analysis of Multivariate Observations

K

Proceedings of 5th Berkeley Symposium, 1967

M

James MacQueen



2.0 Exemplificando um Possível Resultado Classificatório

2.1 Análise Química da Composição de Sucos

Kmeans

Dada 36 x 1000 amostras de sucos (AMO) analisadas segundo suas concentrações médias dos elementos k, P, S, Ca, Cu, Mg, Na, R1, R2, R3:

K = Potássio

P = Fósforo

S = Enxofre

Ca = Cálcio

Cu = Cobre

Mg = Magnésio

Na = Sódio

R1 = Resíduo de Conservante

R2 = Resíduo de Corante

R3 = Resíduo de Amido

mg/l



Elementos Químicos Presentes nos Sucos (mg/l)

Faz-se pertinente do algoritmo KmeanS particionar as próximas 36 AMOstras em k clusters de dados numericamente homogêneas e similares.

Particionamento de Dados

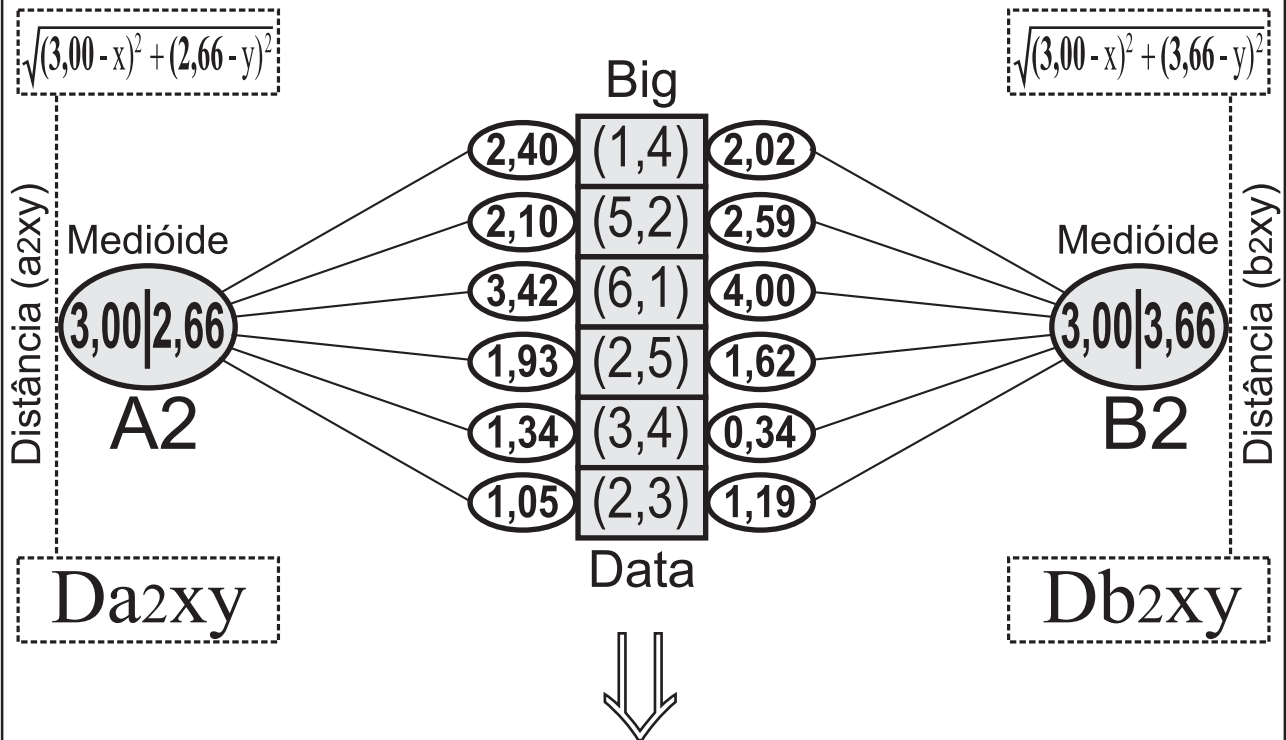


2.3 Amostras do Big Data Particionadas em Nove Clusters

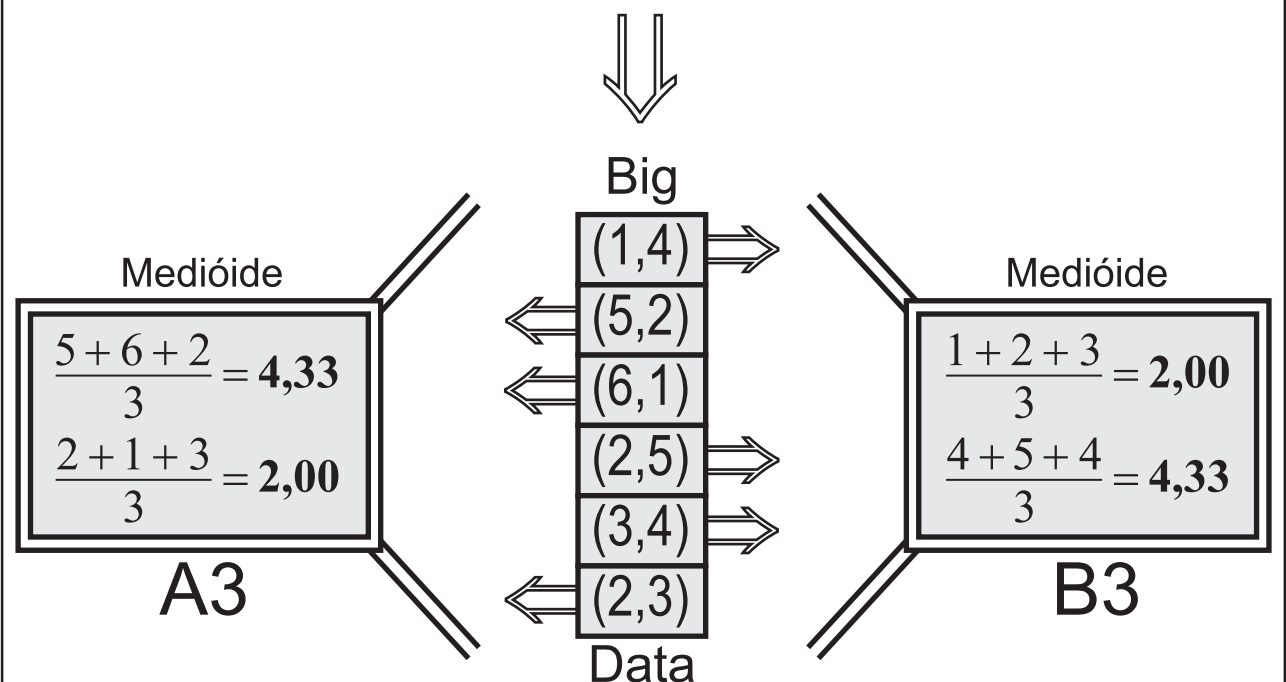
AMO	K	P	S	Ca	Cu	Mg	Na	R1	R2	R3
1	363,60	31,68	7,40	41,16	1,61	31,56	13,08	0,98	0,11	4,12
	368,40	34,20	3,95	37,80	1,56	28,32	14,64	0,95	0,13	4,22
	423,60	36,00	8,04	39,48	1,78	30,60	14,16	0,93	0,10	4,19
	392,30	35,71	6,47	38,72	1,65	29,36	13,84	0,95	0,12	4,20
2	246,00	65,16	33,96	53,16	0,78	33,60	30,00	0,80	0,21	4,82
	278,40	72,84	35,52	53,16	0,76	34,08	32,88	0,82	0,23	4,81
	241,20	76,80	33,96	52,44	0,71	33,00	29,40	0,85	0,22	4,91
	259,30	73,80	34,55	52,87	0,75	34,00	31,82	0,81	0,23	4,82
3	83,16	29,76	11,24	35,88	1,15	26,64	11,16	0,70	0,31	5,62
	87,60	36,48	12,36	35,88	1,37	27,96	12,60	0,71	0,34	5,67
	92,40	37,08	11,32	37,08	1,34	29,40	14,28	0,73	0,33	5,71
	88,35	32,97	12,12	35,70	1,36	27,82	12,58	0,70	0,32	5,59
4	644,40	60,36	18,72	60,84	1,42	44,40	19,92	0,61	0,44	6,89
	704,40	61,20	21,48	61,56	1,66	45,24	20,76	0,64	0,42	6,86
	752,40	64,80	18,07	66,00	1,52	45,60	22,56	0,63	0,41	6,80
	659,37	61,12	20,35	63,14	1,55	45,12	21,19	0,63	0,42	6,85
5	128,00	18,84	9,38	15,60	0,54	11,93	5,83	0,51	0,52	7,25
	135,60	21,96	12,72	19,92	0,56	14,76	7,51	0,52	0,54	7,26
	142,80	24,00	14,28	21,36	0,62	15,96	7,73	0,54	0,51	7,30
	137,91	20,47	11,95	19,82	0,56	13,91	6,87	0,52	0,53	7,12
6	112,00	28,08	19,20	24,24	0,65	19,92	36,24	0,41	0,61	8,80
	100,32	26,88	18,60	23,88	0,66	19,44	37,44	0,44	0,63	8,84
	112,56	30,24	22,32	26,40	0,67	21,60	41,76	0,42	0,61	8,83
	111,17	29,47	20,31	23,27	0,64	20,02	38,45	0,43	0,62	8,82
7	50,64	18,60	9,74	19,44	0,80	14,64	15,96	0,31	0,74	9,42
	53,04	17,40	9,43	19,32	0,73	14,40	16,80	0,33	0,72	9,51
	51,36	18,72	10,30	19,32	0,86	14,52	16,44	0,32	0,73	9,62
	52,07	17,80	10,01	19,30	0,79	14,20	15,99	0,32	0,72	9,48
8	157,00	30,96	99,00	23,28	0,40	12,96	55,20	0,02	1,03	15,12
	162,00	27,36	91,92	19,80	0,35	11,71	45,36	0,05	1,05	15,90
	172,60	36,72	92,18	25,68	0,41	14,64	59,52	0,03	1,06	15,72
	185,20	31,12	95,40	22,40	0,35	12,90	47,30	0,04	1,04	15,82
9	252,00	51,96	69,96	50,76	2,03	44,04	98,40	0,06	1,12	16,90
	229,20	52,80	66,84	48,36	2,17	43,56	91,56	0,07	1,13	16,85
	223,20	48,84	68,40	50,04	2,04	43,56	90,84	0,08	1,14	16,78
	230,00	52,12	67,16	49,18	2,21	42,50	90,51	0,06	1,12	16,80



4.2 Continua as iterações Bidimensionais para KmeanS 2:



O elemento (1,4) é próximo da medióide (B2),
 O elemento (5,2) é próximo da medióide (A2) ...,
 O elemento (2,3) é próximo da medióide (A2).



$DIS(A3) = 5,14$

$DIS(B3) = 2,77$



6.2 Elbow (Cotovelo) Data Chart e Seu K de Virada

