

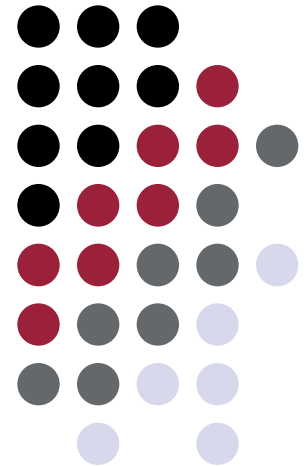
# Mineração de Dados: Agrupamento



Universidade Federal  
de Ouro Preto

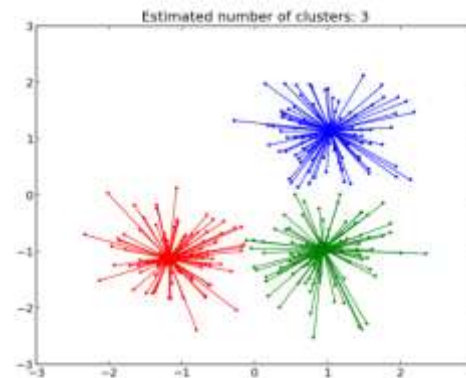
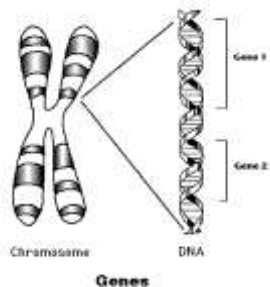
## CEA462 – Sistemas de Apoio à Decisão

Prof. MSc. George H. G. Fonseca  
Universidade Federal de Ouro Preto



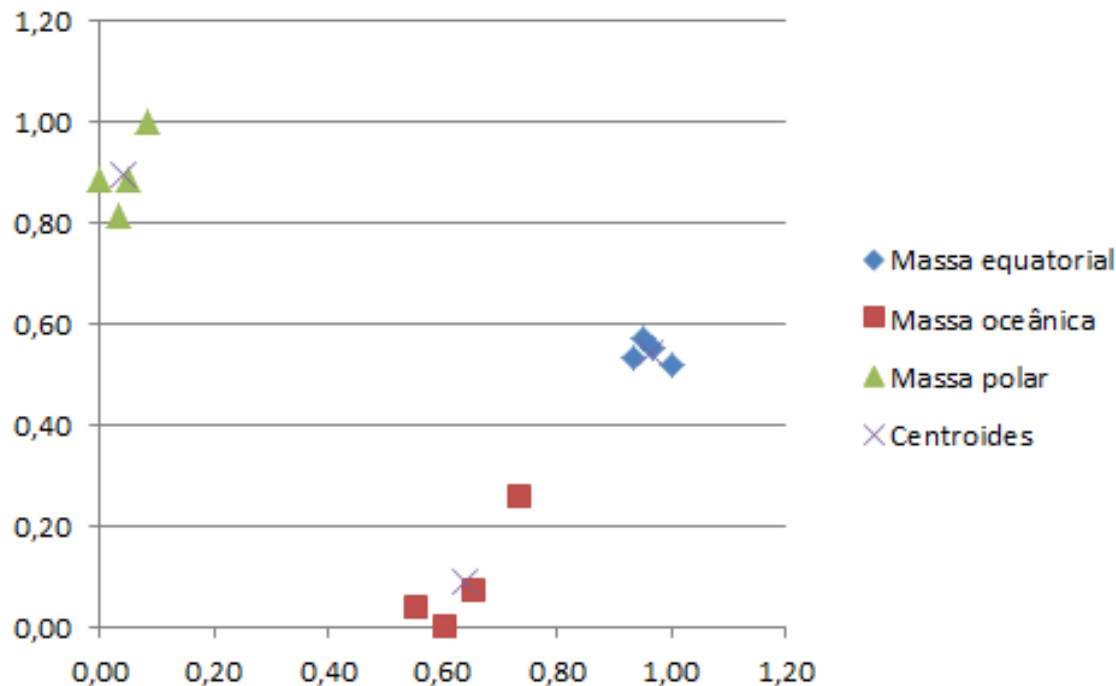


- Agrupamento é a tarefa de dividir os dados em grupos que tenham significado e/ou sejam úteis
- Exemplos de aplicações
  - Encontrar grupos de genes com funções semelhantes
  - Encontrar uma alocação de antenas de telecomunicação
  - Segmentar clientes para atividades de marketing





- Encontrar  $K$  centroides que melhor separam o conjunto de dados
  - $K$  é dado *a priori*



# Algoritmo K-Means



- Define-se K centroides iniciais
  - De forma aleatória
  - Abordagens mais elaboradas
- Repita
  - Atribuir cada ponto a seu centroide mais próximo
  - Recalcular o centroide de cada grupo
- Até que os centroides não mudem

# Algoritmo K-Means



- Como saber qual o centroide mais próximo?
  - Distância euclidiana

- $$d(x, y) = \sqrt{\sum_{a=1}^A (x_a - y_a)^2}$$

- Como saber se um agrupamento está bom?
  - Soma do erro quadrado

- $$SSE = \sum_{i=1}^K \sum_{x \in C_i} d(c_i, x)^2$$

Símb.	Descrição
x / y	Exemplo
a	Atributo
n	Número de atributos
C <sub>i</sub>	Grupo de índice i
c <sub>i</sub>	Centroide do grupo C <sub>i</sub>
m <sub>i</sub>	Núm. objetos no grupo de índice i
m	Núm. objetos no conjunto de dados
K	Núm. grupos

# Algoritmo K-Means



- Como recalcular o centroide?
  - Média dos pontos que representa

- $$c_i = \frac{\sum_{x \in C_i} x}{m_i}$$

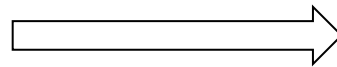
Símb.	Descrição
x / y	Exemplo
a	Atributo
n	Número de atributos
C <sub>i</sub>	Grupo de índice i
c <sub>i</sub>	Centroide do grupo C <sub>i</sub>
m <sub>i</sub>	Núm. objetos no grupo de índice i
m	Núm. objetos no conjunto de dados
K	Núm. grupos

# Algoritmo K-Means



- Exemplo
  - Agrupamento de correntes de ar
  - Normalização

Ponto	Longitude	Latitude
1	160	5
2	150	10
3	140	7
4	145	12
5	25	-60
6	40	-65
7	55	-55
8	80	-30
9	-130	55
10	-120	70
11	-135	45
12	-145	55



Ponto	Longitude	Latitude
1	1,00	0,52
2	0,97	0,56
3	0,93	0,53
4	0,95	0,57
5	0,56	0,04
6	0,61	0,00
7	0,66	0,07
8	0,74	0,26
9	0,05	0,89
10	0,08	1,00
11	0,03	0,81
12	0,00	0,89

# Algoritmo K-Means

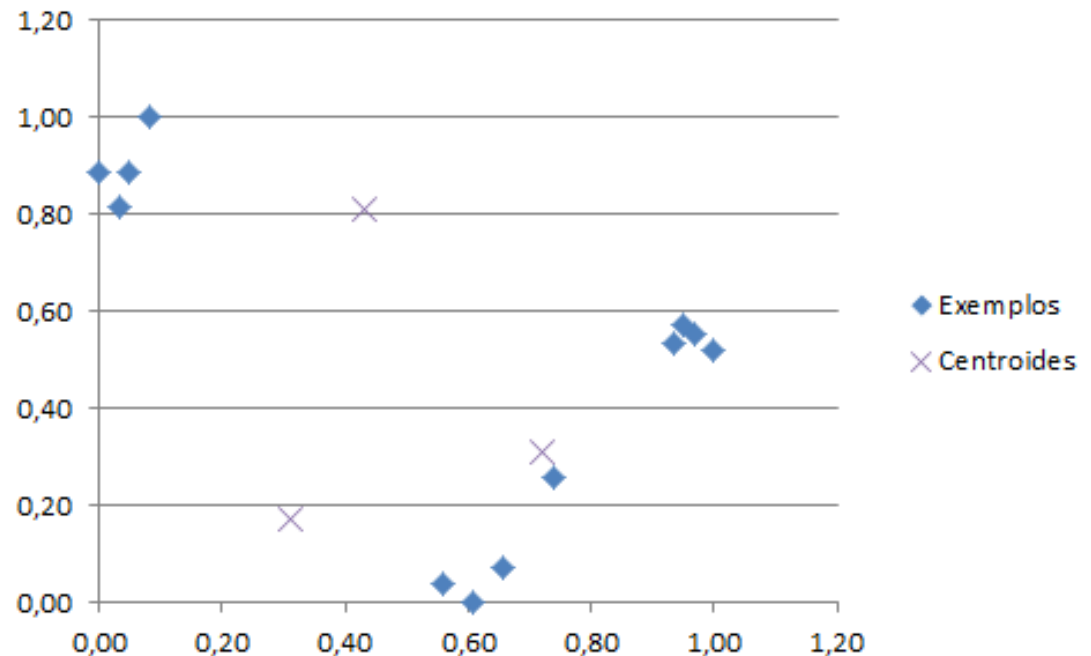


- Exemplo

- Geração de K centroides aleatórios (K = 3)

- $c_1 = (0,72; 0,31)$ ,  $c_2 = (0,43; 0,81)$ ,  $c_3 = (0,31; 0,17)$

Ponto	Longitude	Latitude
1	1,00	0,52
2	0,97	0,56
3	0,93	0,53
4	0,95	0,57
5	0,56	0,04
6	0,61	0,00
7	0,66	0,07
8	0,74	0,26
9	0,05	0,89
10	0,08	1,00
11	0,03	0,81
12	0,00	0,89





# Algoritmo K-Means



- Exemplo

- It.1: Atribuir cada ponto ao centroide mais próximo
  - $c_1 = (0,72; 0,31)$ ,  $c_2 = (0,43; 0,81)$ ,  $c_3 = (0,31; 0,17)$

Ponto	Longitude	Latitude	$d(x, c_1)$	$d(x, c_2)$	$d(x, c_3)$
1	1,00	0,52	<b>0,35</b>	0,64	0,77
2	0,97	0,56	<b>0,35</b>	0,59	0,76
3	0,93	0,53	<b>0,31</b>	0,58	0,72
4	0,95	0,57	<b>0,35</b>	0,57	0,76
5	0,56	0,04	0,32	0,78	<b>0,28</b>
6	0,61	0,00	<b>0,33</b>	0,83	0,34
7	0,66	0,07	<b>0,24</b>	0,77	0,36
8	0,74	0,26	<b>0,05</b>	0,63	0,44
9	0,05	0,89	0,89	<b>0,39</b>	0,76
10	0,08	1,00	0,94	<b>0,40</b>	0,86
11	0,03	0,81	0,85	<b>0,40</b>	0,70
12	0,00	0,89	0,92	<b>0,44</b>	0,78

# Algoritmo K-Means



- Exemplo

- It.1: Calcular custo

- $$SSE = \sum_{i=1}^K \sum_{x \in C_i} d(c_i, x)^2$$
$$= 0,35^2 + 0,35^2 + 0,31^2 + 0,35^2 + 0,28^2 + 0,33^2 + 0,24^2 + 0,05^2$$
$$+ 0,39^2 + 0,40^2 + 0,40^2 + 0,44^2$$
$$= 1,38$$

Ponto	Longitude	Latitude	d(x, c <sub>1</sub> )	d(x, c <sub>2</sub> )	d(x, c <sub>3</sub> )
1	1,00	0,52	<b>0,35</b>	0,64	0,77
2	0,97	0,56	<b>0,35</b>	0,59	0,76
3	0,93	0,53	<b>0,31</b>	0,58	0,72
4	0,95	0,57	<b>0,35</b>	0,57	0,76
5	0,56	0,04	0,32	0,78	<b>0,28</b>
6	0,61	0,00	<b>0,33</b>	0,83	0,34
7	0,66	0,07	<b>0,24</b>	0,77	0,36
8	0,74	0,26	<b>0,05</b>	0,63	0,44
9	0,05	0,89	0,89	<b>0,39</b>	0,76
10	0,08	1,00	0,94	<b>0,40</b>	0,86
11	0,03	0,81	0,85	<b>0,40</b>	0,70
12	0,00	0,89	0,92	<b>0,44</b>	0,78

# Algoritmo K-Means



- Exemplo

- It.1: Recalcular os centroides

- $$c_{1long} = \frac{\sum_{x \in C_i} x}{m_i} = \frac{(1,00+0,97+0,93+0,95+0,61+0,66+0,74)}{7} = 0,84$$

- $$c_{1lat} = \frac{\sum_{x \in C_i} x}{m_i} = \frac{(0,52+0,56+0,53+0,57+0,00+0,07+0,26)}{7} = 0,36$$

Ponto	Longitude	Latitude	d(x, c <sub>1</sub> )	d(x, c <sub>2</sub> )	d(x, c <sub>3</sub> )
1	1,00	0,52	<b>0,35</b>	0,64	0,77
2	0,97	0,56	<b>0,35</b>	0,59	0,76
3	0,93	0,53	<b>0,31</b>	0,58	0,72
4	0,95	0,57	<b>0,35</b>	0,57	0,76
5	0,56	0,04	0,32	0,78	<b>0,28</b>
6	0,61	0,00	<b>0,33</b>	0,83	0,34
7	0,66	0,07	<b>0,24</b>	0,77	0,36
8	0,74	0,26	<b>0,05</b>	0,63	0,44
9	0,05	0,89	0,89	<b>0,39</b>	0,76
10	0,08	1,00	0,94	<b>0,40</b>	0,86
11	0,03	0,81	0,85	<b>0,40</b>	0,70
12	0,00	0,89	0,92	<b>0,44</b>	0,78

# Algoritmo K-Means



- Exemplo

- It.1: Recalcular os centroides

- $$c_{2long} = \frac{\sum_{x \in C_i} x}{m_i} = \frac{(0,05+0,08+0,03+0,00)}{4} = 0,04$$

- $$c_{2lat} = \frac{\sum_{x \in C_i} x}{m_i} = \frac{(0,89+1,00+0,81+0,89)}{4} = 0,90$$

Ponto	Longitude	Latitude	d(x, c <sub>1</sub> )	d(x, c <sub>2</sub> )	d(x, c <sub>3</sub> )
1	1,00	0,52	<b>0,35</b>	0,64	0,77
2	0,97	0,56	<b>0,35</b>	0,59	0,76
3	0,93	0,53	<b>0,31</b>	0,58	0,72
4	0,95	0,57	<b>0,35</b>	0,57	0,76
5	0,56	0,04	0,32	0,78	<b>0,28</b>
6	0,61	0,00	<b>0,33</b>	0,83	0,34
7	0,66	0,07	<b>0,24</b>	0,77	0,36
8	0,74	0,26	<b>0,05</b>	0,63	0,44
9	0,05	0,89	0,89	<b>0,39</b>	0,76
10	0,08	1,00	0,94	<b>0,40</b>	0,86
11	0,03	0,81	0,85	<b>0,40</b>	0,70
12	0,00	0,89	0,92	<b>0,44</b>	0,78

# Algoritmo K-Means



- Exemplo

- It.1: Recalcular os centroides

- $c_{3long} = \frac{\sum_{x \in C_i} x}{m_i} = \frac{(0,56)}{1} = 0,56$

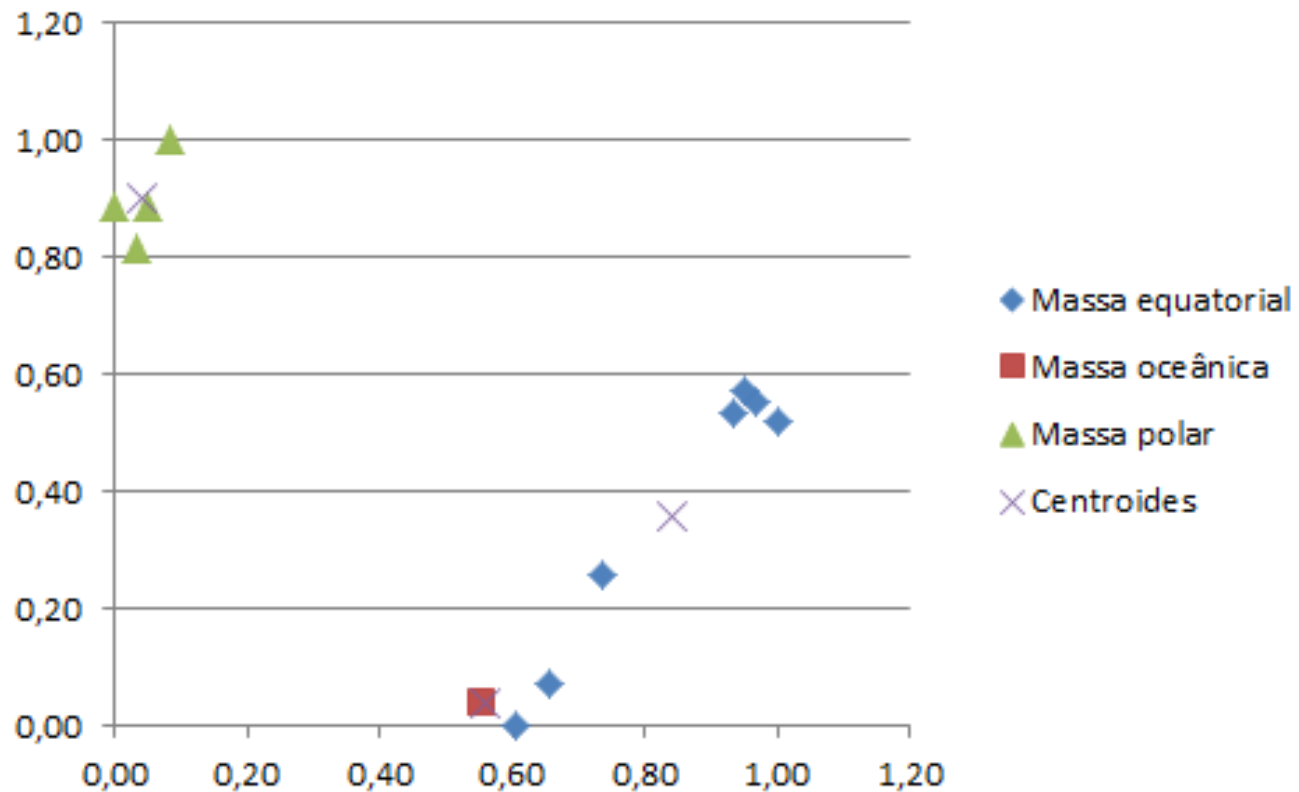
- $c_{3lat} = \frac{\sum_{x \in C_i} x}{m_i} = \frac{(0,04)}{1} = 0,04$

Ponto	Longitude	Latitude	d(x, c <sub>1</sub> )	d(x, c <sub>2</sub> )	d(x, c <sub>3</sub> )
1	1,00	0,52	<b>0,35</b>	0,64	0,77
2	0,97	0,56	<b>0,35</b>	0,59	0,76
3	0,93	0,53	<b>0,31</b>	0,58	0,72
4	0,95	0,57	<b>0,35</b>	0,57	0,76
5	0,56	0,04	0,32	0,78	<b>0,28</b>
6	0,61	0,00	<b>0,33</b>	0,83	0,34
7	0,66	0,07	<b>0,24</b>	0,77	0,36
8	0,74	0,26	<b>0,05</b>	0,63	0,44
9	0,05	0,89	0,89	<b>0,39</b>	0,76
10	0,08	1,00	0,94	<b>0,40</b>	0,86
11	0,03	0,81	0,85	<b>0,40</b>	0,70
12	0,00	0,89	0,92	<b>0,44</b>	0,78

# Algoritmo K-Means



- Exemplo
  - It.1: Representação gráfica



# Algoritmo K-Means



- Exemplo

- It.2: Atribuir cada ponto ao centroide mais próximo
  - $c_1 = (0,84; 0,36)$ ,  $c_2 = (0,04; 0,90)$ ,  $c_3 = (0,56; 0,04)$

Ponto	Longitude	Latitude	$d(x, c_1)$	$d(x, c_2)$	$d(x, c_3)$
1	1,00	0,52	<b>0,23</b>	1,03	0,65
2	0,97	0,56	<b>0,23</b>	0,99	0,66
3	0,93	0,53	<b>0,20</b>	0,97	0,62
4	0,95	0,57	<b>0,24</b>	0,97	0,66
5	0,56	0,04	0,43	1,01	<b>0,00</b>
6	0,61	0,00	0,43	1,06	<b>0,06</b>
7	0,66	0,07	0,34	1,03	<b>0,10</b>
8	0,74	0,26	<b>0,14</b>	0,95	0,28
9	0,05	0,89	0,95	<b>0,01</b>	0,99
10	0,08	1,00	0,99	<b>0,11</b>	1,07
11	0,03	0,81	0,93	<b>0,09</b>	0,94
12	0,00	0,89	0,99	<b>0,04</b>	1,02

# Algoritmo K-Means



- Exemplo

- It.2: Calcular custo

- $$SSE = \sum_{i=1}^K \sum_{x \in C_i} d(c_i, x)^2$$
$$= 0,23^2 + 0,23^2 + 0,20^2 + 0,24^2 + 0,00^2 + 0,06^2 + 0,10^2 + 0,14^2$$
$$+ 0,01^2 + 0,11^2 + 0,09^2 + 0,04^2$$
$$= 0,26$$

Ponto	Longitude	Latitude	d(x, c <sub>1</sub> )	d(x, c <sub>2</sub> )	d(x, c <sub>3</sub> )
1	1,00	0,52	<b>0,23</b>	1,03	0,65
2	0,97	0,56	<b>0,23</b>	0,99	0,66
3	0,93	0,53	<b>0,20</b>	0,97	0,62
4	0,95	0,57	<b>0,24</b>	0,97	0,66
5	0,56	0,04	0,43	1,01	<b>0,00</b>
6	0,61	0,00	0,43	1,06	<b>0,06</b>
7	0,66	0,07	0,34	1,03	<b>0,10</b>
8	0,74	0,26	<b>0,14</b>	0,95	0,28
9	0,05	0,89	0,95	<b>0,01</b>	0,99
10	0,08	1,00	0,99	<b>0,11</b>	1,07
11	0,03	0,81	0,93	<b>0,09</b>	0,94
12	0,00	0,89	0,99	<b>0,04</b>	1,02



# Algoritmo K-Means



- Exemplo

- It.2: Recalcular os centroides

- $$c_{1long} = \frac{\sum_{x \in C_i} x}{m_i} = \frac{(1,00+0,97+0,93+0,95+0,74)}{5} = 0,92$$

- $$c_{1lat} = \frac{\sum_{x \in C_i} x}{m_i} = \frac{(0,52+0,56+0,53+0,57+0,26)}{5} = 0,49$$

Ponto	Longitude	Latitude	d(x, c <sub>1</sub> )	d(x, c <sub>2</sub> )	d(x, c <sub>3</sub> )
1	1,00	0,52	<b>0,23</b>	1,03	0,65
2	0,97	0,56	<b>0,23</b>	0,99	0,66
3	0,93	0,53	<b>0,20</b>	0,97	0,62
4	0,95	0,57	<b>0,24</b>	0,97	0,66
5	0,56	0,04	0,43	1,01	<b>0,00</b>
6	0,61	0,00	0,43	1,06	<b>0,06</b>
7	0,66	0,07	0,34	1,03	<b>0,10</b>
8	0,74	0,26	<b>0,14</b>	0,95	0,28
9	0,05	0,89	0,95	<b>0,01</b>	0,99
10	0,08	1,00	0,99	<b>0,11</b>	1,07
11	0,03	0,81	0,93	<b>0,09</b>	0,94
12	0,00	0,89	0,99	<b>0,04</b>	1,02

# Algoritmo K-Means



- Exemplo

- It.2: Recalcular os centroides

- $$c_{2long} = \frac{\sum_{x \in C_i} x}{m_i} = \frac{(0,01+0,11+0,09+0,04)}{4} = 0,04$$

- $$c_{2lat} = \frac{\sum_{x \in C_i} x}{m_i} = \frac{(0,89+1,00+0,81+0,89)}{4} = 0,90$$

Ponto	Longitude	Latitude	d(x, c <sub>1</sub> )	d(x, c <sub>2</sub> )	d(x, c <sub>3</sub> )
1	1,00	0,52	<b>0,23</b>	1,03	0,65
2	0,97	0,56	<b>0,23</b>	0,99	0,66
3	0,93	0,53	<b>0,20</b>	0,97	0,62
4	0,95	0,57	<b>0,24</b>	0,97	0,66
5	0,56	0,04	0,43	1,01	<b>0,00</b>
6	0,61	0,00	0,43	1,06	<b>0,06</b>
7	0,66	0,07	0,34	1,03	<b>0,10</b>
8	0,74	0,26	<b>0,14</b>	0,95	0,28
9	0,05	0,89	0,95	<b>0,01</b>	0,99
10	0,08	1,00	0,99	<b>0,11</b>	1,07
11	0,03	0,81	0,93	<b>0,09</b>	0,94
12	0,00	0,89	0,99	<b>0,04</b>	1,02

# Algoritmo K-Means



- Exemplo

- It.2: Recalcular os centroides

- $$c_{3long} = \frac{\sum_{x \in C_i} x}{m_i} = \frac{(0,56+0,61+0,66)}{3} = 0,61$$

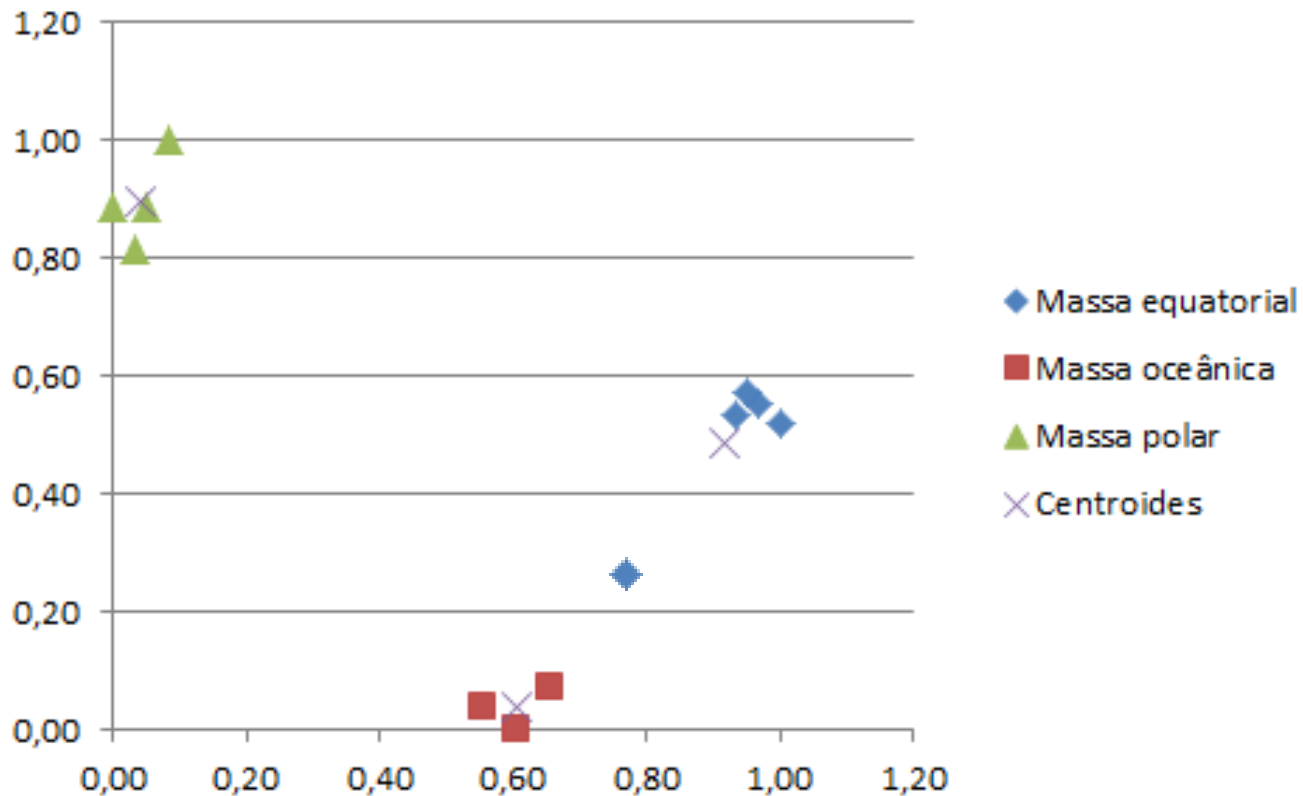
- $$c_{3lat} = \frac{\sum_{x \in C_i} x}{m_i} = \frac{(0,04+0,00+0,07)}{3} = 0,04$$

Ponto	Longitude	Latitude	d(x, c <sub>1</sub> )	d(x, c <sub>2</sub> )	d(x, c <sub>3</sub> )
1	1,00	0,52	<b>0,23</b>	1,03	0,65
2	0,97	0,56	<b>0,23</b>	0,99	0,66
3	0,93	0,53	<b>0,20</b>	0,97	0,62
4	0,95	0,57	<b>0,24</b>	0,97	0,66
5	0,56	0,04	0,43	1,01	<b>0,00</b>
6	0,61	0,00	0,43	1,06	<b>0,06</b>
7	0,66	0,07	0,34	1,03	<b>0,10</b>
8	0,74	0,26	<b>0,14</b>	0,95	0,28
9	0,05	0,89	0,95	<b>0,01</b>	0,99
10	0,08	1,00	0,99	<b>0,11</b>	1,07
11	0,03	0,81	0,93	<b>0,09</b>	0,94
12	0,00	0,89	0,99	<b>0,04</b>	1,02

# Algoritmo K-Means



- Exemplo
  - It.2: Representação gráfica



# Algoritmo K-Means



- Exemplo

- It.3: Atribuir cada ponto ao centroide mais próximo
  - $c_1 = (0,92; 0,49)$ ,  $c_2 = (0,04; 0,90)$ ,  $c_3 = (0,61; 0,04)$

Ponto	Longitude	Latitude	$d(x, c_1)$	$d(x, c_2)$	$d(x, c_3)$
1	1,00	0,52	<b>0,09</b>	1,03	0,62
2	0,97	0,56	<b>0,08</b>	0,99	0,63
3	0,93	0,53	<b>0,05</b>	0,97	0,59
4	0,95	0,57	<b>0,09</b>	0,97	0,63
5	0,56	0,04	0,58	1,00	<b>0,05</b>
6	0,61	0,00	0,58	1,06	<b>0,04</b>
7	0,66	0,07	0,49	1,03	<b>0,06</b>
8	0,74	0,26	0,29	0,95	<b>0,26</b>
9	0,05	0,89	0,96	<b>0,01</b>	1,02
10	0,08	1,00	0,98	<b>0,11</b>	1,10
11	0,03	0,81	0,94	<b>0,08</b>	0,97
12	0,00	0,89	1,00	<b>0,04</b>	1,05

# Algoritmo K-Means



- Exemplo

- It.3: Calcular custo

- $$SSE = \sum_{i=1}^K \sum_{x \in C_i} d(c_i, x)^2$$
$$= 0,09^2 + 0,08^2 + 0,05^2 + 0,09^2 + 0,05^2 + 0,04^2 + 0,06^2 + 0,26^2$$
$$+ 0,01^2 + 0,11^2 + 0,08^2 + 0,04^2$$
$$= 0,11$$

Ponto	Longitude	Latitude	d(x, c <sub>1</sub> )	d(x, c <sub>2</sub> )	d(x, c <sub>3</sub> )
1	1,00	0,52	<b>0,09</b>	1,03	0,62
2	0,97	0,56	<b>0,08</b>	0,99	0,63
3	0,93	0,53	<b>0,05</b>	0,97	0,59
4	0,95	0,57	<b>0,09</b>	0,97	0,63
5	0,56	0,04	0,58	1,00	<b>0,05</b>
6	0,61	0,00	0,58	1,06	<b>0,04</b>
7	0,66	0,07	0,49	1,03	<b>0,06</b>
8	0,74	0,26	0,29	0,95	<b>0,26</b>
9	0,05	0,89	0,96	<b>0,01</b>	1,02
10	0,08	1,00	0,98	<b>0,11</b>	1,10
11	0,03	0,81	0,94	<b>0,08</b>	0,97
12	0,00	0,89	1,00	<b>0,04</b>	1,05

# Algoritmo K-Means



- Exemplo

- It.3: Recalcular os centroides

- $$c_{1long} = \frac{\sum_{x \in C_i} x}{m_i} = \frac{(1,00+0,97+0,93+0,95)}{4} = 0,96$$

- $$c_{1lat} = \frac{\sum_{x \in C_i} x}{m_i} = \frac{(0,52+0,56+0,53+0,57)}{4} = 0,54$$

Ponto	Longitude	Latitude	d(x, c <sub>1</sub> )	d(x, c <sub>2</sub> )	d(x, c <sub>3</sub> )
1	1,00	0,52	<b>0,09</b>	1,03	0,62
2	0,97	0,56	<b>0,08</b>	0,99	0,63
3	0,93	0,53	<b>0,05</b>	0,97	0,59
4	0,95	0,57	<b>0,09</b>	0,97	0,63
5	0,56	0,04	0,58	1,00	<b>0,05</b>
6	0,61	0,00	0,58	1,06	<b>0,04</b>
7	0,66	0,07	0,49	1,03	<b>0,06</b>
8	0,74	0,26	0,29	0,95	<b>0,26</b>
9	0,05	0,89	0,96	<b>0,01</b>	1,02
10	0,08	1,00	0,98	<b>0,11</b>	1,10
11	0,03	0,81	0,94	<b>0,08</b>	0,97
12	0,00	0,89	1,00	<b>0,04</b>	1,05

# Algoritmo K-Means



- Exemplo

- It.3: Recalcular os centroides

- $$c_{2long} = \frac{\sum_{x \in C_i} x}{m_i} = \frac{(0,01+0,11+0,09+0,04)}{4} = 0,04$$

- $$c_{2lat} = \frac{\sum_{x \in C_i} x}{m_i} = \frac{(0,89+1,00+0,81+0,89)}{4} = 0,90$$

Ponto	Longitude	Latitude	d(x, c <sub>1</sub> )	d(x, c <sub>2</sub> )	d(x, c <sub>3</sub> )
1	1,00	0,52	<b>0,09</b>	1,03	0,62
2	0,97	0,56	<b>0,08</b>	0,99	0,63
3	0,93	0,53	<b>0,05</b>	0,97	0,59
4	0,95	0,57	<b>0,09</b>	0,97	0,63
5	0,56	0,04	0,58	1,00	<b>0,05</b>
6	0,61	0,00	0,58	1,06	<b>0,04</b>
7	0,66	0,07	0,49	1,03	<b>0,06</b>
8	0,74	0,26	0,29	0,95	<b>0,26</b>
9	0,05	0,89	0,96	<b>0,01</b>	1,02
10	0,08	1,00	0,98	<b>0,11</b>	1,10
11	0,03	0,81	0,94	<b>0,08</b>	0,97
12	0,00	0,89	1,00	<b>0,04</b>	1,05



# Algoritmo K-Means



- Exemplo

- It.3: Recalcular os centroides

- $c_{3long} = \frac{\sum_{x \in C_i} x}{m_i} = \frac{(0,56+0,61+0,66+0,74)}{4} = 0,64$

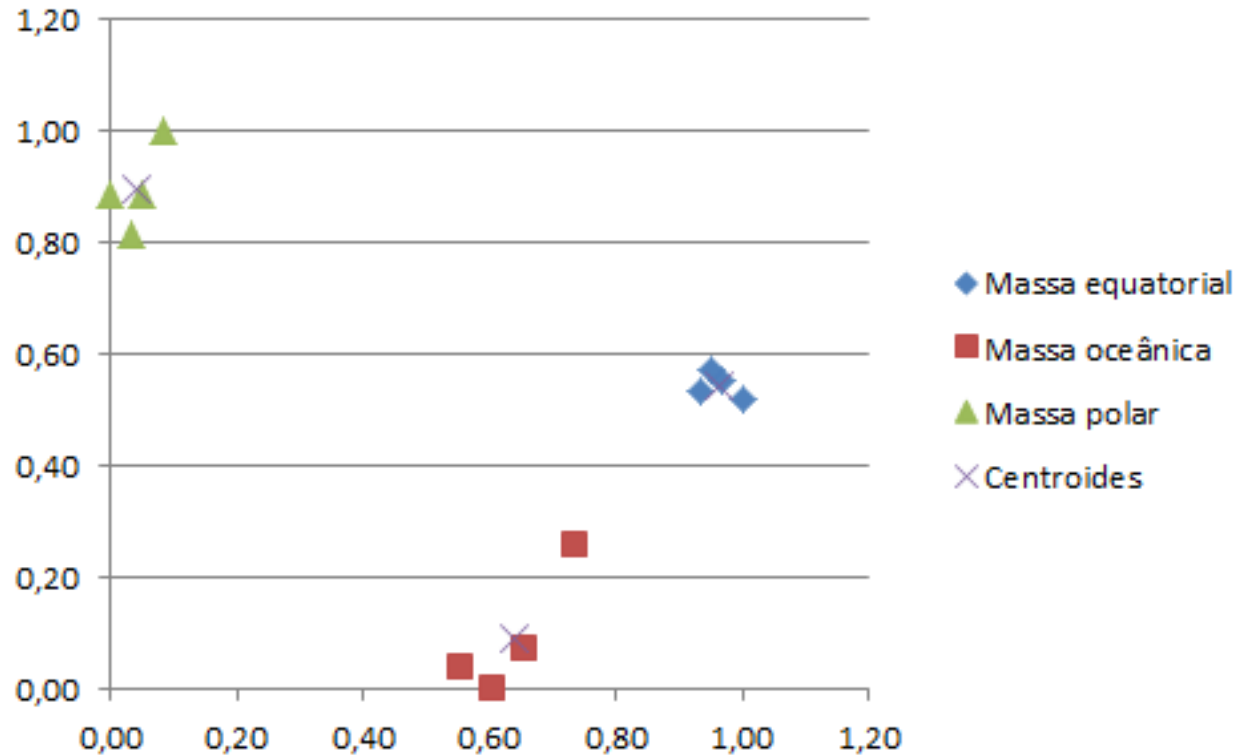
- $c_{3lat} = \frac{\sum_{x \in C_i} x}{m_i} = \frac{(0,04+0,00+0,07+0,26)}{4} = 0,09$

Ponto	Longitude	Latitude	d(x, c <sub>1</sub> )	d(x, c <sub>2</sub> )	d(x, c <sub>3</sub> )
1	1,00	0,52	<b>0,09</b>	1,03	0,62
2	0,97	0,56	<b>0,08</b>	0,99	0,63
3	0,93	0,53	<b>0,05</b>	0,97	0,59
4	0,95	0,57	<b>0,09</b>	0,97	0,63
5	0,56	0,04	0,58	1,00	<b>0,05</b>
6	0,61	0,00	0,58	1,06	<b>0,04</b>
7	0,66	0,07	0,49	1,03	<b>0,06</b>
8	0,74	0,26	0,29	0,95	<b>0,26</b>
9	0,05	0,89	0,96	<b>0,01</b>	1,02
10	0,08	1,00	0,98	<b>0,11</b>	1,10
11	0,03	0,81	0,94	<b>0,08</b>	0,97
12	0,00	0,89	1,00	<b>0,04</b>	1,05

# Algoritmo K-Means



- Exemplo
  - It.3: Representação gráfica



# Algoritmo K-Means



- Exemplo

- It.4: Atribuir cada ponto ao centroide mais próximo
  - $c_1 = (0,96; 0,54)$ ,  $c_2 = (0,04; 0,90)$ ,  $c_3 = (0,64; 0,09)$

Ponto	Longitude	Latitude	$d(x, c_1)$	$d(x, c_2)$	$d(x, c_3)$
1	1,00	0,52	<b>0,09</b>	1,03	0,62
2	0,97	0,56	<b>0,08</b>	0,99	0,63
3	0,93	0,53	<b>0,05</b>	0,97	0,59
4	0,95	0,57	<b>0,09</b>	0,97	0,63
5	0,56	0,04	0,58	1,00	<b>0,05</b>
6	0,61	0,00	0,58	1,06	<b>0,04</b>
7	0,66	0,07	0,49	1,03	<b>0,06</b>
8	0,74	0,26	0,29	0,95	<b>0,26</b>
9	0,05	0,89	0,96	<b>0,01</b>	1,02
10	0,08	1,00	0,98	<b>0,11</b>	1,10
11	0,03	0,81	0,94	<b>0,08</b>	0,97
12	0,00	0,89	1,00	<b>0,04</b>	1,05



- Exemplo

- It.3: Recalcular os centroides

- $c_{1long} = \frac{\sum_{x \in C_i} x}{m_i} = \frac{(1,00+0,97+0,93+0,95)}{4} = 0,96$

- $c_{1lat} = \frac{\sum_{x \in C_i} x}{m_i} = \frac{(0,52+0,56+0,53+0,57)}{4} = 0,54$

- $c_{2long} = \frac{\sum_{x \in C_i} x}{m_i} = \frac{(0,01+0,11+0,09+0,04)}{4} = 0,04$

- $c_{2lat} = \frac{\sum_{x \in C_i} x}{m_i} = \frac{(0,89+1,00+0,81+0,89)}{4} = 0,90$

- $c_{3long} = \frac{\sum_{x \in C_i} x}{m_i} = \frac{(0,56+0,61+0,66+0,74)}{4} = 0,64$

- $c_{3lat} = \frac{\sum_{x \in C_i} x}{m_i} = \frac{(0,04+0,00+0,07+0,26)}{4} = 0,09$

# Algoritmo K-Means



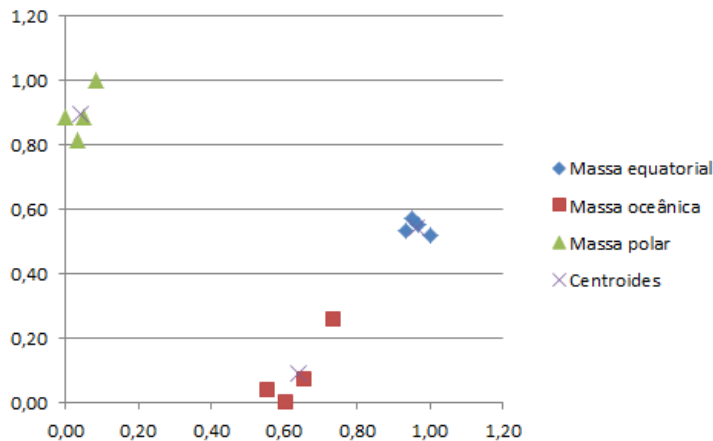
- Exemplo

- Centroides não mudaram!

- Logo a solução final é:

- $c_1 = (0,96; 0,54)$ ,  $c_2 = (0,04; 0,90)$ ,  $c_3 = (0,64; 0,09)$

- Custo = 0,11



Ponto	Longitude	Latitude	$d(x, c_1)$	$d(x, c_2)$	$d(x, c_3)$
1	1,00	0,52	<b>0,09</b>	1,03	0,62
2	0,97	0,56	<b>0,08</b>	0,99	0,63
3	0,93	0,53	<b>0,05</b>	0,97	0,59
4	0,95	0,57	<b>0,09</b>	0,97	0,63
5	0,56	0,04	0,58	1,00	<b>0,05</b>
6	0,61	0,00	0,58	1,06	<b>0,04</b>
7	0,66	0,07	0,49	1,03	<b>0,06</b>
8	0,74	0,26	0,29	0,95	<b>0,26</b>
9	0,05	0,89	0,96	<b>0,01</b>	1,02
10	0,08	1,00	0,98	<b>0,11</b>	1,10
11	0,03	0,81	0,94	<b>0,08</b>	0,97
12	0,00	0,89	1,00	<b>0,04</b>	1,05

# Algoritmo K-Means



- Base de dados para o trabalho
  - Trajetória de caracteres
  - <http://archive.ics.uci.edu/ml/datasets/Character+Trajectories>
  - Dada a trajetória da caneta, agrupar os dados de acordo com o caractere a formar





- *Introdução ao Data Mining*. Steinbach, Michael; Kumar, Vipin; Tan, Pang-ning, Rio de Janeiro: Ed. Ciência Moderna, 2009. Capítulo 8.

