



Aprendizado Automático



João Paulo Pordeus Gomes

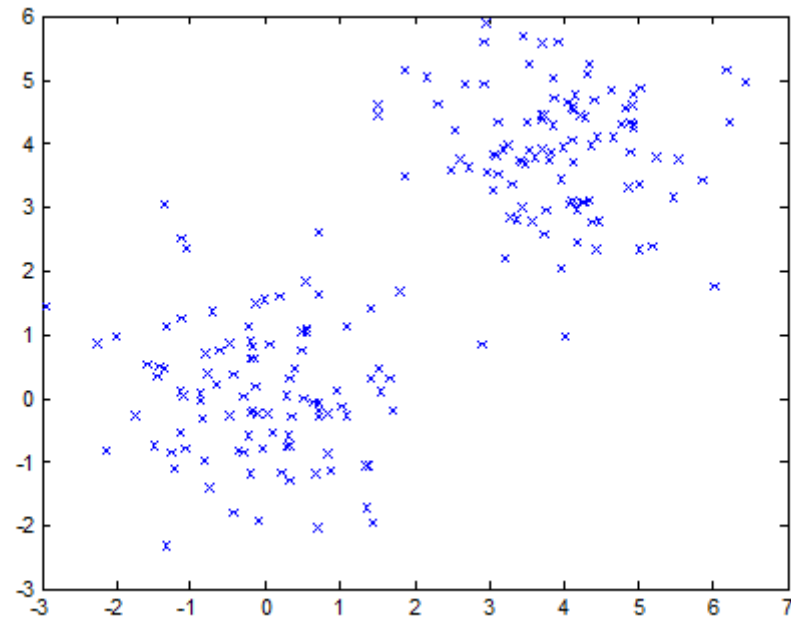
Classificação não supervisionada

Classificação Não Supervisionada

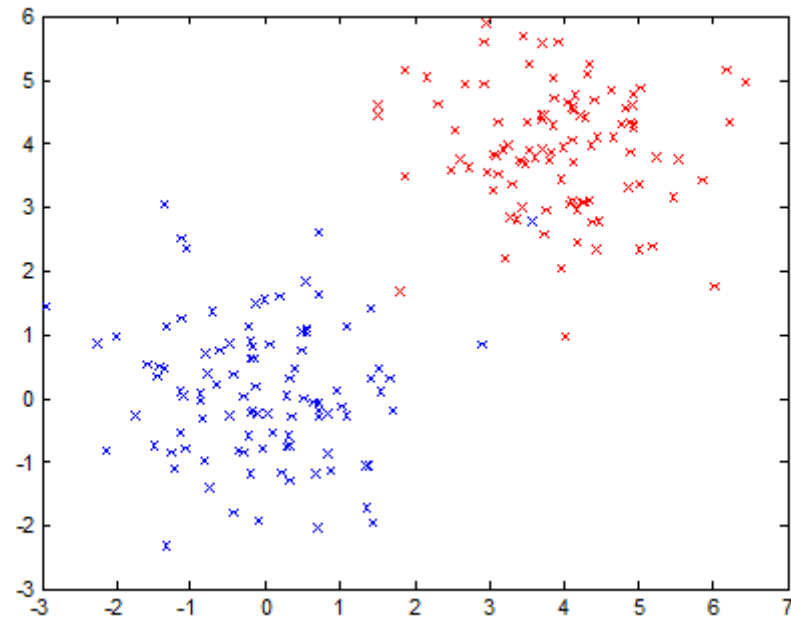
- ▶ Não existe uma saída esperada (y)
- ▶ Existem somente dados de entrada e deseja-se descobrir alguma estrutura neste dados
- ▶ Agrupamento (*clustering*)



Classificação Não Supervisionada



Classificação Não Supervisionada



Classificação Não Supervisionada

- ▶ Quantos grupos existem ?
- ▶ Quais os componentes destes grupos ?



Classificação Não Supervisionada

- ▶ Quantos grupos existem ?
- ▶ Quais os componentes destes grupos ?
- ▶ Métodos de Agrupamento
 - ▶ Hierárquico
 - ▶ Não Hierárquico



Agrupamento Hierárquico

- ▶ Os dados iniciam se grupos definidos
- ▶ Dados similares são agrupados formando pequenos grupos
- ▶ Pequenos grupos são agrupados formando grupos maiores
- ▶ Procedimento é repetido até que todos pertençam a um grupo



Agrupamento Hierárquico

- ▶ Dados são ditos semelhantes de acordo com alguma medida de distância.



Agrupamento Hierárquico

- ▶ Dados são ditos semelhantes de acordo com alguma medida de distância.

- ▶ Euclidiana

- ▶ $D_E(\mathbf{r}, \mathbf{s}) = \sqrt{\sum_{j=1}^p (r_j - s_j)^2}$

- ▶ Manhattan

- ▶ $D_M(\mathbf{r}, \mathbf{s}) = \sum_{j=1}^p |r_j - s_j|$



Agrupamento Hierárquico

- ▶ Dados são ditos semelhantes de acordo com alguma medida de distância.
 - ▶ Euclidiana
 - ▶ $D_E(\mathbf{r}, \mathbf{s}) = \sqrt{\sum_{j=1}^p (r_j - s_j)^2}$
 - ▶ Manhattan
 - ▶ $D_M(\mathbf{r}, \mathbf{s}) = \sum_{j=1}^p |r_j - s_j|$
- ▶ Similaridade entre grupos pode ser medida pela distância entre centróides
- ▶ Gráfico semelhante a uma árvore



Agrupamento Hierárquico

- ▶ Exemplo

- ▶ Dados

- ▶ [1 2],[1 1],[3 3] e [4 3]

- ▶ Calcula-se uma matriz de distâncias (d^2)

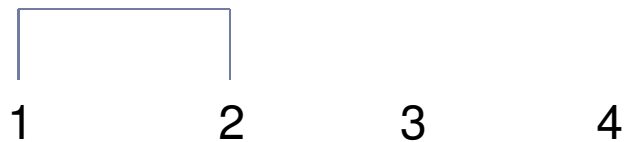
	1	2	3	4
1	0	1	5	10
2	1	0	8	13
3	5	8	0	1
4	10	13	1	0



Agrupamento Hierárquico

► [1 2],[1 1],[3 3] e [4 3]

	1	2	3	4
1	0	1	5	10
2	1	0	8	13
3	5	8	0	1
4	10	13	1	0

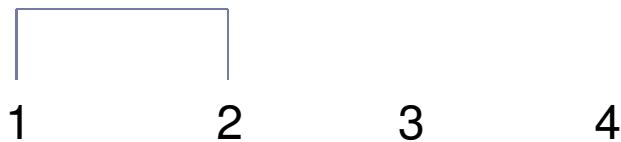


Agrupamento Hierárquico

- [1 2],[1 1],[3 3] e [4 3]

	1,2	3	4
1,2	0		
3		0	1
4		1	0

$$C_{1,2} = [1 \ 1,5]$$



Agrupamento Hierárquico

► [1 1.5],[3 3] e [4 3]

	1,2	3	4
1,2	0	6.25	11.25
3	6.25	0	1
4	11.25	1	0

$$C_{1,2} = [1 \ 1,5]$$

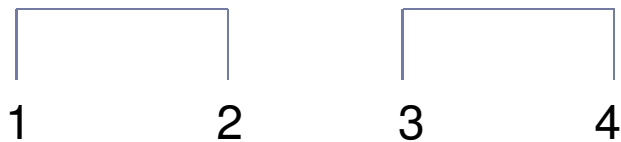


Agrupamento Hierárquico

► [1 1.5],[3 3] e [4 3]

	1,2	3	4
1,2	0	6.25	11.25
3	6.25	0	1
4	11.25	1	0

$$C_{1,2} = [1 \ 1,5]$$

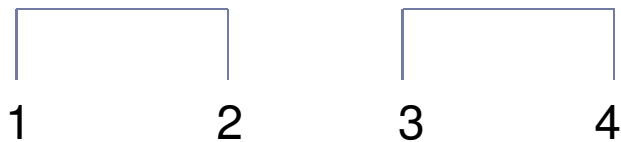


Agrupamento Hierárquico

- [1 1.5],[3 3] e [4 3]

	1,2	3,4
1,2	0	
3,4		0

$$C_{3,4} = [3 \ 3,5]$$

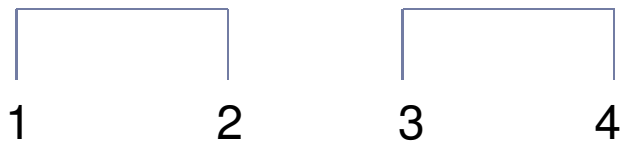


Agrupamento Hierárquico

► [1 1.5] e [3 3.5]

	1,2	3,4
1,2	0	8
3,4	8	0

$$C_{3,4} = [3 \ 3,5]$$

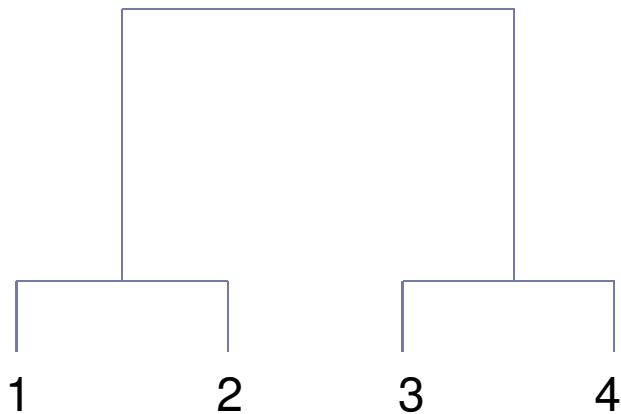


Agrupamento Hierárquico

► [1 1.5] e [3 3.5]

	1,2	3,4
1,2	0	8
3,4	8	0

$$C_{3,4} = [3 \ 3,5]$$

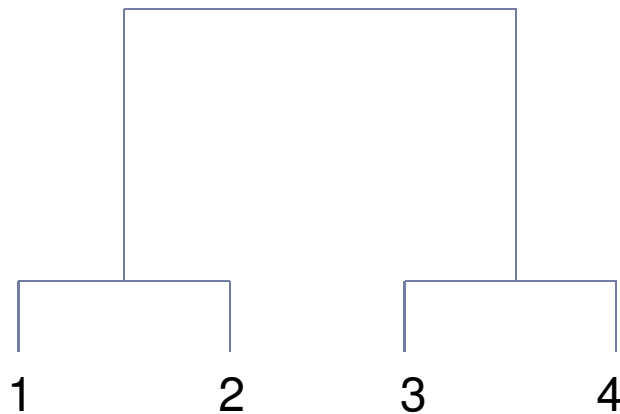


Agrupamento Hierárquico

► [1 1.5] e [3 3.5]

	1,2	3,4
1,2	0	8
3,4	8	0

$$C_{3,4} = [3 \ 3,5]$$



Dendrograma



Agrupamento não Hierárquico

- ▶ Pontos pertencem a algum grupo
- ▶ Pontos mudam de grupo de forma a satisfazer um determinado critério



K-médias

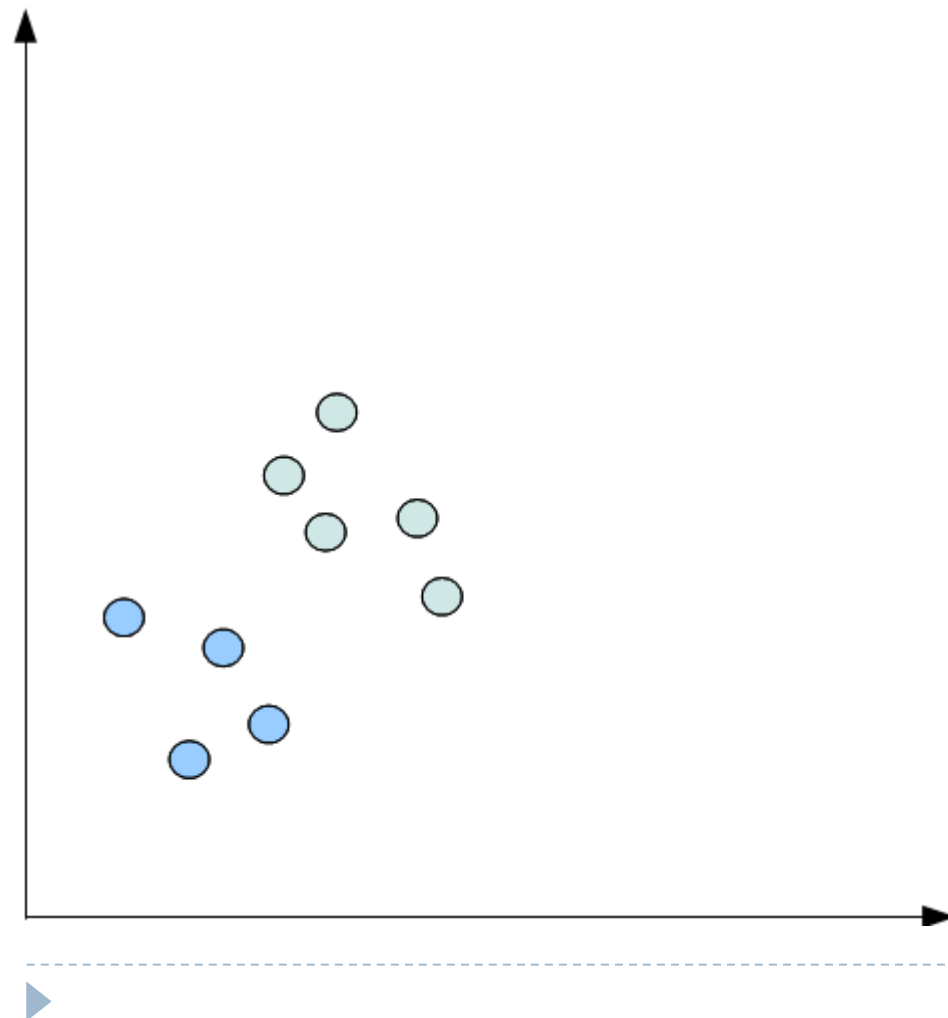


K-médias

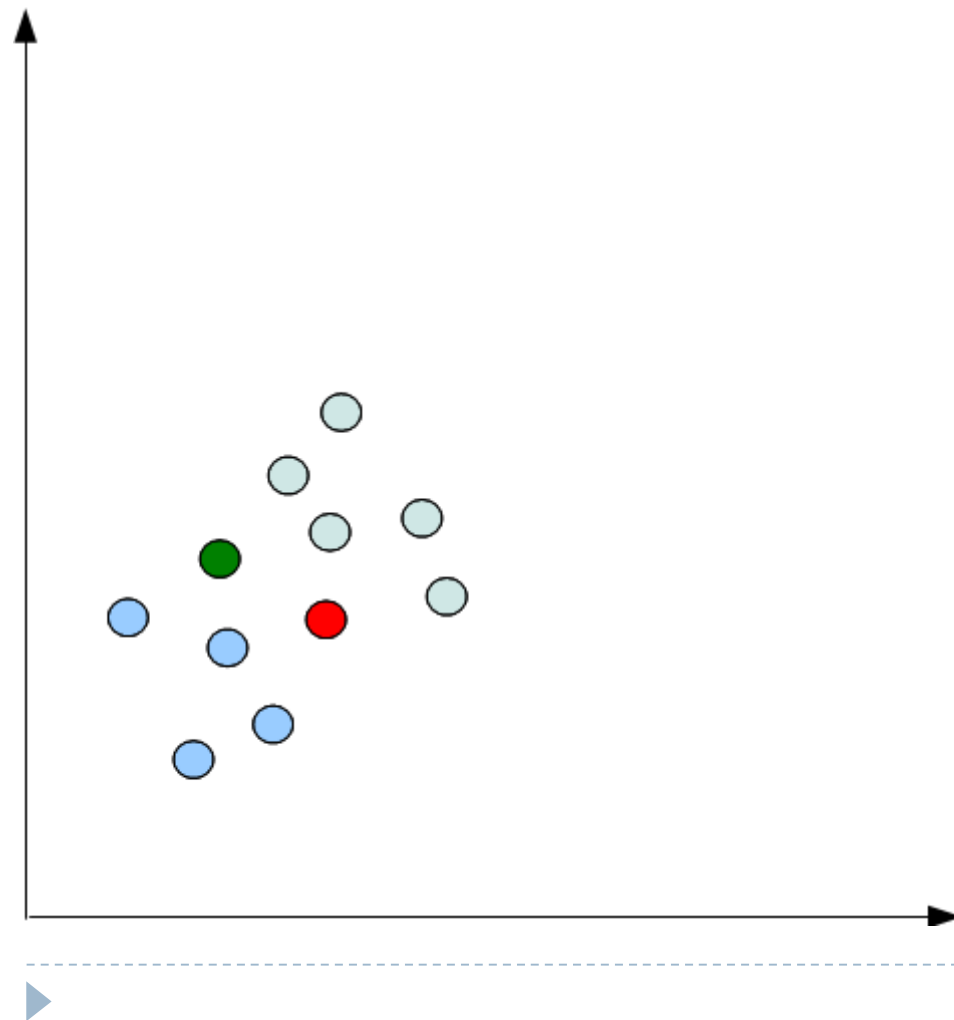
- ▶ É necessário conhecer o número de clusters
- ▶ k centroides são escolhidos aleatoriamente (podem ser escolhidos k membros da população)
- ▶ Calcula-se a distância destes pontos para todos os outros
- ▶ Os pontos passarão a pertencer ao grupo cuja distância é a menor
- ▶ Centróides são recalculados como a média dos pontos do grupo



K-médias

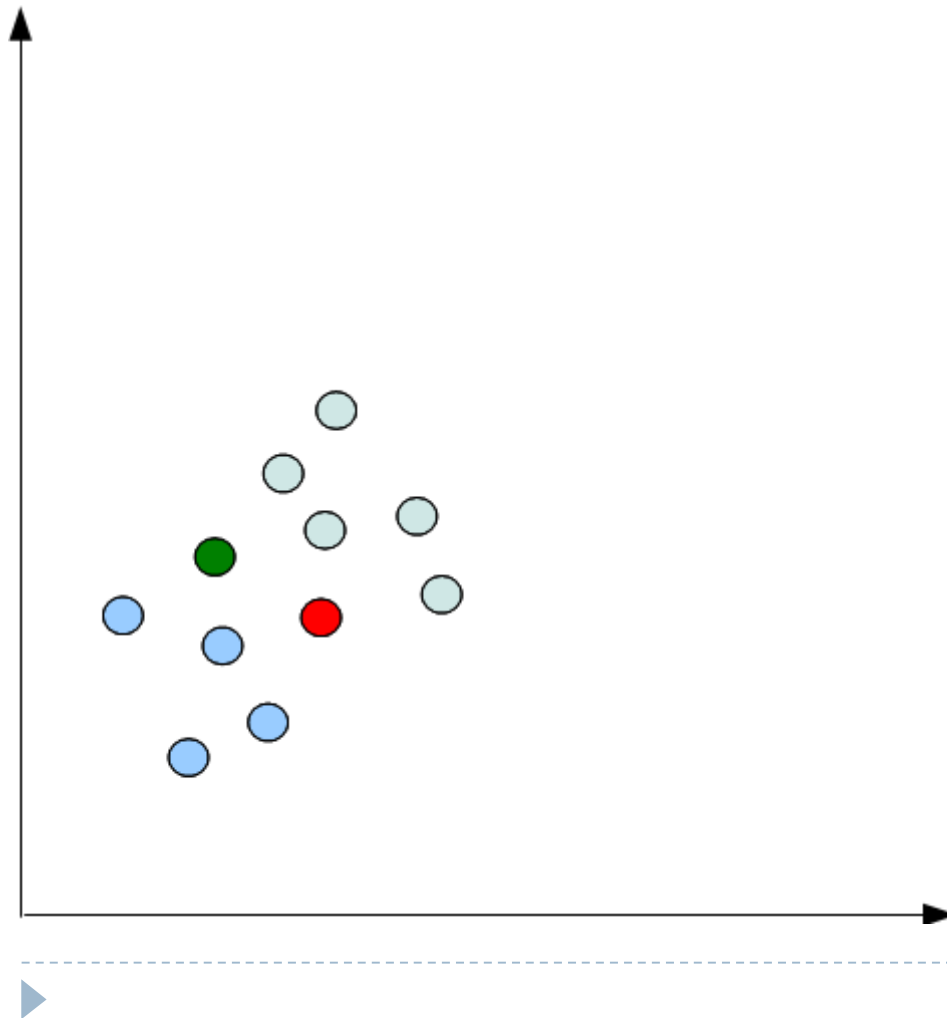


K-médias

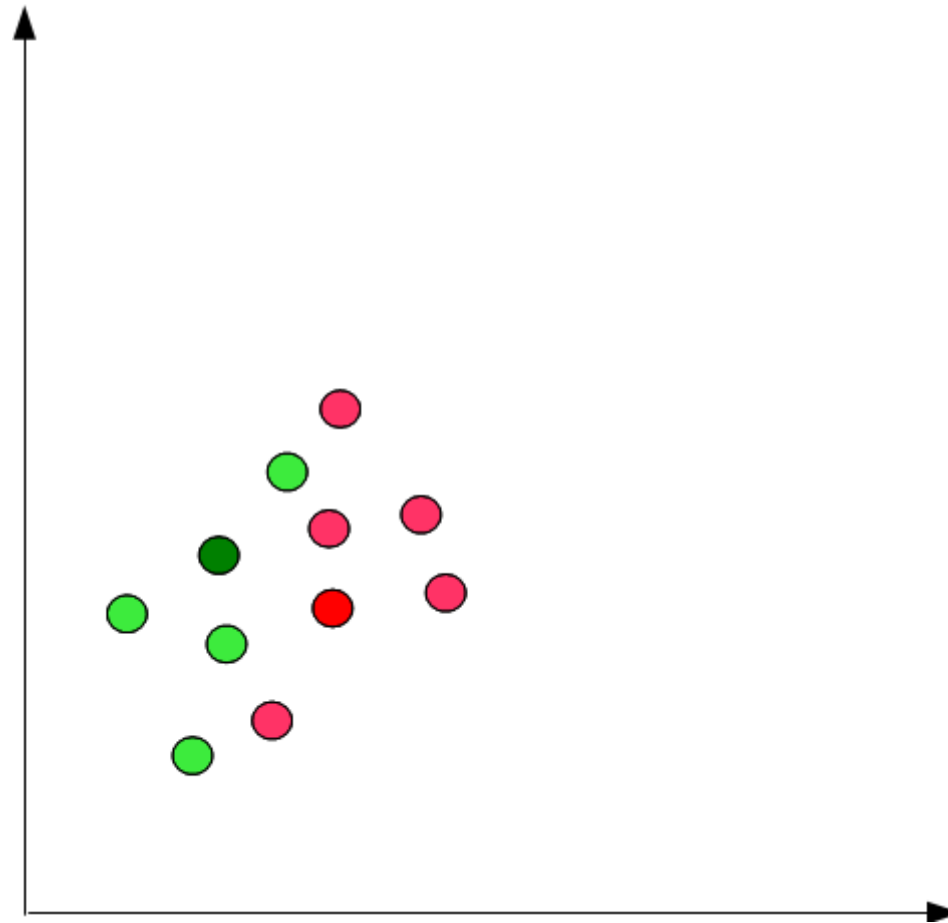


K-médias

1 – Calcular distâncias

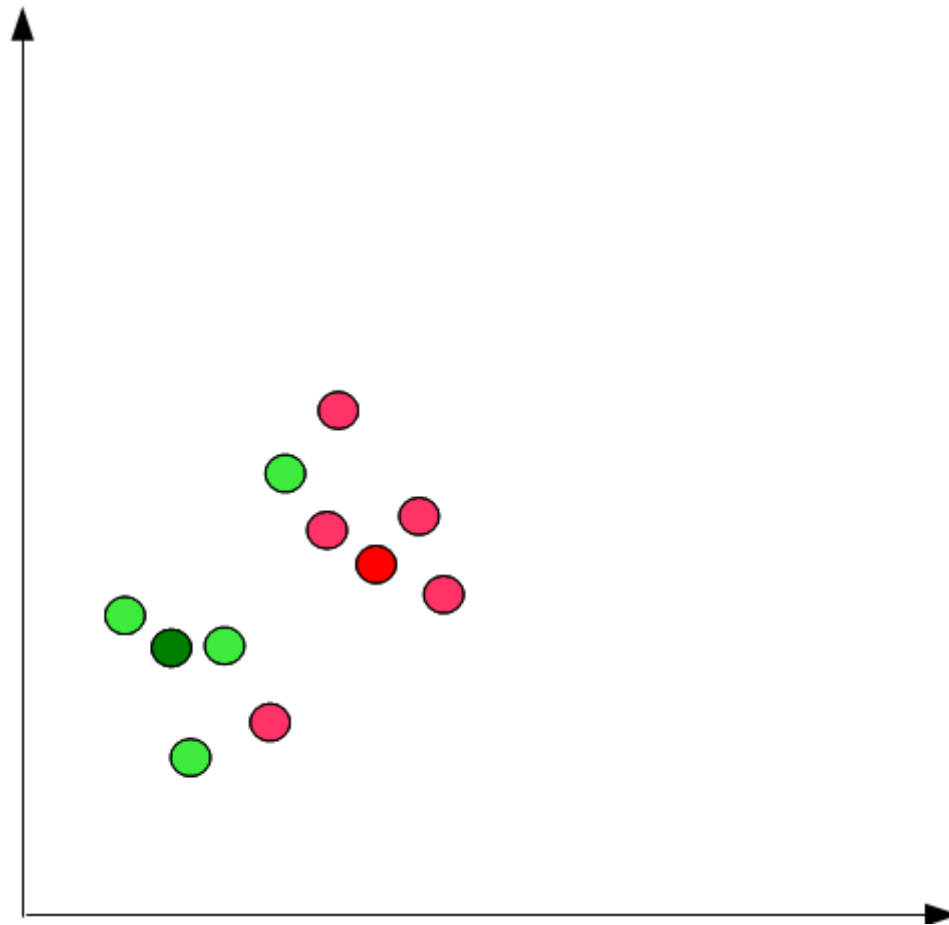


K-médias



- 1 – Calcular distâncias
- 2 – Atribuir Grupos

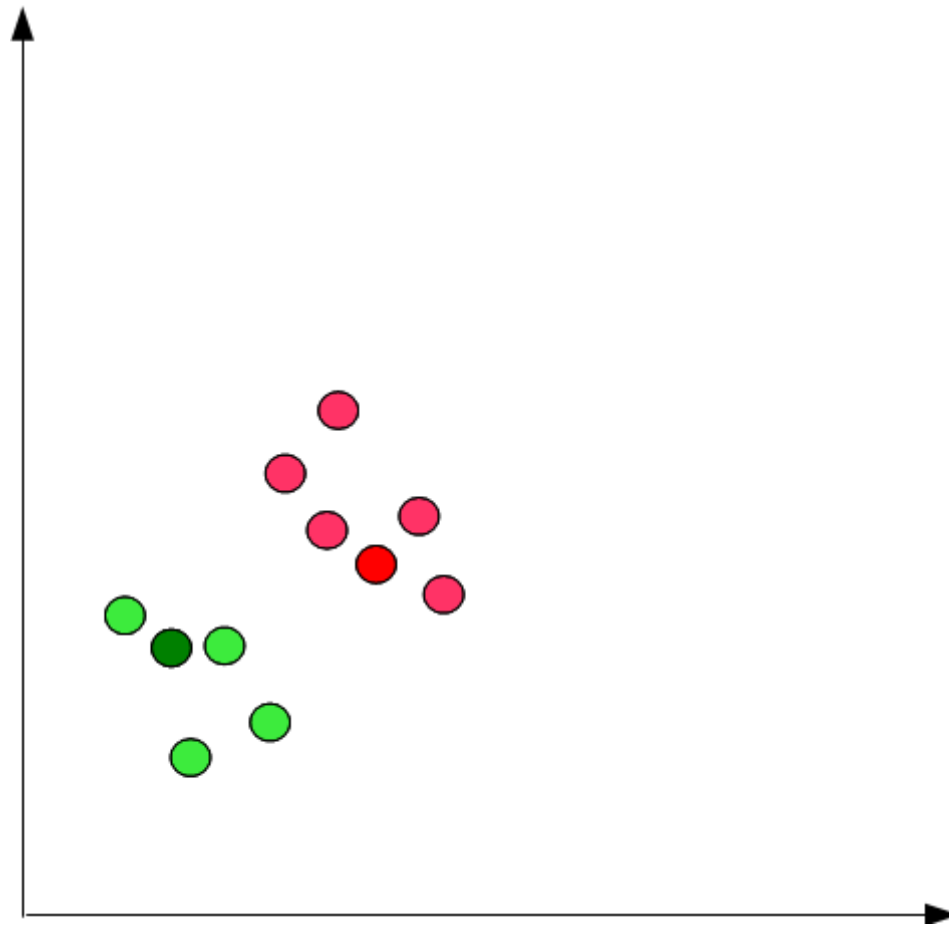
K-médias



- 1 – Calcular distâncias
- 2 – Atribuir Grupos
- 3 – Recalcular centróides



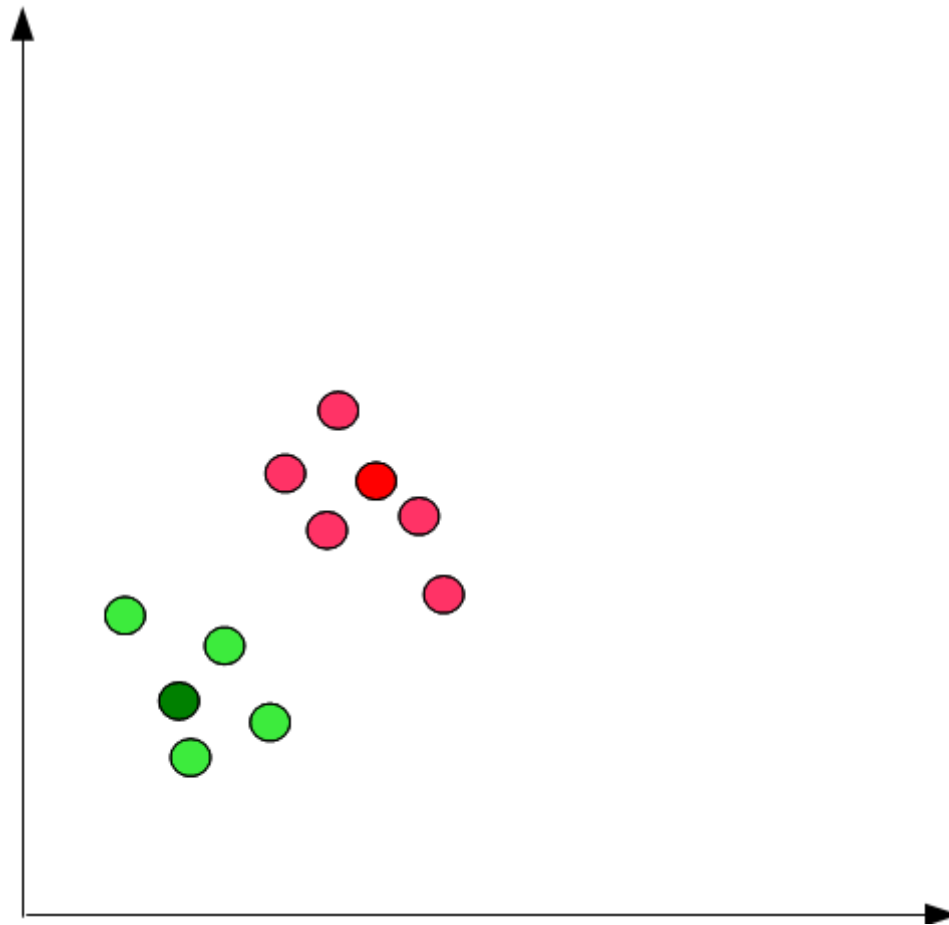
K-médias



- 1 – Calcular distâncias
- 2 – Atribuir Grupos
- 3 – Recalcular centróides



K-médias



- 1 – Calcular distâncias
- 2 – Atribuir Grupos
- 3 – Recalcular centróides



K-médias

► Pseudo-código

Algoritmo 1 Algoritmo convencional das K-médias

Entrada: Conjunto de dados D , Número de clusters k , Dimensão d

Saída: Distribuição dos n pontos de D entre os k clusters

Seja C_i é o i -ésimo cluster

$C_1, C_2, \dots, C_k =$ partição inicial de D

repita

$d_{i,j}$ = distância entre o caso i e o cluster j

para todo $1 \leq j \leq k$ **faça**

$n_i = \arg \min \{d_{i,j} : \forall i, j\}$

 Atribua o caso i ao cluster n_i

 Recalcule o centróide de qualquer cluster modificado acima

fim

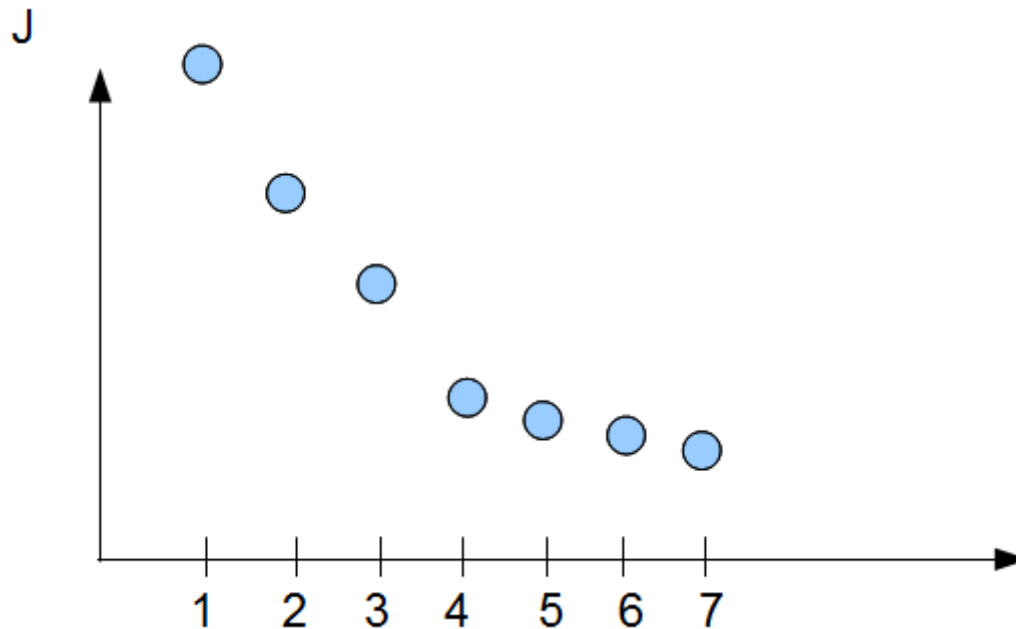
até Nenhum centróide mude de lugar

retorne saída



K-médias

- ▶ Número de grupos
 - ▶ J é a distância entre todos os pontos e seus centróides



Dúvidas ?