

# Conceitos



Tarefas não supervisionadas



Não existem classes



Objetivo é criar grupos a partir de atributos (características) das instâncias

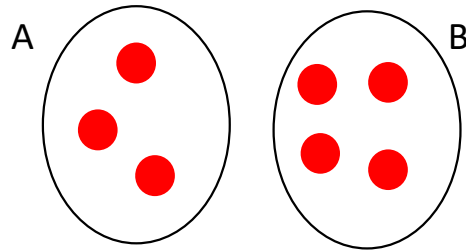
# Aplicações

---

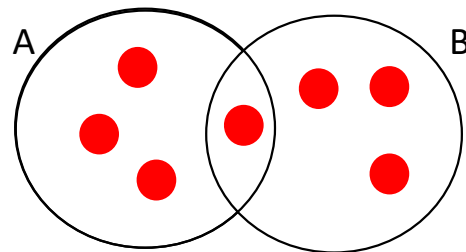
- Particionar clientes em diferentes segmentos
- Reconhecimento de comunidades em análises de redes sociais
- Divisão de imagem em diferentes segmentos
- Detecção de anomalias em dados
- Combate ao crime: identificação de regiões com maior incidência
- Categorização de Produtos, Espécies etc.

# Tipos

- Agrupamento completo: cada elemento é adicionado em um único grupo



- Agrupamento parcial: cada instancia pode pertencer a mais de um grupo

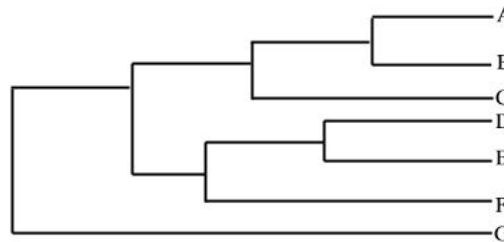


# Tipos

- Modelo Difuso: cada elemento pertence a um grupo segundo uma probabilidade

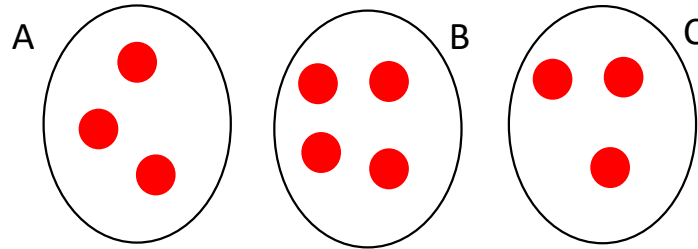
	Grupo A	Grupo B	Grupo C
Elemento A	0.5	0.3	0.2
Elemento B	0.1	0.1	0.8
Elemento C	0.3	0.4	0.3

- Modelo Hierárquico: permite que o grupo tenha subgrupos

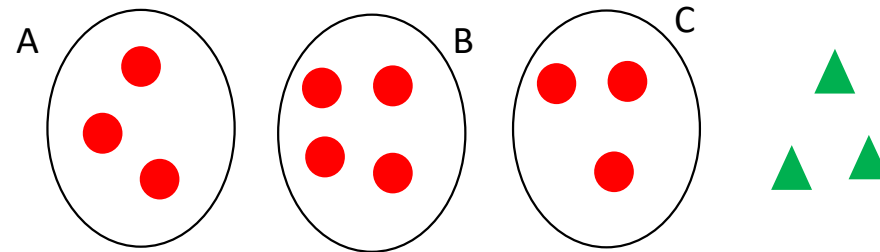


# Tipos

- Agrupa todos os elementos



- Pode deixar elementos sem agrupar (ruído)



# K-means

---

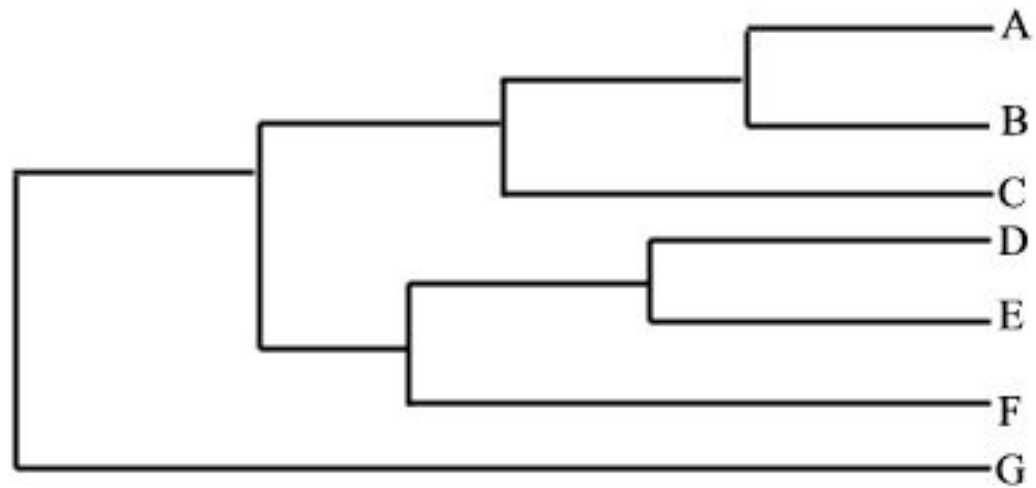
- Simples
- Baseado em protótipo
- Encontra um número de grupos definido pelo usuário
- Agrupa todos os objetos
- Definir os centróides é uma etapa fundamental
- Distância Euclidiana

# DBSCAN

---

- Baseado em Densidade
- Menos afetado por ruído
- Número de grupos definido automaticamente
- Pontos de baixa densidade são definidos com ruído e não agrupados
- A densidade é baseada no raio especificado. Um ponto pode estar no interior, no limite, ou sem classificação (ruído)
- Não é bom em grupos cujas densidades variam muito

## Hierárquico



- Aglomerativa: começa com pontos em grupos individuais e a cada etapa funde os pares mais próximos. Requer uma noção de proximidade. Mais comuns
- Divisiva: Começa incluindo todos, e a cada etapa divide até que reste apenas grupos únicos
- Dendograma



# Clara

- Clustering Large Applications
- Baseado no K-medoids
- Funciona por amostragem

# Dados

ID	MedA	MedB	MedC	MedD
1	5,1	3,5	1,4	0,2
2	4,9	3	1,4	0,2
3	4,7	3,2	1,3	0,2
4	4,6	3,1	1,5	0,2
5	5	3,6	1,4	0,2
6	5,4	3,9	1,7	0,4
7	4,6	3,4	1,4	0,3
8	5	3,4	1,5	0,2
9	4,4	2,9	1,4	0,2
10	4,9	3,1	1,5	0,1
11	5,4	3,7	1,5	0,2
12	4,8	3,4	1,6	0,2
13	4,8	3	1,4	0,1
14	4,3	3	1,1	0,1
15	5,8	4	1,2	0,2
16	5,7	4,4	1,5	0,4
17	5,4	3,9	1,3	0,4
18	5,1	3,5	1,4	0,3
19	5,7	3,8	1,7	0,3
20	5,1	3,8	1,5	0,3
21	5,4	3,4	1,7	0,2
22	5,1	3,7	1,5	0,4
23	4,6	3,6	1	0,2
24	5,1	3,3	1,7	0,5
25	4,8	3,4	1,9	0,2
26	5	3	1,6	0,2