



Agrupamentos

# Conceitos

- Tarefas não supervisionadas
- Não existem classes
- Objetivo é criar grupos a partir de atributos (características) das instâncias



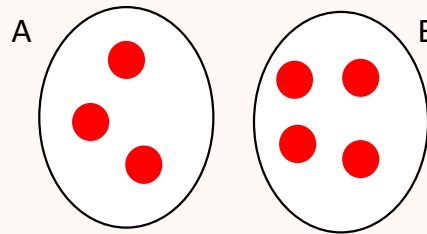
# Aplicações

- 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
- Dividir documentos em grupos conforme o assunto!

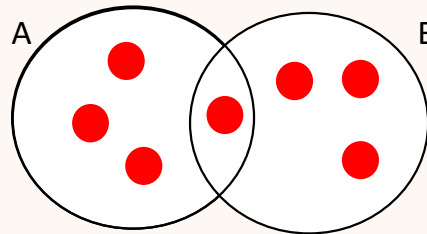


# Tipos

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



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



# Tipos

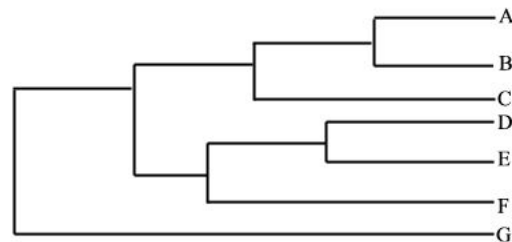
---

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

---

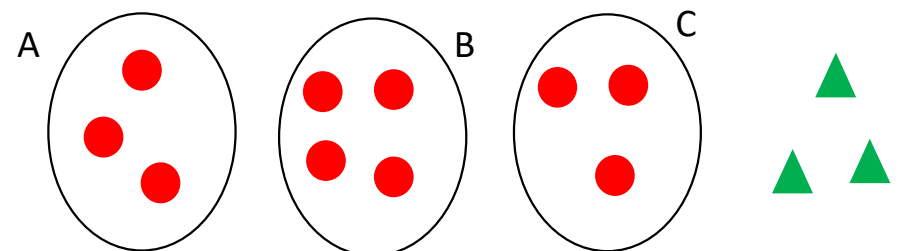
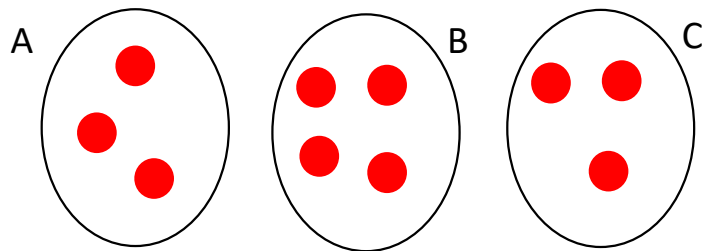
Modelo Hierárquico: permite que o grupo tenha subgrupos



# Tipos

Agrupa todos os  
elementos

Pode deixar elementos  
sem agrupar (ruído)



# K-means e K-medoid

- 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
- 
- K-means: Protótipo é um centróide: média de grupo de pontos. Quase nunca é um ponto real de dados.
  - K-medoid: Protótipo baseado em medóide: ponto mais representativo. É um ponto real de dados.

# K-means e K-medoid

- Tem dificuldade para detectar grupos naturais, não esféricas, de tamanho ou densidades muito diferentes
- Restrito a dados que exista uma noção de centro
- Pode ser melhorado escolhendo os centros





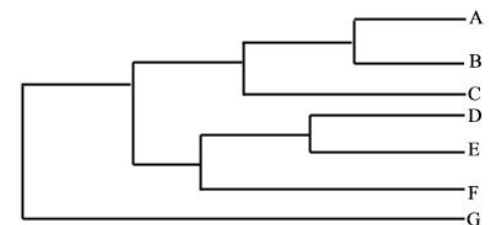
# DBSCAN

- 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





# Sistemas de Recomendação

# Definições

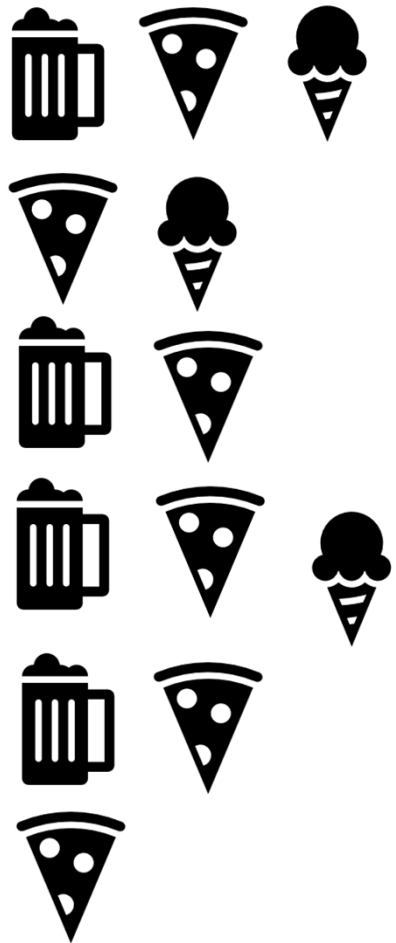
- Busca a relação entre itens
- Aplicações:
  - Cesta de compras
  - Detecção de intrusão
  - Analise de avaliações de consumidores
  - Diagnóstico médico
  - Recomendações de filmes, livros etc.



# Conceitos

- Suporte: numero de eventos que contem todos os itens da transação
- Confiança: indica a proporção de vezes que, em um evento contendo o elemento A, também tem B.





→ Quem compra sorvete também compra pizza

Suporte

Nº Transações: 6

Nº Transações c/ 2 elementos: 3

$$3/6 = 0,5$$

Confiança

Nº Transações com A: 3

Nº Transações com A e B: 3

$$3/3=1$$