

FORMAÇÃO CIENTISTA DE DADOS

AGRUPAMENTOS

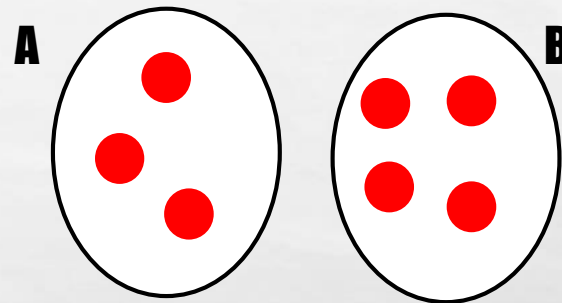


CONCEITOS

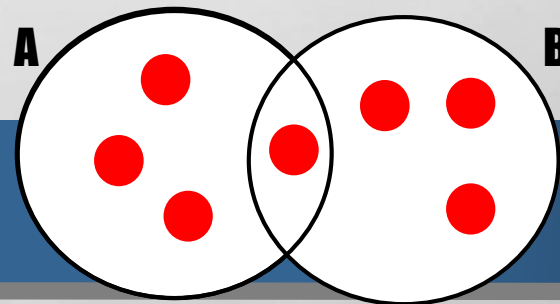
- **TAREFAS NÃO SUPERVISIONADAS**
- **NÃO EXISTEM CLASSES**
- **OBJETIVO É CRIAR GRUPOS A PARTIR DE ATRIBUTOS (CARACTERÍSTICAS) DAS INSTÂNCIAS**

TIPOS

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



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

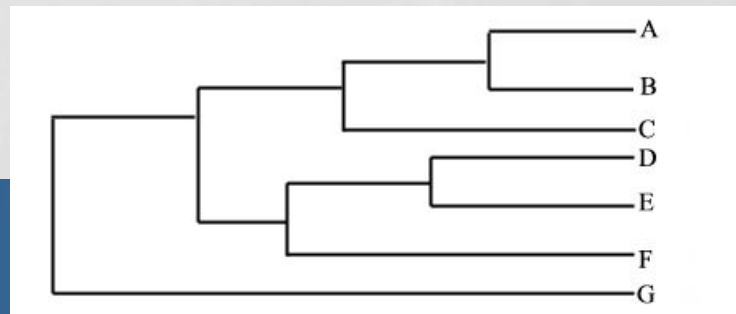


TIPOS

- MODELO DIFUSO: CADA ELEMENTO PERTENCE A UM GRUPO SEGUNDO UMA PROBABILIDADE

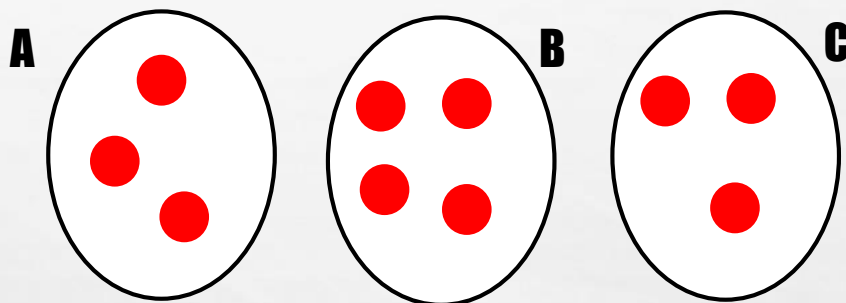
	Grupo A	Grupo B	Grupo C
Elemento A	0.5	0.3	0.2
Elemento B	0.1	0.1	0.8
Elemento C	0.3	0.4	0.3

- MODELO HIERÁRQUICO: PERMITE QUE O GRUPO TENHA SUBGRUPOS

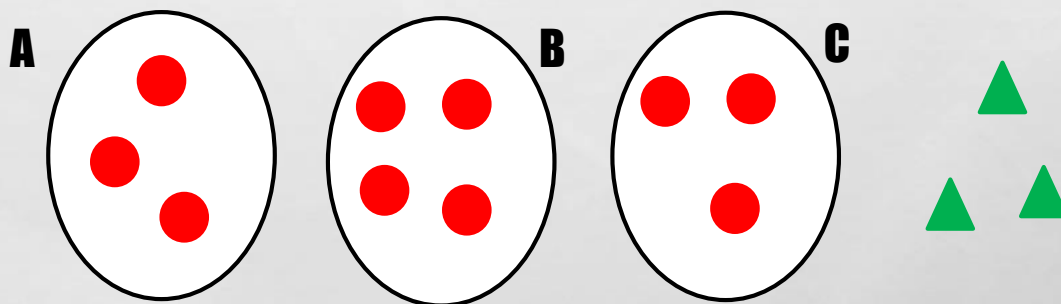


TIPOS

- **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 (USO DE SEEDS)**

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)**

DBSCAN

- **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**