**Aprendizado Não Supervisionado**

Não exige dados rotulados, usa regularidade e estatísticas dos dados

Tarefa: descreve conjuntos de dados

**Sumarização**

Descrição simples e compacta

**Associação**

Relações frequentes entre dados

**Agrupamento**

Divisão baseada em regularidade/similaridade

Organiza os dados quando não há classe para predizer

Etapas:

* **Preparação**: pré processamento e forma de representação
* **Proximidade**: definição de medidas de proximidade apropriadas
* **Agrupamento**: geração de grupos
* **Validação**: verifica se os grupos gerados são significativos
* **Interpretação**: Examina o grupo com objetivo de rotulá-lo

Tipos

* **Particional**: divisão dos dados em grupos tal que cada um pertence a 1 grupo
* **Hierárquico**: conjunto de grupos aninhados na forma de árvore hierárquica

**K-MEANS**

Algoritmo de agrupamento particional:

* Seleciona k objetos centroides
* Calcula a distancia de cada objeto com o centroide de cada cluster
* Objeto será atribuído ao cluster cujo centroide está mais próximo
* Depois de processar todos, recalcula os centroides, considerando os dados em cada cluster

**Agrupamento Hierárquico**

* **Por divisão:** inicia com todos os objetos representando um único grupo, dividindo até que tenha um obj ou k grupos
* **Aglomerativo:** inicia com cada obj individual combinando o par mais próximo

Algoritmos:

* **Single-link:** menor distancia entre 2 clusters
* **Complete-link:** maior distância entre 2 clusters
* **Average-link**: distância média entre 2 clusters
* **Distance Between Centroids:** distância entre centróides de 2 clusters