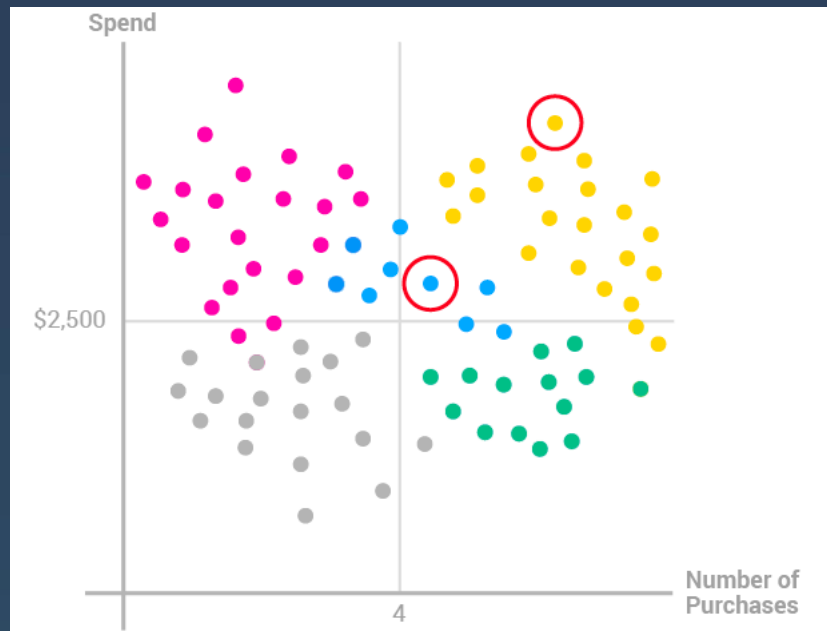




# Clusterização

# Aplicações de Clusterização

- ✓ Segmentar clientes
- ✓ Definir personas
- ✓ Ofertar os produtos mais adequados para cada tipo de cliente
- ✓ Algoritmos de Clusterização +  
Heurísticas por Experiência



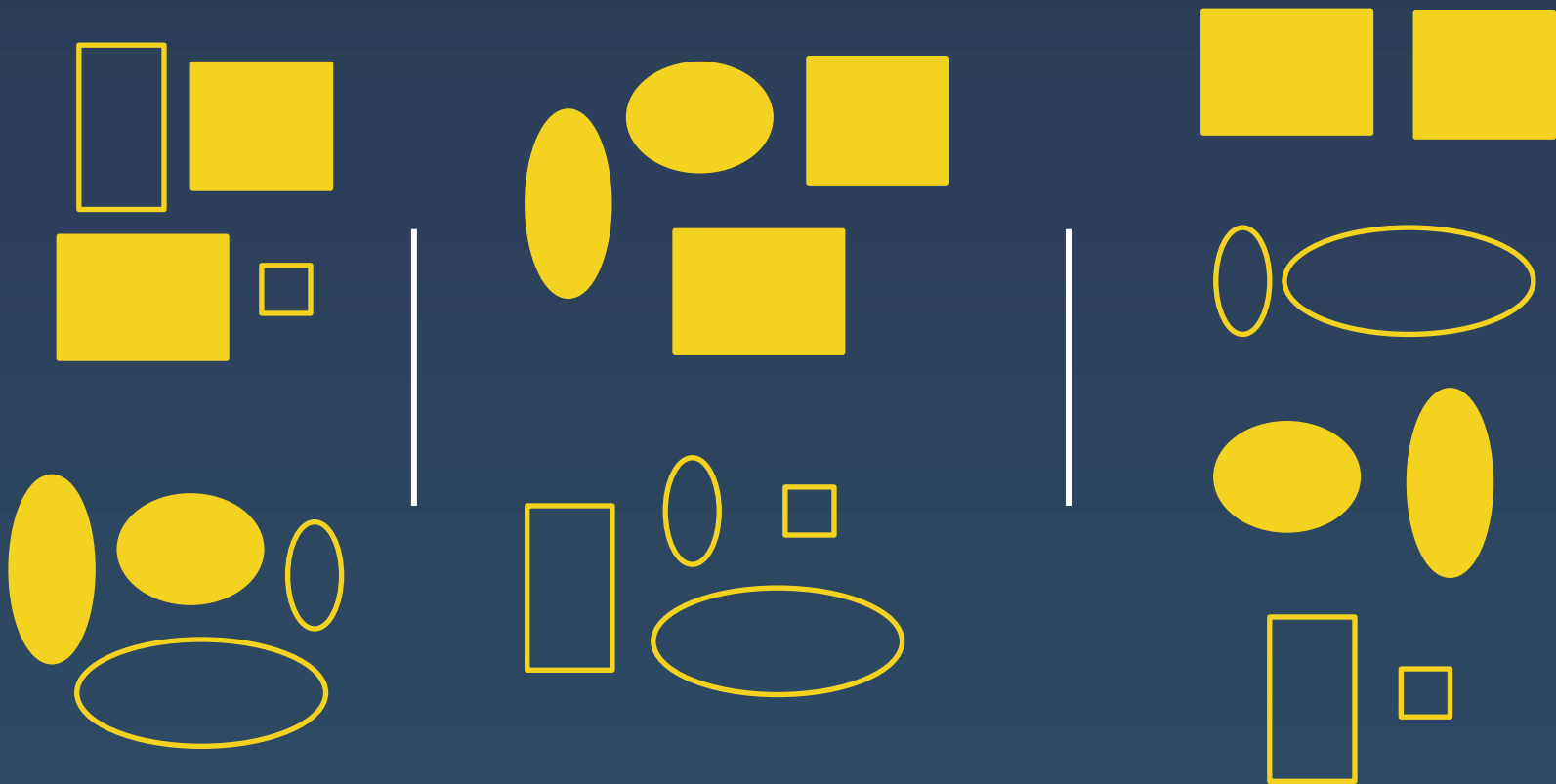
# Introdução a Agrupamento

---



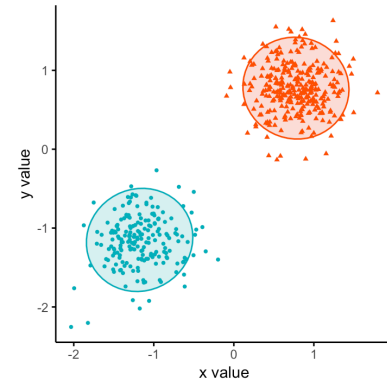
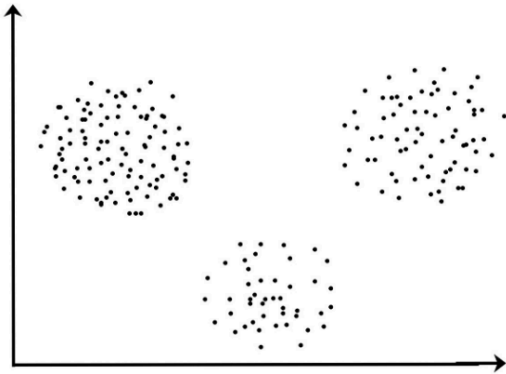
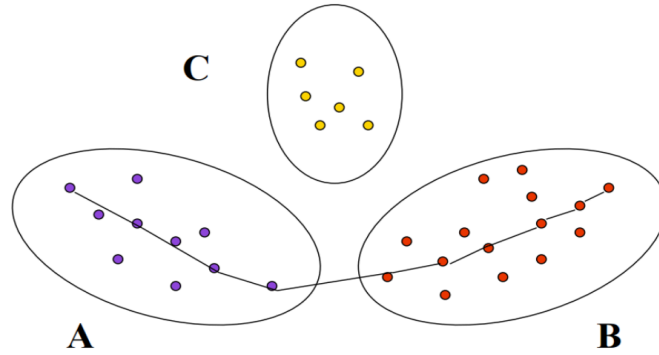
# Introdução a Agrupamento

---



# Introdução a Agrupamento

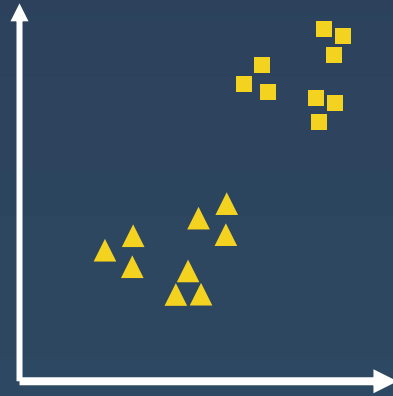
---



# Introdução a Agrupamento

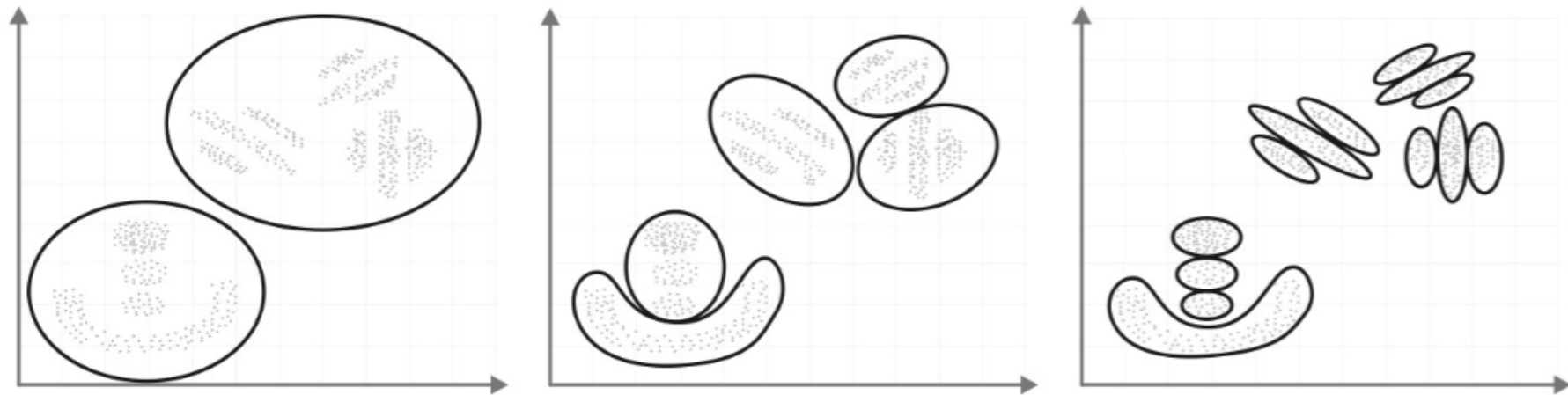
---

- ✓ Os algoritmos de agrupamento podem encontrar diferentes níveis de refinamento.



# Introdução a Teórica

---



# Fluxo de Modelagem

A decorative graphic at the bottom left of the slide, consisting of a series of interconnected lines and circles resembling a circuit board or a stylized city skyline.

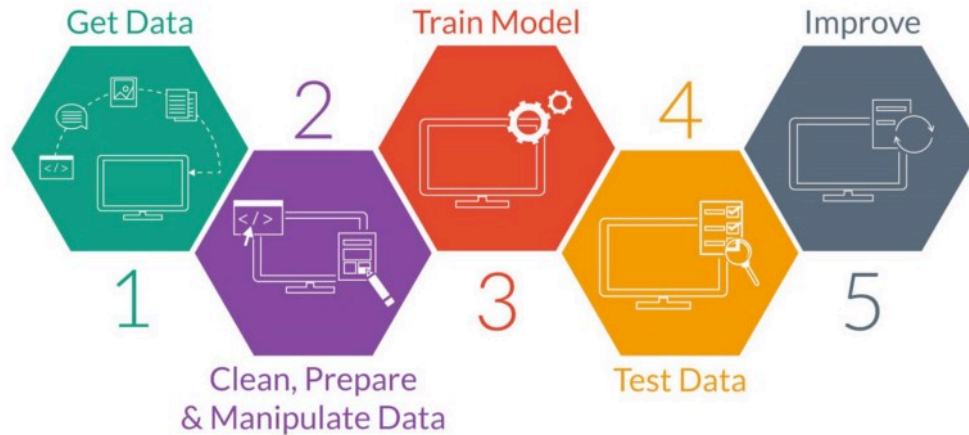
## Objetivos

- ✓ Introdução a Agrupamento;
- ✓ Fluxo de Modelagem.



# Fluxo de Modelagem

- ✓ O processo de modelagem por agrupamento é similar ao processo usado em modelagem de predições;



- ✓ Com um adendo: interpretação do *cluster*.

# Fluxo de Modelagem

---

- ✓ Como *representar* os dados em algoritmos de agrupamentos?
- ✓ Matriz de objetos – similar ao visto em algoritmos preditivos;
- ✓ Matriz de similaridade – cada elemento da matriz  $S$  é dado pela distância  $d(x_i, x_j)$ , ou similaridade  $s(x_i, x_j)$ , entre os objetos  $x_i$  e  $x_j$ ;
- ✓ Grafo de similaridade – diversas topologias, como *minimum spanning tree*.

# Fluxo de Modelagem

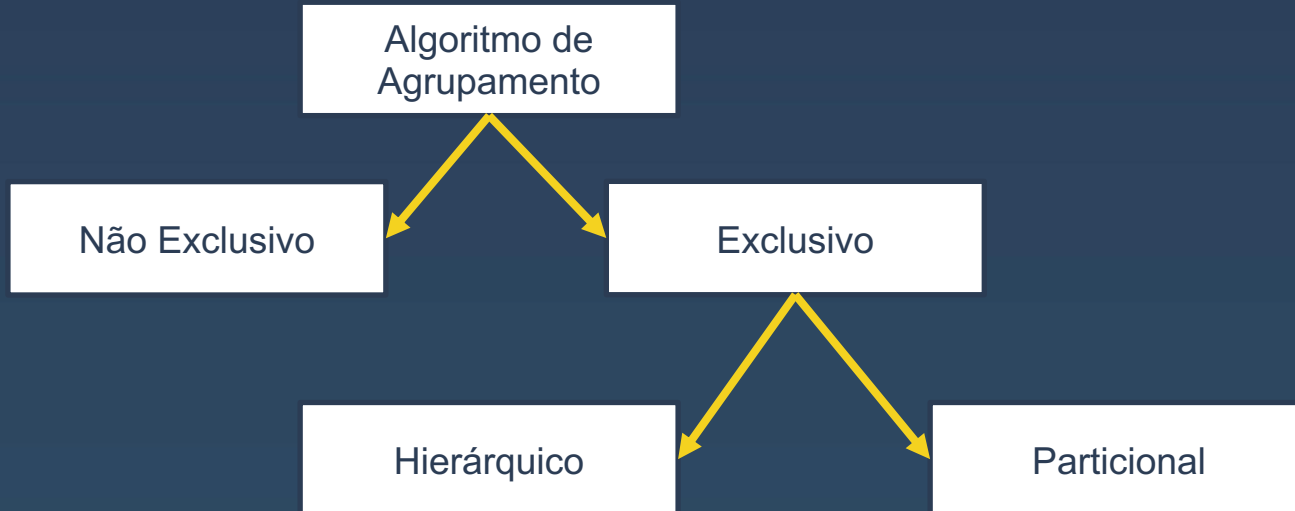
---

- ✓ Possíveis medidas de proximidade/similaridade:
- ✓ Distâncias clássicas (Minkowski, Manhattan, euclidiana);
- ✓ Cosseno;
- ✓ Correlação de Pearson;
- ✓ Distância de Hamming;

# Fluxo de Modelagem

---

- ✓ Na etapa de *agrupamento* (“ajuste”), podemos utilizar dois tipos de estruturas principais.



# Fluxo de Modelagem

---

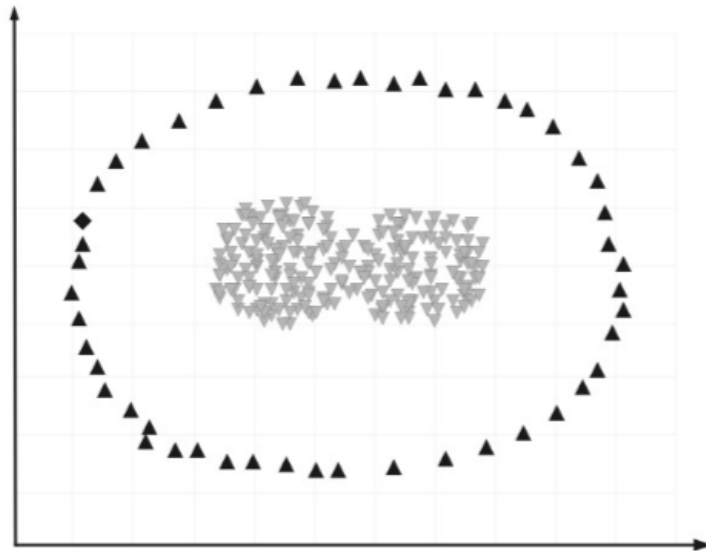
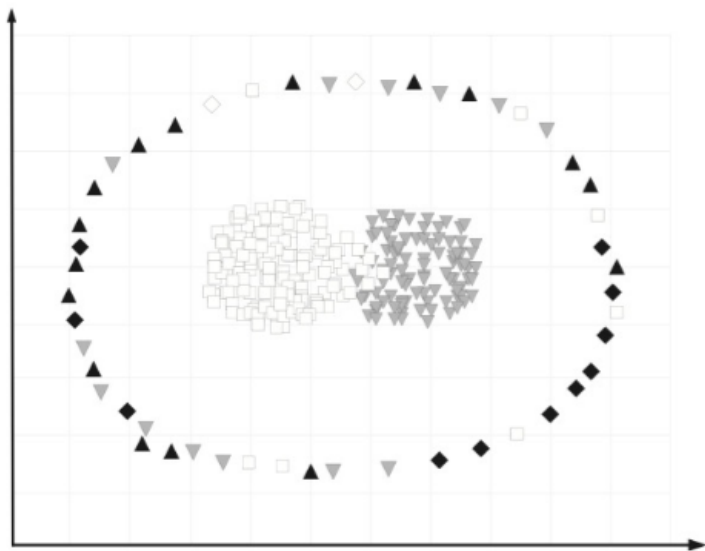
- ✓ Na penúltima etapa, fazemos a *validação* do modelo, determinando se os clusters são significativos;
- ✓ Maioria dos problemas é NP (Non-Deterministic Polynomial time);
- ✓ Todas as abordagens são heurísticas.

*“Diferentes agrupamentos são corretos para diferentes propósitos, assim, não podemos dizer que um agrupamento é melhor.”*

*(Hartigan, 1985).*

# Introdução a Agrupamento

---



# Fluxo de Modelagem

---

- ✓ Finalmente, a *interpretação*;
- ✓ A interpretação de um *cluster* vai além da simples rotulação;
- ✓ O apoio de um especialista é essencial nessa etapa.

# Vamos codar?

---