

# Clustering

Conceitos básicos e K-médias

# Clustering

Aprendizagem não supervisionada

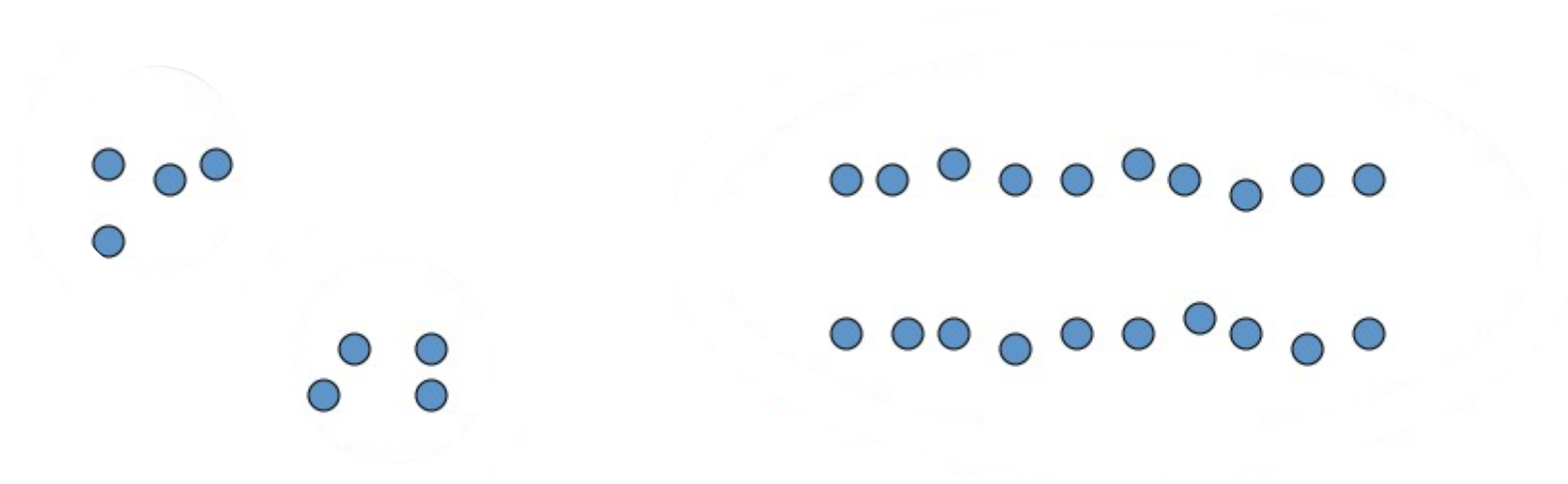
- Requer dados, mas nenhum rótulo
- Detectar padrões, por exemplo em
  - Grupos de e-mails ou resultados de pesquisa
  - padrões de compras do cliente
  - Regiões de imagens
- Útil quando não sabe o que você está procurando

# Clustering

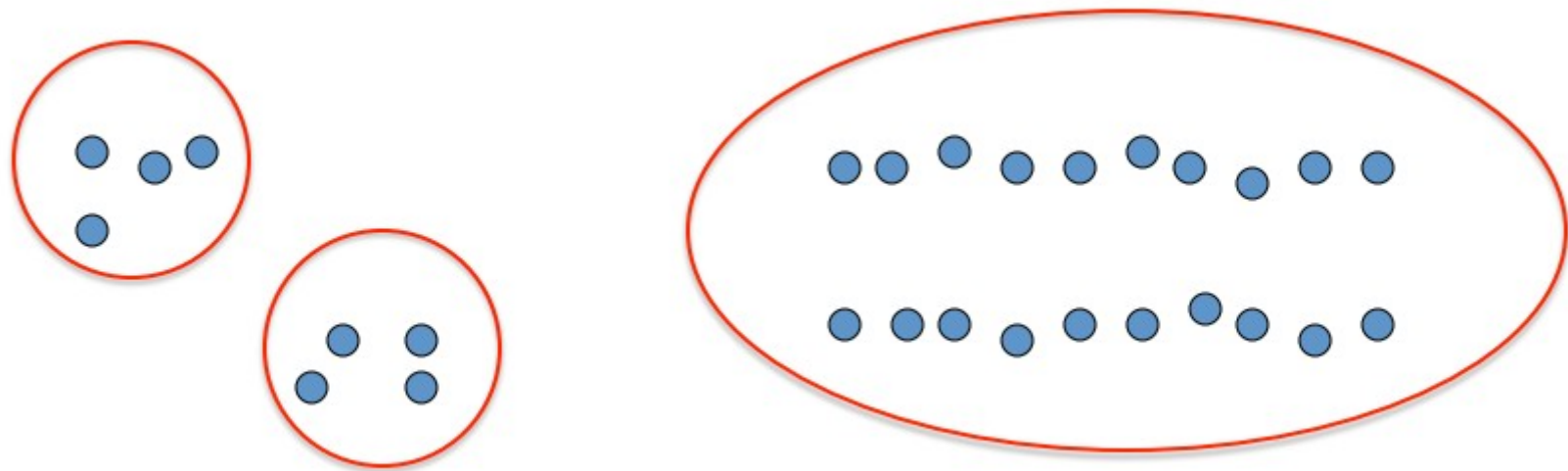
Dado um conjunto de pontos, devemos agrupá-los de maneira que:

- Pontos dentro de um mesmo cluster sejam o mais similar possível;
- Pontos em clusters diferentes sejam o mais diferente possível.

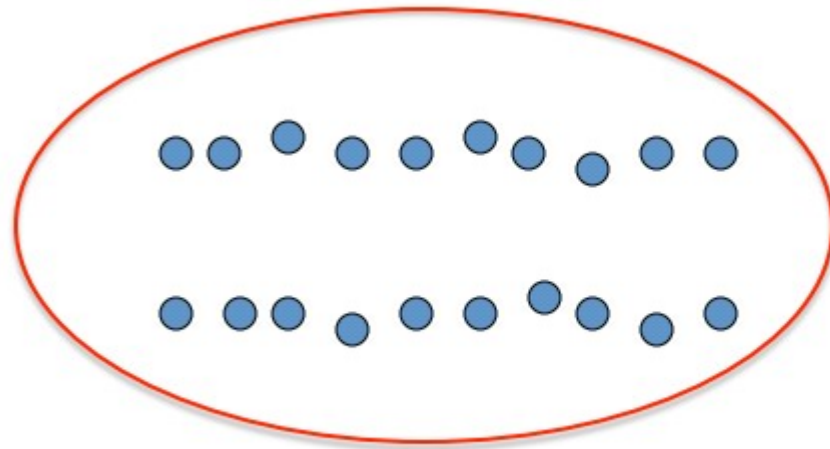
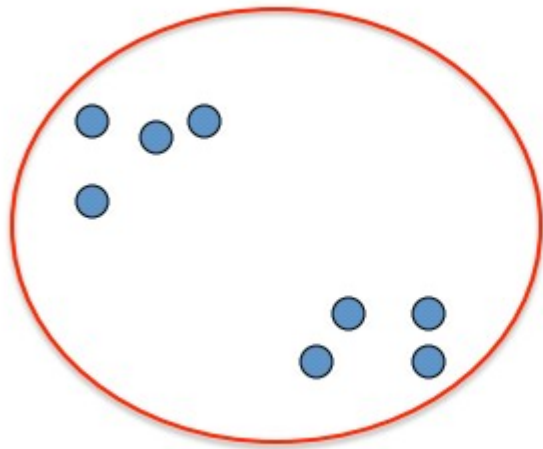
# Clustering - Ideia Básica



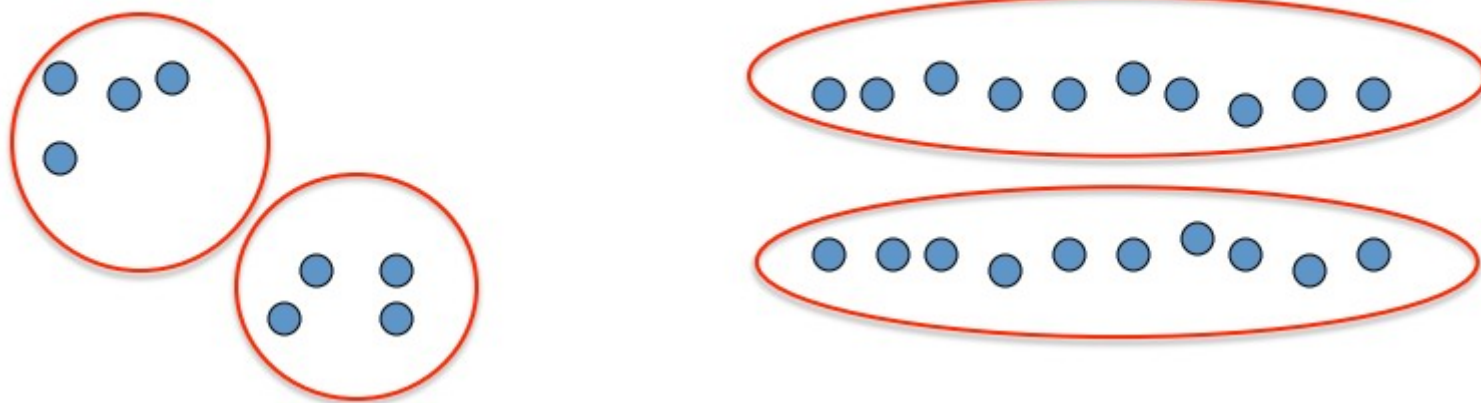
# Clustering - Ideia Básica



# Clustering - Ideia Básica



# Clustering - Ideia Básica



O que significa "semelhante"?

- Uma opção: pequena distância euclidiana (ao quadrado)

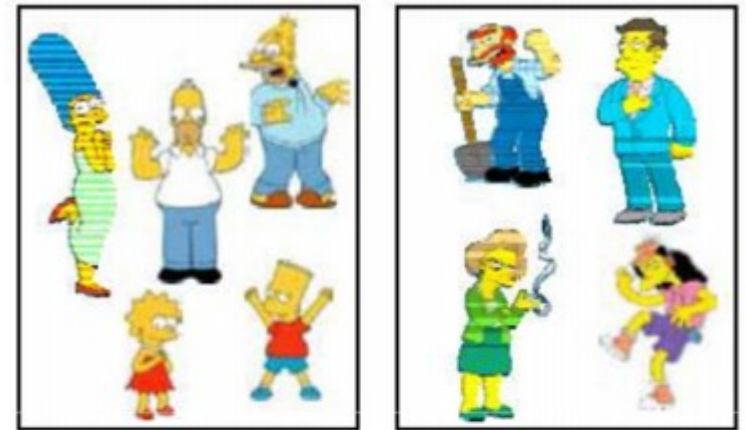
$$\text{dist}(\vec{x}, \vec{y}) = \|\vec{x} - \vec{y}\|_2^2$$

Os resultados de agrupamento dependem crucialmente da medida de similaridade (ou distância) entre "pontos" a serem agrupados

# Clustering - Algoritmos

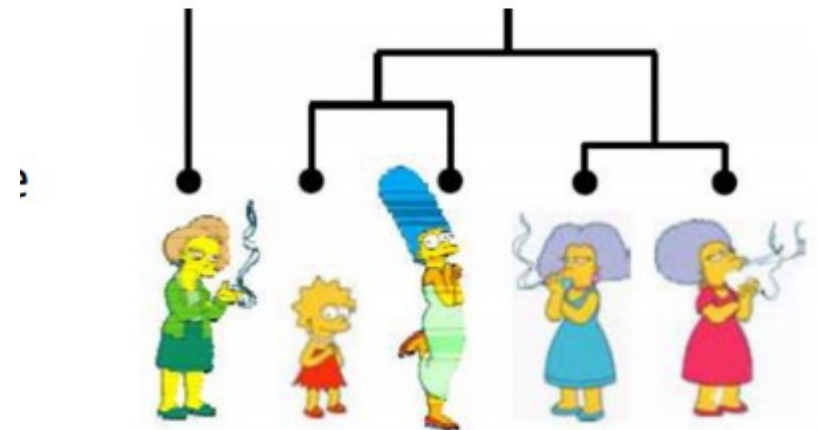
Algoritmos de partição:

- K-means
- Mistura de Gaussianas
- Agrupamento Espectral



Algoritmos Hierárquicos:

- Ward





# Clustering - Exemplos

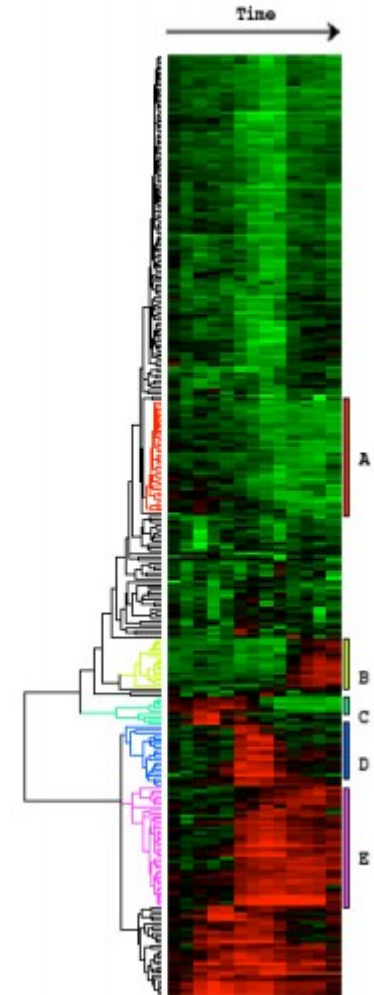
Segmentação de imagens

- Separar a imagem em regiões que tenham algum sentido



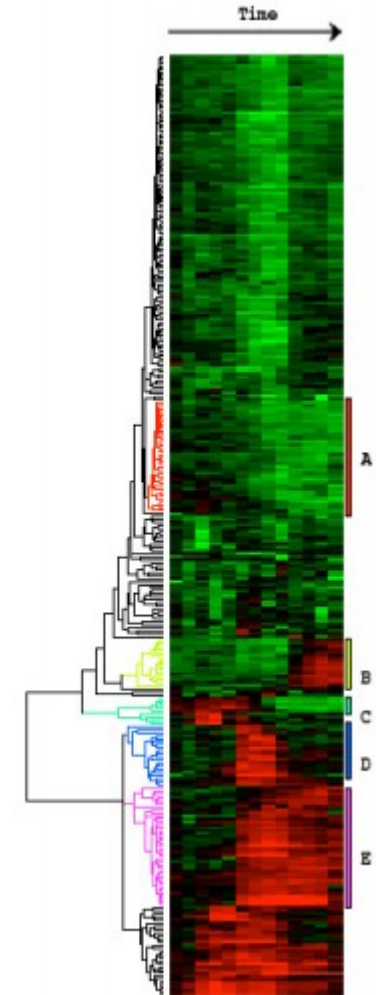
# Clustering - Exemplos

Agrupamento de genes



# Clustering - Exemplos

Agrupamento de genes

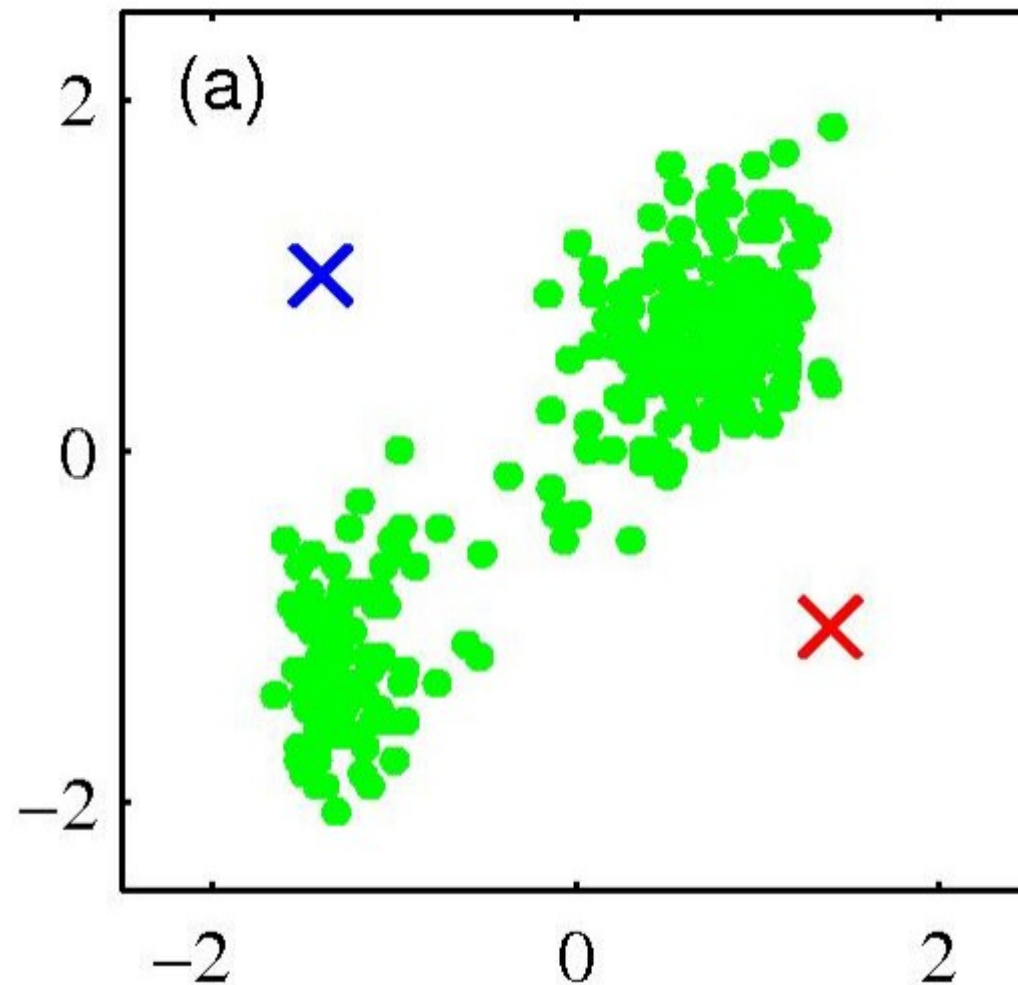


# Clustering - K-médias

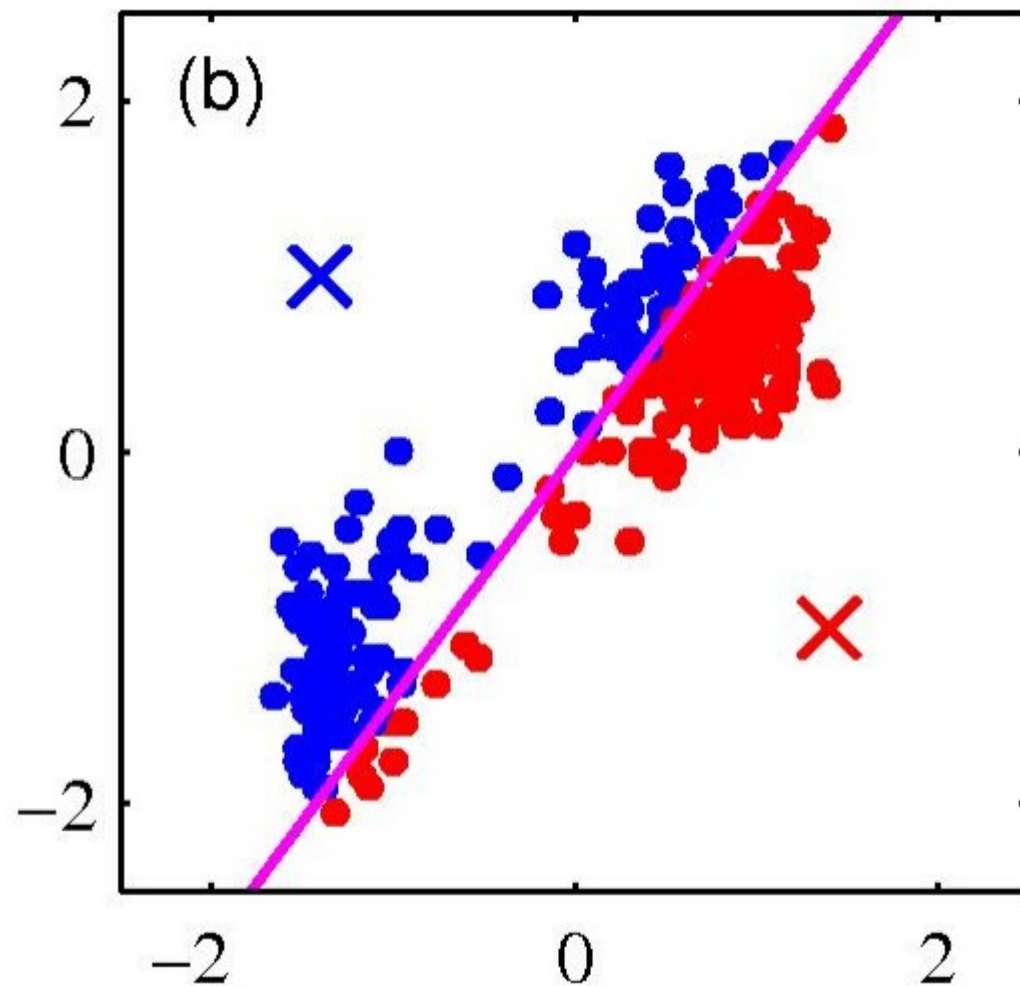
Algoritmo:

- Inicializa K centros
- Associa cada amostra ao centro mais próximo
- Calcula o novo centro do cluster
- Critério de parada: quando os pontos não se alterarem mais

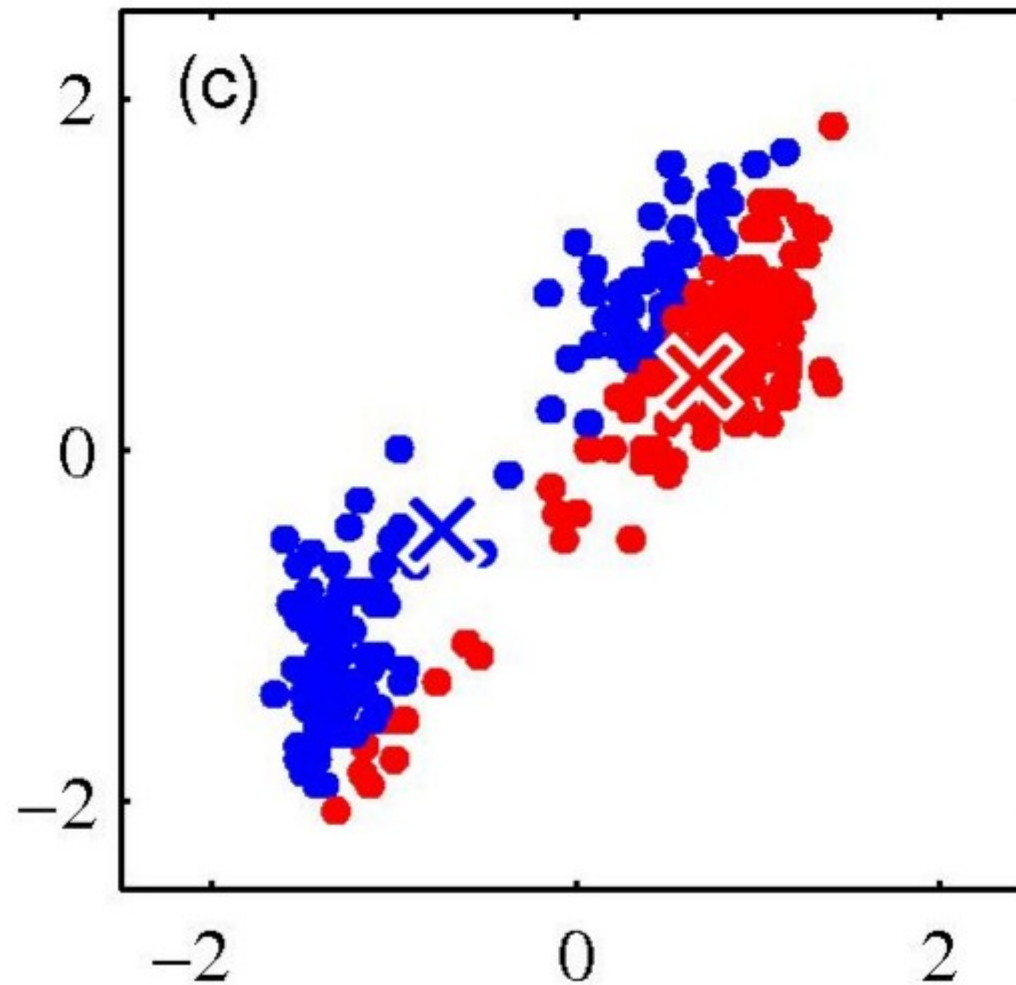
# Clustering - K-médias



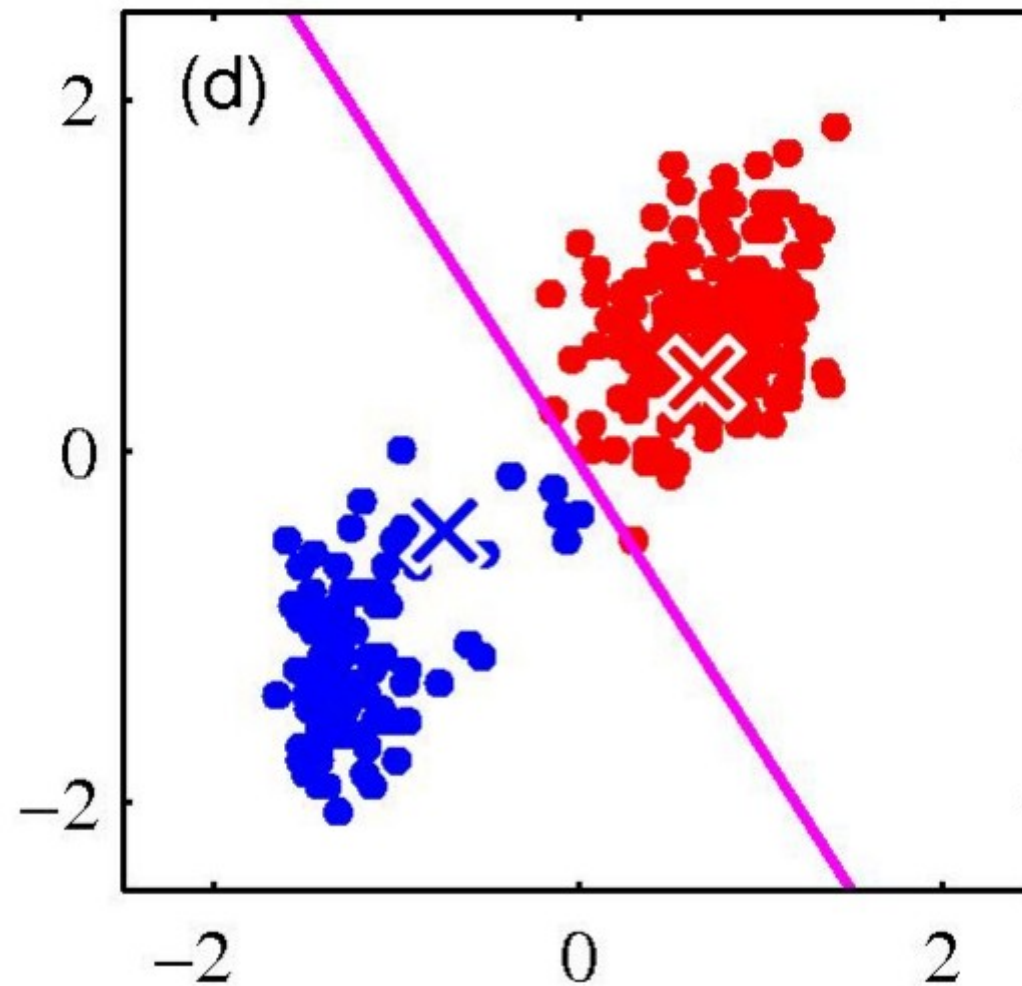
# Clustering - K-médias



# Clustering - K-médias



# Clustering - K-médias





# Clustering - K-médias - Segmentação de imagens

K=2



K=3



K=10



Original

