

Clusterização

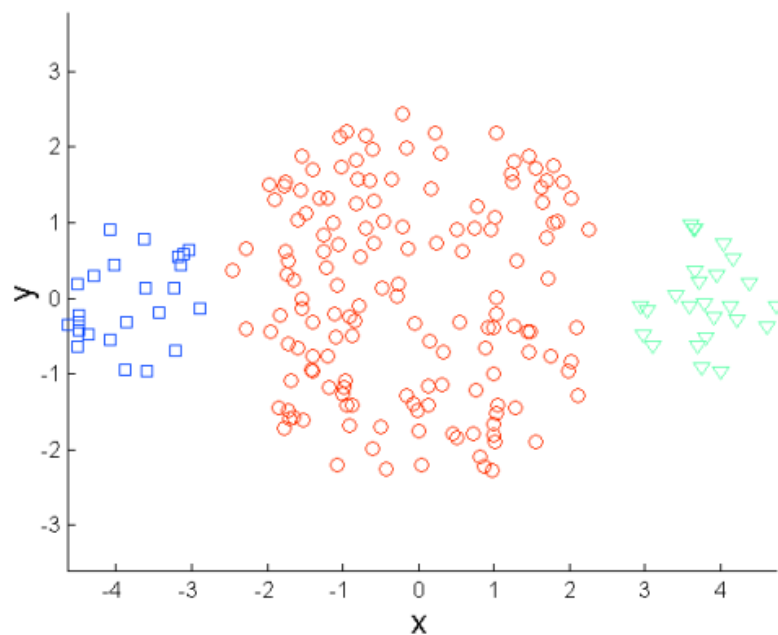
- K-Means:
 - É a técnica mais simples de aprendizagem não supervisionada.
 - Consiste em fixar k centróides (de maneira aleatória), um para cada grupo (clusters).
 - Associar cada indivíduo ao seu centróide mais próximo.
 - Recalcular os centróides com base nos indivíduos classificados.

Clusterização

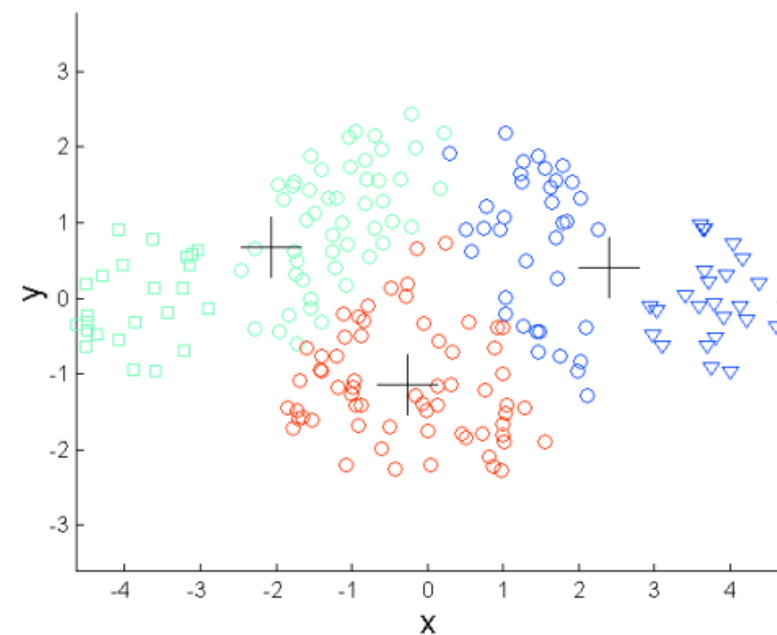
- K-Means:
 - O principal problema do K-Means é a dependência de uma boa inicialização.
 - Algoritmo k-means funciona bem se:
 - Clusters são (hiper)esféricos e bem separados.
 - Clusters de volumes aproximadamente iguais.
 - Cluster com quantidades de pontos semelhantes.

Clusterização

- K-Means:



Estrutura correta



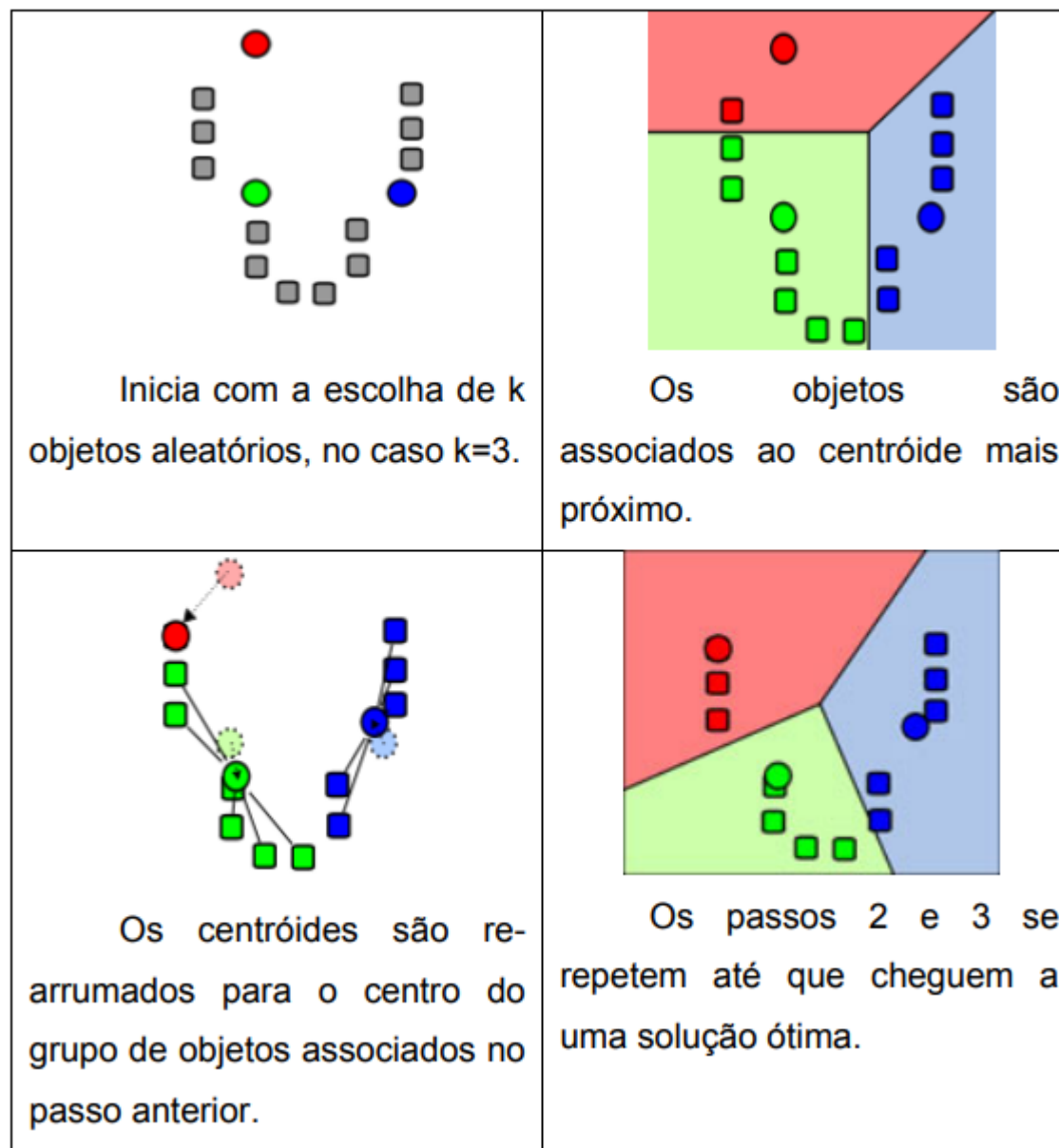
k-means (3 Clusters)

Clusterização

- K-Means:
 - (1) Selecione k centróides iniciais.
 - (2) Forme k clusters associando cada exemplo ao seu centróide mais próximo.
 - (3) Recalcule a posição dos centróides com base no centro de gravidade do cluster.
 - (4) Repita os passos 2 e 3 até que os centróides não sejam mais movimentados.

Clusterização

- K-Means:



Clusterização – Questões de Concurso

- Provas: CESPE / CEBRASPE - 2021 - Polícia Federal - Escrivão de Polícia Federal
- Acerca dos conceitos de mineração de dados, aprendizado de máquina e bigdata, julgue o próximo item.
- A análise de clustering é uma tarefa que consiste em agrupar um conjunto de objetos de tal forma que estes, juntos no mesmo grupo, sejam mais semelhantes entre si que em outros grupos.

Clusterização – Questões de Concurso

- Provas: CESPE / CEBRASPE - 2021 - Polícia Federal - Escrivão de Polícia Federal
- Acerca dos conceitos de mineração de dados, aprendizado de máquina e bigdata, julgue o próximo item.
- A análise de clustering é uma tarefa que consiste em agrupar um conjunto de objetos de tal forma que estes, juntos no mesmo grupo, sejam mais semelhantes entre si que em outros grupos.

Clusterização – Questões de Concurso

- Prova: CESPE - 2013 - SERPRO - Programador de computador
- Julgue os itens que se seguem à luz dos conceitos básicos de datamining e datawarehouse.
- Em algoritmos de clusterização hierárquica, os clusters são formados gradativamente por meio de aglomerações ou divisões de elementos, gerando uma hierarquia de clusters.

Clusterização – Questões de Concurso

- Prova: CESPE - 2013 - SERPRO - Programador de computador
- Julgue os itens que se seguem à luz dos conceitos básicos de datamining e datawarehouse.
- Em algoritmos de clusterização hierárquica, os clusters são formados gradativamente por meio de aglomerações ou divisões de elementos, gerando uma hierarquia de clusters.

Clusterização – Questões de Concurso

- Prova: CESPE - 2013 - SERPRO - Programador de computador
- Julgue os itens que se seguem à luz dos conceitos básicos de datamining e datawarehouse.
- Nos métodos de particionamento para k-clusterização e k-medoids, o elemento que melhor representa o cluster é definido de acordo com seus atributos, sem que haja muita influência dos valores próximos aos limites do cluster.

Clusterização – Questões de Concurso

- Prova: CESPE - 2013 - SERPRO - Programador de computador
- Julgue os itens que se seguem à luz dos conceitos básicos de datamining e datawarehouse.
- Nos métodos de particionamento para k-clusterização e k-medoids, o elemento que melhor representa o cluster é definido de acordo com seus atributos, sem que haja muita influência dos valores próximos aos limites do cluster.

Clusterização – Questões de Concurso

- Prova: CESPE - 2017 - TCE-PE - Analista de Controle Externo - Auditoria de Contas Públicas
- Em relação à análise de agrupamentos (clusterização) em mineração de dados, julgue o item seguinte.
- O método de clustering k-means objetiva particionar 'n' observações entre 'k' grupos; cada observação pertence ao grupo mais próximo da média.

Clusterização – Questões de Concurso

- Prova: CESPE - 2017 - TCE-PE - Analista de Controle Externo - Auditoria de Contas Públicas
- Em relação à análise de agrupamentos (clusterização) em mineração de dados, julgue o item seguinte.
- O método de clustering k-means objetiva particionar 'n' observações entre 'k' grupos; cada observação pertence ao grupo mais próximo da média.

Clusterização – Questões de Concurso

- Prova: CESPE / CEBRASPE - 2013 - SERPRO - Analista - Negócios em Tecnologia da Informação
- Com relação a datamining, datawarehouse, gerenciamento de conteúdo, ambientes colaborativos e portais corporativos, julgue os próximos itens.
- Clusterização é a tarefa preditiva relativa à identificação de um conjunto finito de categorias empregadas para descrever uma informação. Essas categorias nunca poderão ser mutuamente exclusivas.

Clusterização – Questões de Concurso

- Prova: CESPE / CEBRASPE - 2013 - SERPRO - Analista - Negócios em Tecnologia da Informação
- Com relação a datamining, datawarehouse, gerenciamento de conteúdo, ambientes colaborativos e portais corporativos, julgue os próximos itens.
- Clusterização é a tarefa preditiva relativa à identificação de um conjunto finito de categorias empregadas para descrever uma informação. Essas categorias nunca poderão ser mutuamente exclusivas.

Clusterização – Questões de Concurso

- Prova: FGV - 2021 - SEFAZ-ES - Auditor Fiscal da Receita Estadual - Manhã
- Maria está preparando um relatório sobre as empresas de serviços de um município, de modo a identificar e estudar o porte dessas empresas com vistas ao estabelecimento de políticas públicas e previsões de arrecadação. Maria pretende criar nove grupos de empresas, de acordo com os valores de faturamento, e recorreu às técnicas usualmente empregadas em procedimentos de data mining para estabelecer as faixas de valores de cada grupo. Assinale a opção que apresenta a técnica diretamente aplicável a esse tipo de classificação.
- A Algoritmos de associação.
- B Algoritmos de clusterização.
- C Árvores de decisão.
- D Modelagem de dados.
- E Regressão linear.

Clusterização – Questões de Concurso

- Prova: FGV - 2021 - SEFAZ-ES - Auditor Fiscal da Receita Estadual - Manhã
- Maria está preparando um relatório sobre as empresas de serviços de um município, de modo a identificar e estudar o porte dessas empresas com vistas ao estabelecimento de políticas públicas e previsões de arrecadação. Maria pretende criar nove grupos de empresas, de acordo com os valores de faturamento, e recorreu às técnicas usualmente empregadas em procedimentos de data mining para estabelecer as faixas de valores de cada grupo. Assinale a opção que apresenta a técnica diretamente aplicável a esse tipo de classificação.
- A Algoritmos de associação.
- B Algoritmos de clusterização.
- C Árvores de decisão.
- D Modelagem de dados.
- E Regressão linear.

Clusterização – Questões de Concurso

- Prova: CESPE / CEBRASPE - 2022 - DPE-RO - Analista da Defensoria Pública - Administração
- Para a realização de clustering utilizando-se o algoritmo k-médias (k-means), o número mínimo de variáveis ou recursos necessário é
- A 0.
- B 1.
- C 4.
- D 2.
- E 3.

Clusterização – Questões de Concurso

- Prova: CESPE / CEBRASPE - 2022 - DPE-RO - Analista da Defensoria Pública - Administração
- Para a realização de clustering utilizando-se o algoritmo k-médias (k-means), o número mínimo de variáveis ou recursos necessário é
- A 0.
- B 1.
- C 4.
- D 2.
- E 3.