



Machine Learning

K-means

Prof. Hugo de Paula

Algoritmo de Particionamento: K-means

MacQueen'67 – é um dos mais usados.

- Cada cluster é representado por um ponto central.
- Informa-se a quantidade (K) de clusters desejada.
- Variações: k-medóides, k-modas, k-medianas.
- Requer uma medida de distância.

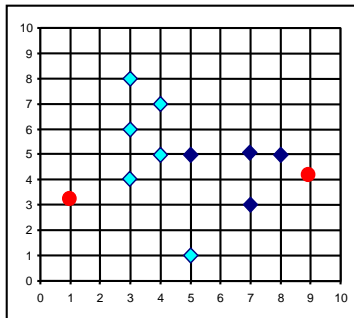
Algoritmo de Particionamento: K-means

- pode encontrar mínimos locais: solução é o *random restart*.
- pode entrar em loop infinito: solução é limitar número de iterações.

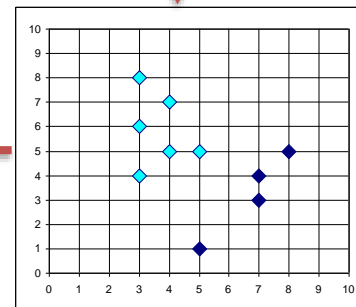
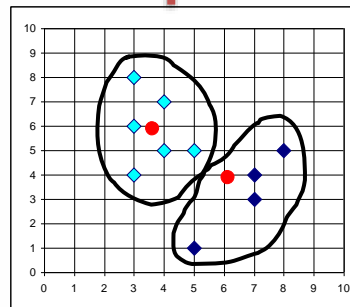
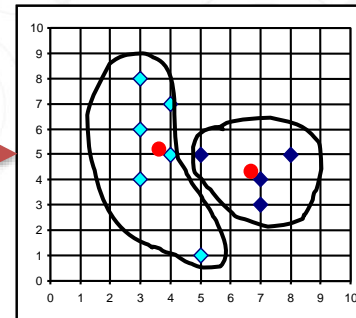
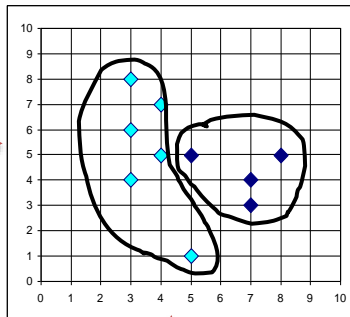
K-means: procedimento

- (1) Escolhe-se arbitrariamente k objetos $\{p_1, \dots, p_k\}$ da base.
 - Estes objetos serão os centros de k clusters
- (2) Para cada objeto **O** diferente da base calcula-se a distância entre **O** e cada um dos p_i 's
- (3) O objeto **O** passa a integrar o cluster representado por p_i mais próximo
- (4) Calcula-se a média dos elementos de cada cluster. Este ponto será o novo representante do cluster.
- (5) Volta para o passo (2) até que nenhum objeto seja realocado para outro cluster.

K-means: Exemplo (loop infinito)



K=2, escolhe-se
arbitrariamente K
elementos para
serem os clusters
iniciais



K-means: Exemplo

Dados = {2, 4, 10, 12, 3, 20, 30, 11, 25}, $k=2$

Centros iniciais, escolhidos aleatoriamente:

$m1 = 3, m2 = 4$

Primeira iteração

– $K1 = \{2, 3\}; m1 = 2.5$

– $K2 = \{4, 10, 12, 20, 30, 11, 25\}; m2 = 16$

K-means: Exemplo

Segunda iteração

- $K1 = \{2, 3, 4\}$; $m1 = 3$;
- $K2 = \{10, 12, 20, 30, 11, 25\}$; $m2 = 18$

Terceira iteração

- $K1 = \{2, 3, 4, 10\}$; $m1 = 4.75$;
- $K2 = \{12, 20, 30, 11, 25\}$; $m2 = 19.6$

K-means: Exemplo

Quarta iteração

- $K1 = \{2, 3, 4, 10, 11, 12\}$; $m1 = 7$; $K2 = \{20, 30, 25\}$; $m2 = 25$

Quinta iteração

- $K1 = \{2, 3, 4, 10, 11, 12\}$; $m1 = 7$; $K2 = \{20, 30, 25\}$; $m2 = 25$
- Sem alteração em relação à quarta iteração, fim do processamento

Aviso legal

O material presente nesta apresentação foi produzido a partir de informações próprias e coletadas de documentos obtidos publicamente a partir da Internet. Este material contém ilustrações adquiridas de bancos de imagens de origem privada ou pública, não possuindo a intenção de violar qualquer direito pertencente à terceiros e sendo voltado para fins acadêmicos ou meramente ilustrativos. Portanto, os textos, fotografias, imagens, logomarcas e sons presentes nesta apresentação se encontram protegidos por direitos autorais ou outros direitos de propriedade intelectual.

Ao usar este material, o usuário deverá respeitar todos os direitos de propriedade intelectual e industrial, os decorrentes da proteção de marcas registradas da mesma, bem como todos os direitos referentes a terceiros que por ventura estejam, ou estiveram, de alguma forma disponíveis nos slides. O simples acesso a este conteúdo não confere ao usuário qualquer direito de uso dos nomes, títulos, palavras, frases, marcas, dentre outras, que nele estejam, ou estiveram, disponíveis.

É vedada sua utilização para finalidades comerciais, publicitárias ou qualquer outra que contrarie a realidade para a qual foi concebido. Sendo que é proibida sua reprodução, distribuição, transmissão, exibição, publicação ou divulgação, total ou parcial, dos textos, figuras, gráficos e demais conteúdos descritos anteriormente, que compõem o presente material, sem prévia e expressa autorização de seu titular, sendo permitida somente a impressão de cópias para uso acadêmico e arquivo pessoal, sem que sejam separadas as partes, permitindo dar o fiel e real entendimento de seu conteúdo e objetivo. Em hipótese alguma o usuário adquirirá quaisquer direitos sobre os mesmos.

O usuário assume toda e qualquer responsabilidade, de caráter civil e/ou criminal, pela utilização indevida das informações, textos, gráficos, marcas, enfim, todo e qualquer direito de propriedade intelectual ou industrial deste material.



PUC Minas
Virtual

© PUC Minas • Todos os direitos reservados, de acordo com o art. 184 do Código Penal e com a lei 9.610 de 19 de fevereiro de 1998.
Proibidas a reprodução, a distribuição, a difusão, a execução pública, a locação e quaisquer outras modalidades de utilização sem a devida autorização da Pontifícia Universidade Católica de Minas Gerais.