



# **Machine Learning**

## **Estruturas de dados para aprendizado não supervisionado**

**Prof. Hugo de Paula**

# Análise de agrupamento: Clustering

## Cluster

- Coleção de objetos que são similares uns aos outros (de acordo com algum critério de similaridade pré-fixado) e dissimilares a objetos pertencentes a outros clusters.

# Análise de agrupamento: Clustering

## **Análise de agrupamento (clustering)**

- Separa os objetos em grupos com base na similaridade, e em seguida atribuir rótulos a cada grupo.

# Aplicações do Clustering

- Distribuição e pré-processamento de dados.
- Proc. de imagens (segmentação);
- Economia e marketing.
- WWW (categorização de documentos, padrões de acesso)
- Agricultura (áreas de uso de terra);
- Planejamento de cidades (agrupar casas de acordo com tipos, valores e localização).

# Análise de agrupamento: Clustering

Qualidade do resultado depende da medida da similaridade usada pelo método.

Requisitos desejáveis:

- Escalabilidade.
- Trata diferentes tipos de atributos.
- Clusters com forma arbitrárias.
- Mínimo conhecimento do domínio.
- Resiliência: valores extremos; ruídos; ordem de processamento.

# Matriz de dados

- Colunas são atributos.
- Linhas são objetos.
- Cada linha é a representação vetorial de um registro.
- $N$  registros e  $P$  atributos:  
matriz  $N \times P$

$$\begin{bmatrix} x_{11} & \dots & x_{1f} & \dots & x_{1p} \\ \dots & \dots & \dots & \dots & \dots \\ x_{i1} & \dots & x_{if} & \dots & x_{ip} \\ \dots & \dots & \dots & \dots & \dots \\ x_{n1} & \dots & x_{nf} & \dots & x_{np} \end{bmatrix}$$

# Matriz de distância (simétrica)

- $N$  registros: matriz  $N \times N$
- Distância entre 2 elementos
- Matriz triangular

$$\begin{bmatrix} 0 & & & & \\ d(2,1) & 0 & & & \\ d(3,1) & d(3,2) & 0 & & \\ \vdots & \vdots & \vdots & & \\ d(n,1) & d(n,2) & \dots & \dots & 0 \end{bmatrix}$$

# 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.