

Inteligência Artificial

Aula 25- Aprendizagem de Máquina: Agrupamento ¹

Sílvia M.W. Moraes

Faculdade de Informática - PUCRS

June 12, 2017

¹Este material não pode ser reproduzido ou utilizado de forma parcial sem a permissão dos autores.

Sinopse

- Nesta aula, continuamos a falar sobre **aprendizagem de máquina**.
- Este material foi construído com base no material sobre Data Mining dos professores Rodrigo Barros, Duncan e Renata de Paris e também nos capítulos:
 - 11 - Inteligência Artificial: Uma abordagem de Aprendizagem de Máquina: Facelli e outros.
 - 10 do livro Inteligência Artificial: Luger
 - 18 do livro Artificial Intelligence – a Modern Approach: Russel & Norvig

Sumário

- 1 O que vimos ...
- 2 Revisando: Paradigmas, Tarefas e Processo de Aprendizagem
- 3 Agrupamento

Aulas anteriores

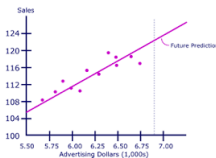
- Agente Reativos e Cognitivos
- Solução de Problemas: Algoritmos de busca
- Planejamento Clássico
- Introdução à Raciocínio Probabilístico
- Introdução à Aprendizagem de Máquina
 - Pré-processamento

Paradigmas e Tarefas de Aprendizagem

- **Paradigma de aprendizagem** é definido pela natureza do problema. Tipo de realimentação usada pelo algoritmo para aprender.
 - Podem ser:
 - **Supervisionado**: aprendizagem de uma função h a partir de exemplos (amostras rotuladas), de entradas (x) e saídas correspondentes ($f(x)$). Com crítica referente ao erro da saída.
 - **Não-supervisionado**: aprendizagem a partir de as amostras não são rotuladas. Essa abordagem não usa os atributos de saída. Sem crítica, usa regularidades e propriedades estatísticas dos dados.
 - **Por reforço**: processo de aprendizagem baseado em punição e recompensa. Reforça uma ação positiva e penaliza, uma negativa. Crítica apenas de desempenho.

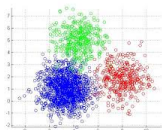
Paradigmas e Tarefas de Aprendizagem

- As **tarefas de aprendizagem** podem ser: **preditivas** ou **descritivas**
 - **preditivas**: tarefa supervisionada, sua meta é encontrar uma função (modelo ou hipótese) a partir dos dados de treino que possa ser usada para prever um rótulo (classe) ou valor de um novo exemplo.
 - Ex: **classificação** (rótulos discretos), **regressão** (rótulos contínuos)



Paradigmas e Tarefas de Aprendizagem

- As tarefas de aprendizagem podem ser: **preditivas** ou **descritivas**
 - descritivas**: tarefa não supervisionada, sua meta é explorar ou descrever um conjunto de dados. (não usam atributos de saída)
 - Ex: **agrupamento** (divisão em grupos baseada em similaridade), **sumarização** (descrição simples e compacta), **associação** (relações frequentes entre dados)

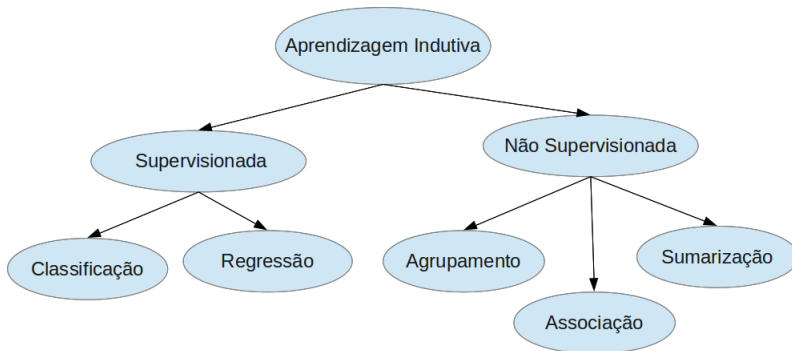


Long Article	
before	after
	Summary
	=====
	TID Produtos Comprados

	1 biscoito, cerveja, chá, salaminho
	2 cerveja, couve, linguiça, pão, queijo
	3 café, brócolis, couve, pão
	4 brócolis, café, cerveja, couve, pão, salaminho
	5 brócolis, café, couve, pão, refrigerante
	6 couve, linguiça

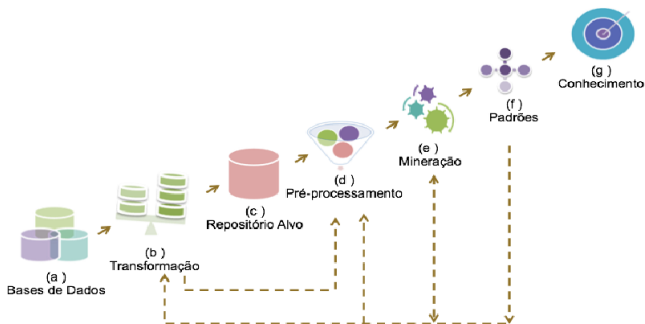
Paradigmas e Tarefas de Aprendizagem

- Resumo:



Processo de Descoberta de Conhecimento

- **Knowledge Discovery in Databases (KDD)**: consiste em uma série de passos bem definida cujo meta é transformar dados em conhecimento.

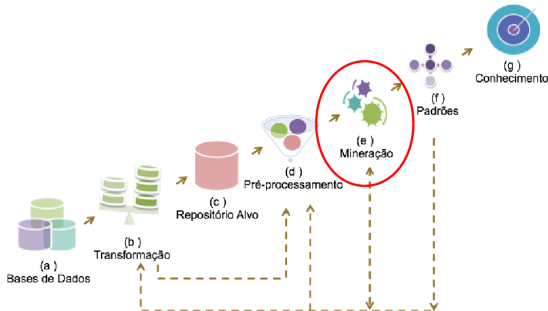


Processo de Descoberta de Conhecimento

- Knowledge Discovery in Databases (KDD):

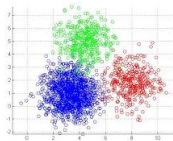
- (e) Mineração :

- Usa Algoritmos de aprendizado de máquina
- Análise de uma séries de dados para compreensão do domínio
- Resultados compreensíveis e especialmente úteis



Conceito

- **Objetivo:** organizar dados (não classificados, sem rótulos) em grupos de acordo com alguma medida de similaridade, tal que exista:
 - Alta similaridade intra-grupo.
 - Baixa similaridade entre grupos.



Características

- **Características:**

- Técnica aplicada para organizar os dados **quando não há classe para prever**.
- **Grupos: formados por dados (objetos) que compartilham características** (podem ser mais genéricos ou mais especializados, diferentes níveis de refinamento).
- **Pode ser** usado como uma **etapa anterior a alguma tarefa**, como por exemplo: sumarização.



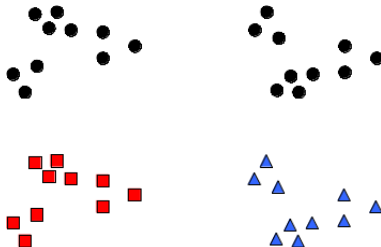
Grupos: Níveis de Refinamento

- Quantos grupos há nessa imagem ?



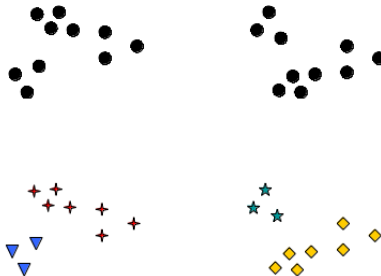
Grupos: Níveis de Refinamento

- Quantos grupos há nessa imagem ? 2.



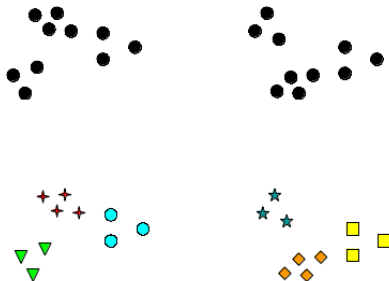
Grupos: Níveis de Refinamento

- Quantos grupos há nessa imagem ? 4.



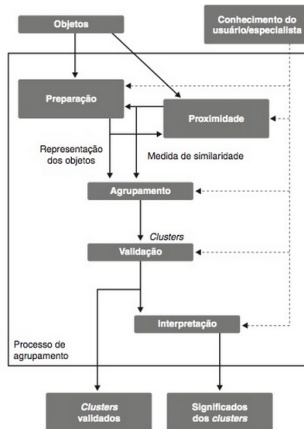
Grupos: Níveis de Refinamento

- Quantos grupos há nessa imagem ? 6.



Etapas do Processo de Agrupamento

- O **processo de agrupamento** inclui as etapas: preparação, proximidade, agrupamento, validação e interpretação.

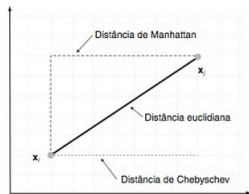


Etapas do Processo de Agrupamento

- O **processo de agrupamento** inclui as etapas: preparação, proximidade, agrupamento, validação e interpretação.
 - **Preparação:** inclui pré-processamento (ex: normalizações, conversão de tipos e redução de dimensionalidade) e forma de representação dos dados (ex: matriz de similaridade) para que o algoritmo de agrupamento possa ser usado.
 - **Proximidade:** definição de medidas de proximidade apropriadas ao domínio e ao tipo de informação que se deseja extrair dos dados.
 - Existem medidas para atributos quantitativos e qualitativos.

Etapas do Processo de Agrupamento: Proximidade

- **Proximidade - Medidas para Atributos Quantitativos:**
 - Medidas de Distância: atributos contínuos e racionais:
 - Manhattan: $d(x_i, x_j) = \sum_{l=1}^d |x_i^l - x_j^l|$ (usual para binários)
 - Euclidiana: $d(x_i, x_j) = \sqrt{\sum_{l=1}^d (x_i^l - x_j^l)^2}$
 - Chebyshev (ou supremum): $d(x_i, x_j) = \max_{1 \leq l \leq d} |x_i^l - x_j^l|$

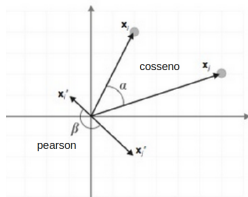


Etapas do Processo de Agrupamento: Proximidade

- **Proximidade** - Medidas para **Atributos Quantitativos**:
 - Medidas de Similaridade:

- Separação angular (cosseno): $\cos(x_i, x_j) = \frac{\sum_{l=1}^d x_i^l x_j^l}{\sqrt{\sum_{l=1}^d x_i^2 \sum_{l=1}^d x_j^2}}$

- Pearson:
$$\rho(x_i, x_j) = \frac{\sum_{l=1}^d (x_l^i - \bar{x}_i)(x_l^j - \bar{x}_j)}{\sqrt{\sum_{l=1}^d (x_l^i - \bar{x}_i)^2 \sum_{l=1}^d (x_l^j - \bar{x}_j)^2}}$$
 (quando magnitude não é importante, mas sim o grau de variação. Ex: Bioinformática)



Etapas do Processo de Agrupamento: Proximidade

- **Proximidade** - Medidas para **Atributos Qualitativos**:

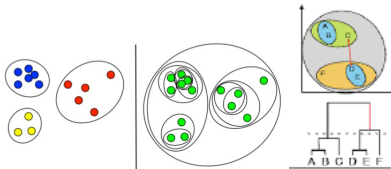
- São obtidas a partir da soma das contribuições individuais de todos os atributos.
- Para atributos nominais, a distância mais usada é a de Hamming.

$$d(x_i, x_j) = \sum_{l=1}^d a(x_i^l, x_j^l) \text{ , onde } a(x_i^l, x_j^l) = \begin{cases} 1 & \text{se } x_i^l \neq x_j^l \\ 0 & \text{c.c} \end{cases}$$

Etapas do Processo de Agrupamento

- O **processo de agrupamento** inclui as etapas: ...
 - **agrupamento**: nessa etapa um ou mais algoritmos de agrupamento são usados para gerar os grupos.
 - **validação**: etapa que verifica se os grupos gerados são significativos. Ajuda a determinar o número adequado de grupos, quando esse número não é conhecido.
 - **interpretação**: processo de examinar o grupo em relação aos outros com o objetivo de rotulá-lo, indicando a natureza do grupo.

Tipos de Agrupamentos



• Agrupamento Particional

- Divisão dos objetos de dados em subconjuntos (grupos) sem sobreposição tal que cada objeto de dados está em exatamente um único grupo.

• Agrupamento Hierárquico

- Um conjunto de grupos aninhados na forma de uma árvore hierárquica.

Agrupamento Particional: Algoritmo k-Means

- **k-means:** Algoritmo de agrupamento particional.
 - Características:
 - Cada grupo está associado a um centroide (objeto central).
 - Cada objeto é atribuído ao grupo com o centroide mais próximo.
 - Número de grupos (k) deve ser especificado

1. Selecione k objetos como centroides iniciais.
2. Repita
- 3 Forme k agrupamentos vinculando todos os objetos aos centroides mais próximos.
- 4 Recalcule o centroide de cada agrupamento.
5. Até que os centroides não mudem.

Agrupamento Particional: Algoritmo k-Means

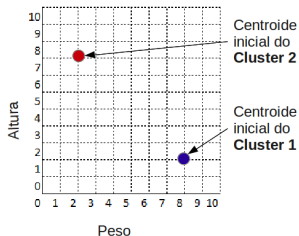
- Características:
 - Centroides iniciais são geralmente aleatórios.
 - Agrupamento varia conforme a inicialização.
 - Centroides são (tipicamente) a média de todos os objetos do grupo.
 - A medida de distância geralmente empregada é a distância Euclidiana.
 - k-means geralmente converge com poucas iterações.
 - Complexidade é $O(n \times k \times i \times d)$, onde n = número de objetos, k = número de grupos, i = número de iterações e d = número de atributos

Agrupamento Particional: Algoritmo k-Means

Exemplo:

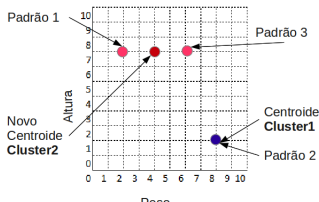
- Considere o conjunto de dados abaixo, o qual possui 6 registros de peso(atributo1) e altura (atributo2) normalizados para o intervalo 0 a 10.
- Supondo que o Padrão1 é o centroide inicial do cluster1 e o Padrão2 é o centroide inicial do cluster2, quais serão os valores dos centroides dos dois clusters, ao final da execução do algoritmo?

Padrão	Peso(Atributo1)	Altura(Atributo2)
1	2	8
2	8	2
3	6	8
4	2	7
5	8	4
6	2	6



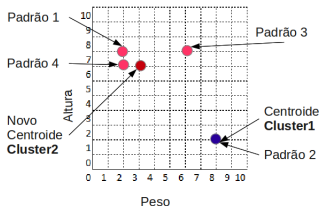
Agrupamento Particional: Algoritmo k-Means

- Considerando, a distância euclidiana a medida de proximidade:
 - E sabendo que o centróide do cluster1 (C1) é definido por (8,2) e do cluster2 (C2) é (2,8)
 - Analisando o **Padrão 3**: (6,8)
 - 1 Calcula a proximidade do padrão em relação aos centróides:
distância de C1: $\sqrt{(8-6)^2 + (2-8)^2} = 6,32$
distância de C2: $\sqrt{(2-6)^2 + (8-8)^2} = 4$
 - 2 Recalcula o centróide do mais próximo ao padrão, usando média $\bar{x}(k) = \frac{1}{n_k} \sum_{x_i \in C_k} x_i$:
 $C2 = (\frac{2+6}{2}, \frac{8+8}{2}) = (4,8)$



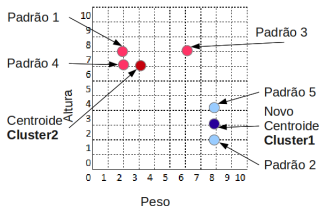
Agrupamento Particional: Algoritmo k-Means

- Considerando, a distância euclidiana a medida de proximidade:
 - E sabendo que novo centróide do cluster1 (C1) é definido por (8,2) e do cluster2 (C2) é (4,8)
 - Analizando o **Padrão 4**: (2,7)
 - 1 Calcula a proximidade do padrão em relação aos centróides:
distância de C1: $\sqrt{(8-2)^2 + (2-7)^2} = 7,81$
distância de C2: $\sqrt{(4-2)^2 + (8-7)^2} = 2,23$
 - 2 Recalcula o centróide do mais próximo ao padrão, usando média:
$$C2 = \left(\frac{2+6+2}{3}, \frac{8+8+7}{3} \right) = (3,7)$$



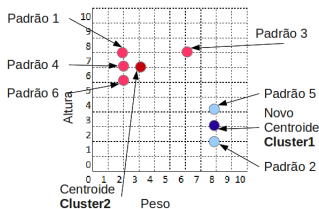
Agrupamento Particional: Algoritmo k-Means

- Considerando, a distância euclidiana a medida de proximidade:
 - E sabendo que novo centróide do cluster1 (C1) é definido por (8,2) e do cluster2 (C2) é (3,7)
 - Analisando o **Padrão 5**: (8,4)
 - Calcula a proximidade do padrão em relação aos centróides:
distância de C1: $\sqrt{(8-8)^2 + (2-4)^2} = 2$
distância de C2: $\sqrt{(3-8)^2 + (7-4)^2} = 5,83$
 - Recalcula o centróide do mais próximo ao padrão, usando média:
 $C1 = (\frac{8+8}{2}, \frac{2+4}{2}) = (8,3)$



Agrupamento Particional: Algoritmo k-Means

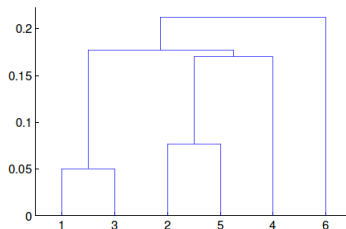
- ...E sabendo que centróide do cluster1 (C1) é (8,3) e do cluster2 (C2) é (3,7).
 - Analisando o **Padrão 6**: (2,6)
 - 1 Calcula a proximidade do padrão em relação aos centróides:
distância de C1: $\sqrt{(2-8)^2 + (6-3)^2} = 6,70$
distância de C2: $\sqrt{(2-3)^2 + (6-7)^2} = 1,41$
 - 2 Recalcula o centróide do mais próximo ao padrão, usando média:
 $C2 = (\frac{2+6+2+2}{4}, \frac{8+8+7+6}{4}) = (3,7)$



Repete o processo até os clusters estabilizarem.

Agrupamento Hierárquico

- Produz um conjunto de **grupos aninhados, organizados como uma árvore**.
- Pode ser visualizado por um **dendograma**.
 - Uma árvore como um diagrama que mostra as sequências de combinações ou partições.



Agrupamento Hierárquico

- **Vantagens** desse tipo de agrupamento:
 - Não é necessário assumir um número particular de grupos
 - Qualquer número de grupos desejado pode ser obtido ao 'cortar' o dendograma no nível apropriado.
 - Podem corresponder a taxonomias úteis. Exemplos em ciências biológicas (e.g., reino animal, reconstrução filogenética, ...)

Agrupamento Hierárquico

- Tipos: **Aglomerativo** ou **por Divisão**
 - **Aglomerativo:**
 - Inicia com cada objeto representando um grupo individual.
 - A cada passo, combina o par mais próximo de grupos, até que somente um grupo (ou k grupos) reste ou algum critério de parada seja atingido.
 - **Por Divisão:**
 - Inicia com todos os objetos representando um único grupo.
 - A cada passo, divide o grupo até que cada grupo contenha um objeto (ou k grupos) ou algum critério de parada seja atingido.
- Algoritmos hierárquicos tradicionais usam a matriz de similaridade ou distância.

Agrupamento Hierárquico

- Os algoritmos aglomerativos são mais usados

Algoritmo Básico

Calcular a matriz de distância entre os dados

Cada ponto inicialmente é considerado um cluster

Repetir

 Juntar os dois clusters mais próximos

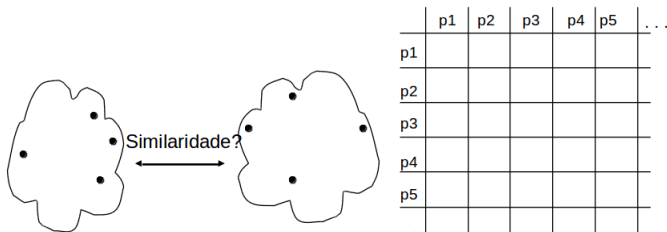
 Atualizar a matriz de distâncias

Até que reste apenas um cluster

- A operação-chave é a distância entre os dois clusters
- A diferença entre os algoritmos que seguem essa abordagem está justamente no cálculo dessa distância.

Agrupamento Hierárquico

- Similaridade baseada nas distâncias entre os elementos ou centróides dos clusters.
- Em geral usam uma **matriz de similaridade**.

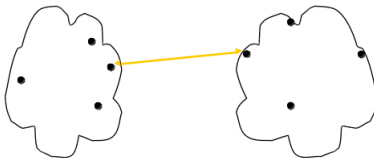


- **Tipos:**
 - Single-link , Complete-link, Average-link, Distance Between Centroids e outros.

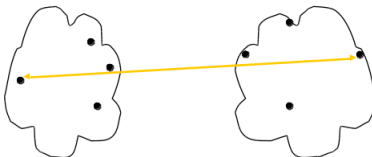
Agrupamento Hierárquico

- Tipos:

- Single-link** : baseado na **menor distância** entre dois clusters C1 e C2 (usa os dados mais próximos, um de cada cluster).



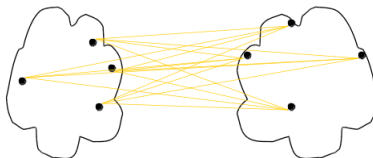
- Complete-Link**: baseado na **maior distância** entre dois clusters C1 e C2 (usa os dados mais afastados, um de cada cluster).



Agrupamento Hierárquico

- Tipos:

- **Average-link:** baseado na **distância média** entre dois clusters C1 e C2.



- **Distance Between Centroids:** baseado na **distância** entre os centróides de dois clusters C1 e C2.

