

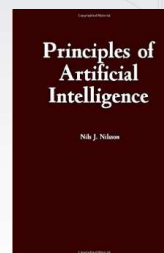
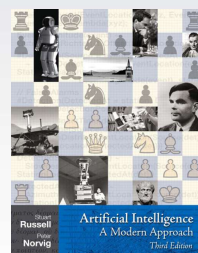
Inteligência Artificial

Sistemas de Informação

Cleiton Garcia <cleitonsg@catolicasc.org.br>



Referências



10/04/2018

Prof. Msc. Cleiton Garcia

2



Agenda

- Relembrar: Aprendizagem supervisionada
 - Árvore de decisão
 - K-NN
- Aprendizagem não supervisionada

10/04/2018

Prof. Msc. Cleiton Garcia

3



Introdução

- No aprendizado **supervisionado**, todos os exemplos de treinamento são **rotulados (classe definida por especialista)**.

0.51 0.14 0.12 0.04 0.65 0.01 0.08 2

Vetor de Atributos

Classe

- Estes exemplos são ditos “supervisionados”, pois, contém tanto a **entrada** (atributos), quanto a **saída** (classe).
- Muito utilizado em *data mining*, *buscar agrupamentos (clusters) consistentes*.

10/04/2018

Prof. Msc. Cleiton Garcia

4



Introdução

- Os **algoritmos de agrupamento** buscam identificar padrões existentes em conjuntos de dados.
- Os algoritmos de agrupamento podem ser divididos em categorias, tais como:
 - Baseados em Hierarquia
 - pairwise distance matrix
 - Baseados em partição / otimização funções custo
 - K-Means / SOM
 - Baseados em densidade
 - Outros: Fuzzy, LVQ...

10/04/2018

Prof. Msc. Cleiton Garcia

5



K-Means

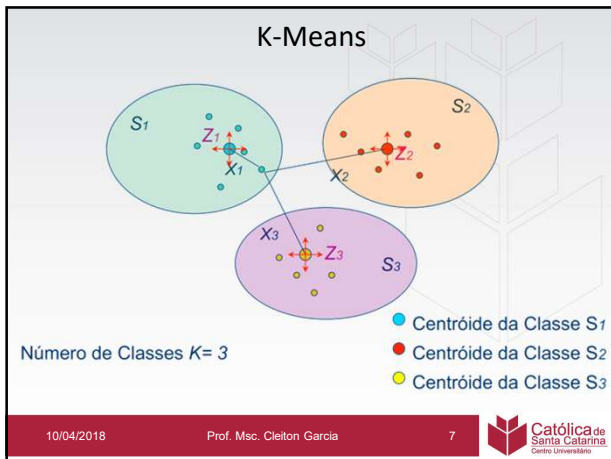
- O algoritmo k-means baseia-se na minimização de uma medida de custo, a distância interna entre os padrões de um agrupamento;
- A minimização do custo garante encontrar um mínimo local da função objetivo, que dependerá do ponto inicial do algoritmo;
- Segue uma abordagem não supervisionada baseada em partições

10/04/2018

Prof. Msc. Cleiton Garcia

6





K-means – meta

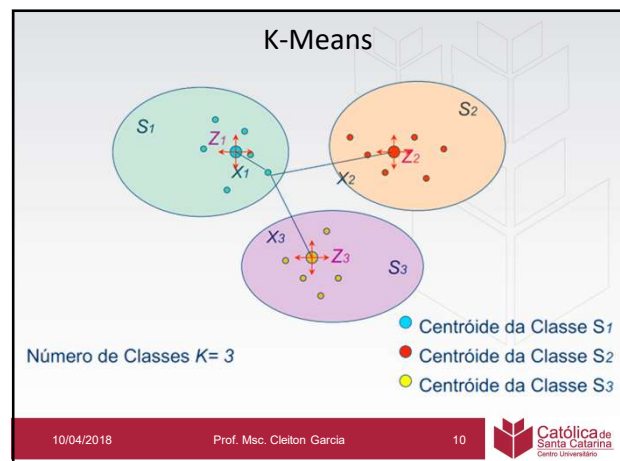
- A classificação de uma amostra desconhecida é feita com base nos *centroid*
 - Para cada *centroid* que forma as classe W_k
 - Calcule a distância entre a amostras e o centroid
 - Escolher a de menor distância

10/04/2018 Prof. Msc. Cleiton Garcia 8

K-means - passo a passo

- Inicialização - Selecionar arbitrariamente os k centros iniciais de agrupamentos (*centroids*), Z_1, Z_2, \dots, Z_k (sementes);
- Determinar a distância X_i , com $i = 1, 2, \dots, k$, entre cada *centroid* e as demais instâncias;
- Identificar o *centroid* mais próximo a cada padrão (*instância*);
- Atualizar os centros de agrupamento (*mean* dos mais próximos);
- Enquanto centroid possuir mudança significativa, repetir a partir do passo 2.

10/04/2018 Prof. Msc. Cleiton Garcia 9



K-means – variações

- Existem variações deste algoritmo que empregam operações de particionamento, união e descarte dos *clusters* resultantes.
- Exemplos de variações na inicialização:
 - Inicializar com sementes aleatórias;
 - Calcular todas as distâncias e buscar os K mais distantes;
 - Semi-supervisionado onde você determina as sementes iniciais. Exemplo clicar em pontos da imagem.

10/04/2018 Prof. Msc. Cleiton Garcia 11

K-means - comparativo

- Vantagens**
 - Simplicidade e eficiência
 - Não demanda uma base rotulada o que pode custar muito tempo, esforço, dinheiro;
 - Podemos utilizar dados **não rotulados** para encontrar padrões existentes e depois avaliar/corrigir os rótulos (semi-supervisionado...);
- Desvantagens**
 - Dependência dos valores iniciais de k (classes);
 - Escolha dos primeiros centros de agrupamento;
 - Dependência da geometria das amostras disponíveis para análise

10/04/2018 Prof. Msc. Cleiton Garcia 12

Exercícios

1. Dado o conjunto de exemplos, referente aos valores RGB de uma imagem. Efetue o agrupamento utilizando o K-means.

R	G	B
1	10	200
2	20	230
6	25	150
7	45	100
10	50	125
3	24	111
100	4	10
250	7	50
243	5	68
210	2	90
200	1	95
215	0	68
56	200	1
79	234	3
80	210	8
95	200	10
80	210	4
49	207	1

10/04/2018

Prof. Msc. Cleiton Garcia

13

