

## Ciência da Computação


### Exercícios de Inteligência Artificial

Cristiane Neri Nobre

Data de entrega: até dia 29/09

1. Considere o conjunto de 8 instâncias da tabela abaixo, caracterizadas por dois atributos (X e Y). Aplique o algoritmo K-means para determinar uma partição adequada dessas instâncias em 2 grupos (K=2). Realize a primeira execução do algoritmo considerando como centroides as instâncias 3 (Ex3 -> X=0 e Y=2) e 4 (Ex4 -> X=0 e Y = -2). Utilize a distância Manhattan para o cálculo de distância.

Exemplo	X	Y
Ex1	-5	2
Ex2	5	-2
Ex3	0	2
Ex4	0	-2
Ex5	-5	1
Ex6	-5	-2
Ex7	5	2
Ex8	5	-1

2. Abra a base de dados Iris.arff no WEKA (ou em qualquer outra ferramenta de aprendizado de máquina) e rode o Kmeans. A partir disso, pede-se:
  - a) Rode o algoritmo K-means com K=3
  - b) Veja visualmente os grupos obtidos. O que aconteceu com as instâncias? Houve erro de agrupamento? Por quê?
  - c) Agora marque a opção  **Classes to clusters evaluation** (lado esquerdo da tela), desconsidere o atributo de classe e rode o Kmeans novamente. Avalie novamente os agrupamentos encontrados. Comente sobre os resultados encontrados.
3. Abra a base de dados Iris.arff no WEKA (ou em qualquer outra ferramenta de aprendizado de máquina) e rode o Kmeans. A partir disso, pede-se:

- a) Rode o algoritmo K-means com K=3
- b) O que significam estes valores exibidos na tela de resultados?

```
Initial starting points (random):

Cluster 0: 6.1,2.9,4.7,1.4,Iris-versicolor
Cluster 1: 6.2,2.9,4.3,1.3,Iris-versicolor
Cluster 2: 6.9,3.1,5.1,2.3,Iris-virginica
```

- c) O que significam estes valores abaixo?

Final cluster centroids:				
Attribute	Full Data (150.0)	Cluster#		
		0 (50.0)	1 (50.0)	2 (50.0)
sepalength	5.8433	5.936	5.006	6.588
sepalwidth	3.054	2.77	3.418	2.974
petallength	3.7587	4.26	1.464	5.552
petalwidth	1.1987	1.326	0.244	2.026
class	Iris-setosa	Iris-versicolor	Iris-setosa	Iris-virginica

4. Assista o vídeo:

[https://www.youtube.com/watch?v=E2M\\_yTulcmU](https://www.youtube.com/watch?v=E2M_yTulcmU)

E analise as limitações deste algoritmo.

5. Sobre **métricas de qualidade** de agrupamento, pede-se:

- a) Cite e explique o funcionamento de **duas métricas** para avaliação de agrupamento.
- b) Instale o pacote **KVALID.ZIP** que está no CANVAS e avalie os agrupamentos realizados com a base de dados da IRIS.

6. Conforme conversamos, os algoritmos de agrupamento agrupam as instâncias com base em similaridades. Esta similaridade é calculada usando o cálculo de distâncias (euclidiana, Manhattan, etc) a partir dos atributos.

Após o agrupamento é importante que a gente adicione o rótulo a estes grupos. Isso não é realizado pelo algoritmo

**Pergunta:** o que faria para rotular estes grupos?