Aula Prática 8

Tema: Agrupamento

Murielly Oliveira Nascimento – 11921BSI22

**Perguntas:**

1. Diferencie aprendizado supervisionado de não-supervisionado. Exemplifique.

No aprendizado supervisionado, novos objetos não rotulados são atribuídos um rótulo de classe usando um modelo desenvolvido a partir de objetos com rótulos de classe conhecido, por exemplo, classificar se pacientes estão com COVID ou não com base nos dados de pacientes previamente diagnosticados. Já no aprendizado não-supervisionado os rótulos são derivados apenas dos dados, sem que haja o uso de objetos com rótulos pré-determinados, por exemplo segmentar clientes em menos grupos para análise adicional e atividades de marketing.

1. Execute o K-Means na seguinte base de dados. Mostre o passo a passo da execução. Considere o valor de K = 2. Considere as duas primeiras instâncias da base com os centroides iniciais. Lembre-se, o atributo classe não será usado no agrupamento.

**Tabela

Descrição gerada automaticamente**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Nome | Febre | Enjoo | Manchas | Dores |  |
| João | 1 | 1 | 0 | 1 | Centróide1 |
| Pedro | 0 | 0 | 1 | 0 | Centróide2 |
| Maria | 1 | 1 | 0 | 0 |  |
| José | 1 | 0 | 1 | 1 |  |
| Ana | 1 | 0 | 0 | 1 |  |
| Leila | 0 | 0 | 1 | 1 |  |

Coeficiente de Casamento Simples:

1. João -> Maria

X = 1,1,0,1

Y = 1,1,0,1

F0,1 = 0; F0,0 = 1; F1,0 = 1; F1,1 = 2

SMC = 2+1/0+1+1+2 = ¾ = 0,75

|  |  |  |  |
| --- | --- | --- | --- |
|  | João | Pedro | Classificação |
| Maria | 0,75 | 0,25 | doente |
| José | 0,5 | 0,5 | ? |
| Ana | 0,75 | 0,25 | doente |
| Leila | 0,25 | 0,75 | saudável |

1. Procure na internet por um código que implemente o K-Means. Mostre onde cada um dos passos do K-Means, discutido nos slides, estão nesta implementação. Tente também responder às seguintes questões:
2. Como o código seleciona os centroides iniciais?

Os centroides são selecionados aleatoriamente na primeira iteração. Depois eles são ajustados a cada iteração. Sendo escolhidos com base na média dos pontos pertencentes a eles.

1. Qual a medida de distância usada?

Distância Euclidiana

1. Qual o critério de convergência?

Um número máximo de iterações é determinado e enquanto os centroides não se repetirem.