Lista #7

Curso: Ciência da Computação

Disciplina: Inteligência Artificial

Profa. Cristiane Neri Nobre

Pedro Henrique Lima Carvalho

Matrícula: 651230

1-

Kmeans é um algoritmo de agrupamento que busca separar as instâncias de uma base de dados em *k* grupos chamados clusters.

O algoritmo vincula cada instância a um cluster com base na sua distância ao centro de cada cluster(centroide). A instância pertencerá ao cluster cuja centroide estiver a menor distância, em um espaço *n* dimensional, sendo *n* o número de atributos (numéricos) de cada instância.

Para o cálculo da distância podem ser usadas diversas fórmulas, como a distância euclidiana, de Manhattan ou Mahanalobis. Algoritmo:

1- São escolhidos *k* pontos aleatórios, centroides iniciais, no espaço.

2- Cada instância será atribuída a um cluster com base na menor distância.

3- As centroides de cada cluster são recalculadas de forma a minimizar a distância entre a centroide e todas a instâncias do cluster.

Repete-se os passos 2 e 3 até uma cláusula de parada.

A cláusula de parada pode ser a ausência de troca de clusters por qualquer instância, alcance de um percentual máximo de trocas, número de iterações etc.

2- Para análise dos outliers foi realizado o boxplot para cada atributo:

Chart, box and whisker chart

Description automatically generated

3- Valores Mínimos e Máximos de cada atributo:

Table

Description automatically generated