

Approach

Digra Murtaza Izham – 1313621010

Pengantar Kecerdasan Buatan – UAS

1. Open and Deserialize data

```
1 using DataFrames, Serialization, Statistics, Random
2
3 binary_data = open("data_9m.mat", "r")
4 data = deserialize(binary_data)
5 close(binary_data)
6
7 if isa(data, Matrix)
8     col_names = [:sepalength, :sepalwidth, :petallength, :petalwidth, :x5]
9     data = DataFrame(convert(Matrix{Float64}, data), col_names)
10 end
```

Pertama saya membaca data lalu desirialize data tersebut dan tutup data, setelahnya saya membuat dataframe dari data yang sudah ada juga mengonversi data tersebut ke float64

2. Lalu membuat Centroids

```
1 class1 = data[data[:, :x5].==1.0, :]
2 class2 = data[data[:, :x5].==2.0, :]
3 class3 = data[data[:, :x5].==3.0, :]
4
5 means_class1 = [mean(skipmissing(class1[:, i])) for i in names(class1)[1:end-1]]
6 means_class2 = [mean(skipmissing(class2[:, i])) for i in names(class2)[1:end-1]]
7 means_class3 = [mean(skipmissing(class3[:, i])) for i in names(class3)[1:end-1]]
8
9 sepal1 = (means_class1[1] + means_class1[2]) / 2
10 sepal2 = (means_class2[1] + means_class2[2]) / 2
11 sepal3 = (means_class3[1] + means_class3[2]) / 2
12
13 petal1 = (means_class1[3] + means_class1[4]) / 2
14 petal2 = (means_class2[3] + means_class2[4]) / 2
15 petal3 = (means_class3[3] + means_class3[4]) / 2
16
17 sepal = vcat(sepal1, sepal2, sepal3)
18 petal = vcat(petal1, petal2, petal3)
19
20 # display mean
21 println("\nMean values :\n$sepal")
22 println("\nMean values :\n$petal")
```

Disini saya membuat 3 centroids dari 3 class dengan membagi 3 class berbeda berdasarkan kolom :x5 lalu menghitung rata rata dari setiap feature yang ada, setelahnya

menghitung rata rata dari sepalwidth sepallength dan petallength petalwidth, dengan hasilnya digabung dalam satu variable, setelahnya diprint untuk memastikan hasilnya sebagai berikut

```
Mean values :  
[2.502464045912493, 2.830968245440163, 3.337584402132779]  
  
Mean values :  
[-0.8124572482844815, 1.6397853253060022, 1.8046086382633075]
```

3. Membagi antara train dan test data

```
1 function train_test_split(data, train_ratio)  
2     n = nrow(data)  
3     indices = Random.shuffle(1:n)  
4     train_size = Int(floor(train_ratio * n))  
5  
6     train_indices = indices[1:train_size]  
7     test_indices = indices[train_size+1:end]  
8  
9     global train_data = data[train_indices, :]  
10    global test_data = data[test_indices, :]  
11  
12    return train_data, test_data  
13 end  
14  
15 train_ratio = 0.5  
16 train_data, test_data = train_test_split(data, train_ratio)  
17  
18 # Display the training and testing sets  
19 display(nrow(train_data))  
20 display(nrow(test_data))
```

Disini saya membagi data yang ada menjadi train data dan test data dengan rasio 50% sesuai dengan yang diminta soal.

```
4915200  
4915200
```

4.