University of Information Technology and Sciences (UITS)
Department of Information Technology

Lab Report No.: 4

IT-452 : Machine Learning

K means Clustering

Submitted To:

Sumaiya Akhtar Mitu
Lecturer,
Department of IT ,
UITS
Email:sumaiya.akhtar@uits.edu.bd

Submitted By:

Name:Nazmul Zaman
Student ID:2014755055
Department of IT, UITS

04 April 2023 Department of IT, UITS © All rights reserved.

Contents

1 Abstract

The abstract highlights the goal, approach, outcomes, and conclusions of a lab report on K-means clustering, summarizing its major points. Here is an illustration of how an abstraction for a lab report on K-means clustering may appear:

The purpose of this study was to look into the use of K-means clustering as a method for classifying data into distinct groupings. The method involves putting the K-means algorithm into practice and using it to analyze a dataset with a variety of attributes. Iterative clustering was used, starting with an initial centroids assignment and finishing with updated cluster assignments. According to the outcomes, K-means clustering successfully divided the data into several groups based on similarity, enabling the discovery of underlying patterns. Moreover, the experiment

2 Working Procedure

The working of the K-Means algorithm is explained in the below steps:

Step-1: Select the number K to decide the number of clusters.

Step-2: Select random K points or centroids. (It can be other from the input dataset).

Step-3: Assign each data point to their closest centroid, which will form the predefined K clusters.

Step-4: Calculate the variance and place a new centroid of each cluster.

Step-5: Repeat the third steps, which means reassign each datapoint to the new closest centroid of each cluster.

Step-6: If any reassignment occurs, then go to step-4 else go to FINISH.

Step-7: The model is ready

3 Applications of Naïve Bayes Classifier:

kmeans algorithm is very popular and used in a variety of applications such as
1.market segmentation
2.document clustering,
3.image segmentation and
4.image compression,

4 Code

```
# -*- coding: utf-8 -*-
"""Untiimport numpy as np import pandas as pdtled4.ipynb

Automatically generated by Colaboratory.

Original file is located at
    https://colab.research.google.com/drive/1qakkWXpI2_2vK9VGbpGYN-XvsUvVUlXw
"""

import numpy as np
import pandas as pd

df = pd.read_csv('https://raw.githubusercontent.com/campusx-official/100-days-of-machine-learning/main/kmeans/student_clustering.csv')
print("The shape of data is",df.shape)
df.head()

import matplotlib.pyplot as plt
plt.scatter(df['cgpa'],df['iq'])

from sklearn.cluster import KMeans

wcss = []

for i in range(1,11):
    km = KMeans(n_clusters=i)
    km.fit_predict(df)
    wcss.append(km.inertia_)

wcss

plt.plot(range(1,11),wcss)

X = df.iloc[:,:].values
km = KMeans(n_clusters=4)
y_means = km.fit_predict(X)

y_means

X[y_means == 3,1]

plt.scatter(X[y_means == 0,0],X[y_means == 0,1],color='blue')
plt.scatter(X[y_means == 1,0],X[y_means == 1,1],color='red')
plt.scatter(X[y_means == 2,0],X[y_means == 2,1],color='green')
plt.scatter(X[y_means == 3,0],X[y_means == 3,1],color='yellow')

from sklearn.datasets import make_blobs

centroids = [(-5,-5,5),(5,5,-5),(3.5,-2.5,4),(-2.5,2.5,-4)]
cluster_std = [1,1,1,1]
```

```
50
51 X,y = make_blobs(n_samples=200,cluster_std=cluster_std,centers=
       centroids,n_features=3,random_state=1)
52
53 X
54
55 import plotly.express as px
56 fig = px.scatter_3d(x=X[:,0], y=X[:,1], z=X[:,2])
57 fig.show()
58
59 wcss = []
60 for i in range(1,21):
61     km = KMeans(n_clusters=i)
62     km.fit_predict(X)
63     wcss.append(km.inertia_)
64
65 plt.plot(range(1,21),wcss)
66
67 km = KMeans(n_clusters=4)
68 y_pred = km.fit_predict(X)
69
70 df = pd.DataFrame()
71
72 df['col1'] = X[:,0]
73 df['col2'] = X[:,1]
74 df['col3'] = X[:,2]
75 df['label'] = y_pred
76
77 fig = px.scatter_3d(df,x='col1', y='col2', z='col3',color='
       label')
78 fig.show()
```

Figure 1:

5 Output

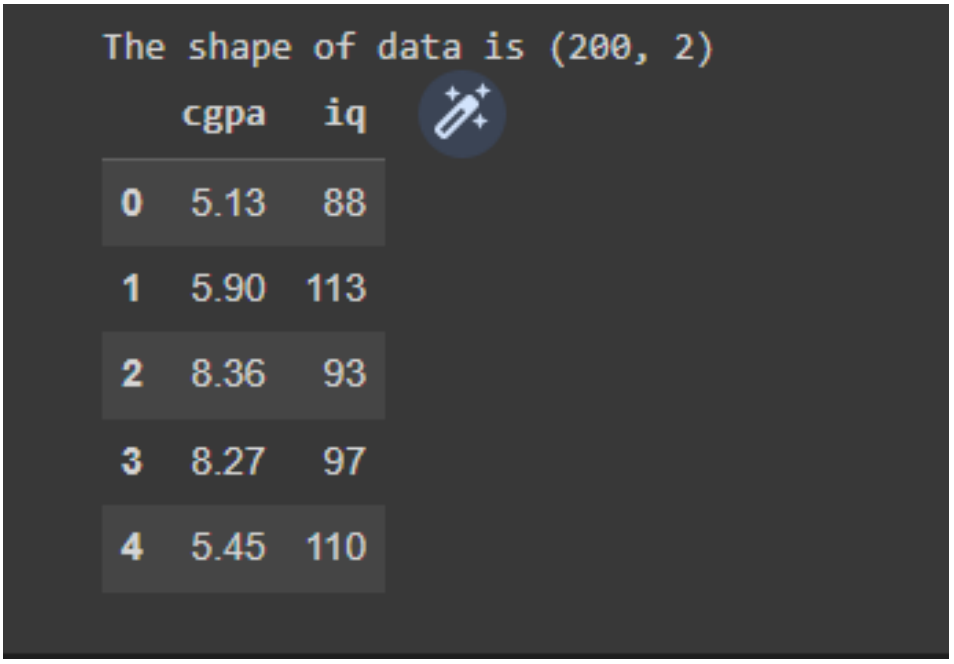

Figure 2:

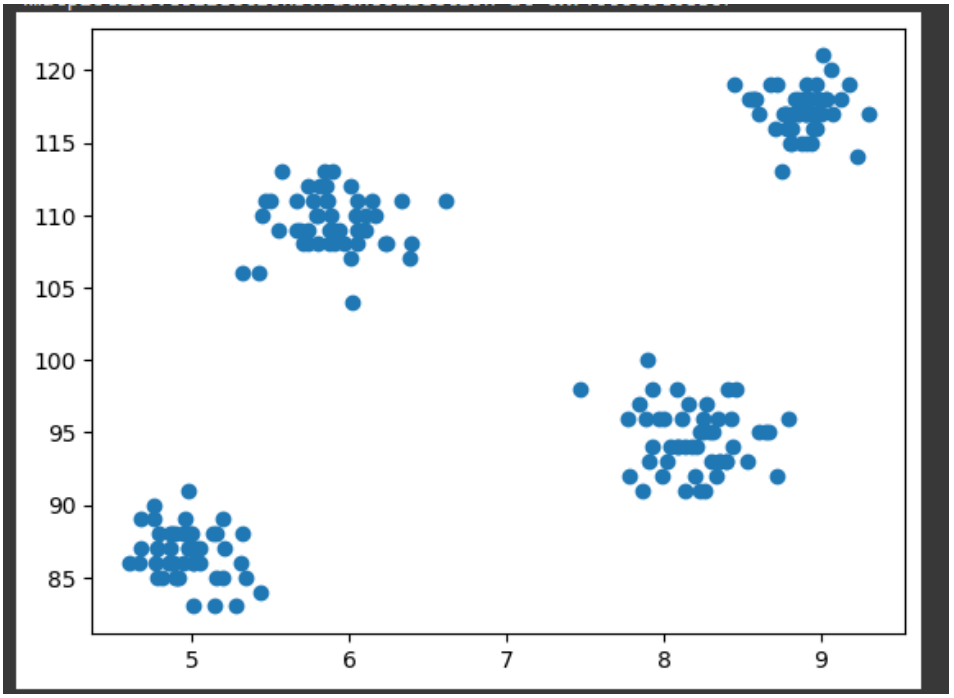

Figure 3:

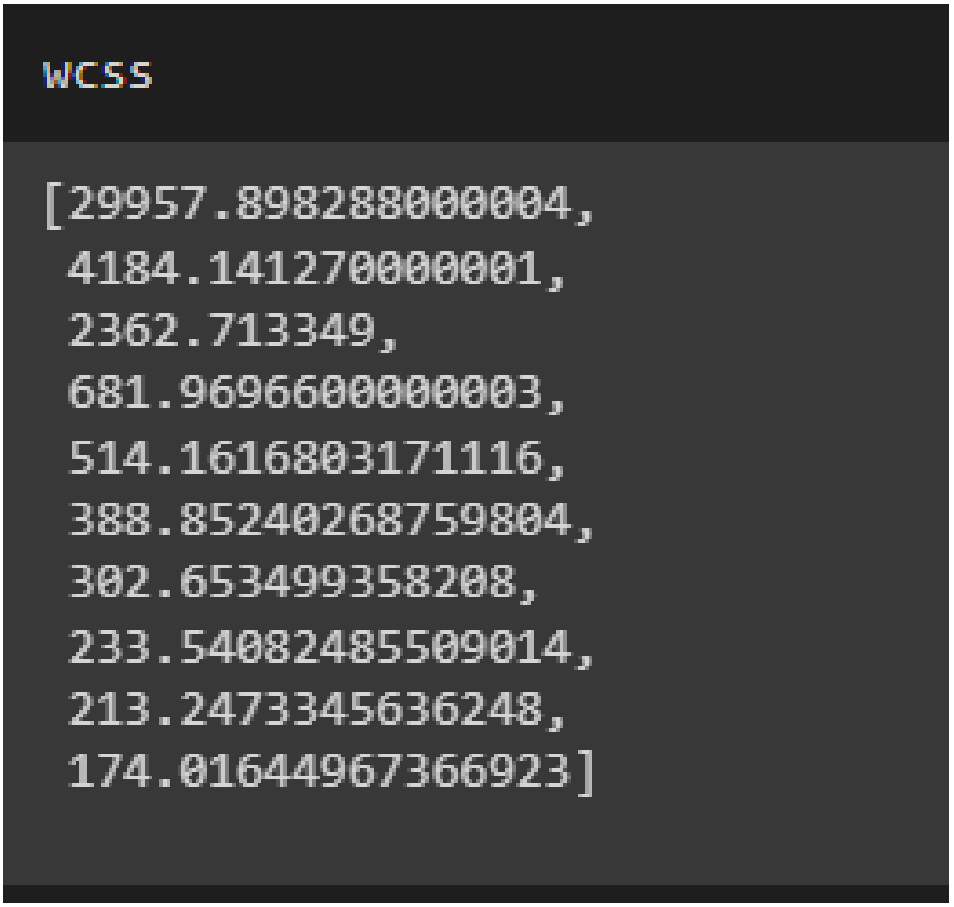
WCSS
[29957.898288000004,
4184.141270000001,
2362.713349,
681.9696600000003,
514.1616803171116,
388.85240268759804,
302.653499358208,
233.54082485509014,
213.2473345636248,
174.01644967366923]

Figure 4:

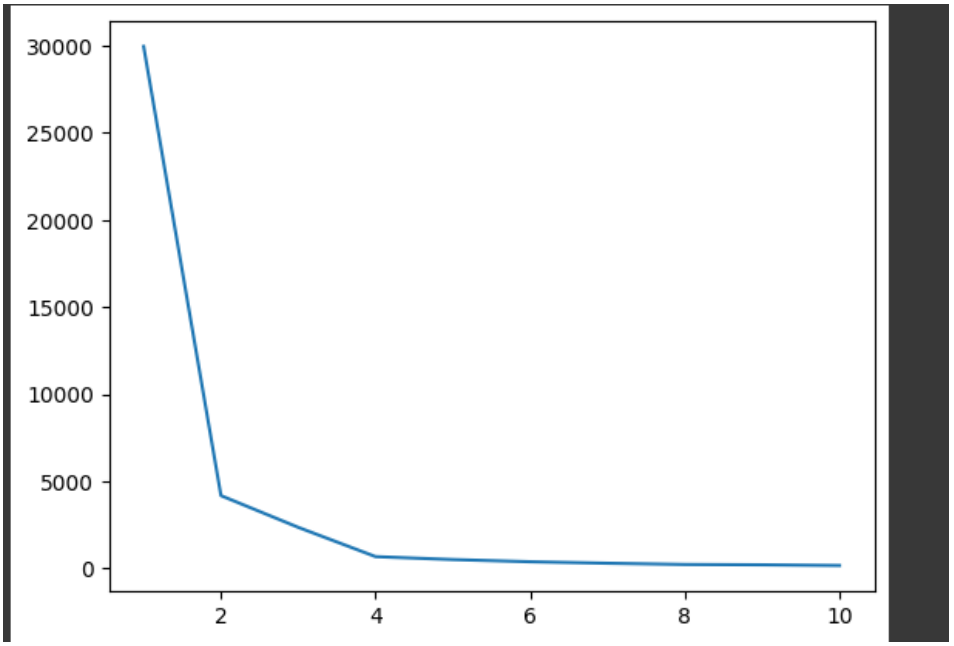

Figure 5:

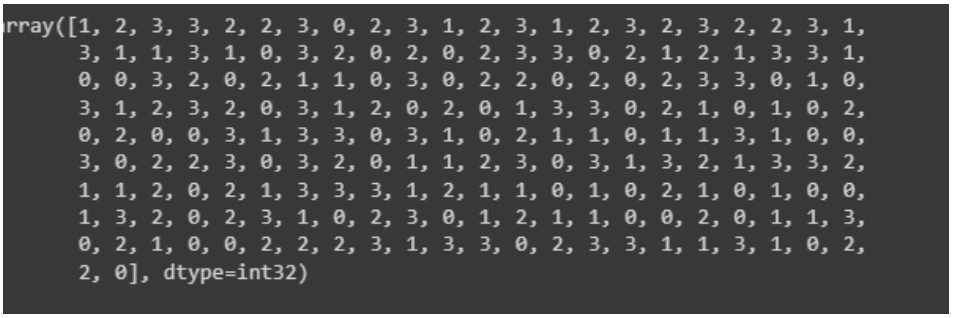
rray([1, 2, 3, 3, 2, 2, 3, 0, 2, 3, 1, 2, 3, 1, 2, 3, 2, 3, 2, 2, 3, 1,
3, 1, 1, 3, 1, 0, 3, 2, 0, 2, 0, 2, 3, 3, 0, 2, 1, 2, 1, 3, 3, 1,
0, 0, 3, 2, 0, 2, 1, 1, 0, 3, 0, 2, 2, 0, 2, 0, 2, 3, 3, 0, 1, 0,
3, 1, 2, 3, 2, 0, 3, 1, 2, 0, 2, 0, 1, 3, 3, 0, 2, 1, 0, 1, 0, 2,
0, 2, 0, 0, 3, 1, 3, 3, 0, 3, 1, 0, 2, 1, 1, 0, 1, 1, 3, 1, 0, 0,
3, 0, 2, 2, 3, 0, 3, 2, 0, 1, 1, 2, 3, 0, 3, 1, 3, 2, 1, 3, 3, 2,
1, 1, 2, 0, 2, 1, 3, 3, 3, 1, 2, 1, 1, 0, 1, 0, 2, 1, 0, 1, 0, 0,
1, 3, 2, 0, 2, 3, 1, 0, 2, 3, 0, 1, 2, 1, 1, 0, 0, 2, 0, 1, 1, 3,
0, 2, 1, 0, 0, 2, 2, 2, 3, 1, 3, 3, 0, 2, 3, 3, 1, 1, 3, 1, 0, 2,
2, 0], dtype=int32)

Figure 6:

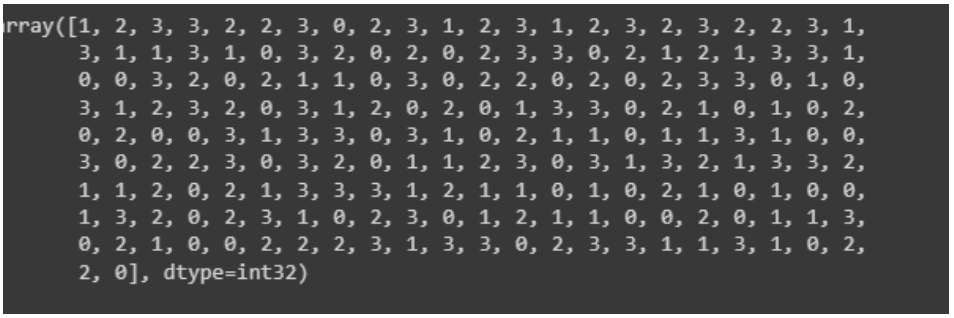
rray([1, 2, 3, 3, 2, 2, 3, 0, 2, 3, 1, 2, 3, 1, 2, 3, 2, 3, 2, 2, 3, 1,
3, 1, 1, 3, 1, 0, 3, 2, 0, 2, 0, 2, 3, 3, 0, 2, 1, 2, 1, 3, 3, 1,
0, 0, 3, 2, 0, 2, 1, 1, 0, 3, 0, 2, 2, 0, 2, 0, 2, 3, 3, 0, 1, 0,
3, 1, 2, 3, 2, 0, 3, 1, 2, 0, 2, 0, 1, 3, 3, 0, 2, 1, 0, 1, 0, 2,
0, 2, 0, 0, 3, 1, 3, 3, 0, 3, 1, 0, 2, 1, 1, 0, 1, 1, 3, 1, 0, 0,
3, 0, 2, 2, 3, 0, 3, 2, 0, 1, 1, 2, 3, 0, 3, 1, 3, 2, 1, 3, 3, 2,
1, 1, 2, 0, 2, 1, 3, 3, 3, 1, 2, 1, 1, 0, 1, 0, 2, 1, 0, 1, 0, 0,
1, 3, 2, 0, 2, 3, 1, 0, 2, 3, 0, 1, 2, 1, 1, 0, 0, 2, 0, 1, 1, 3,
0, 2, 1, 0, 0, 2, 2, 2, 3, 1, 3, 3, 0, 2, 3, 3, 1, 1, 3, 1, 0, 2,
2, 0], dtype=int32)

Figure 7:

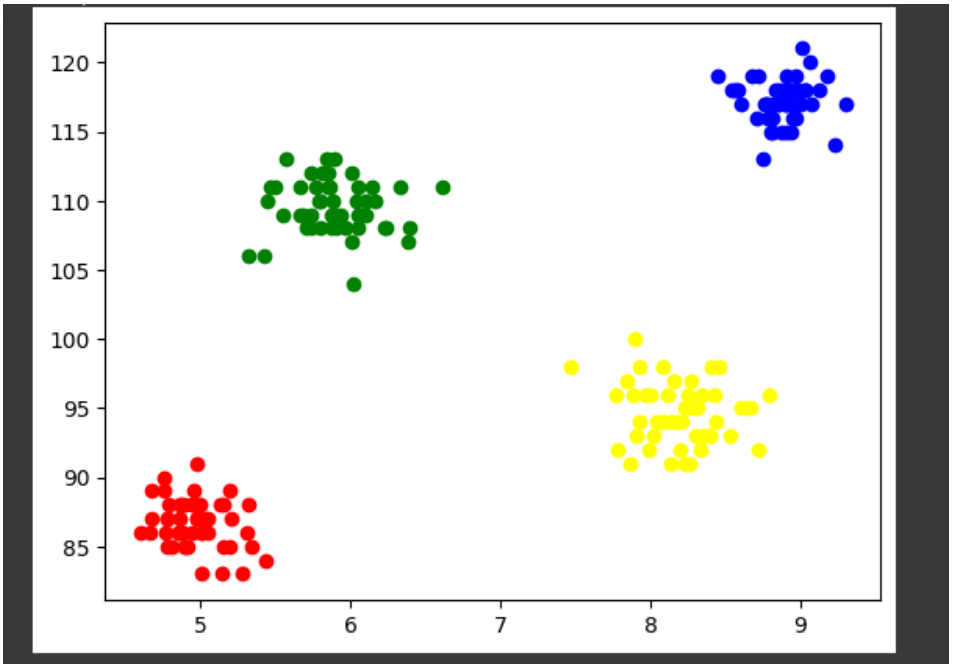

Figure 8:

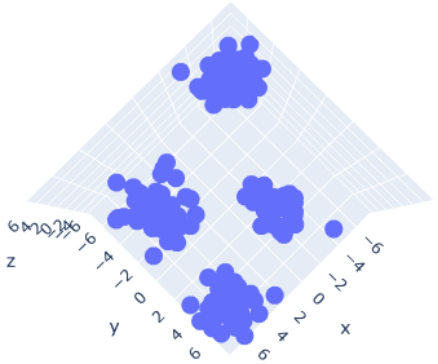

Figure 9:

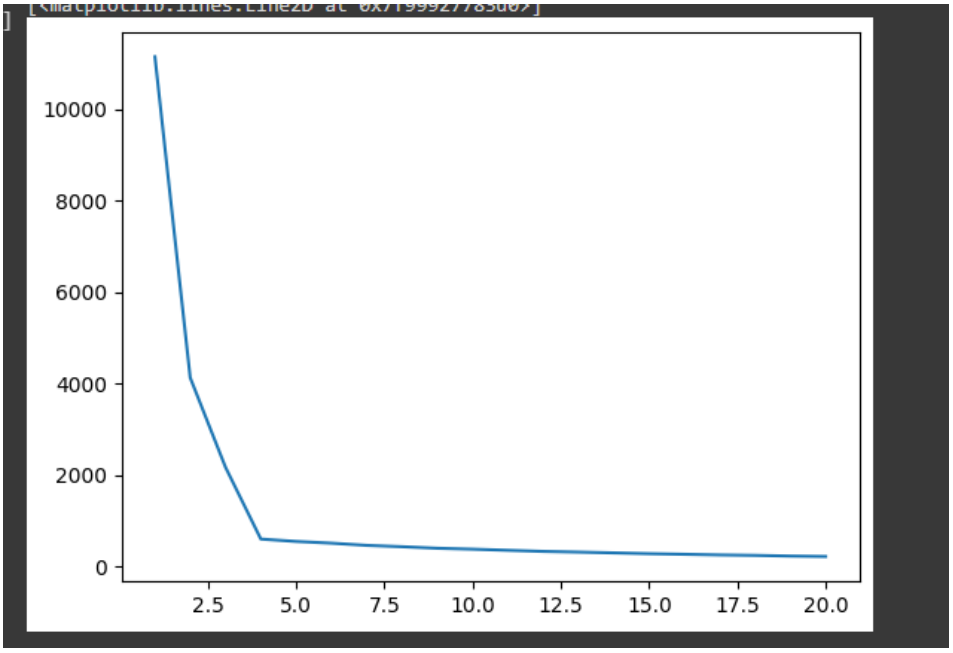

Figure 10:

6 Conclusion

In this lab report I learn how to implement kmeans using python sklearn library. We will use Kmeans to cluster students in a dataset and find the final output of this progeam.

7 References

2. https://www.javatpoint.com/linear-regression-in-machine-learning
3. https://www.javatpoint.com/machine-learning-polynomial-regression