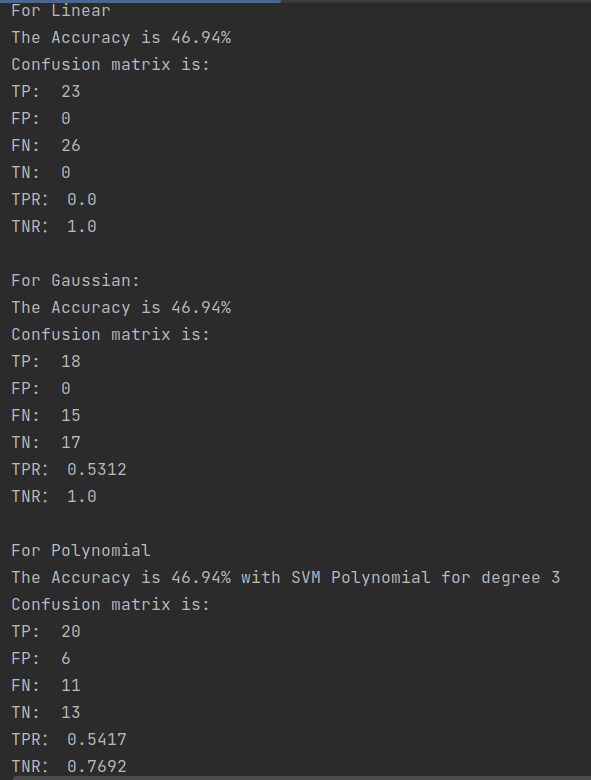
**Assignment 12 seeds**

# **Question 1:**



## **Take the subset of the dataset containing your two class labels. You will use random 50/50 splits for training and testing data.**

## **1. implement a linear kernel SVM. What is your accuracy and confusion matrix?**

Accuracy: 46.94%

Confusion matrix

TP:23

FP:0

FN:26

TN:0

## **2. implement a Gaussian kernel SVM. What is your accuracy and confusion matrix?**

Accuracy: 46.94%

Confusion matrix

TP:18

FP:0

FN:15

TN:17

## **3. implement a polynomial kernel SVM of degree 3. What is your accuracy and confusion matrix?**

Accuracy: 46.94%

Confusion matrix

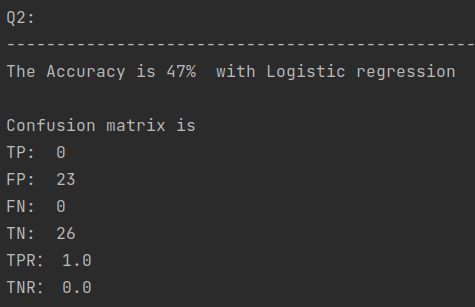
TP:20

FP:6

FN:11

TN:13

# **Question 2:**



## **Pick up any classifier for supervised learning (e.g. kNN, logistic regression, Naive Bayesian, etc)**

## **1. use this classifier to your dataset. What is your accuracy and confusion matrix?**

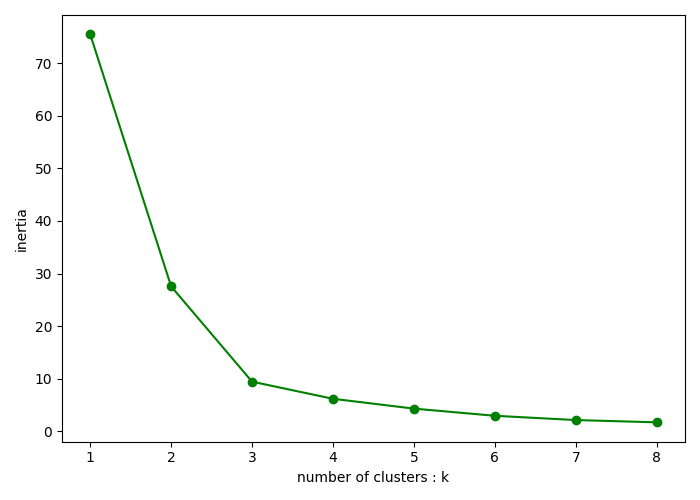
## **2. summarize your findings in a table below and discuss your results**

| Model | TP | FP | TN | FN | accuracy | TPR | TNR |
| --- | --- | --- | --- | --- | --- | --- | --- |
| linear SVM | 23 | 0 | 26 | 0 | 46.94% | 0 | 1 |
| Gaussian SVM | 18 | 0 | 15 | 17 | 46.94% | 0.5312 | 1 |
| polynomial SVM | 20 | 6 | 11 | 13 | 46.94% | 0.5417 | 0.7692 |
| Logistic Regression | 0 | 23 | 0 | 26 | 47% | 1 | 0 |

# **Question 3:**

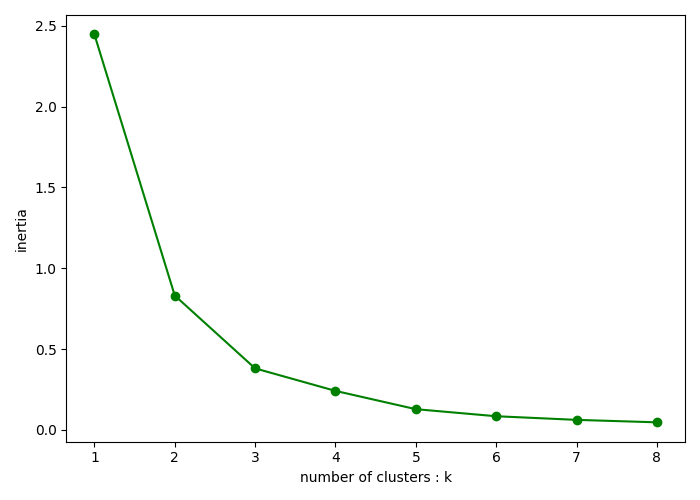
## **Take the original dataset with all 3 class labels.**

## **1. for k = 1, 2, . . . , 8 use k-means clustering with random initialization and defaults. Compute and plot distortion vs k. Use the ”knee” method to find the best k.**



According to the “knee” method, the best k is 3

## **2. re-run your clustering with best k clusters. Pick two features fi and fj at random (using python, of course) and plot your datapoints (different color for each class and centroids) using fi and fj as axis. Examine your plot. Are there any interesting patterns?**



Yes, their graphs are very similar.