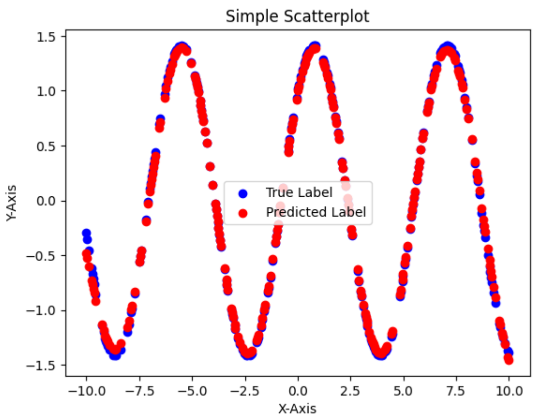
Introduction to ML (CS771), Autumn 2023
Indian Institute of Technology Kanpur
Homework Assignment Number 2

QUESTION 1

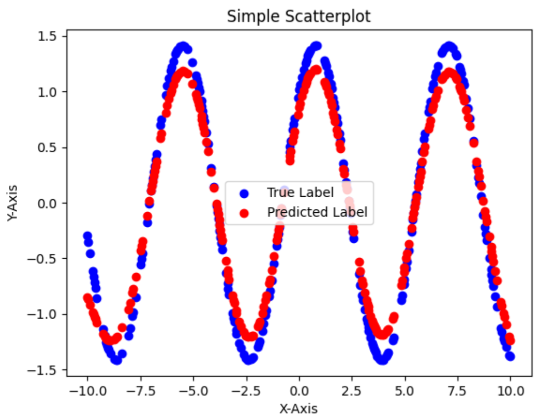
Student Name: Dhiraj Pareek
Roll Number: 231110012
Date: August 30, 2024

(i)
From the plots we can observe that the Root Mean Square Error (RMSE) value increases as we increase the value of λ, because when we increase the value of λ in Ridge Regression, we increase the strength of the regularization. Which means that the optimization process will try to keep the model coefficients as small as possible, penalizing large coefficients more heavily.

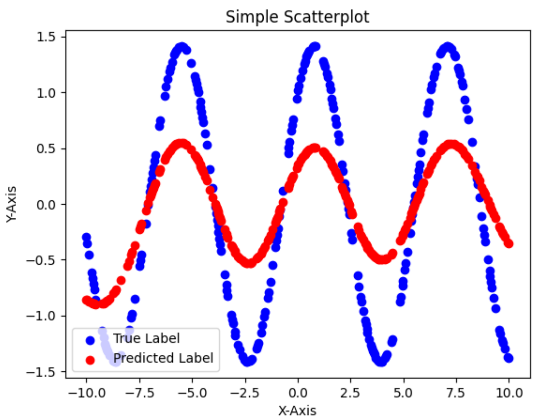
The RMSE values are:
RMSE VALUE FOR LAMBDA = 0.1 is : 0.03257767029357659
RMSE VALUE FOR LAMBDA = 1 is : 0.17030390344202528
RMSE VALUE FOR LAMBDA = 10 is : 0.6092671596540065
RMSE VALUE FOR LAMBDA = 100 is : 0.9110858052767243



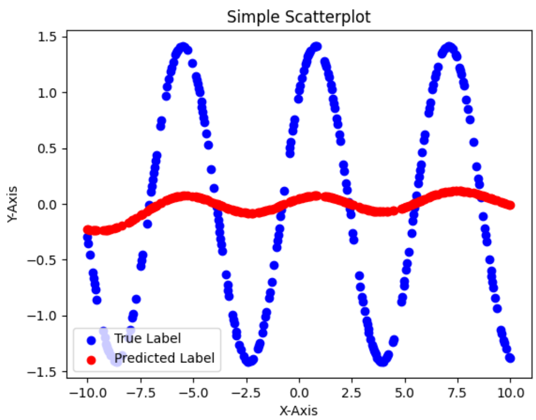
(a) lambda = 0.1



(b) lambda = 1



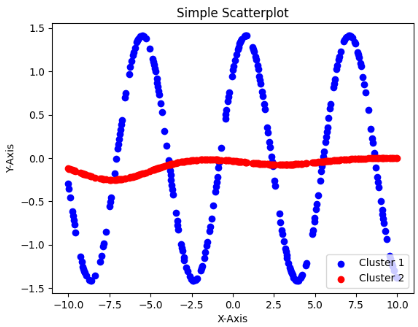
(c) lambda = 10



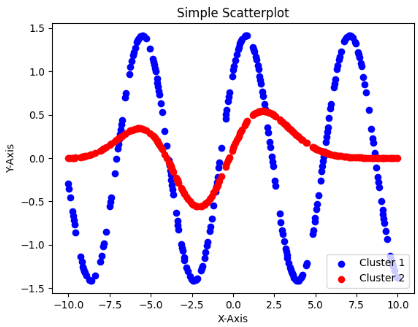
(d) lambda = 100

(ii)
Observations made from the plots of landmark ridge is that RMSE value decreases because adding more random landmarks effectively increases the dimensionality of the feature space. This can result in a higher capacity for the model to capture complex relationships in the data. With more landmarks, the model may generalize better to unseen data.

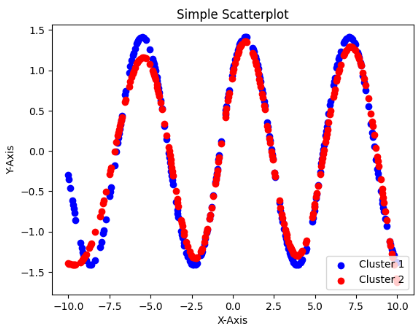
The RMSE values are:
RMSE VALUE FOR LAMBDA = 2 is : 0.9726299346666142
RMSE VALUE FOR LAMBDA = 5 is : 0.9616907013072213
RMSE VALUE FOR LAMBDA = 20 is : 0.3468804486859466
RMSE VALUE FOR LAMBDA = 50 is : 0.07848641688567466
RMSE VALUE FOR LAMBDA = 100 is : 0.060036454866626759



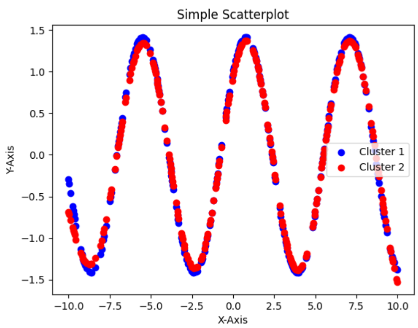
(e) L=2



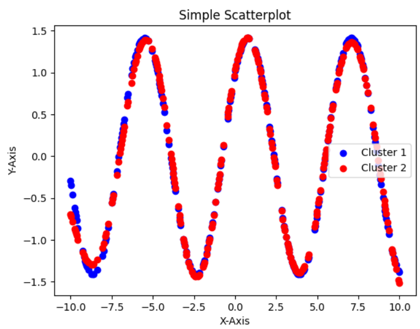
(f) L = 5



(g) lambda = 20



(h) lambda = 50



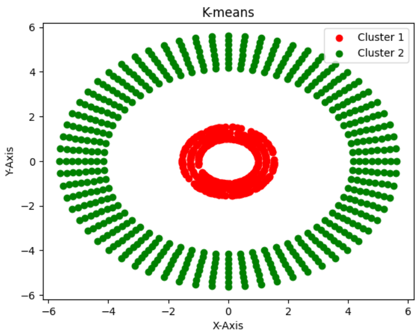
(i) lambda = 100

Introduction to ML (CS771), Autumn 2023
Indian Institute of Technology Kanpur
Homework Assignment Number 2

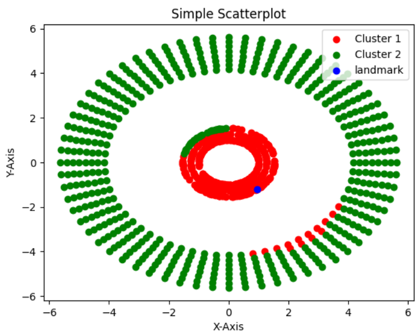
Student Name: Dhiraj Pareek
Roll Number: 231110012
Date: August 30, 2024

QUESTION 2

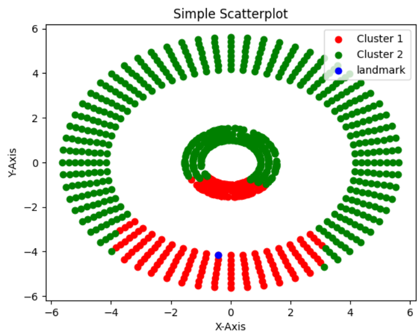
(i) Since the data is not linearly separable so to apply k-means I have transformed the data points using $x^2 + y^2$ as feature transformation.



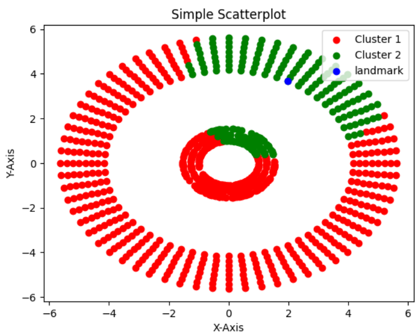
(ii) On plotting graphs for different landmark point (selected randomly), it can be observed that the clusters are better separated when the landmark selected is close to the center as the data points of cluster 1 lies around it.



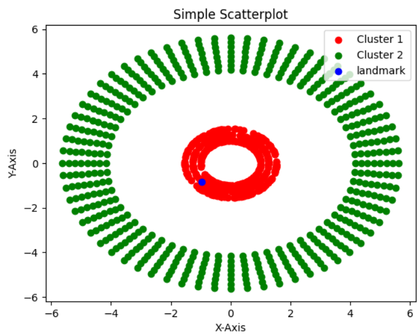
(j)



(k)



(l)



(m)

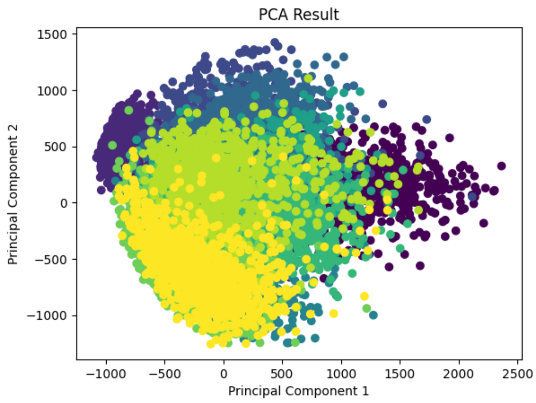
Introduction to ML (CS771), Autumn 2023
Indian Institute of Technology Kanpur
Homework Assignment Number 2

Student Name: Dhiraj Pareek
Roll Number: 231110012
Date: August 30, 2024

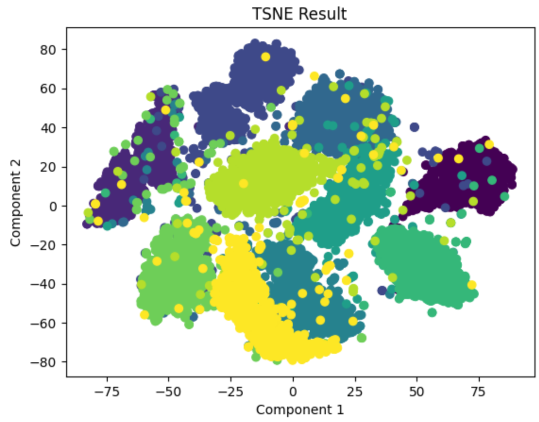
QUESTION 3

tSNE is a non-linear dimensionality reduction technique that focuses on preserving the pair-wise similarities between data points in the high-dimensional space. It is particularly effective at capturing complex, non-linear relationships in the data. While PCA is a linear technique that aims to capture the maximum variance in the data along orthogonal axes. PCA may not perform as well when the relationships in the data are highly non-linear.

So, according to the plot we can infer that tSNE shows better clustering than PCA, as data clusters in PCA plot is overlapping while in tSNE where gloal structure is preserved we are not getting overlapping.



(n) PCA



(o) tSNE