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

QUESTION 1

Student Name: Arpit Nigam
Roll Number: 231110008
Date: August 30, 2024

(i)

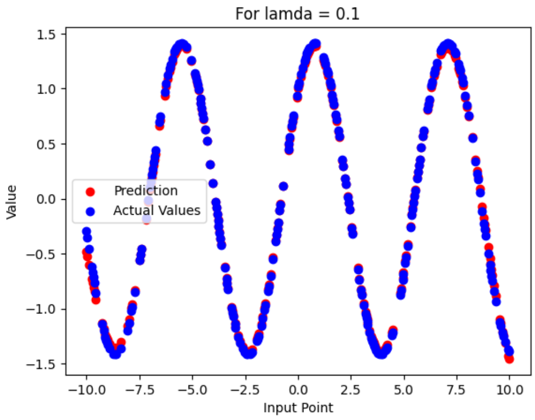
From the plots we can infer that the value of Root Mean Square Error (RMSE) tends to increase as we increase the value of regularization hypermeter λ that is prediction degrades with increase in λ. As the regularization hyperparameter controls the trade-off between fitting the training data well and keeping the model's coefficients small to avoid overfitting. When we increase the value of the regularization hyperparameter, it tends to penalize the magnitude of the coefficients more strongly. Excessive regularization can lead to underfitting, where the model is too constrained and fails to capture the underlying patterns in the data. As a result, the model might struggle to make accurate predictions, causing the RMSE to increase.

RMSE values for each λ is:
Root Mean Square Error for lambda = 0.1 is 0.032577670293571295
Root Mean Square Error for lambda = 1 is 0.17030390344202051
Root Mean Square Error for lambda = 10 is 0.6092671596540067
Root Mean Square Error for lambda = 100 is 0.9110858052767243

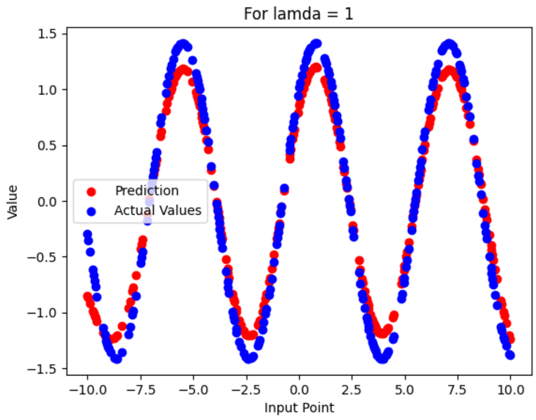
(ii)

From the plots we can infer that as we increase the number of landmarks, we are able to achieve lower RMSE value that is our prediction gets better with increasing number of landmarks. Because landmarks serve as representative points, and having more of them enables the model to better represent the variability and patterns in the dataset. The increased number of landmarks provides the model with more basis functions to express the relationship between input features and the target variable. This increased flexibility can result in a better fit to the training data.

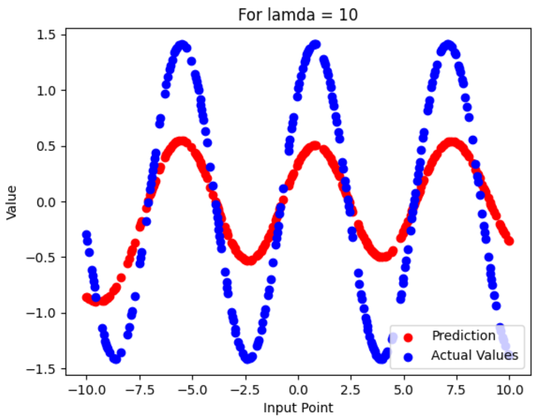
RMSE value for each L:
Root Mean Square Error value for L= 2 is 0.9735983043074795
Root Mean Square Error value for L= 5 is 0.9467226860998962
Root Mean Square Error value for L= 20 is 0.1508923905242473
Root Mean Square Error value for L= 50 is 0.09310328478293449
Root Mean Square Error value for L= 100 is 0.05952480191323201



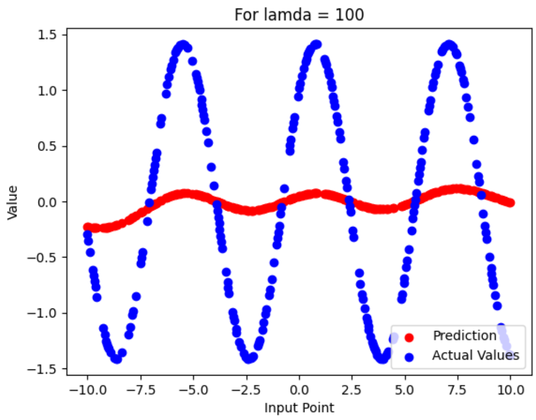
(a) lambda = 0.1



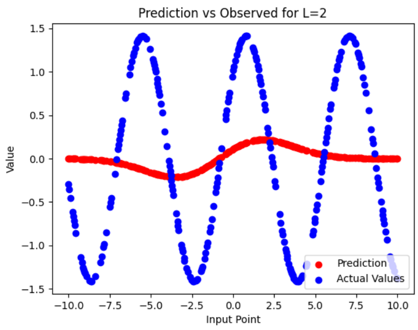
(b) lambda = 1



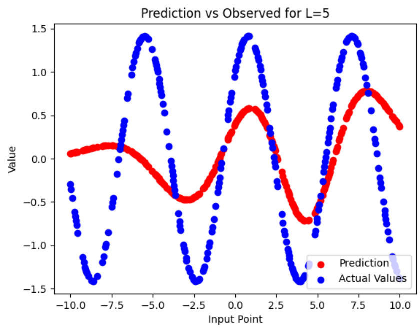
(c) lambda = 10



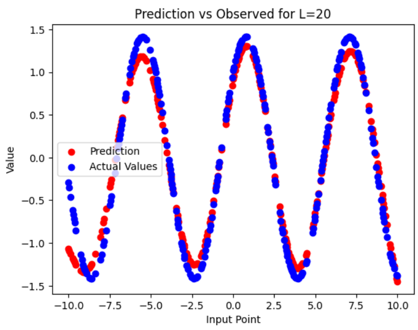
(d) lambda = 100



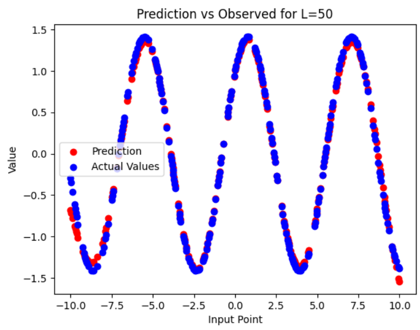
(e) L=2



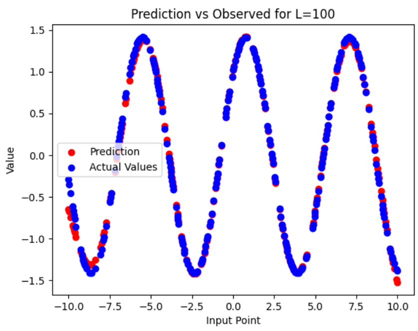
(f) L = 5



(g) L = 20



(h) L = 50



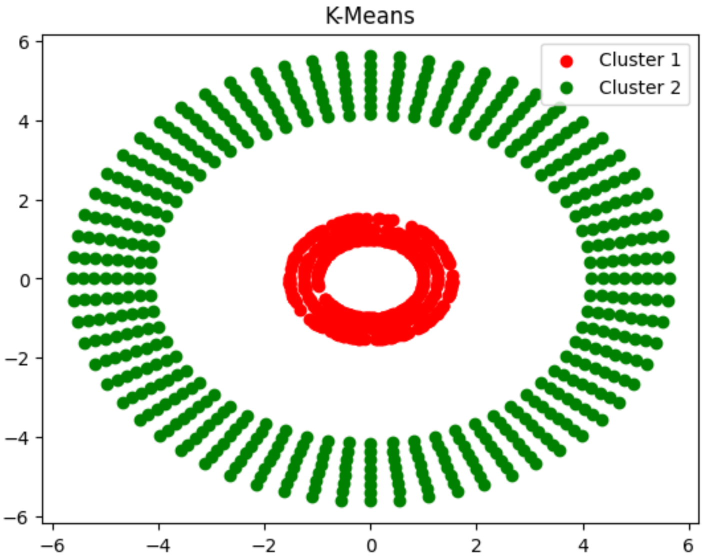
(i) L = 100

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

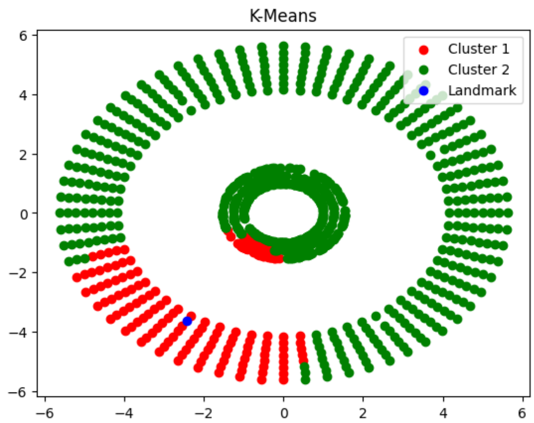
Student Name: Arpit Nigam
Roll Number: 231110008
Date: August 30, 2024

QUESTION 2

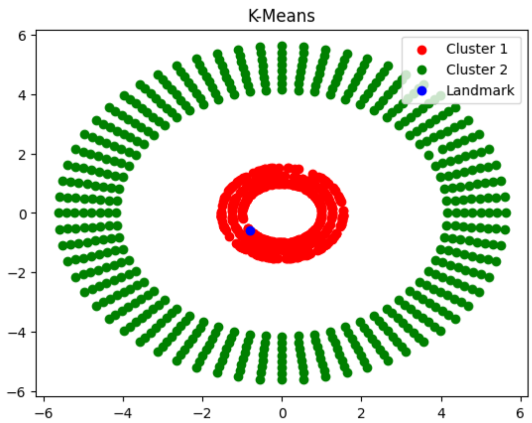
(i) Here, The transformation that i have used is $x^2 + y^2$ for feature transformation to make the data linearly separable so that k-means can be applied to it.



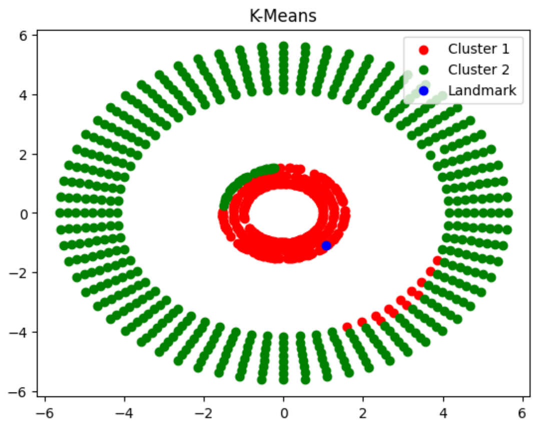
(ii) Here, we are given a kernel (i.e rbf) which we have to use to transform our dataset before applying k-means on it.Here we are randomly choosing landmarks and we get correct clustering whenever the random landmark is near the center of the inner concentric circle and not so correct clustering in other cases. Since the points are arranged in concentric circles, when the landmark is near the center, we majorly get two distances for all the points which transform the data into linearly separable plane and therefore we get correct clustering for some landmarks and not so correct clustering for other landmarks.


K-Means
Cluster 1
Cluster 2
Landmark

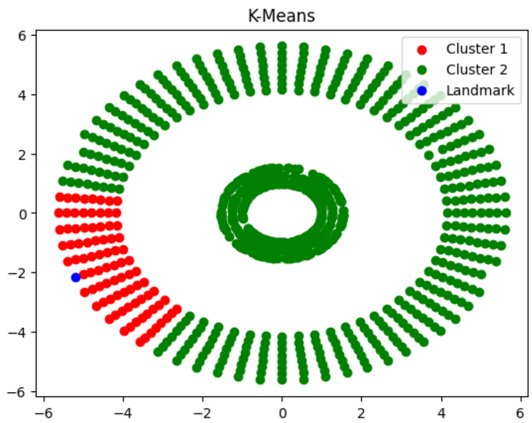
(j)


K-Means
Cluster 1
Cluster 2
Landmark

(k)


K-Means
Cluster 1
Cluster 2
Landmark

(l)


K-Means
Cluster 1
Cluster 2
Landmark

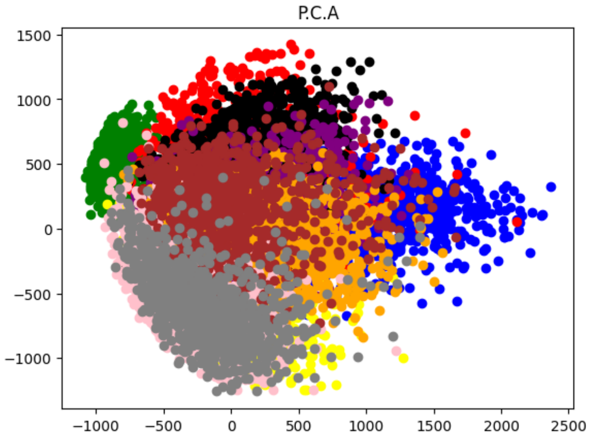
(m)

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

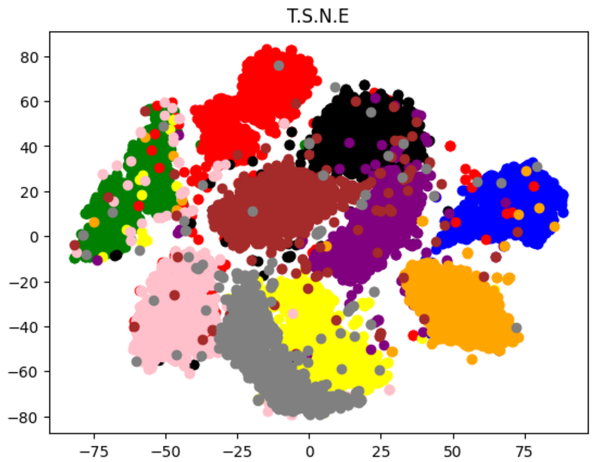
Student Name: Arpit Nigam
Roll Number: 231110008
Date: August 30, 2024

QUESTION 3

From both the plots it can be observed that in PCA clusters are overlapping but not in tSNE. This is because PCA is a linear method, meaning it captures linear relationships in the data. t-SNE, on the other hand, is a non-linear technique. It can reveal complex non-linear structures in the data, making it more suitable for capturing intricate patterns and relationships. Moreover, t-SNE is particularly good at preserving local structures in the data. It tends to group similar data points together in the low-dimensional space, which can be useful for visualizing clusters or neighborhoods of points with similar properties. Outliers might have less impact on the overall structure of the t-SNE representation compared to PCA, which is sensitive to extreme values.



(n) PCA



(o) tSNE