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

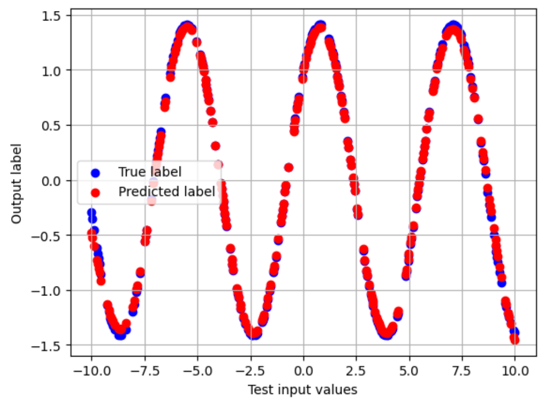
*Student Name:* Pratibha Gupta
*Roll Number:* 231110038
*Date:* November 17, 2023

# QUESTION 5

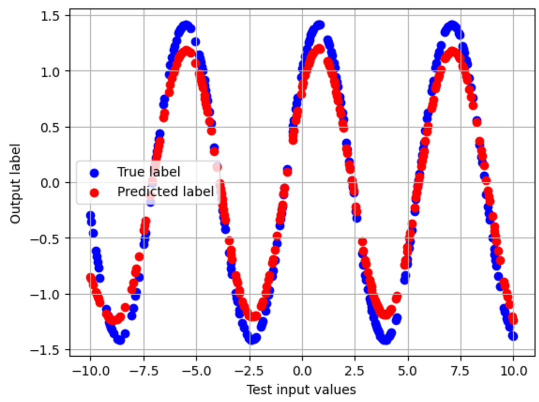
(i)

From the plots we can observe that prediction tends to degrade as we increase the value of $\lambda$ because as the $\lambda$ increases, the regularization effect dominates the weight matrix making the weights much smaller so when the actual value is large enough our prediction will not be accurate due to dominating regularization effect.

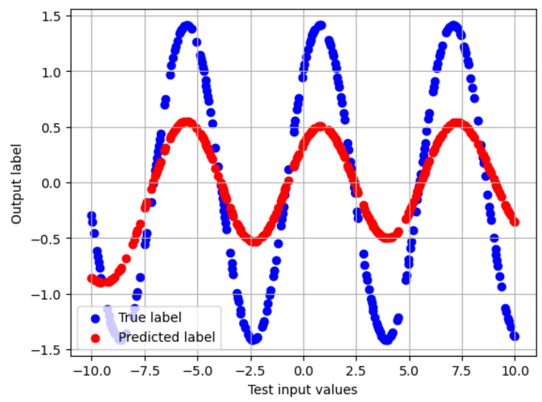
The Root Mean Square Error (RMSE) values increases as we in increase $\lambda$ :
ROOT MEAN SQUARE ERROR FOR (LAMBDA = 0.1 ) is : 0.032577670293573044
ROOT MEAN SQUARE ERROR FOR (LAMBDA = 1 ) is : 0.17030390344202528
ROOT MEAN SQUARE ERROR FOR (LAMBDA = 10 ) is : 0.6092671596540066
ROOT MEAN SQUARE ERROR FOR (LAMBDA = 100 ) is : 0.9110858052767243



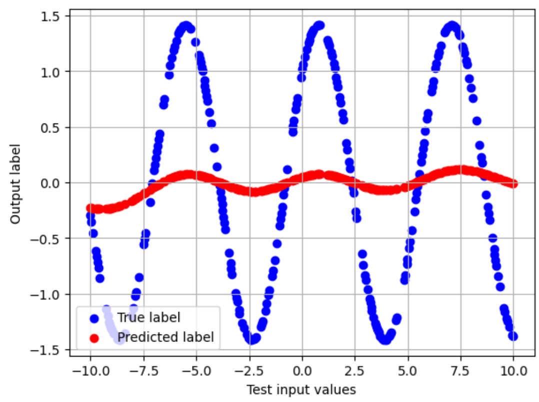
(a) lambda = 0.1



(b) lambda = 1



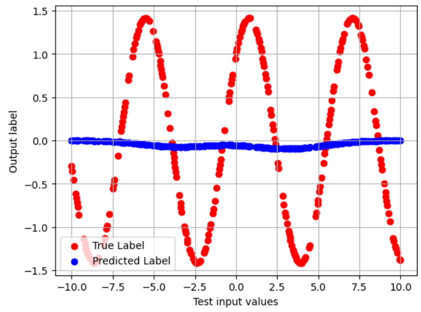
(c) lambda = 10



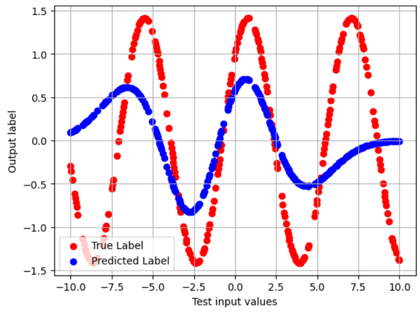
(d) lambda = 100

(ii)
From the plots for Landmark Ridge we can note that as we increase the number of landmarks we tend to get better predictions as By increasing the number of landmarks, the model gains a more detailed representation of the data set, allowing it to better capture complex patterns and relationships. A model with a larger set of landmarks is likely to generalize better to new, unseen data.

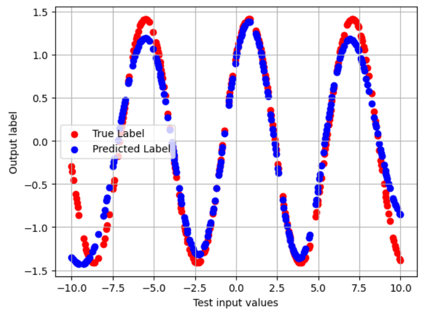
The RMSE value decreases as we increase the number of landmarks considered:
ROOT MEAN SQUARE ERROR FOR (L = 2 ) IS 0.9593499469146843
ROOT MEAN SQUARE ERROR FOR (L = 5 ) IS 0.947444461595836
ROOT MEAN SQUARE ERROR FOR (L = 20 ) IS 0.21965201241854837
ROOT MEAN SQUARE ERROR FOR (L = 50 ) IS 0.08363317007267723
ROOT MEAN SQUARE ERROR FOR (L = 100 ) IS 0.07830743665325934
Note: the RMSE changes in each run as random landmarks are selected.
Larger value of L is a better choice.



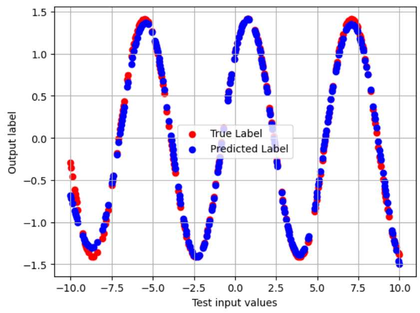

(e) L=2



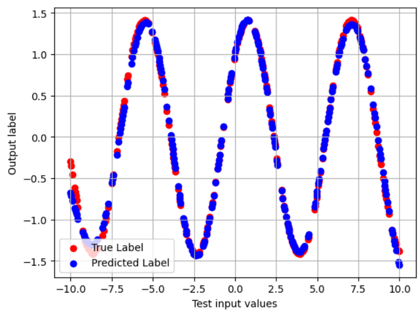

(f) L = 5




(g) lambda = 20




(h) lambda = 50



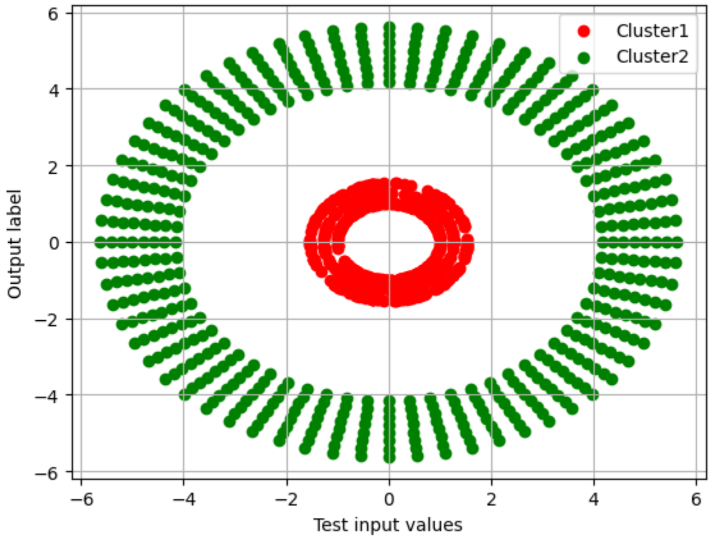

(i) lambda = 100

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

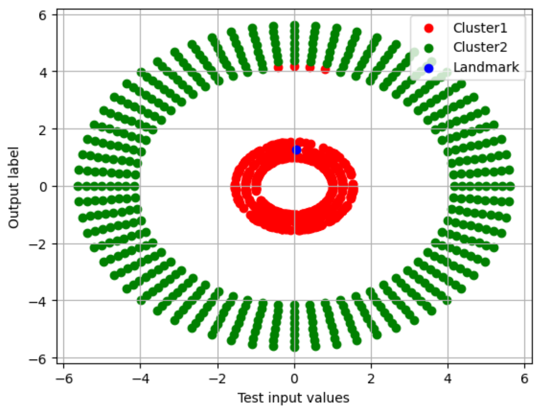
*Student Name:* Pratibha Gupta
*Roll Number:* 231110038
*Date:* November 17, 2023

# QUESTION 6

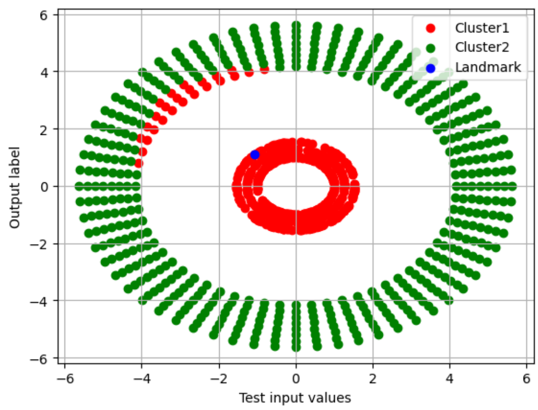
(i)
Here, I have used $\sqrt{a^2 + b^2}$ as feature transformation to make data linearly separable so that k-means can be applied on transformed features.



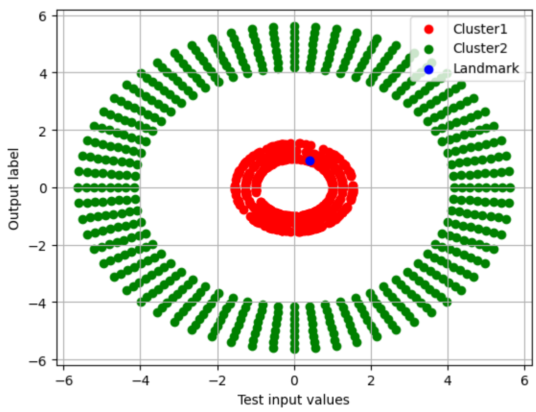

(ii)
In this question we are supposed to use RBF-kernel for feature transformation with random landmarks selected from train set. We are getting correct clustering when the landmark chosen is closer to the center as the data points are clustered around it. And when the landmark lies away from center it does not gives better clustering.


Cluster1
Cluster2
Landmark
Output label
Test input values

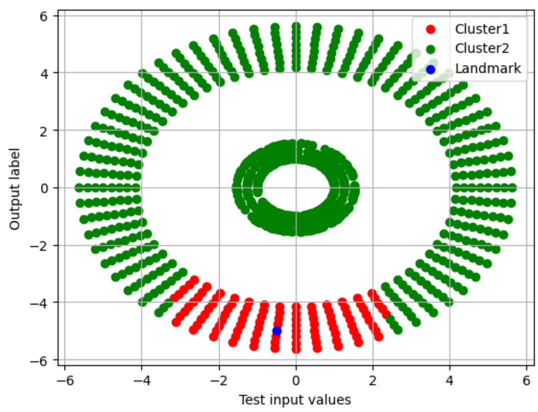
(j)


Cluster1
Cluster2
Landmark
Output label
Test input values

(k)


Cluster1
Cluster2
Landmark
Output label
Test input values

(l)


Cluster1
Cluster2
Landmark
Output label
Test input values

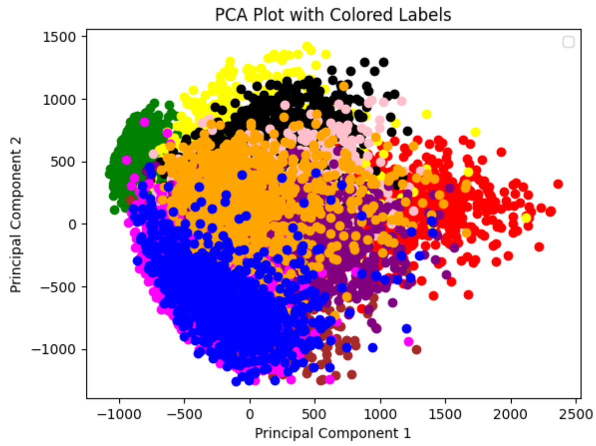
(m)

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

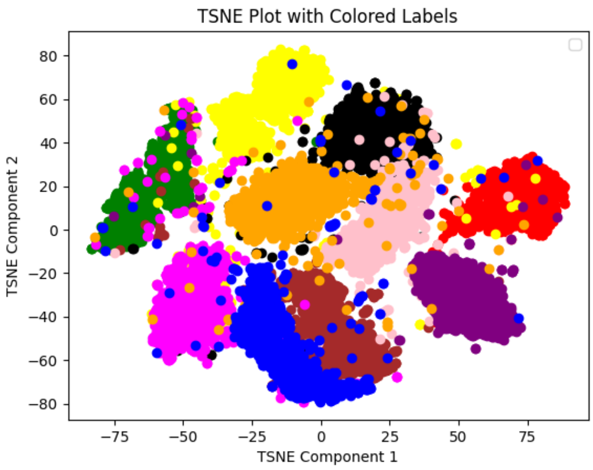
*Student Name:* Pratibha Gupta
*Roll Number:* 231110038
*Date:* November 17, 2023

# QUESTION 7

The plot for tSNE shows better clustering of data points than in PCA. In PCA we witness overlapping of clusters because it tries to preserve global structure of data and gets affected by outliers. Instead tSNE preserves local structure (cluster) of data and can also handle outliers. It embeds the points from a higher dimension to a lower dimension trying to preserve the neighborhood of that point.




(n) PCA




(o) tSNE