ASSIGNMENT 02

PIYUSH KUMAR,SRNO-23801

28 March 2025

1 PROBLEM:

Support Vector Machine and Perceptron (10 marks)

For this question, you are given with CIFAR-100 dataset. You will validate its separability by testing the convergence of the perceptron, before isolating (if they exist) sources of non-separability through the support vector machine.

Tasks

1. Perceptron: Run the perceptron algorithm on your data. Report whether it converges, or appears not to. If it doesn't seem to converge, make certain that you are reasonably sure.

2. Linear SVM: Construct a slack support vector machine with the linear kernel. Solve both the primal and dual versions of the SVM quadratic programs using cvxopt. Use the SVM's solution to isolate the sources of non-separability.

3. Kernelized SVM: Repeat the previous construction, but with the Gaussian kernel this time. Choose your hyperparameters such that non-separability is no longer an issue, and your decision boundary is consistent with the training data's labels.

4. Perceptron, again: Retrain the perceptron, after removing the sources of non-separability isolated by the linear SVM. Verify that it converges.

TASKS 1. A plot between misclassification rate and number of iterations

for the perceptron algorithm as defined in Task 1. (2 marks)

2. Which, between the primal and dual, is solved faster for Task 2. Report the times taken for running both, and justify any patterns you see. (2 marks)

3. The images that cause non-separability. (2 marks)

4. The final misclassification rate for the kernelized SVM for Task 3.(2 marks)

SOLUTION:- TASKS 1.1 We are taking misclassification loss on the train

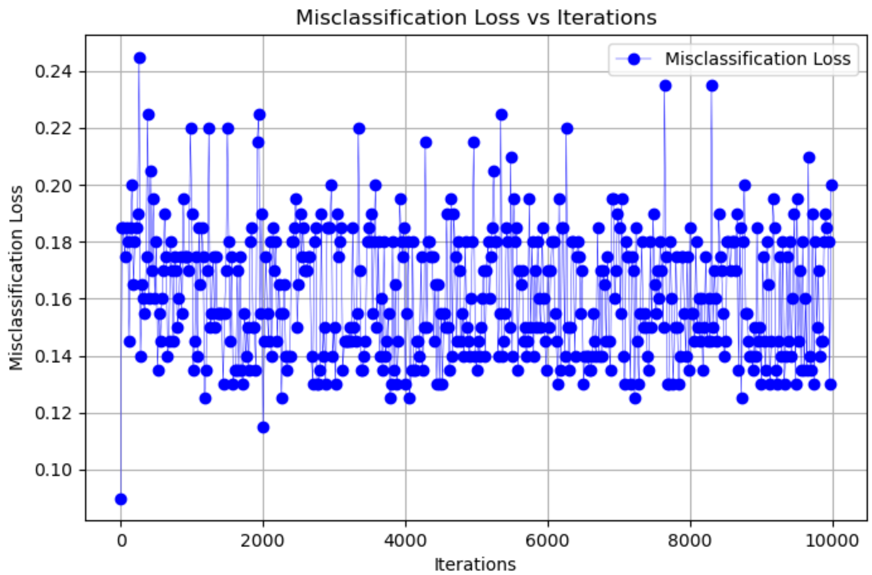


Figure 1: misclassification rate vs iterations count for 10k iterations on the train data.

dataset. Since we need Train error to be zero for algorithm to converge. Hence we can say that the SVM doesn't seem to converge.

SOLUTION:-

TASK:2- Theory of Linear Support Vector Machines (SVM)

Given a training dataset

$$\{(\mathbf{x}_i, y_i)\}_{i=1}^n, \quad \mathbf{x}_i \in R^d,\ y_i \in \{-1, 1\},$$

the goal of a linear SVM is to find a hyperplane that best separates the two classes with the maximum margin.

A hyperplane in R^d is defined as:

$$\mathbf{w}^\top \mathbf{x} + b = 0,$$

where $\mathbf{w} \in R^d$ is the weight vector and $b \in R$ is the bias.

For a data point $(\mathbf{x}_i, y_i)$, the *functional margin* is given by:

$$\hat{\gamma}_i = y_i(\mathbf{w}^\top \mathbf{x}_i + b).$$

The *geometric margin* is:

$$\gamma_i = \frac{\hat{\gamma}_i}{\|\mathbf{w}\|}.$$

The objective is to maximize the minimum geometric margin among all samples.

For linearly separable data, the optimization problem is formulated as:

$$\begin{aligned} \min_{\mathbf{w},b} \quad & \frac{1}{2}\|\mathbf{w}\|^2, \\ \text{subject to} \quad & y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1, \quad i = 1, \dots, n. \end{aligned}$$

To address non-separable data, slack variables $\xi_i \geq 0$ are introduced:

$$\begin{aligned} \min_{\mathbf{w},b,\boldsymbol{\xi}} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n} \xi_i, \\ \text{subject to} \quad & y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0, \quad i = 1, \dots, n, \end{aligned}$$

where $C > 0$ is a regularization parameter balancing margin maximization and classification error.

The dual problem corresponding to the primal is:

$$\max_{\boldsymbol{\lambda}} \quad \sum_{i=1}^{n} \lambda_i - \frac{1}{2}\sum_{i,j=1}^{n} \lambda_i \lambda_j y_i y_j \mathbf{x}_i^\top \mathbf{x}_j,$$

$$\text{subject to} \quad 0 \leq \lambda_i \leq C, \quad \sum_{i=1}^{n} \lambda_i y_i = 0,$$

where $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \dots, \lambda_n]$ are the Lagrange multipliers.

After solving for $\mathbf{w}$ and b, or recovering them from the dual variables, the decision function for classifying a new point $\mathbf{x}$ is:

$$f(\mathbf{x}) = \text{sign}(\mathbf{w}^\top \mathbf{x} + b).$$

Which are calculate using the support vectors only.

- *The Time Taken to solve the optimization problem using the dual form of the SVM is -2.746834183249297.*

- Dual SVM time: 3.0025 seconds

- The Time Taken to solve the optimization problem using the primal form of the SVM is -2.7465648687052386.

- Primal SVM Time: 3.7959 seconds

Number of support vector for the given dataset . this is support vector indices in the dataset.

```
[5, 7, 14, 18, 39, 43, 45, 54, 55, 59, 64, 66, 74, 83, 85, 89, 100,
101, 103, 104, 105, 111, 114, 128, 131, 140, 145, 147, 152, 157, 160,
162, 166, 172, 179, 180, 184, 194, 196, 208, 215, 218, 219, 225, 234,
235, 246, 252, 253, 255, 257, 259, 261, 271, 272, 277, 278, 279, 283,
284, 288, 289, 304, 309, 322, 325, 341, 349, 352, 357, 365, 366, 371,
372, 374, 375, 377, 379, 380, 388, 391, 392, 400, 403, 420, 424, 429,
434, 443, 452, 455, 460, 485, 487, 494, 495, 502, 517, 534, 538, 544,
556, 557, 564, 565, 566, 570, 576, 580, 584, 592, 600, 601, 609, 610,
614, 616, 617, 624, 629, 633, 634, 636, 638, 657, 659, 665, 675, 676,
680, 696, 708, 712, 713, 717, 719, 722, 729, 745, 747, 749, 751, 754,
755, 760, 761, 764, 769, 772, 774, 811, 827, 833, 861, 867, 869, 872,
877, 878, 881, 883, 886, 888, 889, 890, 891, 896, 901, 903, 904, 910,
911, 921, 922, 923, 931, 939, 941, 942, 958, 961, 963, 971, 979, 987,
989, 993].
```

SOLUTION:- TASK4-

C=100, gamma=5, Best Accuracy: 100.00Test Misclassification Loss: 18.50

We choose these values of C and gamma as they were consistent with decision boundary for train dataset.

SOLUTION:- TASK5-

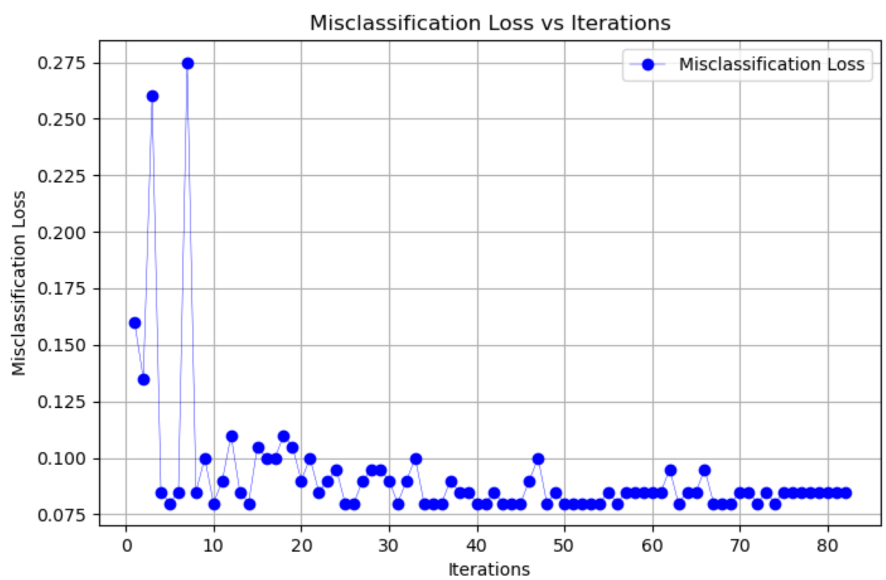


Figure 2: Misclassification rate vs iterations count for 10k iterations on the train data.

QUESTION 2

For the Task Part Please check the code

Deliverables

1:

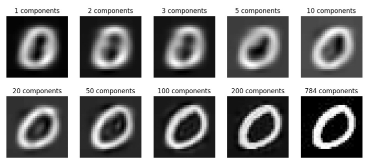
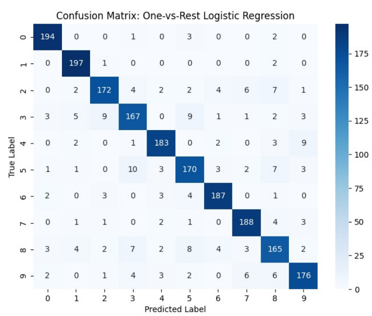
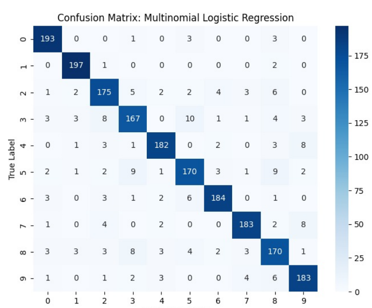


image reconstruction

2:

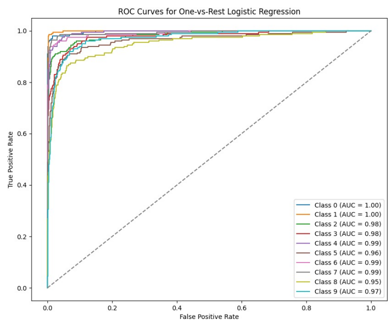


one vs rest binary



multinomial linear regression

3:



ROC graphs of binary classifiers

Average AUC score: 0.981.

QUESTION 3

Deliverables

1: No because we cannot find the inverse of the X matrix.

2:

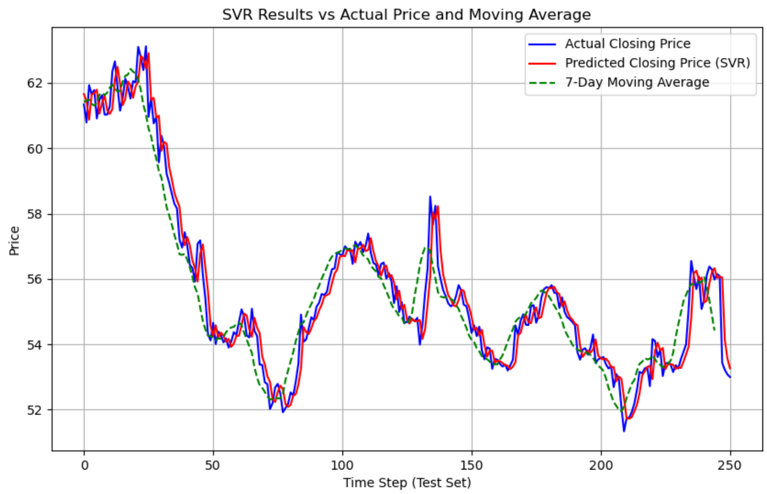
The Mean Squared Error (MSE) loss values for Ordinary Least Squares (OLS) and Ridge Regression (RR) on the two datasets are as follows:

$$\text{MSE loss for OLS on dataset 1} = 0.026559482003597135$$
$$\text{MSE loss for OLS on dataset 2} = 2453.1462230648613$$
$$\text{MSE loss for RR on dataset 1} = 1.20106215$$
$$\text{MSE loss for RR on dataset 2} = 13.25460649$$

The weight vectors obtained for OLS and RR on dataset 1 are:

$$w_{\text{OLS},1} = \begin{bmatrix} 0.36896147 \\ 0.27090286 \\ 0.20636517 \\ 0.96683252 \\ 0.48897033 \end{bmatrix}$$

$$w_{\text{RR},1} = \begin{bmatrix} 0.33022161 \\ 0.24860689 \\ 0.19816869 \\ 0.85295097 \\ 0.48041067 \end{bmatrix}$$



Dual problem with linear kernel