CS:5101 Machine Learning

Term 3 (Dec 2020 - Feb 2021)

Practice Assignment

SVM with Kernels

Due Date:

Follow the instructions given below carefully:

1. You must submit your code in a **single** python .ipynb notebook with naming format as follows: Firstname_Lastname_PracticeAssignment.ipynb
2. Your code must be properly commented explaining each step clearly.
3. If any of the above instructions are not followed, penalty will be there for the same.
4. Your code and answers will be checked for plagiarism and if found plagiarised, zero marks will be provided for this assignment. So make sure you actually code and solve the questions rather than noting down the answers.
5. NOTE: The total marks for this assignment is 5 marks.
6. It is NOT mandatory to submit this practice assignment.

Question 3 : Multi-class classification with SVM

In this exercise you will do handwritten digit classification using multi-class SVM with a Gaussian (RBF) kernel.

- (1 mark) Write a function "getKernelSVMSolution()" which
  - Takes Inputs:
    - X_train
    - Y_train
    - C [C is tradeoff hyperparameter]
    - λ [λ is kernel width]
    - X_test
  - Returns Output:
    - Y_prediction (for X_test)
- The problem deals with the classification of handwritten digits (10 classes, i.e., digits from 0-9). You are supposed to use the SVM with the Gaussian (RBF) kernel:

$$k(x, y) = \exp -\lambda \|x - y\|_2^2$$

. The training and test data is provided in the files USPSTrain.csv and USPSTest.csv and labels are provided in the files USPSTrainLabel.csv and USPSTestLabel.csv respectively.

- (1 mark) Write a function "OneVsOne()" which implements the multi class classification by using OnevsOne scheme.
  - Convert both the train and test data from USPSTrain.csv, USPSTrainLabel.csv and USPSTest.csv to SVM compatible format for each of the binary classification problems in the OneVsOne scheme.
    - Create appropriate vector for class label "Y" which contains only 1 and -1.
  - Then execute the binary SVMs with modified data.
  - Predict multi-class class label from your binary-class predictions.

  - Calculate an appropriate classification error for your multi-class classification task,i.e, F1 score.

- (1 mark) Following the similar scheme also write a function "OneVsRest()" which implements the multi-class classification by using the OnevsRest scheme.
- In both cases use $C = 100$ and $\lambda = 3/\gamma$, where γ is the median of all squared distances between training points, as parameters for the binary SVMs.
- (2 marks) Write a report "USPSreport_FirstnameLastname.pdf' containing following
  - Test errors for both the cases, i.e., OneVsOne and OneVsRest schemes.
  - Visually inspect the digits which have been misclassified using confusion matrix for multi-class classification.
  - How do you judge the result? Compare the quality of the classification obtained by the two multi-class schemes.
  - How do the two multi-class schemes compare in terms of runtime?
  - Also generate for both cases a figure (ErrorsOneVersusOne.png and ErrorsOneVersusRest.png) containing the misclassified images in the test set.
- Save your prediction on the test set in two files named as PredOneVersusOne.txt and PredOneVersusRest.txt.