

CS:5101 Machine Learning

Term 3 (Dec 2020 - Feb 2021)

Programming Assignment - 4

SVM with Kernels

Due Date: 17th Dec 2020 11:59 PM

Follow the instructions given below carefully:

1. You are allowed to use ONLY those python libraries (libraries means numpy etc.) which were shown in today's demo.
2. You must submit your code in a **single** python .ipynb notebook with naming format as follows:
Firstname.Lastname_assignment4.ipynb
3. For each question, submit the .csv file mentioned in the following naming format:
Firstname.Lastname_test< *question_number* >.csv
For example, for question 1, the format is John.Doe_test1.csv
4. For each question, create a separate text block containing the question followed by a code block containing the solution.
5. Your code must be properly commented explaining each step clearly.
6. If any of the above instructions are not followed, penalty will be there for the same.
7. Your code and answers will be checked for plagiarism and if found plagiarised, zero marks will be provided for assignment 4. So make sure you actually code and solve the questions rather than noting down the answers.
8. NOTE: The total marks for this assignment is 5 marks.
9. You are provided with following four files as part of this assignment:
 - data_linear_train.csv
 - data_linear_test.csv
 - data_nonlinear_train.csv
 - data_nonlinear_test.csv

Questions

1. (2.5 marks) Learn an SVM classifier using the dataset provided in data_linear_train.csv. Plot the datapoints in a 2D plot with different colors for the two classes. Print which kernel type is the best for this dataset. Also print the values for the best hyperparameters that you have selected depending on the kernel type. Plot the decision region plots learnt by the classifier (plots must have train data superimposed in the figure). Predict the labels Y for the datapoints provided in data_linear_test.csv and store them in a new column named 'predictions' in this csv file. Print the confusion matrix, accuracy scores for both the training and the test datapoints given in the two csv files. Submit the updated test csv file after renaming it according to the format in instruction 3.
2. (2.5 marks) Repeat question 1 using the files data_nonlinear_train.csv and data_nonlinear_test.csv.