CS5101 Machine Learning Lab (Aug - Dec 2021)
Assignment 5

(SVM Kernels)

Rules:

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 the 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`
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, the 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 mark for this assignment is 5 marks.
9. Deadline for submission is 18-Sep 23:00 IST

Questions:

1. You are provided with the following data files:
    a. data_linear_train.csv
    b. data_linear_test.csv
    c. data_nonlinear_train.csv
    d. data_nonlinear_test.csv

2. [2 marks] Learn an SVM classifier using the dataset provided in data_linear_train.csv(you can use scikit learn) and report the results with observation as mentioned below.
    - Plot the data points in a 2D plot with different colors for the two classes.
    - Print which kernel type is the best for this dataset.
    - Also, perform hyperparameter tunning and print the values for the best hyperparameters that you have selected depending on the kernel type.
    - Plot the decision region plots learned by the classifier.
    - Predict the labels y for the data points 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 data points given in the two CSV files In the last cell of the linear dataset. Submit the updated test csv file after renaming it according to the format in instruction 3.

3. [2 marks] Repeat question 1 with the files data_nonlinear_train.csv and data_nonlinear_test.csv.
4. [1 mark] Write your observations with comparisons in the last cell

[Bonus 1 mark]
Implement the below very basic algorithm for svm and compare your score

---

Pseudocode:

Step 1: Start with a random line of equation ax + by + c = 0.
Draw parallel lines with equations:
- ax + by + c = 1 and
- ax + by + c = -1

Step 2: Pick a large number as the number of iterations (niter)

Step 3: Pick a learning rate. (lr) remember, the learning rate should be small. Ideally 0 < lr < 1.

Step 4: Pick an expanding rate (λ)(For Parallel Hyperplanes)

Step 5: (repeat niter times)
- Pick a random point (x1, x2)
- If a point is correctly classified
  - Do nothing
- if a point above the hyperplanes is misclassified
  - Subtract lr*x1 to a, lr*x2 to b, and lr to c
- if a point below the hyperplanes is misclassified
  - Add lr*x1 to a, lr*x2 to b, and lr to c
- Multiply a, b, c by expanding rate(λ)
