CS:5101 Machine Learning

Term 3 (Dec 2020 - Feb 2021)

End Semester Exam

Due Date: 18th Jan 2021 11:59 PM

Follow the instructions given below carefully:

1. **Use only the models and the techniques that you have learned in this ML course.**

2. You must submit your code in a **single** python .ipynb notebook with naming format as follows:
   Firstname_Lastname_endsem.ipynb

3. For each question, create a separate text block containing the question followed by a code block containing the solution.

4. Follow each and every instruction given in each question carefully.

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 end semester exam. So make sure you actually code and solve the questions rather than noting down the answers.

Questions

1. Image Classification

   (a) Task: You are given a dataset of eye images divided into three folders namely, Open, Partially Closed and Closed which constitute the labels (y) for the images. This dataset is a modified version of the original dataset from the github repository: `https://github.com/hamzahsaleem/eye-blink-detection`. Perform multiclass classification using any technique of your choice, which you have learned till now in the machine learning course, to classify the eye images into open, partially closed or closed. Submission has to be as following:

      i. (1 point) Correct code which we will be able to run.
      ii. (1 point) Apply a technique to find out the best generalization accuracy and report it.
      iii. (2 points) Write a Text section in a separate text block describing what you have done and why.
      iv. (1 point) Generate a figure containing the misclassified images and write down your observations in the text block regarding why these images were misclassified.

2. Multivariate Time Series Forecasting

   (a) Task: You are provided with a Multivariate time series data containing observations for 400 timestamps. Each column in data file corresponds to one timestamp with 5 observations (corresponding to 5 variables/dimensions). Read this multivariate time series data and predict the future possible observations for 100 timestamps, using any of the suitable machine learning model you have learnt in this machine learning course. Evaluation scheme will be as follows:

      i. (1.5 point) Correct code which we will be able to run (do not include code parts which are not relevant for the task).
      ii. (1.5 point) Check with multiple hyperparameter values and choose best possible values for the model and report the mean squared error (MSE) for train data using the best model chosen.

iii. (1 point) Predict values for 100 more time stamps using the best model you have obtained after training and write your predictions to a csv file (in the same format of train data) named as follows $< Firstname\_Lastname\_prediction.csv >$.

iv. (1 point) Explain in a text block whether the task is supervised or unsupervised and why?; whether there are any interdependencies among the 5 variables and if yes, then explain which interdependencies you observed; how you chose the best hyperparameter value for the model, the criteria which you have checked for hyperparameter selection.

(b) (**1 point bonus mark**) We will be comparing your prediction values with actual values and top 5 good predictions based on MSE criteria will be provided with bonus marks.