CS564 - Foundation of Machine Learning
(Read all the instructions carefully and adhere to them.)

Assignment - 5: Stacking and Adaboost

Deadline: 21/11/2019 **Date: 13/11/2019**

Submission guidelines:

1.Markings will be based on the correctness and soundness of the outputs. Marks will be deducted in case of plagiarism.
2. Proper indentation and appropriate comments (if necessary) are mandatory.
3. You should zip all the required files and name the zip file as roll_no_of_all_group_members.zip, eg. 1601cs11_1601cs03_1621cs05.zip.
4. Upload your assignment (the zip file) in the following link:
https://www.dropbox.com/request/A8zfxy0ClFnzGkd74LKj

Q.1 Stacking is a meta-learner, it learns in two steps. The first model learns from the input data, and the second model learns from the predictions of model-I. Steps:

- Design a classification model using stacking based learning on the below given data. Use 5-cross validation for reporting the performance of the model.
- Apply different available Machine Learning based classifiers (such as Decision Tree, KNN, Random Forest, MLP, SVM) on the given dataset, change the categorical data to numerical(wherever needed).
- Save the learnt ML models.
- Load the saved models, and save their predictions.
- Use these predictions to do the final classification using Different ML classifiers.

Report the performance(Precision, Recall, f-measure, Accuracy) of different ML classifiers for step-1 and step-2.

Q2. In case of Boosting ensemble learning algorithm, in each iteration, a new model is created and the base model is being updated from the errors of the previous models.

Task 2.1 : Design Decision tree algorithm and report its Precision, Recall and F-measure. Task 2.2 : Design Boosting based ensemble model and report its Precision, Recall and F-measure. Task 2.3 : Give a comparative study between the above two tasks, i.e., plot a graph which will indicate the performance between the above two tasks.

Find the attached corpus for the above questions, from the below link :

https://drive.google.com/open?id=1bVwDpVzhUkNXkKxceF7xHDf6AWcQ6YIm

Data Description:

1. Number of Instances: 8124
2. Number of Attributes: 22
3. Dependent feature: type (the first column)
4. Independent features: the rest of the columns in the dataset

('cap-shape','cap-surface','cap-color','bruises','odor','gill-attachmen’,'gill-spacing','gill-size','gill-color','stalk-shape','stalk-root','stalk-surface-above-ring','stalk-surface-below-ring','stalk-color-above-ring','stalk-color-below-ring','veil-type','veil-color','ring-number','ring-type','spore-print-color','population','habitat')

Notes:

1. No predefined libraries are allowed to use for question no. 2 (i.e., Decision tree and Boosting ensemble model).
2. Cross-validation has to be performed.