ELL409/ELL791 II Semester 2017-2018

Assignment 1: Survey of classification algorithms and various model selection methods.

Max Marks : 100 Deadline : 15th Feb 2018 12 PM (After Noon)

January 24, 2018

1 Problem Statement

1.1 Survey of various classification algorithms

Search for various existing classification algorithms and write a comprehensive comparison of those methods. Their comparisons must include objective function, assumptions, advantages, draw-backs etc.

1.2 Summary of various scoring methods for classification

Explain in detail the various scoring methods that exist for classification i.e. accuracy, precision, recall, F1 Score, Area Under Curve(AUC), Receiver Operating Characteristics(ROC) etc.

1.3 Multiclass classification strategies

Comparison of **One-vs-One** and **One-vs-all** methods.
Extra Credit: Devise your own approach for this task.

1.4 Model selection Techniques

Exhaustive study of various model selection techniques. In addition, describe difference between hyper-parameters/tunable-parameters and learn-able parameters of any model.

1.5 Implementation of Logistic Regression Algorithm

Implement Logistic Regression on the dataset provided. Code everything on your own in **Python (v2 or v3)**, none of the existing machine libraries should be used for this task. (If you are using **Numpy**, only use the most basic matrix functions from the library only.)
Submit your train/validation/test score-plots over the iterations. In addition, mention train accuracy and test accuracy for given data explicitly. Report will be evaluated on the basis of various experiments performed and included in the report. Give proper justification of your choice of score.
Extra Credit: Experiment with various versions of Stochastic Gradient Descent (SGD) algorithms, like momentum method, Adagrad etc. Show the statistics of training time, accuracy and other aspects of algorithm. Credit will be given only if proper comparison of algorithms is provided.

1.6 Comparison

Compare your implementation of logistic regression algorithm with other classification algorithms that you have surveyed on the given dataset. For other classification algorithm you may use their existing implementations. In report, mention the results of these algorithms on the given dataset.

2 Submission guidelines

- Logistic regression algorithm should be implemented only in Python v2 or Python v3.
- Include proper comments in your code.
- Submit your implementation as .py file only.
- Code should be runnable. Else it will not be evaluated.
- Name your code as <EntryNumber_PythonVersion>.py
- Report should be written in LaTeX in the provided format.
- Name your report as <EntryNumber>.pdf
- **Report and Code will not be accepted beyond the given deadline. Should only be submitted via the specified medium**.

3 Evaluation

- Both code and report will be checked for plagiarism.
- If the similarity index for the code is found to be more than 25%, you will be awarded negative marks same as the weightage of that assignment.