Machine Learning (2019 Fall Semester)

Programming Assignment: Classification of Credit Approval Data Set

1. Benchmark Dataset : The machine learning model solves the problem of classifying the credit approval data set from the UCI data repository and determining whether the credit is approved or not. This data set can be downloaded from the following web page: http://archive.ics.uci.edu/ml/datasets/Credit+Approval

2. Machine Learning Models : The following three types of machine learning models are learned and their performances are evaluated.

2-1 Canonical Models : performing experiments by selecting two or more learning models. For example, Decision Tree, MLP, SVM.

2-2. Committee Machines : performing experiments with committee machines such as Random Forest and Ada Boost.

2-3. Deep Learning Models : performing experiments with deep learning models such as CNN and RBM. For the implementation of the learning models, you can use Scikit-Learn's packages or other possible sources.

3. Evaluation Methods : The learning and evaluation of each learning model uses a 10-fold evaluation method; That is, after dividing the entire data into 10 parts, the 9 parts are used for the evaluation and the remaining part is used for the evaluation. This procedure is repeated 10 times and then averaged to evaluate each learning model. Here, the following performance measures are used for the evaluation of models:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}, \; Precision = \frac{TP}{TP + FP},$$

$$Recall = \frac{TP}{TP + FN}, \; F_1 = \frac{2 \cdot Precision \cdot Recall}{Precison + Recall},$$

where

$$TP = true \; positive, \; TN = true \; negative, \; FP = false \; positive, \; FN = false \; negative.$$

4. Report Format : (전체 80점)
  1. Describe the input and functional structure of the machine learning model for the Liver Disorders data set. (10 points)
  2. Describe how and why learning parameters were determined for the selected learning model. (10 points)
  3. Using the methods of evaluation of the learning models described above, the measures of evaluation for each learning model are summarized. (20 points)
  4. Compare and analyze the evaluation results of the summarized learning models and explain the reasons for experiment results. (30 points)
  5. Summarize references to the learning models and sources used in this assignment. (10 points)

5. The assignment should be submitted by 19th of December.