

Machine Learning (Fall Semester, 2018)

2018.11.06

Programming Assignment: Classification of Credit Approval Data

1. **Benchmark Dataset:** In the UCI data repository, choose the credit approval data set and solve the problem of approving or rejecting credit card applications using machine learning models. Here, download the material from the following web page:

<http://archive.ics.uci.edu/ml/datasets/Credit+Approval>

2. **Machine Learning Models:** Investigate the data on the credit approval data set for three types of machine learning models and evaluate their performances.

2-1 Canonical Models:

Experiment with two or more learning models. For example, Decision Tree, MLP, SVM, and so on.

2-2. Committee Machines:

Experiment with one of committee models such as Random Forest, Ada Boost, etc.

2-3. Deep Learning Models:

Experiment with one of deep learning models such as CNN, DAE, DBN, etc.

Here, you can use the Skit-Learn package or other possible sources for the implementation of the learning models.

3. **Evaluation Methods:** Learning and evaluation of each learning model uses a 10-fold evaluation method; that is, after dividing the whole data by 10, 9 is used for learning and the remaining one is used for evaluation. Then, after repeating this process 10 times, each learning model is evaluated by its average performances. Here, we use the following performance measures:

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

$$Recall = \frac{TP}{TP + FN}, \quad F_1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}, \quad \text{where}$$

TP = true positive, TN = true negative, FP = false positive, FN = false negative.

4. Report Format: (Total 60 points)

1. Explain the functional structure of the input features and machine learning models for credit approval data classification. (10 points)
2. Explains the method of determining the learning parameters for the selected learning model and its reasons. (10 points)
3. Using the evaluation method of the learning model described above, describe the performance of each learning model. (15 points)
4. Compare the experiment results of machine learning models and describe the proper reasons for these results. (20 points)
5. Organize references of learning models and sources used in this assignment. (5 points)

5. This assignment is due on 20th, December.