Computer Exercise 4

Oct. 28, 2021

Task:

Classify patients' survival (0: survived; 1: dead) by selecting a subset of features from the original 108 features.

Goal:

1. Select a subset of features or extract new features from original 108 features (age, BMI, height, weight, heart rate, blood pressure, etc.) for the classification task. Observe the effect of feature selection on classification performance.
2. Find a public package of Random Forest (RF) (in scikit-learn, etc.) for the task. Study the method used for evaluating the relative contribution of the features in the random forest. Make observations on the performances and effects of optional choices on the performance.

Data:

Please check the "data1forEx" folder for the following datasets.

Datasets	Sample size	Feature data file	Class label file
TrainingSet-1	5000	train1_icu_data.csv	train1_icu_label.csv
TestSet-1	1097	test1_icu_data.csv	test1_icu_label.csv

Experiment 8 (Feature Selection):

1) Design your own feature selection experiment with publicly available packages or with your own codes. Describe the method for feature selection, the method for classification and the scheme you use for evaluating the performance. Report the selected features.
2) Compare the training error, cross validation error and test error before and after the feature selection.
3) Analyze and discuss observations in the experiment.

Experiment 9 (Lasso):

1）Learn the principle of Lasso and find a package to do the feature selection and classification task using Lasso. Compare the results with Experiment 8 and discuss your understanding and observations on Lasso.

Experiment 10 (Random Forest):

1) Find a free RF packages and learn its usage. Describe the algorithm(s) it used to assess the feature contributions, and the hyper-parameters and options users need to choose.
2) Design your own experiments with multiple parameter settings to train RF classifiers on TrainingSet-1. Explain why you design your experiments in this way. Apply the trained RF classifiers on the TestSet-1.
3) Discuss the observations in the experiment. Analyze the feature contributions for the classification results. Compare the evaluation of features with the selection results in Experiments 8 and 9.

Experiment Report:

- Write an experiment report to describe and analyze the experiment observations. The report should also include the short essay on parameter choices.
- Provide detailed supplementary materials that should include at least the following:

- A readme file containing information on all supplementary files, programming environment and parameters used in the experiments (if any)
- Source codes (should let TAs to be able to run the code and reproduce your experiments)
- Experiment result files

Due date: Nov. 10 (Wed.) 23:00 Beijing time