Homework 5

Oct. 26, 2023

Task:

Classify patients' survival (0: survived; 1: dead) using 108 features (a mixture of numeric and binary variables) from their Intensive Care Unit (ICU) records, such as age, BMI, height, weight, heart rate, blood pressure, etc.

Goal:

- Find a public package of Random Forest (RF) 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 "data1forEx1to4" 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

Requirements:

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. 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. Discuss the observations in the experiment. Analyze the feature contributions for the classification results.

2) 【Optional】 Apply some of the feature selection methods we covered in Chapter 8, compare and discuss the evaluation of features obtained by RF with the selection results.

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. 1 (Wed.) 23:00 Beijing time