

Computer Exercise 3

Oct. 14, 2021

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:

Use a publicly available Python package of SVM (e.g. scikit-learn) for the classification task. Make observations on the learning procedures, 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 |

Experiment 6 (SVM):

- 1) Read the document of SVC function in the scikit-learn package (<https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html>). Check the parameters about the kernel function and the training process. Write a short essay (within 1 page) about the default and optional values of these parameters and explain how to select parameters for different tasks.
- 2) Use TrainingSet-1 to train the SVM with linear, Gaussian and polynomial kernel with at least 3 different sets of parameters (kernel parameters, C, etc.). Calculate the training error and the cross-validation error on the training set.
- 3) Apply the trained SVM classifiers on TestSet-1. Calculate the test error.
- 4) Discuss and analyze your observations in the experiments.

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: Oct. 27 (Wed.) 23:00 Beijing time