

Homework 2

Sep. 28, 2023

Task:

Classify patients' survival (0: survived; 1: dead) using 108 features (a mixture of numeric or binary variables) from their Intensive Care Unit (ICU) records, such as age, BMI, height, weight, heart rate, blood pressure, etc. (Detailed descriptions available in the data folder.) Please find a public package of Logistic Regression (LR) or write your own codes to do experiments on the data. Observe the learning procedures, performances and effects of optional choices on the performances.

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
TrainingSet-2	1475	train2_icu_data.csv	train2_icu_label.csv
TestSet-1	1097	test1_icu_data.csv	test1_icu_label.csv
TestSet-2	450	test2_icu_data.csv	test2_icu_label.csv

Note: You may need to scale features to the same reasonable range before training. Meaning of each feature can be found in “feature_description.csv”. Original data were from Kaggle (<https://www.kaggle.com/c/widsdatathon2020/data>).

Requirements:

- 1) Use TrainingSet-1 to train the classifier with Logistic Regression. Calculate the training error and cross validation error on the training set. Apply the trained classifier on TestSet-1. Calculate the test error. Using the Python package *matplotlib* to draw the ROC curve according to the test results.
- 2) **【optional】** Analyze the significance of association between each feature and patients' survival. You may need to study some materials beyond the course content by yourself. (Hints: you may refer to available Python packages like *statsmodels*).
- 3) **【optional】** For the logistic function $\theta(s) = \frac{1}{1+e^{-s}}$, show the following relations:

$$\theta(s) = \frac{1}{1+e^{-s}} = \frac{e^s}{e^s + 1},$$

$$\theta(-s) = 1 - \theta(s),$$

$$\theta'(s) = \theta(s)(1 - \theta(s)).$$

(Note: Optional requirements will not gain you extra score in your homework, but will help you to gain useful skills.)

Experiment Report:

- Write an experiment report to describe and analyze the experiment observations (no more than 4 pages).
- Provide detailed supplementary materials that should include at least the following:
 - A readme file with information on all supplementary files, programming environment and parameters used in the experiments (if any),
 - Links to the original source of the packages you used, or the source codes of your own program (TAs should be able to run the code and reproduce your experiments), and
 - all experiment result files.

Due date: Oct. 11 (Wed.) 23:59 Beijing time