

## Computer Exercise 1

Sep. 23, 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. (Detailed descriptions available in the data folder.)

**Goal:**

Write your own computer programs of Fisher Linear Discriminant (FLD), Perceptron and Logistic Regression (LR) algorithms in Python, without directly importing packages of these algorithms. Make observations 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
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>).

**Experiment 1 (FLD):**

Use TrainingSet-1 to calculate the discriminant function using FLD. Apply the discriminant function on TestSet-1. Calculate the error rate.

**Experiment 2 (Perceptron):**

- 1) Use TrainingSet-1 and TrainingSet-2 to train the Perceptron with both the “Fixed increment rule” and the “Variable increment rule” for each training set. Check the convergence of all the training processes.
- 2) If the training process does not converge, take certain strategy to force it to converge after a reasonable period of training. Compare the number of iterations needed for the two increment rules. Draw the learning curve for each training process.
- 3) Apply the trained perceptrons using TrainingSet-2 to TestSet-1 and TestSet-2, respectively. Compare the error rates.
- 4) Apply the trained perceptrons using TrainingSet-1 to TestSet-1. Compare the error rates.
- 5) Discuss your observations on the results.

**Experiment 3 (LR):**

- 1) Use TrainingSet-1 to train the classifier with Logistic Regression. Calculate the training error and cross validation error on the training set.
- 2) Apply the trained classifier on TestSet-1. Calculate the test error.
- 3) Using Python package matplotlib to draw the ROC curve according to the test results.
- 4) 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).

### **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 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. 6 (Wed.) 23:00 Beijing time**