Computer Exercise 1

Sep. 17, 2020

Task:

Classify the occurrence of patients' heart failures (1: occur; 0: no heart failure) using 12 clinical features such as age, sex, anaemia, diabetes, platelets in the blood, level of sodium in the blood, etc.

Goal:

Write your own computer programs of Perceptron and FLD algorithms in Python for the classification task. Make observations the learning procedures, performances and effects of optional choices on the performance.

Data:

Please check the "data1" folder for the following datasets.

Datasets	Sample size	Feature data file	Class label file
TrainingSet-1	100	train_hf_data_sub.csv	train_hf_label_sub.csv
TrainingSet-2	240	train_hf_data.csv	train_hf_label.csv
TestSet-1	59	test_hf_data.csv	test_hf_label.csv
TestSet-2	199	test_hf_data_add.csv	test_hf_label_add.csv

Note: You may need to scale features to the same reasonable range before training. Meanings of the features can be found in "meaning_of_features.csv". Data were downloaded from Kaggle (https://www.kaggle.com/andrewmvd/heart-failure-clinical-data).

Experiment 1 (Perceptron):

1) Use TrainingSet-1 and TrainingSet-2 to train two perceptrons, respectively, with both the "Fixed increment rule" and the "Variable increment rule". Observe on the convergence of the training processes.

2) If the training process does not converge, adopt certain strategy to force the algorithm 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-1 to TestSet-1 and TestSet-2, respectively. Compare the error rates.

4) Apply the trained perceptrons using TrainingSet-2 to TestSet-1. Compare error rates with Perceptron

Experiment 2 (FLD):

5) Use TrainingSet-2 to calculate the discriminant function using FLD. Apply the discriminant function on TestSet-1. Calculate the error rate.

Experiment Report:

- Write an experiment report to describe and analyze the experiment observations.
- Provide detailed supplementary materials that includes at least the following materials:

   - A readme file with information on all supplementary files, programming environment and parameters used in the experiments (if any)

   - Source code (It should let TAs be able to run the code and reproduce your experiments.)

   - Experiment result files

All files should be packed in one file for submission. Acceptable formats are .zip, .rar, and .7z.

Report due date: Sept. 30 (Wednesday), 22:00 Beijing time, 2020.