Computer Exercise 2

Sep. 30, 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:

1. Use publicly available Python packages of KNN (from scikit-learn, etc.) for the classification task. Make observations on the performances and effects of optional choices on the performance.
2. Write your own computer programs of MLP in Python 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 4 (KNN):

1) Download a free KNN package and learn the usage. Describe the algorithm(s) used to search nearest neighbors, the hyper-parameters and options users need to choose.
2) Design your own experiments with multiple parameter settings to train the KNN classifier on TrainingSet-1. Explain why you design your experiments in this way. Apply the trained KNN classifiers on TestSet-1.
3) Discuss the observations in the experiment. Analyze how the choice of K influence the training process and test performance.

Experiment 5 (MLP):

4) Design an MLP classifier and use TrainingSet-1 to train the MLP. Show the learning curve. Calculate the training error and the cross-validation error on the training set.
5) Apply the trained MLP classifier on TestSet-1. Calculate the test error.
6) Try some strategies suggested in the lecture slides or in reference books to improve the model performance. Study their effects on the results.

Experiment Report:

- Write an experiment report to describe and analyze the experiment observations.
- 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. 13 (Wed.) 23:00 Beijing time