Homework 6

Nov. 2, 2023

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:

- Find a package (or write your own computer program) of Naïve Bayes Classifier in Python for the classification task. Practice modeling data with certain distribution. Gain better understanding on the Bayesian decision.

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

Requirements:

1) Choose proper data distribution for each clinical feature and explain your reasons.
2) Use TrainingSet-1 to build a Naïve Bayes classifier. Calculate the training error and cross-validation error on the training set.
3) Apply the trained classifier on TestSet-1. Calculate the test error.
4) Consider the risk of making wrong decision and propose a table of decision risks. Set up a minimal-risk Bayesian decision criterion using the trained Naïve Bayes classifier. Apply the criterion on TestSet-1 and calculate the test error.
5) Analyze and discuss 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: Nov. 8 (Wed.) 23:59 Beijing time