Diagnose Heart Anomaly

Dajun Gu

November 19, 2019

In this homework three, the Naive Bayesian classifier is used for identify people who have heart anomaly. The dataset is binary files and come from instructor Bart. The programming language in this project is Python3.

According to the Bayes's rule, the probability of abnormal or normal can be represent by $L_i = log(Pr(H_i)) + \sum_j logPr(F_j = f_j(c)|H_i)$, the Naive Bayesian algorithm has been implement.

For three different data sets, the accuracy of abnormal instances is more important. Because if the algorithm identify heart disease patient as a normal person, the possibility of death will increase. However, if the program classify a normal personal as a heart disease patient, may be this person should pay money to treat, but in the further treatment, the likelihood of this patient being misdiagnosed will decrease.

Averagely, the SPECTF data set gave the highest accurate rate. Always, this data set have the highest true negative rate. It classified all abnormal instances correctly.