CS 411 Assignment 4 Report

Guangyu Zhou gzhou6

1. **Introduction:**

In this machine problem, I implemented two classification methods. Both of them are supervised learning algorithm which requires a training set in advance.

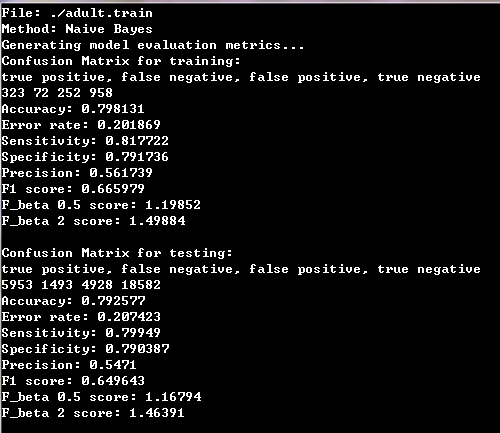
The first classification methods I implemented is Naïve Bayes: Based on the Bayes’ Probability Theorem, and assume that attributes are conditionally independent, the classifier first run on the training set and then test on the test set.

The second classification methods is Ada boost by utilizing the weights and the above Naïve Bayes model several times to generate an ensemble method.

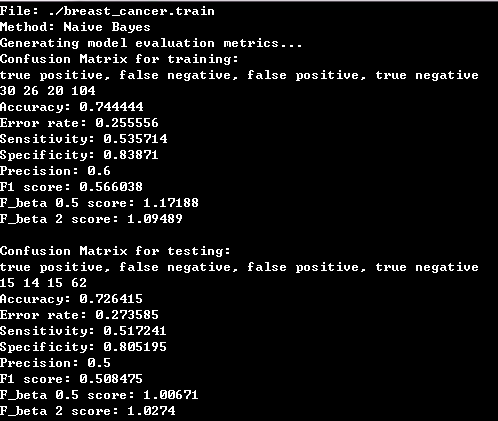
1. **Evaluation measurement result**

Naïve Bayes:

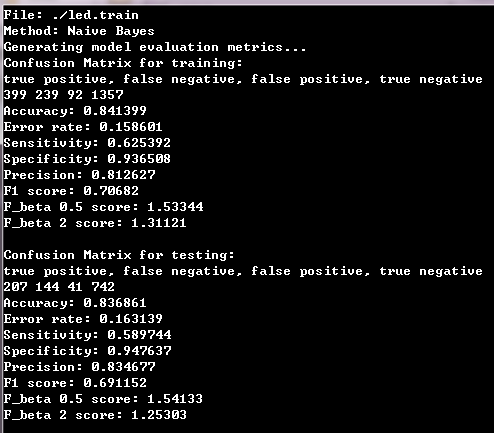
1. Adult



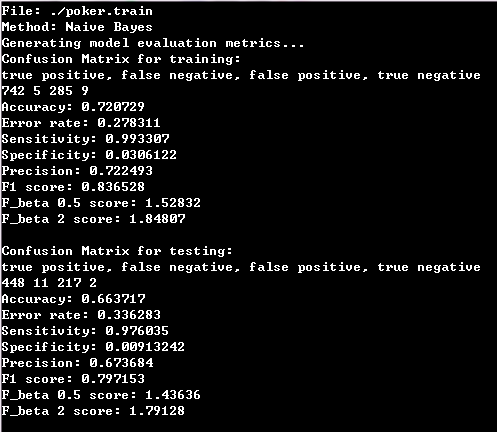
1. Breast cancer



1. Led

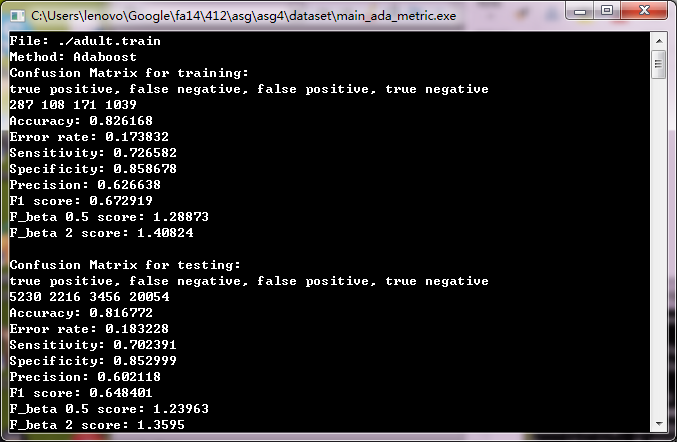


1. Poker

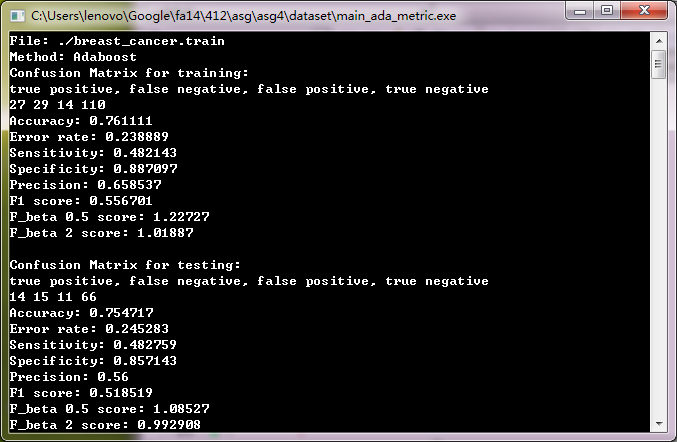


Ada boost:

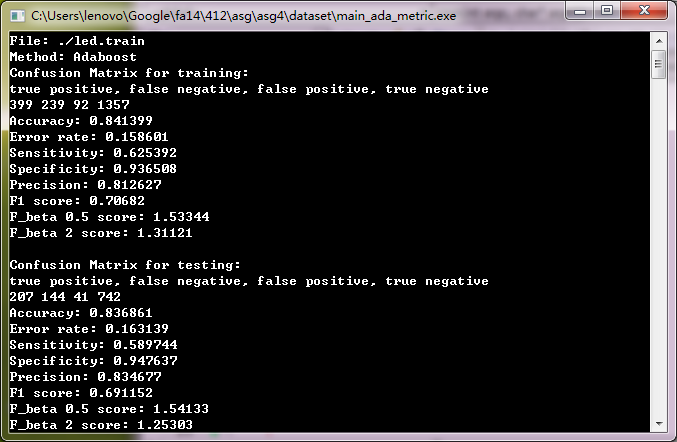
1. Adult



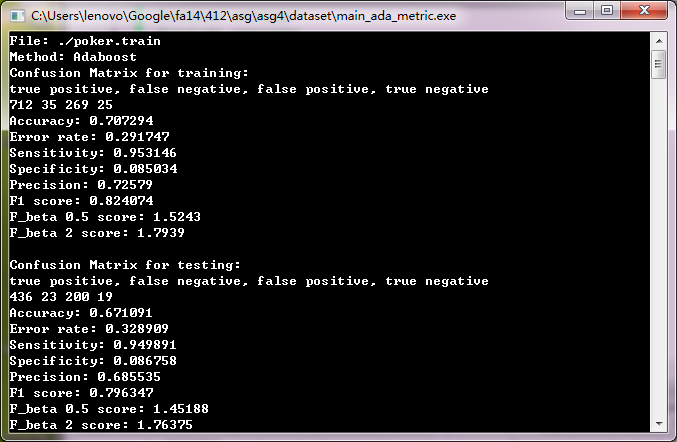
1. Breast cancer



1. Led



1. Poker



1. **Chosen of Parameters**
2. First of all, I use Laplacian correction to smooth the dataset by adding 1 to each 0 element to avoid zero probability and minimize the error.
3. Secondly, I assign number of weak learner to be 6, after trying different rounds from 4-10, it turns out that 6 is the best.
4. **Conclusion**

After careful chosen of parameters and comparison of the basic classification and the ensemble method, it can be concluded that there is an improvement.

Below are the ratio of improvement for the accuracy of the test file:

1. Led +0%
2. Breast Cancer + 3%
3. Poker 1%
4. Adult + 2%

The reason behind this improvement is that by assigning different weight for each round of weak learner, those false predications will have a greater probability being modified, thus the result turns out to be more precise.