CSE 5522 Homework #3

Cheng Zhang

Dept. Computer Science and Engineering, The Ohio State University
zhang.7804@osu.edu

1 Problem 1

(a)

$$\begin{aligned} Entropy(P(c), P(v), P(s)) &= -P(c)\log_2 P(c) - P(v)\log_2 P(v) - P(s)\log_2 P(s) \\ &= -0.7 \times \log_2 0.7 - 0.2 \times \log_2 0.2 - 0.1 \times \log_2 0.1 \\ &= 1.1567 \end{aligned}$$

(b)

$$Entropy_{max} = -\frac{1}{3} \times \log_2 \frac{1}{3} \times 3 = 1.585$$

(c) Testing top-level splits: Gender

$Entropy(female) = -0.67 \times \log_2 0.67 - 0.22 \times \log_2 0.22 - 0.11 \times \log_2 0.11 = 1.2179$

$Entropy(male) = -0.80 \times \log_2 0.80 - 0.12 \times \log_2 0.12 - 0.08 \times \log_2 0.08 = 0.9161$

$Gain = Entropy(P(c), P(v), P(s)) - [P(female) * Entropy(female) + P(male) * Entropy(male)]$

$= 1.1567 - 0.75 \times 1.2179 - 0.25 \times 0.9161 = 0.01425$

Testing top-level splits: StudentType

$Entropy(grad) = -0.76 \times \log_2 0.76 - 0.16 \times \log_2 0.16 - 0.08 \times \log_2 0.08 = 1.0154$

$Entropy(undergrad) = -0.64 \times \log_2 0.64 - 0.24 \times \log_2 0.24 - 0.12 \times \log_2 0.12 = 1.2732$

$Gain = Entropy(P(c), P(v), P(s)) - [P(grad) * Entropy(grad) + P(undergrad) * Entropy(undergrad)]$

$= 1.1567 - 0.5 \times 1.0154 - 0.5 \times 1.2732 = 0.0124$

Therefore, gender is an appropriate first decision be for the decision tree algorithm.

2 Problem 2

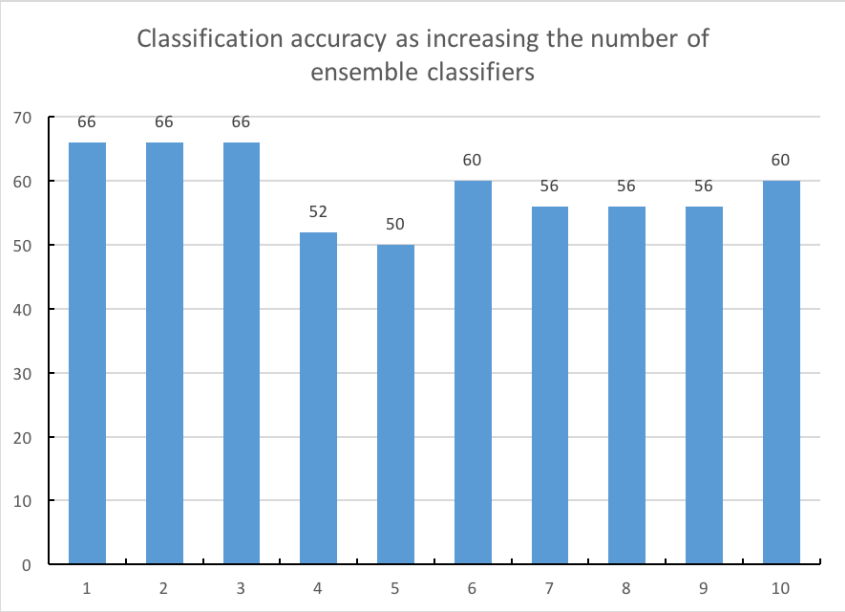

Figure 1: Classification accuracy as increasing the number of ensemble classifiers. The number of ensemble classifiers is from 1 to 10.

(a) The best first question is Question#4 (mood: silly, happy, tired). The accuracy of that classifier on the test set is 66%. And the average probability is 65%.

(b) Fig. 1 shows the classification accuracy as increasing the number of ensemble classifiers. The number of ensemble classifiers is from 1 to 10. Please refer to source code and README.md file for more details.

Extra credit

In this situation, adaptive boosting is from decision stumps to decision trees. In other words, we could implement Random Forest algorithm on this dataset. From the observations in (b), boosting method does not help to the improve the performance. To this end, it is hardly to enhance the classification result by using more decision trees.

3 Problem 3

(a) Single layer perceptron.

The function can be represented as $x_1 + x_2 + x_3 + ... + x_7 - 3.5 > 0$

(b) Multi-layer perceptron.

This is a multi-layer perceptron because it is not lineraly seperable. It is higher dimensional XOR problem.

(c) Single layer perceptron.

The function can be represented as $x_2 - 3x_1 > 0$

(d) Multiple layer perceptron.

The boundaries functions are represented as $x_1 + x_2 - 10 < 0$, $x_1 > 0$, and $x_2 < 0$.

4 Problem 4

For each iteration, the majority label of the training data will be different from the label of the validation data samples. For example, if the validation data instance is negative, the majority label of the training data (100 positives and 99 negatives) will be positive. Therefore, the accuracy will always be 0% for all runs.