**Introduction to Machine Learning (Spring 2019)**

**Homework #6 (50 Pts, June 23)**

**Student ID**

**Name**

**Instruction:** We provide all codes and datasets in zip files. Please write your code to complete Decision Tree and Gaussian Naïve Bayesian models. Compress only 4 Files ‘Decision\_Tree\_Answer.py’, ‘NaïveBayesian\_Answer.py’ ‘utils\_Answer.py’ & your report and submit with the filename ‘HW6\_STUDENT\_ ID.zip’.

**1 – (1) [10 pts]** Implement Decision Tree in ‘Decision\_Tree\_Answer.py’, some parts of ‘utils\_Answer.py’

1. **[Gini index]** Implement Gini\_index in ‘utils\_Answer.py’
2. **[Entropy]** Implement Entropy in ‘utils\_Answer.py’
3. **[Find Best feature]** Implement a ‘Find\_Best\_Feature’ in ‘Answer.py’.

**Answer: Fill your code here. You also have to submit your code to i-campus.**

**NOTE 1**: **You should write your codes in ‘EDIT HERE’ signs.** It is not recommended to edit other parts. Once you complete your implementation, run the check codes (‘code\_validation.py’) to check if it is done correctly.

**NOTE 2**: **Read the instructions in template codes VERY CAREFULLY.** Funcionality and input, output format of any function must be the same as what is written.

**1 – (2) [10 Pts]** Experiment results

1. you are given 2 dataset (Heart, Carseats) with Binary classification(Yes or No). Measure the performance of Decision tree given setting environments.

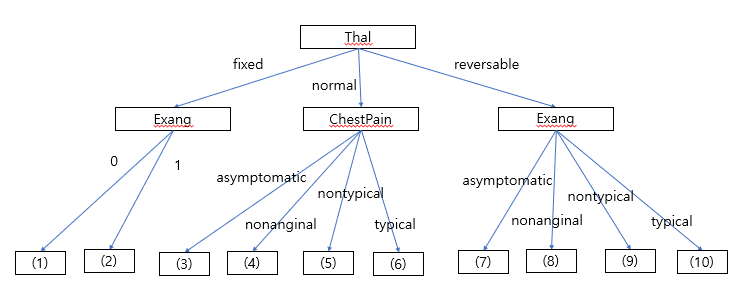
**Answer: Fill the blank in the table. Show the plot of training & test accuracy with a brief explanation.**

**[Decision Tree]**

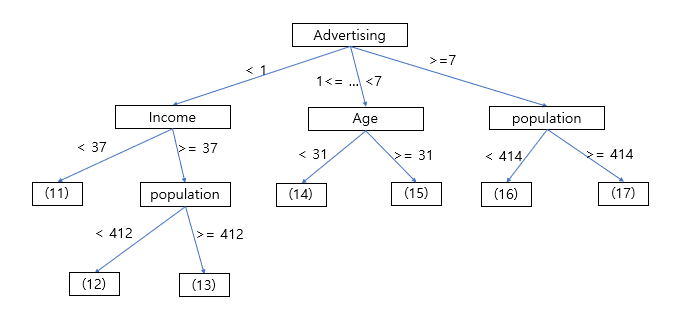
|  |  |  |  |
| --- | --- | --- | --- |
| Dataset | Impurity function | Max depth | Accuracy |
| Heart | Entropy | 2 | 0.6842 |
| Heart | Gini\_index | 3 | 0.7544 |
| Carseats | Entropy | 3 | 0.9125 |
| Carseats | Gini\_index | 3 | 0.8875 |

**1 – (3). [10 pts]** Analysis

Heart\_dataset, (Impurity function = entropy, max\_depth = 2)



Carseat\_dataset, (Impurity function = entropy, max\_depth = 3)



Above figure is the result of Decision Tree for the given condition. Please fill the label of leaf node. And Explain which node can be pruned.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| (1)No | (2)Yes | (3)No | (4)No | (5)No |
| (6)No | (7)Yes | (8)Yes | (9)No | (10)No |
| (11) No | (12) No | (13) No | (14) No | (15) Yes |
| (16) Yes | (17) Yes |  |  |  |

2 - (1). **[10 pts]** Implement Gaussian Naïve Bayesian in ‘NaiveBayesian\_Answer.py’ and the rest parts of ‘utils \_Answer.py’ (To prevent that likelihood becomes zero, please apply the Laplacian smoothing technique.).

1. **[likelihood, prior]** Implement ‘fit’ in ‘NaïveBayesian\_Answer.py’ for the following function: (likelihood, prior).
2. **[Posterior]** Implement ‘predict’ in ‘NaiveBayesian\_Answer.py’ for the posterior probability.
3. **[Gaussian probability]** Implement Gaussian\_prob in ‘utils\_Answer.py’. for the following function:

2 - (2). **[10 pts]** Experimental results.

1. you are given 2 dataset (Heart, Carseats) with Binary classification(Yes or No). Measure the performance of Decision tree and fill the blank.

**[Naïve Bayesian]**

|  |  |
| --- | --- |
| Dataset | Accuracy |
| Heart | 0.7368 |
| Carseats | 0.6125 |