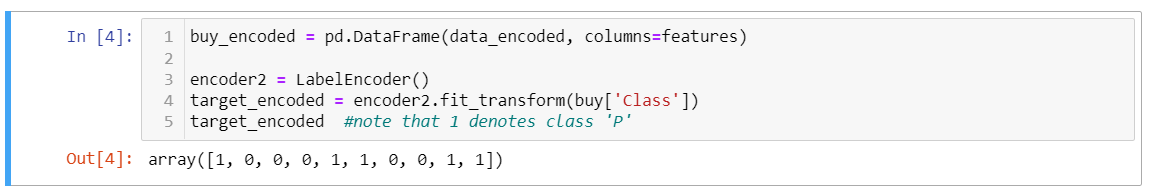
Data mining (2022/11/14)

1. (20%) With the following training dataset, use the Naïve Bayes classification algorithm to predict the class label of (A=0, B=1, C=0). Please set the Laplacian smoothing parameter to 1 to prevent the “0-probability problem”. Please give all the calculation steps, not just give a random guess for the answer. Please note that the attributes of the dataset include “A”, “B”,” C” and “Class”. Furthermore, “Class” is the class label (the response variable)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Index | A | B | C | Class |
| 1 | 0 | 0 | 0 | P |
| 2 | 1 | 0 | 1 | N |
| 3 | 0 | 1 | 1 | N |
| 4 | 0 | 1 | 1 | N |
| 5 | 0 | 0 | 1 | P |
| 6 | 1 | 0 | 1 | P |
| 7 | 1 | 0 | 1 | N |
| 8 | 1 | 0 | 1 | N |
| 9 | 1 | 1 | 1 | P |
| 10 | 1 | 0 | 1 | P |

Ans: Positive class ‘P’





1. (20%)With the following dataset, answer the following questions.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Instance | a1 | a2 | a3 | Target Class |
| 1 | T | T | 1.0 | + |
| 2 | T | T | 6.0 | + |
| 3 | T | F | 5.0 | - |
| 4 | F | F | 4.0 | + |
| 5 | F | T | 7.0 | - |
| 6 | F | T | 3.0 | - |
| 7 | F | F | 8.0 | - |
| 8 | T | F | 7.0 | + |
| 9 | F | T | 5.0 | - |

1. Using Gini function as attribute selection measure; select the better splitting attribute among attributes a1 and a2 for the root. (5%)
2. For a3, which is a continuous attribute, compute the Gini value for every possible split. What is the best splitting condition for this attribute? (10%)
3. According to the results of questions a and b, draw a two-level decision tree and calculate its classification accuracy based on the training dataset. (5%)

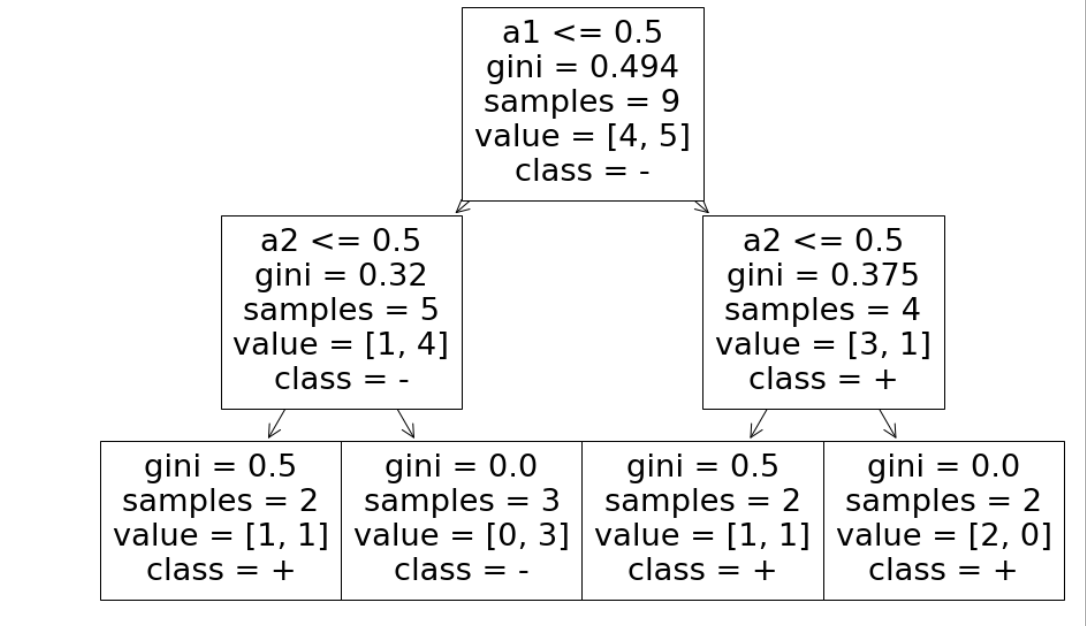
An example for splitting a continuous attribute is shown in the following:

Gini = =2p1 \*p2 (for two-class case)

scn0001.tif

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| class | + | - | + | - | - | + | + | - | - |
| a3 | 1 | 3 | 4 | 5 | 5 | 6 | 7 | 7 | 8 |

Split at a3≤2, Gini=0.468



**Accuracy: 0.78**

1. (20%) Given a simple MLP with two inputs a and b, one hidden node c and one output node d, please find the updated weight of **Wcd** and **Wbc** after training the network once with the following training example. Assume that the initial values of the weights of wac, wbc, wcd are 0.1, 0.2, 0.2, respectively, and the learning rate is 0.9.

The training example:

a b d

1 0 1

(Notes: 1. Let not consider the biases (i.e., no biases).

2. Wac denotes the weight of the link from node a to node c. Similarly, Wbc and Wcd denote the weight of the link from node b to node c and the weight of the link from node c to node d, respectively.)

Appendix：

**Output nodes**

**Input nodes**

**Hidden nodes**

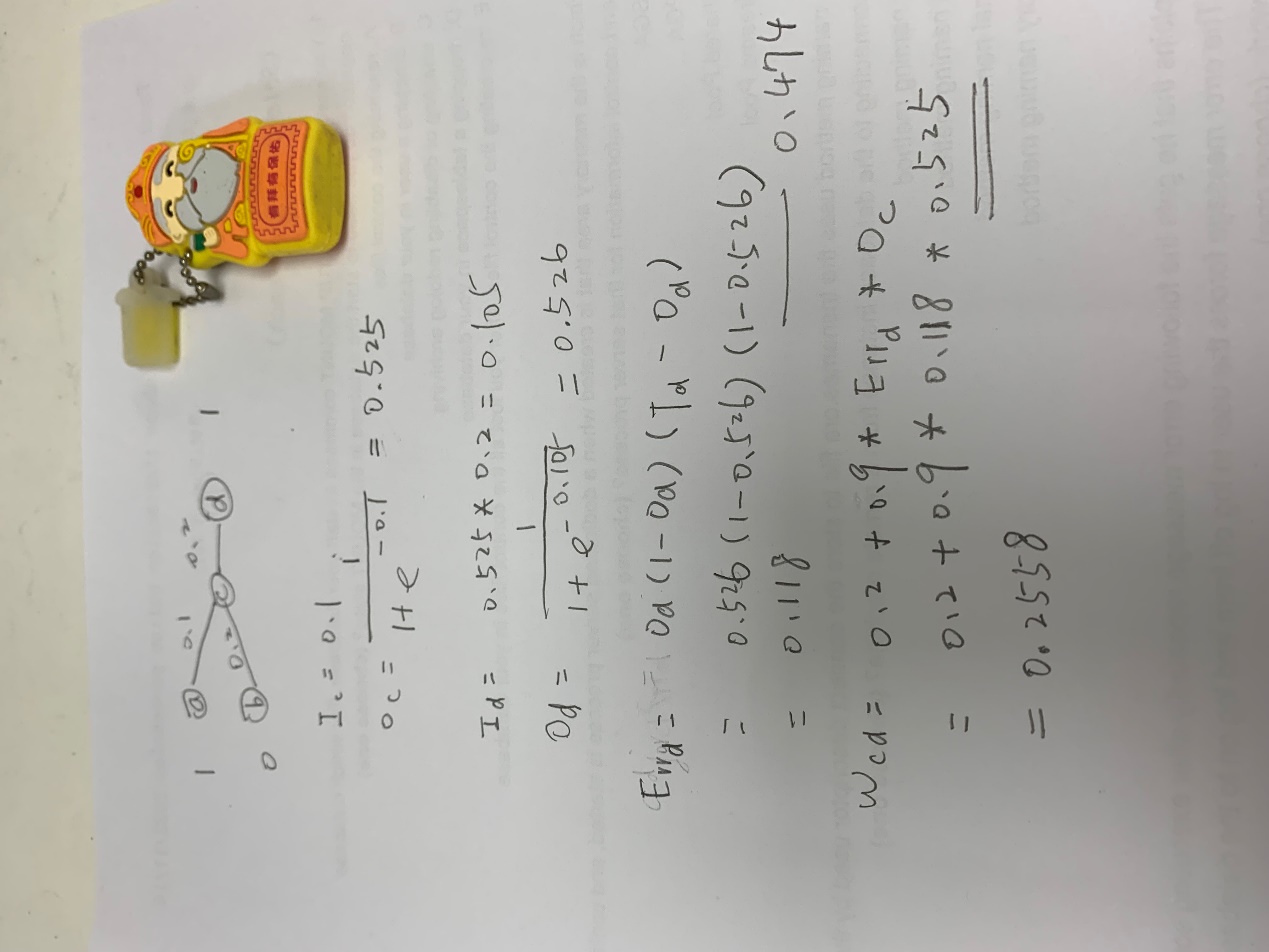
**Output vector**

**Input vector: *xi***

*wij*

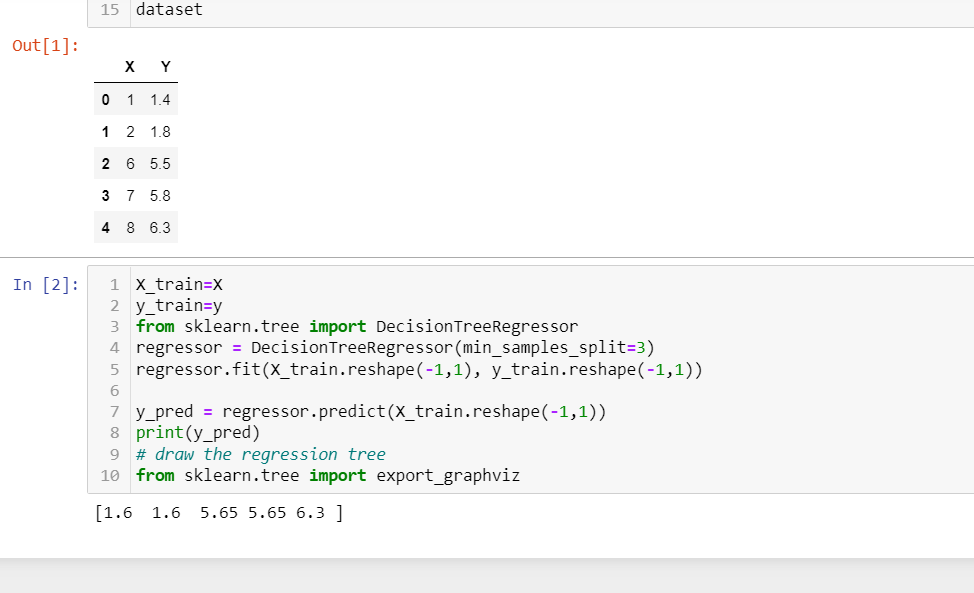


Ans: 1. Wbc does not change since input of node ‘b’ is zero



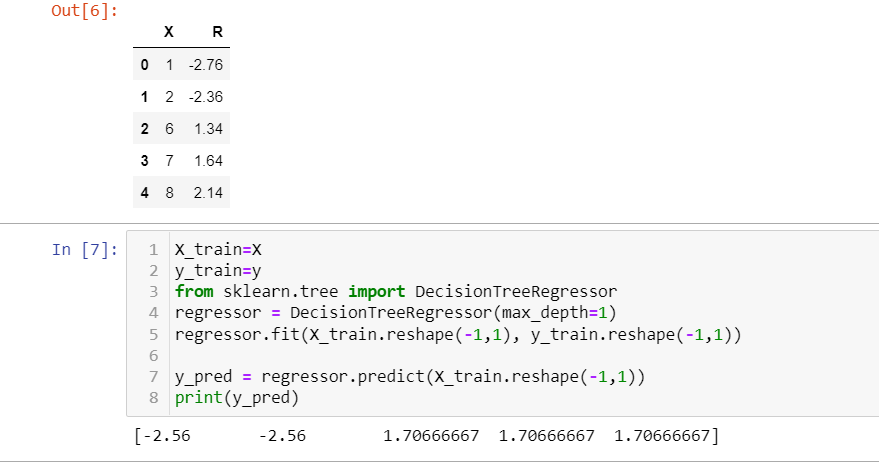
1. (20%) The following dataset contains two columns, X and Y. Please build a regression tree with X as the input variable and Y as the output variable (10%). Then, fill in the predicted value for each X using the regression tree. (10%)

**Note: Set the minimum number of samples in a node to three to allow splitting the node in building the regression tree. In other words, if the number of samples in a node is less than or equal to two (≤2), no splitting of this node is allowed.**



|  |  |  |
| --- | --- | --- |
| **X** | **Y** | **Predicted** |
| **1** | **1.4** | **1.6** |
| **2** | **1.8** | **1.6** |
| **6** | **5.5** | **5.65** |
| **7** | **5.8** | **5.65** |
| **8** | **6.3** | **6.3** |

1. (20%, Gradient Boosting, with learning rate of 0.5) With the same dataset for problem 4, use gradient boosting to construct a model for predicting Y values. To reduce the required computation, use only one one-level regression tree. That is, use F0 = 4.16 as the first prediction for all Y’s; construct a one-level regression tree h1 to predict the residuals. Then, compute the predicted values for all Y’s using the **learning rate of 0.5**



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **X** | **Y** | **y-F0** | **Predicted Residuals** | **Predicted y** |
| **1** | **1.4** | **-2.76** | **-2.56** | **2.88** |
| **2** | **1.8** | **-2.36** | **-2.56** | **2.88** |
| **6** | **5.5** | **1.34** | **1.707** | **5.014** |
| **7** | **5.8** | **1.64** | **1.707** | **5.014** |
| **8** | **6.3** | **2.14** | **1.707** | **5.014** |

1.4-F0=1.4-4.16=-2.76

1.8-F0=1.8-4.16=-2.36

5.5-F0=5.5-4.16=1.34

5.8-F0=5.8-4.16=1.64

6.3-F0=6.3-4.16=2.14

Y1=F0+0.5\*h(x1) =4.16 +0.5\*(-2.56)=2.88