**HW2 – ML**

**Assignment Authors and ID’s :**

Abed Shogan, 212112106

Noam Shani,

Q.1)

A screenshot of a computer

Description automatically generated

1. We will calculate the IG ( Information Gain) associated with the Position attribute, by using the formula given in practice number 5:

Total Information Gain From Splitting The Decision Tree

)

Total Entropy before split:

Calculate Probability of Each position:

Calculate Entropy for each Position:

)

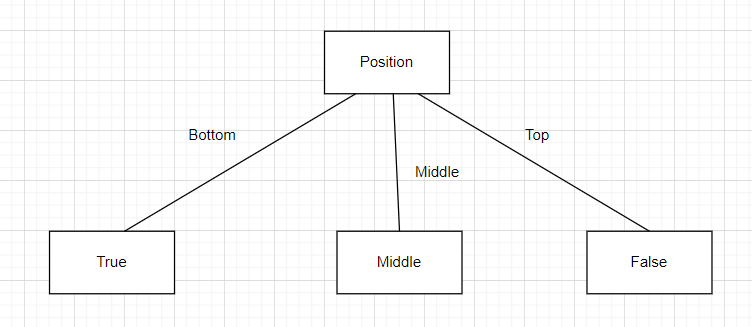
=

The Information Gain associated with splitting the Tree based on Position is 0.6.

b. Building the Full Decision Tree

We will use the Gini Impurity (GI) Score for each Feature, to select which nodes will be set in what order.

Step 1) Place Position as Root and, Place its Values outputs as Leaves.





We can see that Bottom and Top are pure Leaves, Bottom got True (Clicked) in all instances and Top got False (Not Clicked) in all its instances.

Middle is impure, therefore we will calculate the GI (Gini Impurity) for each other Feature and select the best one (The Minimum Value assigned to feature).

We will only examine the rows which have the attribute “Middle” , calculate the GI based on the new sub Table.

|  |  |  |  |
| --- | --- | --- | --- |
| Clicked | Size | Position | Sound |
| False | Small | Middle | Yes |
| False | Small | Middle | Yes |
| True | Small | Middle | No |
| True | Small | Middle | No |

Gini Impurity for Size:

) = 0/4

) = 4/4

) = 2/4

) = 2/4

) = 0/4

) = 0/4

Gini Impurity for Sound:

) = 2/4

) = 2/4

) = 2/2

) = 0/2

) = 0/2

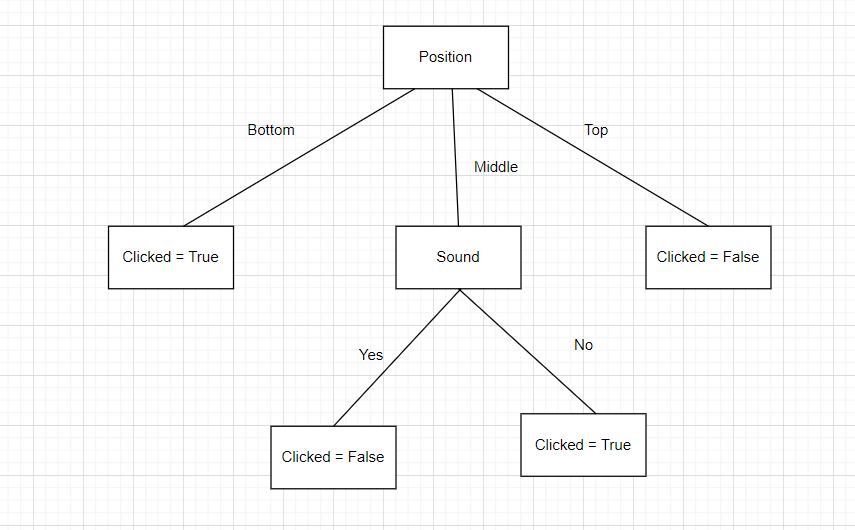
) = 2/2

We can see that

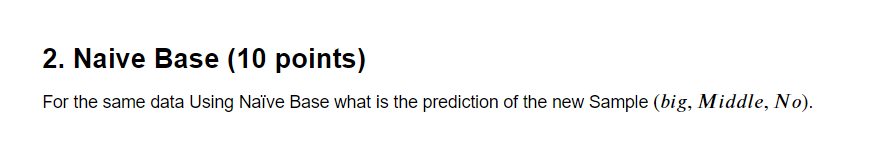
We can see the GI of size is 0.5 which is the worst possible value, and Sound has a value of 0 which is the best.

So we will choose the Sound as our next feature.

Our Final Decision Tree:



**Question 2)**



We will use the formula for Naïve Bayes to calculate whether our new Sample, which we will assign to the Variable X will be evaluated as Clicked=True or Clicked=False.

So, we will calculate for

We will assign the Variable X to our new sample Data:

,

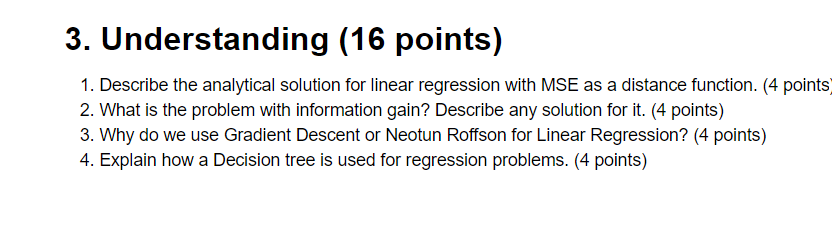
For Clicked = True

**For Clicked = False**

The denominator is the same so we can ignore it.

So, Naïve Bayes predicts that the new Sample (X = [Big,Middle,No]) will be classified as Clicked = True (Meaning the Ad was clicked).

**Question 3)**



**3.1)**

**TODO**

**3.2)** The problem with Information Gain is that attributes with a large number of values.

Therefore Information Gain is biased towards choosing attributes with a large number of values which may result is Over-Fitting(selecting an attribute which is non-optimal for prediction).

An Example would be a credit card number.

We will not include it in the decision tree.

One solution would be - **Feature Engineering**:

Instead of using raw attributes like a credit card number, we can derive more meaningful features from them. For instance, instead of using the entire credit card number, we can extract features like the issuer bank, credit card type, or even perform dimensionality reduction techniques to represent the data more effectively.

**3.3)**

Gradient Descent and Newton-Raphson are Numeric Methods which are used while performing Linear Regression.

because they efficiently find the optimal parameters of the linear regression model.

Both Gradient Descent and Newton-Raphson are versatile optimization algorithms that can be applied to various optimization problems, including linear regression. The choice between them often depends on factors such as the size of the dataset, the complexity of the cost function, and computational resources available. Gradient descent is commonly used for large datasets and when memory constraints are a concern, while Newton-Raphson may be preferred for smaller datasets or when faster convergence is desired.

**3.4)**

A Decision Tree is used for Regression Problems where Linear Regression will not work effectively, such as clustered data where you cannot fit a line that divides the data properly.

It is used for Numerical data types and not categorical ones.

First, we start by building the Tree for the Data set:

you split the dataset based on the feature that provides the best separation according to some criterion. Instead of Gini impurity or entropy, regression decision trees commonly use criteria like mean squared error (MSE) or mean absolute error (MAE). The goal is to minimize the variance of the target variable within each split.

Each leaf node in the decision tree represents a region of the feature space. In regression decision trees, the prediction for a new data point falling into a particular leaf node is typically the mean (average) of the target variable values of the training instances within that node.

The decision tree is constructed recursively, with splits made based on features and thresholds that minimize the chosen error metric (e.g., MSE or MAE) within each node. This process continues until a stopping criterion is met (e.g., a maximum depth is reached, a minimum number of samples per leaf node is reached).

To predict the target variable for a new data point, you traverse the decision tree based on the values of its features until you reach a leaf node. The prediction for that data point is then the mean of the target variable values of the training instances within that leaf node.

Regression trees suffer from a few problems, Notably Over-Fitting.

Some solutions include Tree Pruning, and setting a Max Depth Threshold for the tree.

We can measure how good a Decision Tree is, by Regression Metrics such as MSE and MAE or R-Squared.