DWDM

Assignment-4

18121a0116001
Vishwas Acharya

Page No. :
Date : / /

1) What is classification? What are the methods of it?

Ans - It is Data analysis task, i.e. the process of finding a model that describes and distinguishes data classes and concepts.

- Classification is the problem of identifying to which of a set of categories (subpopulations), a new observation belongs to, on the basis of a training set of data containing observations and whose categories membership is known.

- **Methods**

↳ Decision Trees
↳ Logistic Regression
↳ Naive Bayes Classification
↳ k-nearest neighbors
↳ Support Vector machines

⇒ Hence, this are the methods of classification in data mining.

Page No.:
Date: / /

2) Explain decision tree with example.

Ans - Decision tree algorithm falls under the category of supervised learning.

- They can be used to solve both regression and classification problems.

- Decision tree uses the tree representation to solve the problem in which each leaf node corresponds to a class label and attributes are represented on the internal node of tree.

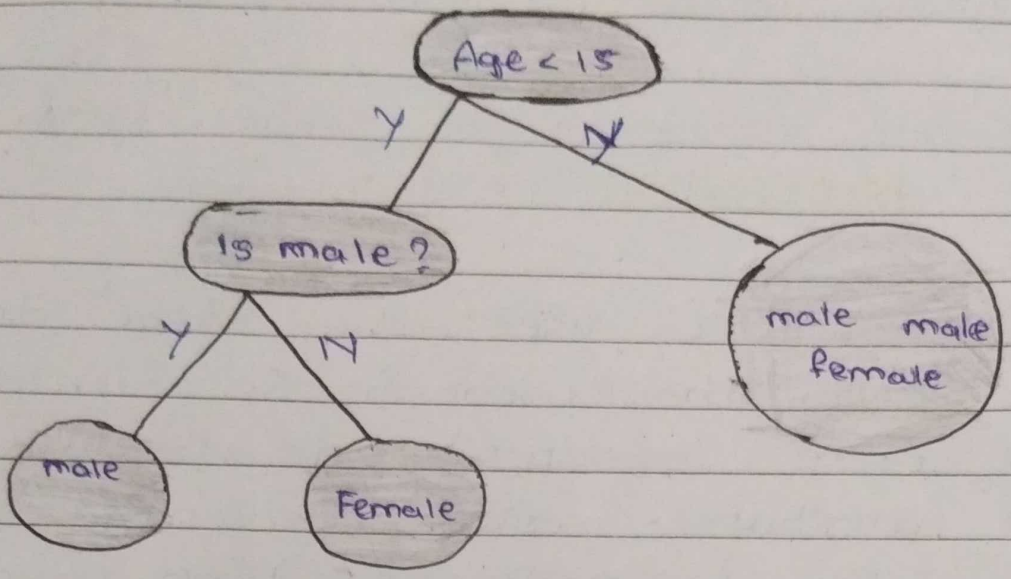
- We can represent any boolean function on discrete attributes using the decision tree.

- In decision tree the major challenge is to identification of the attribute for the root node in each level. This process is known as attribute selection.

- We have two popular attribute selection measures

1. Information Gain
2. Gini Index

- Example:



3) Explain 2 steps process for classification

Ans - Before starting any project, we need to check it's feasibility.
- In this case, a classifier is required to predict class labels such as 'safe' and 'Risky' for adopting the project and to further approve it.
- It is two-step process such as

1. Learning Step (Training Phase): Construction of classification model different algorithms are used to build a classifier by making the model learn using the training set available.
- The model has to be trained for the prediction of accurate results

2) **Classification Step:** Model used to predict class labels and testing the constructed model on test data and hence estimate the accuracy of the classification rules.

4) What is prediction? Explain logistics regression.

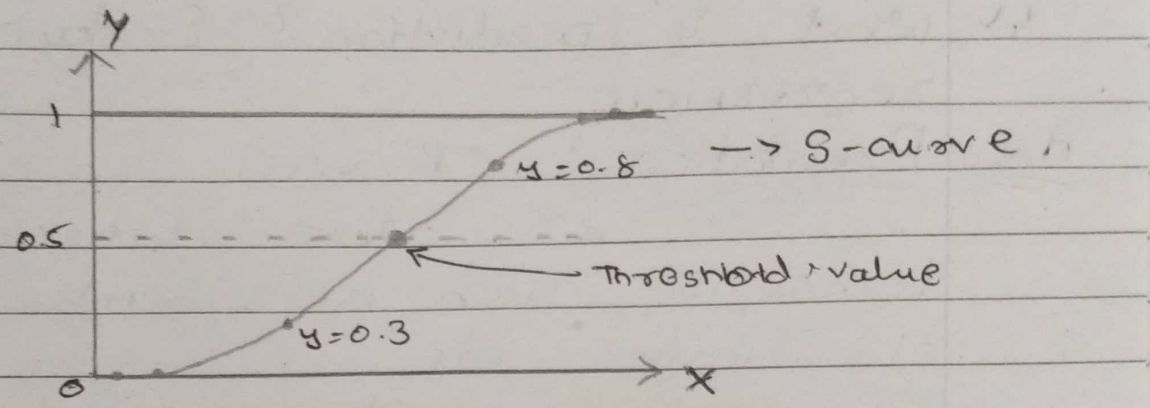
Ans • **Prediction** :- It used a combination of other data mining techniques such as trends, clustering, classification etc.

- It analyzes past events or instances in the right sequence to predict a future event.
- It is data mining techniques that discover relationship between independent variables & relationship between dependent variables.

• **Logistics regression** :- It is supervised learning technique.

- It is used for predicting the categorial dependent variable using a given set of independent variables
- It predicts the output of a categrial dependent variable.
- It can Therefore the outcome must be a categoriesal or discrete value.
- It can be either yes or no, 0 or 1, true or False, etc. but instead

- of giving the exact value as 0 & 1, it gives the probabilistics values which lie between 0 and 1
- It is used for solving the classification problems



- Logistic Regression Equation :-

$$P = \frac{1}{1 + e^{-(b_0 + b_1 x)}}$$

- **Assumptions** :- The dependent variable must be categorical in nature.
- The independent variable should not have multi-collinearity.

- **Advantages** :- It is easier to implement, interpret & very efficient to train.
- It can easily extend to multiple classes
- It is very fast at classifying unknown records
- It can interpret model coefficients as

indicates of feature importance.

- **Disadvantages:-** If the numbers of observations is lesses than the number of features, logistic regression should not be used, otherwise it may lead to overfitting.
  - It constructs linear boundaries
  - Non-linear problems can't be solve with logistic regression because it has a linear decision surface.