Classification##

Basic concepts:-

Data Mining: Data mining in general terms means mining or digging deep into data that is in different forms to gain patterns, and to gain knowledge on that pattern. In the process of data mining, large data sets are first sorted, then patterns are identified and relationships are established to perform data analysis and solve problems.

Classification: It is a 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.

Example: Before starting any project, we need to check its 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 a 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.

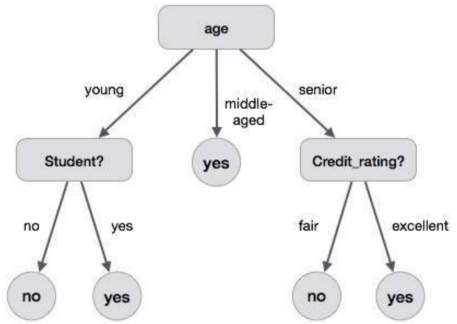
Classifiers can be categorized into two major types:

Discriminative: It is a very basic classifier and determines just one class for each row of data. It tries to model just by depending on the observed data, depends heavily on the quality of data rather than on distributions.
Example: Logistic Regression
Generative: It models the distribution of individual classes and tries to learn the model that generates the data behind the scenes by estimating assumptions and distributions of the model. Used to predict the unseen data.
Example: Naive Bayes Classifier
Detecting Spam emails by looking at the previous data. Suppose 100 emails and that too divided in 1:4 i.e. Class A: 25%(Spam emails) and Class B: 75%(Non-Spam emails). Now if a user wants to check that if an email contains the word cheap, then that may be termed as Spam.

Decision Tree Induction :-

A decision tree is a structure that includes a root node, branches, and leaf nodes. Each internal node denotes a test on an attribute, each branch denotes the outcome of a test, and each leaf node holds a class label. The topmost node in the tree is the root node.

The following decision tree is for the concept buy_computer that indicates whether a customer at a company is likely to buy a computer or not. Each internal node represents a test on an attribute. Each leaf node represents a class.

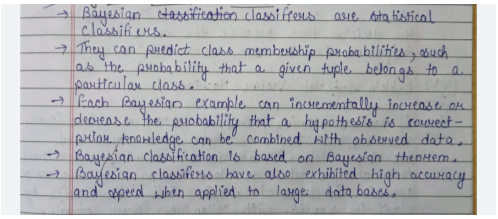


The benefits of having a decision tree are as follows –

- It does not require any domain knowledge.
- It is easy to comprehend.
- The learning and classification steps of a decision tree are simple and fast.

Bayesian Classification methods:-

Bayesian classification is based on Bayes' Theorem.



Baye's Theorem

Bayes' Theorem is named after Thomas Bayes. There are two types of probabilities –

- Posterior Probability [P(H/X)]
- Prior Probability [P(H)]

where X is data tuple and H is some hypothesis.

According to Bayes' Theorem,

$$P(H/X)= P(X/H)P(H) / P(X)$$

Bayesian Belief Network

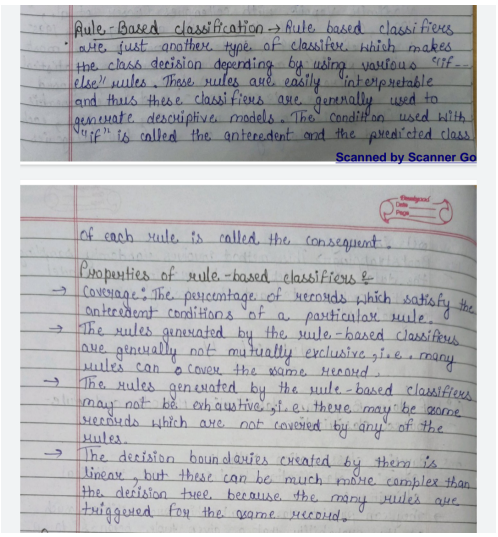
Bayesian Belief Networks specify joint conditional probability distributions. They are also known as Belief Networks, Bayesian Networks, or Probabilistic Networks.

- A Belief Network allows class conditional independencies to be defined between subsets of variables.
- It provides a graphical model of causal relationship on which learning can be performed.
- We can use a trained Bayesian Network for classification.

There are two components that define a Bayesian Belief Network –

- Directed acyclic graph
- A set of conditional probability tables

Rule - Based Classification:-



IF-THEN Rules

Rule-based classifier makes use of a set of IF-THEN rules for classification. We can express a rule in the following from −

IF condition THEN conclusion

Let us consider a rule R1,

```
R1: IF age = youth AND student = yes
   THEN buy_computer = yes
```

Points to remember −

- The IF part of the rule is called **rule antecedent** or **precondition**.
- The THEN part of the rule is called **rule consequent**.
- The antecedent part the condition consist of one or more attribute tests and these tests are logically ANDed.
- The consequent part consists of class prediction.

Note − We can also write rule R1 as follows −

```
R1: (age = youth) ^ (student = yes))(buy
```

If the condition holds true for a given tuple, then the antecedent is satisfied.

Rule Extraction

Here we will learn how to build a rule-based classifier by extracting IF-THEN rules from a decision tree.

Points to remember −

To extract a rule from a decision tree −

- One rule is created for each path from the root to the leaf node.
- To form a rule antecedent, each splitting criterion is logically ANDed.
- The leaf node holds the class prediction, forming the rule consequent.

Model Evaluation and selection:-

Model Selection is the process of choosing between the different learning algorithms for modelling our data, for solving a classification problem the choices could be made between Logistic Regression, SVM, Tree-based algorithms etc. And for a regression problem decisions also need to be made for the degree of linear regression algorithms.

Model Evaluation aims to check the generalization ability of our model, i.e ability of our model to perform well on an unseen dataset.

Model evaluation is the process of checking the model performance to see how much our model is able to explain the data whereas model selection is the process of seeing the level of flexibility we need for describing the data.

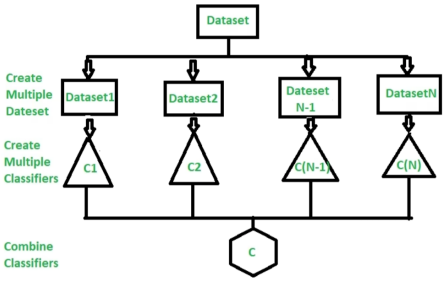
Techniques to improve classification accuracy:-

 # Ensemble methods :-

Ensemble Classifier | Data Mining

Ensemble learning helps improve machine learning results by combining several models. This approach allows the production of better predictive performance compared to a single model. Basic idea is to learn a set of classifiers (experts) and to allow them to vote.

> **Advantage :** Improvement in predictive accuracy.
> **Disadvantage :** It is difficult to understand an ensemble of classifiers.



 # Handling Different kinds of cases in classification:-

Classification is a two-step process involving,

Learning Step: It is a step where the Classification model is to be constructed. In this phase, training data are analyzed by a classification Algorithm.

Classification Step: it's a step where the model is employed to predict class labels for given data. In this phase, test data are wont to estimate the accuracy of classification rules.

 # Classification of Neural Networks :-

Q.8 Explain classification by Neural networks.

Ans Neural networks are the most efficient way (yes, you read it right) to solve real-world problems in Artificial Intelligence. Currently, it is also one of the much extensively researched areas in computer science that a new form of neural network would have been developed while you are reading this article. There are hundreds of neural networks to solve problems specific to different domains. Here we are going to walk you through different types of basic neural networks in the order of increasing complexity.

Different types of Basics in classification of neural networks –

1. Shallow Neural Networks (collaborative filtering) – Neural networks are made of groups of perceptron to simulate the neural structure of

the human brain. Shallow neural networks have a single hidden layer of the perceptron.

2. Multilayer Perceptron (Deep Neural networks) – Neural Networks with more than one hidden layer is called Deep Neural Networks. Spoiler Alert! All following neural networks are a form of deep neural network tweaked / improved to tackle domain-specific problems.

3. Convolutional Neural Network (CNN) – CNN's are the most mature form of deep neural networks to produce the most accurate i.e. better than human results in computer vision.

4. Recurrent Neural Network (RNN) – RNNs are the most recent form of deep neural networks for solving problems in NLP.

5. Long short Term Memory (LSTM) – LSTMs are designed especially to address the vanishing gradients problem with the RNN.

6. Attention-based Networks – Attention models are slowly taking over even the new RNNs in practice.

7. Generative Adversarial Network (GAN) – Although deep learning models provide state of the art results, they can be fooled by far more intelligent human counterparts by adding noise to the real-world data. GANs are the latest development in deep learning to tackle such scenarios.

Support Vector Machines :-

Q.2 Explain Support Vector Machines.

Ans In machine learning, support vector machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data for classification and regression analysis. Developed at AT & T Bell Laboratories by Vladimir Vapnik with colleagues (Boser et al., 1992, Guyon et al., 1993, Cortes and Vapnik, 1995, Vapnik et al., 1997) SVMs are one of the most robust prediction methods, being based on statistical learning frameworks or VC theory proposed by Vapnik (1982, 1995) and Chervonenkis (1974). Given a set of training examples, each marked as belonging to one of two categories,

an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier (although methods such as Platt scaling exist to use SVM in a probabilistic classification setting). SVM maps training examples to point in space so as to maximise the width of the gap between the two categories. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall.

 # Pattern Based Classification :-

Pattern-based classification is an approach of classification that makes use of "patterns" for deciding the label of each instance. It is related to rule-based classification, as it relies on the extraction of characteristics from a database to classify unseen instances [15,16,24]. In pattern-based classification, instead of extracting rules, a series of features that apply to various instances in a class are obtained.

The use of patterns in predictive models is a topic that has received a lot of attention in recent years. Pattern mining can help to obtain models for structured domains, such as graphs and sequences, and has been proposed as a means to obtain more accurate and more interpretable models. Despite the large amount of publications devoted to this topic, we believe however that an overview of what has been accomplished in this area is missing. This paper presents our perspective on this evolving area. We identify the principles of pattern mining that are important when mining patterns for models and provide an overview of pattern-based classification methods. We categorize these methods along the following dimensions: (1) whether they post-process a pre-computed set of patterns or iteratively execute pattern mining algorithms; (2) whether they select patterns model-independently or whether the pattern selection is guided by a model. We summarize the results that have been obtained for each of these methods.

 # Lazy Learners :-