Unit-2

Q.1 Explain the prediction methods and models for business intelligence.
Q.2 **Write a short note on Neural network.**
The process of building a prediction model usually consists of a few steps:

Data preparation.
To avoid the situation of "garbage in, garbage out," the relevant data must be "prepared." This step includes data transformation, normalization, creation of derived attributes, variable selection, elimination of noisy data, supplying missing values, and data cleaning. This stage is often augmented by preliminary data analysis to identify the most relevant variables and to determine the complexity of the underlying problem. The data preparation step can be the most laborious, and many people believe that it constitutes 80% of any data mining effort.

Model building.
This step includes a complete analysis of the data (i. e., the data mining stage), the selection of the best prediction method on the basis of (a) explaining the variability in question, and (b) producing consistent results, and the development of one or more prediction models.

Deployment and evaluation
This step includes implementing the best prediction model, and applying it to new data to generate predictions. However, because new data arrive on a continuous basis, it is essential to measure the prediction model's performance and tune it accordingly.

Different Prediction Methods:
Prediction is the process of choosing the best possible outcomes based on historical data and requirement of the prediction.
We can group these different prediction methods into a few broad categories:
Mathematical (e. g., linear regression, statistical methods).
Distance (e. g., instance-based learning, clustering).
Logic (e. g., decision tables, decision trees, classification rules).
Modern heuristic (e. g., neural networks, evolutionary algorithms, fuzzy logic).
• local optimization technique
• stochastic hill climber

Mathematical Methods

1) Regression

- The expected output exhibits some explanatory relationship with some other variables.
- For example, someone's (predicted) salary might be a function of education, experience, industry, and location.
- In such cases, an explanatory method would be used to find the relationship between these variables and make a prediction.

a) Linear Regression

- Probably the most popular explanatory method is linear regression.
- If the predicted outcome is numeric and all the variables in the prediction model are numeric, then linear regression is the classic choice.
- In this method, we build a linear expression that uses the values of different variables to produce a predicted value for a "new" variable (i.e., a variable not used in the model).
- Saleprice = a+(b*mileage)+(c*year)+(d*color)….
- Challenge is to find value of a, b, c, d that gives model best performance
- straight-line relationship – Form: y=mx+b

b) Nonlinear Regression

- Non-linear implies curved relationships logarithmic relationships

2) Time Series

- The goal of time series models, on the other hand, is not to discover or explain the relationships between variables; their goal is purely one of prediction.
- A time series is just collection of past values of the variable being predicted. Also known as naïve methods. Goal is to isolate patterns in past data.
- Neural networks are a good example of this.
- Neural network may not understand the connection between the weights yet provide quite accurate prediction.
- Data collected over time (generally equi distant)
- Price of stock everyday,
- heart beat recorded every minute
- Plotted data will have a pattern

Time Series Problem

• Given a series of historical visits in a previous time section, and a context of the latest visit location with the time of the next visit, the location prediction problem can be described as finding the probability

• p(vi = l, ti = t, vi-1 = lk), (1)

• where vi = l indicates the i-th visit at location l,

• ti = t indicates the i-th visit happens at time t,

• vi -1 = lk indicates the (i – 1)th visit happened at location lk.

• Note that the variable t here is a periodic time indicating the time stamp of the visit, such as a specific hour (e.g., 23:00pm), a day of the week (e.g., Monday), a month (e.g., January) or even a year. The candidate location l with the highest probability would be the prediction of the i-th visit location

Time Series Method

• Many statistical time series models have been proposed during the last few decades, including exponential smoothing model, autoregressive/integrated /moving average model, transfer function models, state space model etc

Distance Method:

Another method for building prediction models is based on the concept of "distance between cases." Any two cases in a data set can be compared for similarity, and this similarity measure (called "distance") is assigned some value: the more similar the cases, the smaller the value. Using a distance measure within a data set would allow us to compare a new case with the most "similar" existing case. The outcome of the most similar case (e. g., the loan was repaid, the transaction was fraudulent) would be the prediction for the new case.

Type of Distance Method

Instance based learning (KNN method)

• The examples are stored verbatim, and a distance function is used to determine which members of the database are closest to a new example with a desirable prediction.

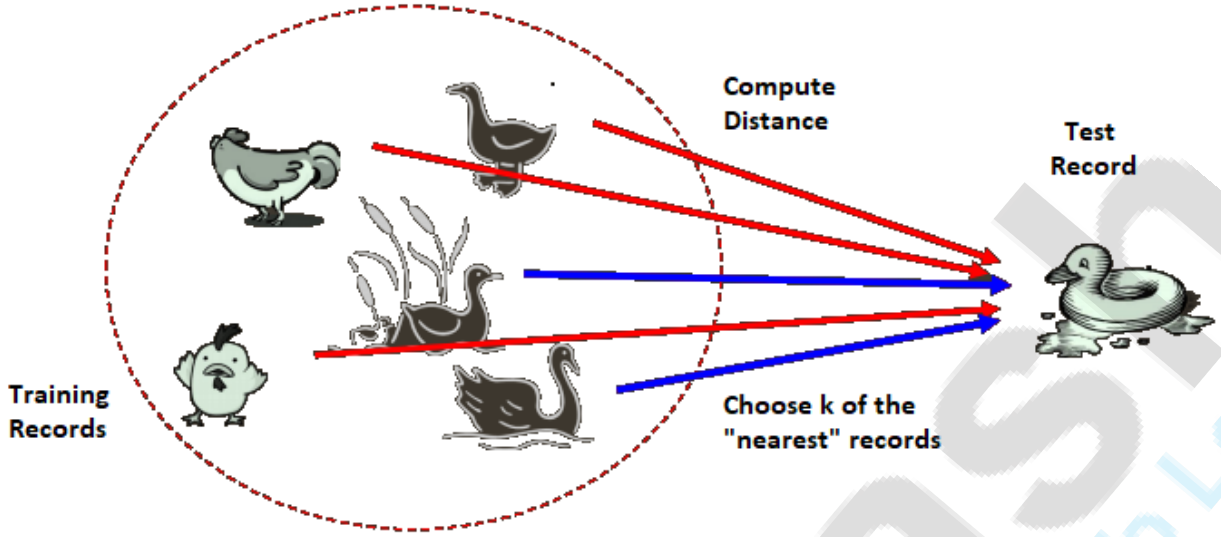
• The K-Nearest Neighbor (KNN) is the most representative method.

• They are good candidates to be improved through data reduction procedures.

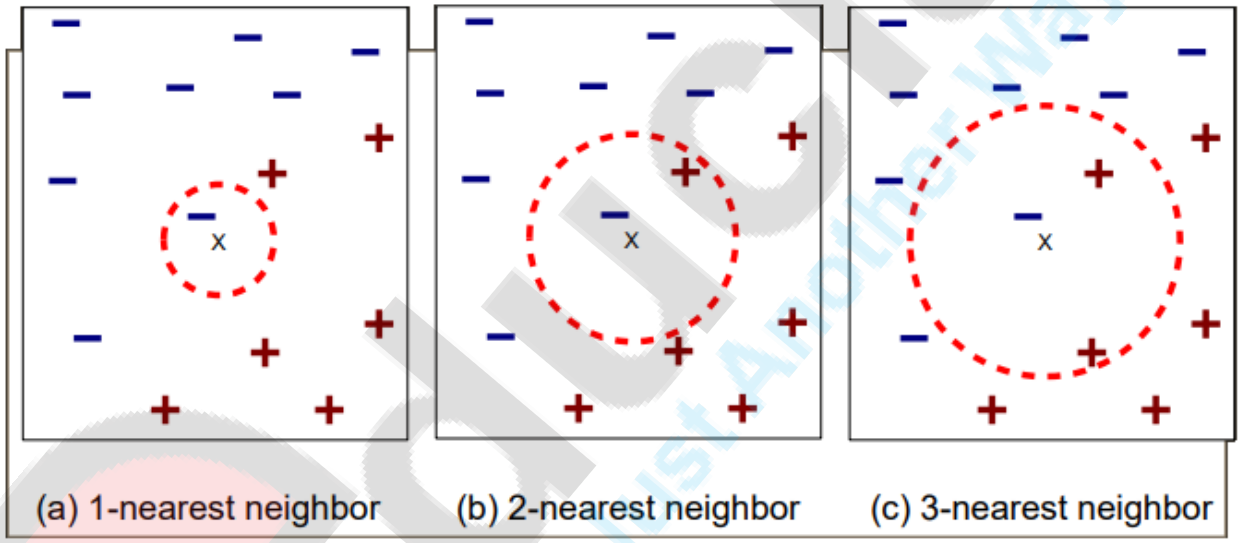
Nearest Neighbor Classifiers

• Basic idea:

• If it walks like a duck, quacks like a duck, then it's probably a duck



Definition of Nearest Neighbor



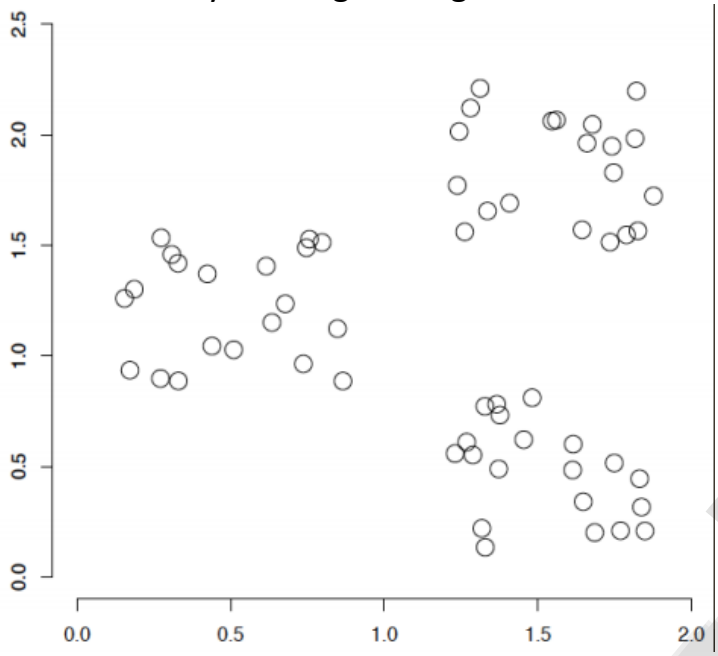
K-nearest neighbors of a record x are data points that have the k smallest distance to x

Clustering (K-mean method)

- The method of identifying similar groups of data in a data set is called clustering.
- Documents within a cluster should be similar.
- Documents from different clusters should be dissimilar.
- The commonest form of unsupervised learning
- Unsupervised learning = learning from raw data, as opposed to supervised data where a classification of examples is given
- A common and important task that finds many applications in IR and other places.

A data set with clear cluster structure

• How would you design an algorithm for finding the three clusters in this case?



Logic Method:

Decision Tables

• It is also known as lookup table is simplest logic based method for prediction

• There are many such table for estimation.

• E.g. used sold car in auction (make, model, year, body style, average, mile used, color, damage level, price).

• Anyone can locate the appropriate car and price from the given table.

• So, this type of table giving decision factor in purchasing used car in auction.

Decision Trees:

• Decision trees are powerful and popular tools for classification and prediction.

• Decision trees represent rules, which can be understood by humans and used in knowledge system such as database.

• CART, C4.5 and PUBLIC are good examples of this family.

classification rules:

A decision rule for a classification problem is often called a classification rule.

Association rules, which describe some regularity present in the data and can "predict" any variable (rather than just the class).

For example, an associate rule may state that

if Make = Porsche & Model = Carrera,

then Location in {Jacksonville, Tampa, Los Angeles, San Francisco, San Diego}

as Porsche Carreras are sold only at auction sites in Florida and California.

Rule-based systems, which consist of a collection of rules and an inference system,12 are quite popular, because each rule specifies a small piece of knowledge and people are good at handling small pieces of knowledge! Separate rules can be discovered from data mining activities or interviewing experts, and instead of specifying the overall model only the decision rules and inference system are needed. Note that the rule-based system will try to behave like an expert, performing some reasoning on the basis of the knowledge present in the system.

Modern Heuristic Methods:
A heuristic technique, often called simply a heuristic, is any approach to problem solving, learning, or discovery that employs a practical method not guaranteed to be optimal or perfect, but sufficient for the immediate goals.

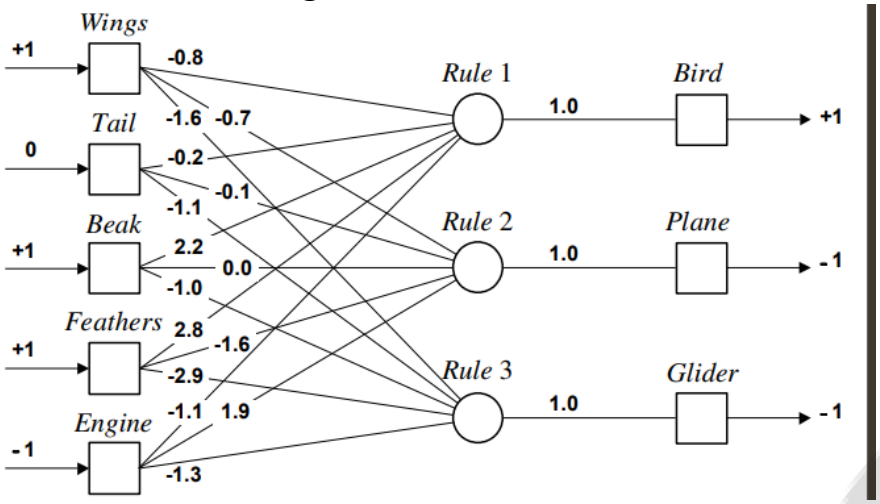
Neural Networks:
Neural expert systems:

- Expert systems rely on logical inferences and decision trees and focus on modelling human reasoning. Neural networks rely on parallel data processing and focus on modelling a human brain. ν Expert systems treat the brain as a black-box. Neural networks look at its structure and functions, particularly at its ability to learn. ν Knowledge in a rule-based expert system is represented by IF-THEN production rules. Knowledge in neural networks is stored as synaptic weights between neurons.
- In expert systems, knowledge can be divided into individual rules and the user can see and understand the piece of knowledge applied by the system. ν In neural networks, one cannot select a single synaptic weight as a discrete piece of knowledge. Here knowledge is embedded in the entire network; it cannot be broken into individual pieces, and any change of a synaptic weight may lead to unpredictable results. A neural network is, in fact, a black-box for its user.

Rule Extraction:

- Neurons in the network are connected by links, each of which has a numerical weight attached to it. ν The weights in a trained neural network determine the strength or importance of the associated neuron inputs.

The neural knowledge base:



If we set each input of the input layer to either +1 (true), -1 (false), or 0 (unknown), we can give a semantic interpretation for the activation of any output neuron.
For example, if the object has Wings (+1), Beak (+1) and Feathers (+1), but does not have Engine (-1), then we can conclude that this object is Bird (+1):

XRule1 = 1*(-0.8) + 0*(-0.2) + 1*2.2 + 1*2.8 + (-1)*(-1.1) = 5.3 > 0
YRule1 = YBird = +1

We can similarly conclude that this object is not Plane:
X Rule 2 =1·(-0.7) + 0·(-0.1) +1·0.0 +1·(-1.6) + (-1)·1.9 = -4.2 < 0
YRule 2 = YPlane = -1

and not Glider:

XRule 3 =1·(-0.6) +0·(- .1 1) +1·(-1.0) +1·(-2.9) +(-1)·(-1.3) = -4.2 < 0
YRule 3 = YGlider = -1

Evolutionary algorithms

Fuzzy logic

- A form of knowledge representation suitable for notions that cannot be defined precisely, but which depend upon their contexts.
- Fuzzy logic provides an alternative way to represent linguistic and subjective attributes of the real world in computing. Humans say things like "If it is sunny and warm today, I will drive fast"

Heuristic method

Optimization:

Is the process of getting the best result under a given circumstances
Eg. Work done should be max in min time
Optimization can be finding max or min of a function.

Three things always need to be specified before searching for any solution:
Representation of the solution : determines the search space and its size (travelling
sales man problem in Mumbai) Objective : task to be achieved (minimize total **distance of the route) Evaluation function**: allows you to compare the quality of different solution (distance is the evaluation function)

Global optimum:

The goal is to find a solution that is feasible and better than any other solution present in the entire search space. The solution that satisfies these two conditions is called a global optimum. Finding global optimum is difficult , a much easier approach is to search the neighborhood of the that solution.

Local optimization technique

Finding global optimum is difficult , a much easier approach is to search the neighborhood of the that solution. The problem of finding a solution with the highest quality measure score is similar to searching for a peak in a foggy mountain range.

Hill climbing:

it is a graph search algorithm where the current path is extended with a successor node which is closer to the solution than the end of the current path.
Hill climbing is used widely in artificial intelligence fields, for reaching a goal state from a starting node

stochastic hill climber

Proper choice is always dependent on the problem.
May accept a inferior solution as a new current solution
A new solution is accepted with some probability p.

The neighborhood of a current solution Vc consist from only one solution Vn.
The probability of acceptance of the solution Vn depends on:
Difference in merit between Vc and Vn

Parameter T

$$P=\frac{1}{1+e^{\frac{eval(Vc)-rval(Vn)}{T}}}$$

T remains constant during the execution of algorithm.

Role of Parameter T:

- Example:
  - eval(v_c)=107, eval(v_n)=120
  - maximization problem

$$p=\frac{1}{1+e^{\frac{-13}{T}}}$$

T	p
1	1.00
5	0.93
10	0.78
20	0.66
50	0.56
10^{10}	0.5...

$$p=\frac{1}{1+e^{\frac{-13}{T}}}$$

The greater the parameter T, the smaller the importance of the relative merit of the competing points v_c and v_n

T	p
1	1.00
5	0.93
10	0.78
20	0.66
50	0.56
10^{10}	0.5...

- If T is huge -> search becomes random
- T is very small -> stochastic hill-climber reverts into ordinary hill climber

Q.3 Explain the process of evaluation of models.

Evaluation of Models:

Although it is possible to use a variety of different prediction methods to build a variety of different prediction models, the key issue is which method should be applied to a particular problem. To answer this question, it is necessary to evaluate and compare different models. Because the comparisons have to be unbiased, the evaluation methodology should be fair and just. At first blush, this may seem easy. After all, after we complete and train a few models, we can test them on the data and measure the prediction error. The best model would then be selected for implementation.

Unfortunately, it is not that simple.

Following are the issue to select best model:

First of all, the amount of available data might not be that large.

Secondly, the performance of a prediction model on the training data might be very different from the performance of the same model on an independent set of data.

Thirdly, prediction models that provide different outcomes require different techniques for error measurement.

Finally, we have to take into account the cost of a potential error.

Once a prediction model is created on the basis of the training data set, it can be fairly evaluated for performance on the test data set.

Data set is split

Training set : to build the model

Validation set : tuning the parameters of the model

Test data set : evaluate the performance of the model