PART-1

Learning, Types of Learning.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 1.1. Define the term learning. What are the components of a learning system ?

Answer

1. Learning refers to the change in a subject's behaviour to a given situation brought by repeated experiences in that situation, provided that the behaviour changes cannot be explained on the basis of native response tendencies, matriculation or temporary states of the subject.
2. Learning agent can be thought of as containing a performance element that decides what actions to take and a learning element that modifies the performance element so that it makes better decisions.
3. The design of a learning element is affected by three major issues :
   a. Components of the performance element.
   b. Feedback of components.
   c. Representation of the components.

The important components of learning are :

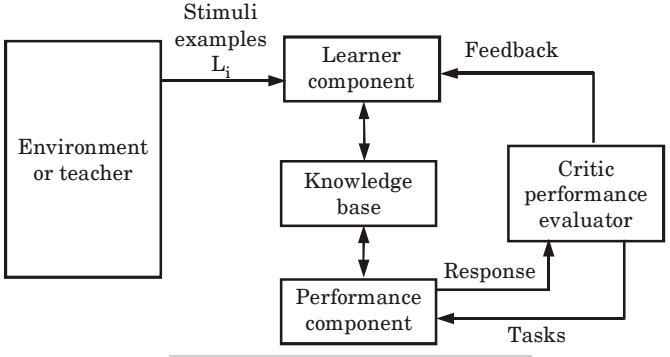


Fig. 1.1.1. General learning model.

1. Acquisition of new knowledge :

a. One component of learning is the acquisition of new knowledge.

b. Simple data acquisition is easy for computers, even though it is difficult for people.

2. Problem solving :

The other component of learning is the problem solving that is required for both to integrate into the system, new knowledge that is presented to it and to deduce new information when required facts are not been presented.

Que 1.2. Write down the performance measures for learning.

Answer

Following are the performance measures for learning are :

1. Generality :

a. The most important performance measure for learning methods is the generality or scope of the method.

b. Generality is a measure of the case with which the method can be adapted to different domains of application.

c. A completely general algorithm is one which is a fixed or self adjusting configuration that can learn or adapt in any environment or application domain.

2. Efficiency :

a. The efficiency of a method is a measure of the average time required to construct the target knowledge structures from some specified initial structures.

b. Since this measure is often difficult to determine and is meaningless without some standard comparison time, a relative efficiency index can be used instead.

3. Robustness :

a. Robustness is the ability of a learning system to function with unreliable feedback and with a variety of training examples, including noisy ones.

b. A robust system must be able to build tentative structures which are subjected to modification or withdrawal if later found to be inconsistent with statistically sound structures.

4. Efficacy :

a. The efficacy of a system is a measure of the overall power of the system. It is a combination of the factors generality, efficiency, and robustness.

5. Ease of implementation :

a. Ease of implementation relates to the complexity of the programs and data structures, and the resources required to develop the given learning system.

b. Lacking good complexity metrics, this measure will often be somewhat subjective.

Que 1.3. Discuss supervised and unsupervised learning.

Answer

Supervised learning :

1. Supervised learning is also known as associative learning, in which the network is trained by providing it with input and matching output patterns.
2. Supervised training requires the pairing of each input vector with a target vector representing the desired output.
3. The input vector together with the corresponding target vector is called training pair.

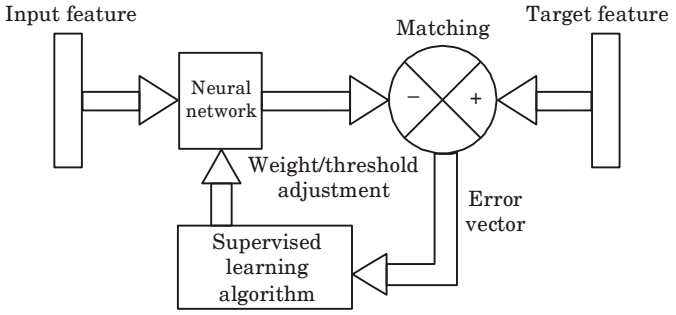


Fig. 1.3.1.

4. During the training session an input vector is applied to the network, and it results in an output vector.
5. This response is compared with the target response.
6. If the actual response differs from the target response, the network will generate an error signal.
7. This error signal is then used to calculate the adjustment that should be made in the synaptic weights so that the actual output matches the target output.
8. The error minimization in this kind of training requires a supervisor or teacher.
9. These input-output pairs can be provided by an external teacher, or by the system which contains the neural network (self-supervised).
10. Supervised training methods are used to perform non-linear mapping in pattern classification networks, pattern association networks and multilayer neural networks.

11. Supervised learning generates a global model that maps input objects to desired outputs.
12. In some cases, the map is implemented as a set of local models such as in case-based reasoning or the nearest neighbour algorithm.
13. In order to solve problem of supervised learning following steps are considered :
    i. Determine the type of training examples.
    ii. Gathering a training set.
    iii. Determine the input feature representation of the learned function.
    iv. Determine the structure of the learned function and corresponding learning algorithm.
    v. Complete the design.

Unsupervised learning :

1. It is a learning in which an output unit is trained to respond to clusters of pattern within the input.
2. Unsupervised training is employed in self-organizing neural networks.
3. This training does not require a teacher.
4. In this method of training, the input vectors of similar types are grouped without the use of training data to specify how a typical member of each group looks or to which group a member belongs.
5. During training the neural network receives input patterns and organizes these patterns into categories.
6. When new input pattern is applied, the neural network provides an output response indicating the class to which the input pattern belongs.
7. If a class cannot be found for the input pattern, a new class is generated.
8. Though unsupervised training does not require a teacher, it requires certain guidelines to form groups.
9. Grouping can be done based on color, shape and any other property of the object.
10. It is a method of machine learning where a model is fit to observations.
11. It is distinguished from supervised learning by the fact that there is no priori output.
12. In this, a data set of input objects is gathered.
13. It treats input objects as a set of random variables. It can be used in conjunction with Bayesian inference to produce conditional probabilities.

14. Unsupervised learning is useful for data compression and clustering.

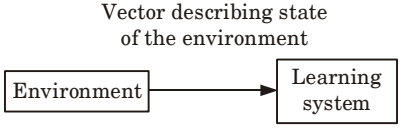


Fig. 1.3.2. Block diagram of unsupervised learning.

15. In unsupervised learning, system is supposed to discover statistically salient features of the input population.
16. Unlike the supervised learning paradigm, there is not a priori set of categories into which the patterns are to be classified; rather the system must develop its own representation of the input stimuli.

Que 1.4. Describe briefly reinforcement learning ?

Answer

1. Reinforcement learning is the study of how artificial system can learn to optimize their behaviour in the face of rewards and punishments.
2. Reinforcement learning algorithms have been developed that are closely related to methods of dynamic programming which is a general approach to optimal control.
3. Reinforcement learning phenomena have been observed in psychological studies of animal behaviour, and in neurobiological investigations of neuromodulation and addiction.

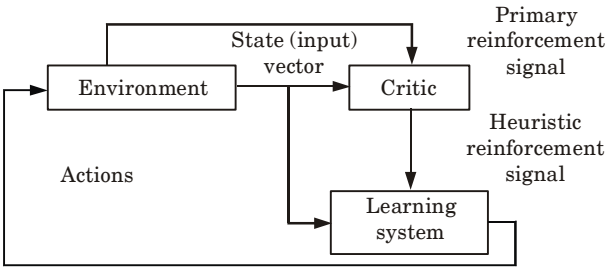


Fig. 1.4.1. Block diagram of reinforcement learning.

4. The task of reinforcement learning is to use observed rewards to learn an optimal policy for the environment.
5. An optimal policy is a policy that maximizes the expected total reward.
6. Without some feedback about what is good and what is bad, the agent will have no grounds for deciding which move to make.
7. The agents need to know that something good has happened when it wins and that something bad has happened when it loses.
8. This kind of feedback is called a reward or reinforcement.

9. Reinforcement learning is very valuable in the field of robotics, where the tasks to be performed are frequently complex enough to defy encoding as programs and no training data is available.
10. The robot's task consists of finding out, through trial and error (or success), which actions are good in a certain situation and which are not.
11. In many cases humans learn in a very similar way.
12. For example, when a child learns to walk, this usually happens without instruction, rather simply through reinforcement.
13. Successful attempts at working are rewarded by forward progress, and unsuccessful attempts are penalized by often painful falls.
14. Positive and negative reinforcement are also important factors in successful learning in school and in many sports.
15. In many complex domains, reinforcement learning is the only feasible way to train a program to perform at high levels.

Que 1.5. **What are the steps used to design a learning system ?**

Answer

Steps used to design a learning system are :

1. Specify the learning task.
2. Choose a suitable set of training data to serve as the training experience.
3. Divide the training data into groups or classes and label accordingly.
4. Determine the type of knowledge representation to be learned from the training experience.
5. Choose a learner classifier that can generate general hypotheses from the training data.
6. Apply the learner classifier to test data.
7. Compare the performance of the system with that of an expert human.

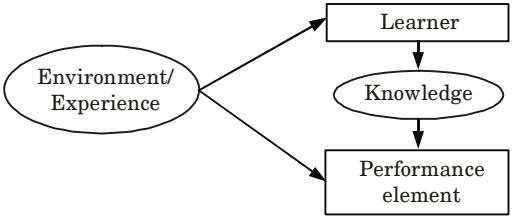


Fig. 1.5.1.

PART-2

Well Defined Learning Problems, Designing a Learning System.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 1.6. Write short note on well defined learning problem with example.

Answer

Well defined learning problem :

A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.

Three features in learning problems :

1. The class of tasks (T)
2. The measure of performance to be improved (P)
3. The source of experience (E)

For example :

1. **A checkers learning problem :**
   a. **Task (T) :** Playing checkers.
   b. **Performance measure (P) :** Percent of games won against opponents.
   c. **Training experience (E) :** Playing practice games against itself.
2. **A handwriting recognition learning problem :**
   a. **Task (T) :** Recognizing and classifying handwritten words within images.
   b. **Performance measure (P) :** Percent of words correctly classified.
   c. **Training experience (E) :** A database of handwritten words with given classifications.
3. **A robot driving learning problem :**
   a. **Task (T) :** Driving on public four-lane highways using vision sensors.
   b. **Performance measure (P) :** Average distance travelled before an error (as judged by human overseer).
   c. **Training experience (E) :** A sequence of images and steering commands recorded while observing a human driver.

Que 1.7. Describe well defined learning problems role's in machine learning.

Answer

Well defined learning problems role's in machine learning :

1. Learning to recognize spoken words :

a. Successful speech recognition systems employ machine learning in some form.

b. For example, the SPHINX system learns speaker-specific strategies for recognizing the primitive sounds (phonemes) and words from the observed speech signal.

c. Neural network learning methods and methods for learning hidden Markov models are effective for automatically customizing to individual speakers, vocabularies, microphone characteristics, background noise, etc.

2. Learning to drive an autonomous vehicle :

a. Machine learning methods have been used to train computer controlled vehicles to steer correctly when driving on a variety of road types.

b. For example, the ALYINN system has used its learned strategies to drive unassisted at 70 miles per hour for 90 miles on public highways among other cars.

3. Learning to classify new astronomical structures :

a. Machine learning methods have been applied to a variety of large databases to learn general regularities implicit in the data.

b. For example, decision tree learning algorithms have been used by NASA to learn how to classify celestial objects from the second Palomar Observatory Sky Survey.

c. This system is used to automatically classify all objects in the Sky Survey, which consists of three terabytes of image data.

4. Learning to play world class backgammon :

a. The most successful computer programs for playing games such as backgammon are based on machine learning algorithms.

b. For example, the world's top computer program for backgammon, TD-GAMMON learned its strategy by playing over one million practice games against itself.

PART-3

History of ML, Introduction of Machine Learning Approaches - (Artificial Neural Network, Clustering, Reinforcement Learning, Decision Tree Learning, Bayesian Network, Support Vector Machine, Genetic Algorithm).

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 1.8. Describe briefly the history of machine learning.

Answer

A. Early history of machine learning :

1. In 1943, neurophysiologist Warren McCulloch and mathematician Walter Pitts wrote a paper about neurons, and how they work. They created a model of neurons using an electrical circuit, and thus the neural network was created.
2. In 1952, Arthur Samuel created the first computer program which could learn as it ran.
3. Frank Rosenblatt designed the first artificial neural network in 1958, called Perceptron. The main goal of this was pattern and shape recognition.
4. In 1959, Bernard Widrow and Marcian Hoff created two models of neural network. The first was called ADELINE, and it could detect binary patterns. For example, in a stream of bits, it could predict what the next one would be. The second was called MADELINE, and it could eliminate echo on phone lines.

B. 1980s and 1990s :

1. In 1982, John Hopfield suggested creating a network which had bidirectional lines, similar to how neurons actually work.
2. Use of back propagation in neural networks came in 1986, when researchers from the Stanford psychology department decided to extend an algorithm created by Widrow and Hoff in 1962. This allowed multiple layers to be used in a neural network, creating what are known as 'slow learners', which will learn over a long period of time.
3. In 1997, the IBM computer Deep Blue, which was a chess-playing computer, beat the world chess champion.
4. In 1998, research at AT&T Bell Laboratories on digit recognition resulted in good accuracy in detecting handwritten postcodes from the US Postal Service.

C. 21st Century :

1. Since the start of the 21st century, many businesses have realised that machine learning will increase calculation potential. This is why they are researching more heavily in it, in order to stay ahead of the competition.

2. Some large projects include :
   i. GoogleBrain (2012)
   ii. AlexNet (2012)
   iii. DeepFace (2014)
   iv. DeepMind (2014)
   v. OpenAI (2015)
   vi. ResNet (2015)
   vii. U-net (2015)

Que 1.9. Explain briefly the term machine learning.

Answer

1. Machine learning is an application of Artificial Intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed.
2. Machine learning focuses on the development of computer programs that can access data.
3. The primary aim is to allow the computers to learn automatically without human intervention or assistance and adjust actions accordingly.
4. Machine learning enables analysis of massive quantities of data.
5. It generally delivers faster and more accurate results in order to identify profitable opportunities or dangerous risks.
6. Combining machine learning with AI and cognitive technologies can make it even more effective in processing large volumes of information.

Que 1.10. What are the applications of machine learning ?

Answer

Following are the applications of machine learning :

1. Image recognition :
   a. Image recognition is the process of identifying and detecting an object or a feature in a digital image or video.
   b. This is used in many applications like systems for factory automation, toll booth monitoring, and security surveillance.

2. Speech recognition :
   a. Speech Recognition (SR) is the translation of spoken words into text.
   b. It is also known as Automatic Speech Recognition (ASR), computer speech recognition, or Speech To Text (STT).

c. In speech recognition, a software application recognizes spoken words.

3. Medical diagnosis :

a. ML provides methods, techniques, and tools that can help in solving diagnostic and prognostic problems in a variety of medical domains.

b. It is being used for the analysis of the importance of clinical parameters and their combinations for prognosis.

4. Statistical arbitrage :

a. In finance, statistical arbitrage refers to automated trading strategies that are typical of a short-term and involve a large number of securities.

b. In such strategies, the user tries to implement a trading algorithm for a set of securities on the basis of quantities such as historical correlations and general economic variables.

5. Learning associations : Learning association is the process for discovering relations between variables in large data base.

6. Extraction :

a. Information Extraction (IE) is another application of machine learning.

b. It is the process of extracting structured information from unstructured data.

Que 1.11. What are the advantages and disadvantages of machine learning ?

Answer

Advantages of machine learning are :

1. Easily identifies trends and patterns :

a. Machine learning can review large volumes of data and discover specific trends and patterns that would not be apparent to humans.

b. For an e-commerce website like Flipkart, it serves to understand the browsing behaviours and purchase histories of its users to help cater to the right products, deals, and reminders relevant to them.

c. It uses the results to reveal relevant advertisements to them.

2. No human intervention needed (automation) : Machine learning does not require physical force *i.e.*, no human intervention is needed.

3. Continuous improvement :

a. ML algorithms gain experience, they keep improving in accuracy and efficiency.

b. As the amount of data keeps growing, algorithms learn to make accurate predictions faster.

4. Handling multi-dimensional and multi-variety data :

a. Machine learning algorithms are good at handling data that are multi-dimensional and multi-variety, and they can do this in dynamic or uncertain environments.

Disadvantages of machine learning are :

1. Data acquisition :

a. Machine learning requires massive data sets to train on, and these should be inclusive/unbiased, and of good quality.

2. Time and resources :

a. ML needs enough time to let the algorithms learn and develop enough to fulfill their purpose with a considerable amount of accuracy and relevancy.

b. It also needs massive resources to function.

3. Interpretation of results :

a. To accurately interpret results generated by the algorithms. We must carefully choose the algorithms for our purpose.

4. High error-susceptibility :

a. Machine learning is autonomous but highly susceptible to errors.

b. It takes time to recognize the source of the issue, and even longer to correct it.

Que 1.12. What are the advantages and disadvantages of different types of machine learning algorithm ?

Answer

Advantages of supervised machine learning algorithm :

1. Classes represent the features on the ground.
2. Training data is reusable unless features change.

Disadvantages of supervised machine learning algorithm :

1. Classes may not match spectral classes.
2. Varying consistency in classes.
3. Cost and time are involved in selecting training data.

Advantages of unsupervised machine learning algorithm :

1. No previous knowledge of the image area is required.
2. The opportunity for human error is minimised.
3. It produces unique spectral classes.
4. Relatively easy and fast to carry out.

Disadvantages of unsupervised machine learning algorithm :

1. The spectral classes do not necessarily represent the features on the ground.
2. It does not consider spatial relationships in the data.
3. It can take time to interpret the spectral classes.

Advantages of semi-supervised machine learning algorithm :

1. It is easy to understand.
2. It reduces the amount of annotated data used.
3. It is stable, fast convergent.
4. It is simple.
5. It has high efficiency.

Disadvantages of semi-supervised machine learning algorithm :

1. Iteration results are not stable.
2. It is not applicable to network level data.
3. It has low accuracy.

Advantages of reinforcement learning algorithm :

1. Reinforcement learning is used to solve complex problems that cannot be solved by conventional techniques.
2. This technique is preferred to achieve long-term results which are very difficult to achieve.
3. This learning model is very similar to the learning of human beings. Hence, it is close to achieving perfection.

Disadvantages of reinforcement learning algorithm :

1. Too much reinforcement learning can lead to an overload of states which can diminish the results.
2. Reinforcement learning is not preferable for solving simple problems.
3. Reinforcement learning needs a lot of data and a lot of computation.
4. The curse of dimensionality limits reinforcement learning for real physical systems.

Que 1.13. Write short note on Artificial Neural Network (ANN).

Answer

1. Artificial Neural Networks (ANN) or neural networks are computational algorithms that intended to simulate the behaviour of biological systems composed of neurons.

2. ANNs are computational models inspired by an animal's central nervous systems.
3. It is capable of machine learning as well as pattern recognition.
4. A neural network is an oriented graph. It consists of nodes which in the biological analogy represent neurons, connected by arcs.
5. It corresponds to dendrites and synapses. Each arc associated with a weight at each node.
6. A neural network is a machine learning algorithm based on the model of a human neuron. The human brain consists of millions of neurons.
7. It sends and process signals in the form of electrical and chemical signals.
8. These neurons are connected with a special structure known as synapses. Synapses allow neurons to pass signals.
9. An Artificial Neural Network is an information processing technique. It works like the way human brain processes information.
10. ANN includes a large number of connected processing units that work together to process information. They also generate meaningful results from it.

Que 1.14. **Write short note on clustering.**

Answer

1. Clustering is a division of data into groups of similar objects.
2. Each group or cluster consists of objects that are similar among themselves and dissimilar to objects of other groups as shown in Fig. 1.14.1.

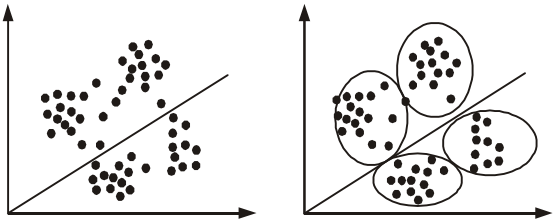


Fig. 1.14.1. Clusters.

3. A cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the object in the other cluster.
4. Clusters may be described as connected regions of a multidimensional space containing relatively high density points, separated from each other by a region containing a relatively low density points.
5. From the machine learning perspective, clustering can be viewed as unsupervised learning of concepts.
6. Clustering analyzes data objects without help of known class label.

7. In clustering, the class labels are not present in training data simply because they are not known to cluster the data objects.
8. Hence, it is the type of unsupervised learning.
9. For this reason, clustering is a form of learning by observation rather than learning by examples.
10. There are certain situations where clustering is useful. These include :
    a. The collection and classification of training data can be costly and time consuming. Therefore it is difficult to collect a training data set. A large number of training samples are not all labelled. Then it is useful to train a supervised classifier with a small portion of training data and then use clustering procedures to tune the classifier based on the large, unclassified dataset.
    b. For data mining, it can be useful to search for grouping among the data and then recognize the cluster.
    c. The properties of feature vectors can change over time. Then, supervised classification is not reasonable. Because the test feature vectors may have completely different properties.
    d. The clustering can be useful when it is required to search for good parametric families for the class conditional densities, in case of supervised classification.

Que 1.15. What are the applications of clustering ?

Answer

Following are the applications of clustering :

1. Data reduction :
    a. In many cases, the amount of available data is very large and its processing becomes complicated.
    b. Cluster analysis can be used to group the data into a number of clusters and then process each cluster as a single entity.
    c. In this way, data compression is achieved.

2. Hypothesis generation :
    a. In this case, cluster analysis is applied to a data set to infer hypothesis that concerns about the nature of the data.
    b. Clustering is used here to suggest hypothesis that must be verified using other data sets.

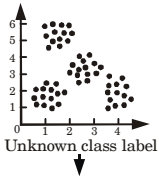
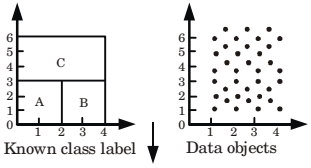
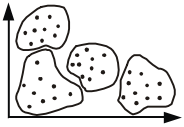
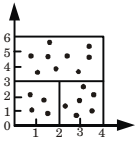
3. Hypothesis testing : In this context, cluster analysis is used for the verification of the validity of a specific hypothesis.

4. Prediction based on groups :
    a. In this case, cluster analysis is applied to the available data set and then the resulting clusters are characterized based on the characteristics of the patterns by which they are formed.

b. In this sequence, if an unknown pattern is given, we can determine the cluster to which it is more likely to belong and characterize it based on the characterization of the respective cluster.

Que 1.16. Differentiate between clustering and classification.

Answer

S.No.	Clustering	Classification
1.	Clustering analyzes data objects without known class label.	In classification, data are grouped by analyzing the data objects whose class label is known.
2.	There is no prior knowledge of the attributes of the data to form clusters.	There is some prior knowledge of the attributes of each classification.
3.	It is done by grouping only the input data because output is not predefined.	It is done by classifying output based on the values of the input data.
4.	The number of clusters is not known before clustering. These are identified after the completion of clustering. 	The number of classes is known before classification as there is predefined output based on input data. 
5.		
6.	It is considered as unsupervised learning because there is no prior knowledge of the class labels.	It is considered as the supervised learning because class labels are known before.

Que 1.17. What are the various clustering techniques ?

Answer

1. Clustering techniques are used for combining observed examples into clusters or groups which satisfy two following main criteria :
   a. Each group or cluster is homogeneous *i.e.*, examples belong to the same group are similar to each other.
   b. Each group or cluster should be different from other clusters *i.e.*, examples that belong to one cluster should be different from the examples of the other clusters.
2. Depending on the clustering techniques, clusters can be expressed in different ways :
   a. Identified clusters may be exclusive, so that any example belongs to only one cluster.
   b. They may be overlapping *i.e.*, an example may belong to several clusters.
   c. They may be probabilistic *i.e.*, an example belongs to each cluster with a certain probability.
   d. Clusters might have hierarchical structure.

Major classifications of clustering techniques are :

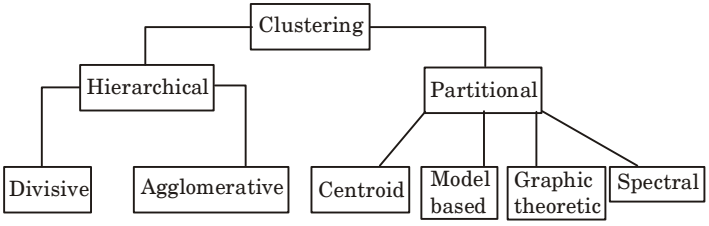


Fig. 1.17.1. Types of clustering.

   a. Once a criterion function has been selected, clustering becomes a well-defined problem in discrete optimization. We find those partitions of the set of samples that extremize the criterion function.
   c. The sample set is finite, there are only a finite number of possible partitions.
   d. The clustering problem can always be solved by exhaustive enumeration.

1. Hierarchical clustering :

   a. This method works by grouping data object into a tree of clusters.
   b. This method can be further classified depending on whether the hierarchical decomposition is formed in bottom up (merging) or top down (splitting) fashion.

Following are the two types of hierarchical clustering :

a. Agglomerative hierarchical clustering : This bottom up strategy starts by placing each object in its own cluster and then merges these atomic clusters into larger and larger clusters, until all of the objects are in a single cluster.

b. Divisive hierarchical clustering :

i. This top down strategy does the reverse of agglomerative strategy by starting with all objects in one cluster.

ii. It subdivides the cluster into smaller and smaller pieces until each object forms a cluster on its own.

2. Partitional clustering :

a. This method first creates an initial set of number of partitions where each partition represents a cluster.

b. The clusters are formed to optimize an objective partition criterion such as a dissimilarity function based on distance so that the objects within a cluster are similar whereas the objects of different clusters are dissimilar.

Following are the types of partitioning methods :

a. Centroid based clustering :

i. In this, it takes the input parameter and partitions a set of object into a number of clusters so that resulting intracluster similarity is high but the intercluster similarity is low.

ii. Cluster similarity is measured in terms of the mean value of the objects in the cluster, which can be viewed as the cluster's centroid or center of gravity.

b. Model-based clustering : This method hypothesizes a model for each of the cluster and finds the best fit of the data to that model.

Que 1.18. Describe reinforcement learning.

Answer

1. Reinforcement learning is the study of how animals and artificial systems can learn to optimize their behaviour in the face of rewards and punishments.
2. Reinforcement learning algorithms related to methods of dynamic programming which is a general approach to optimal control.
3. Reinforcement learning phenomena have been observed in psychological studies of animal behaviour, and in neurobiological investigations of neuromodulation and addiction.
4. The task of reinforcement learning is to use observed rewards to learn an optimal policy for the environment. An optimal policy is a policy that maximizes the expected total reward.

Que 1.19. Explain decision tree in detail.

Answer

1. A decision tree is a flowchart structure in which each internal node represents a test on a feature, each leaf node represents a class label and branches represent conjunctions of features that lead to those class labels.
2. The paths from root to leaf represent classification rules.
3. Fig 1.19.1, illustrate the basic flow of decision tree for decision making with labels (Rain(Yes), Rain(No)).

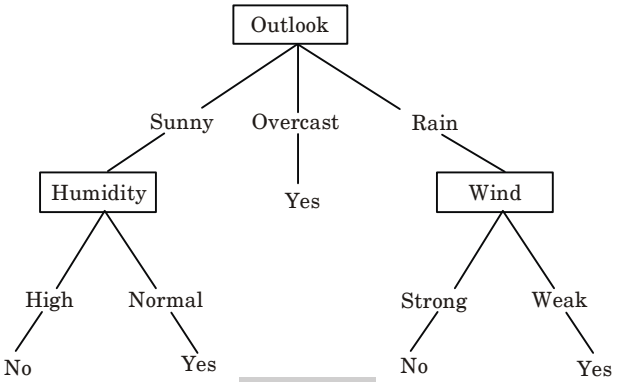


Fig. 1.19.1.

4. Decision tree is the predictive modelling approach used in statistics, data mining and machine learning.
5. Decision trees are constructed via an algorithmic approach that identifies the ways to split a data set based on different conditions.
6. Decision trees are a non-parametric supervised learning method used for both classification and regression tasks.
7. Classification trees are the tree models where the target variable can take a discrete set of values.
8. Regression trees are the decision trees where the target variable can take continuous set of values.

Que 1.20. What are the steps used for making decision tree ?

Answer

Steps used for making decision tree are :

1. Get list of rows (dataset) which are taken into consideration for making decision tree (recursively at each node).

2. Calculate uncertainty of our dataset or Gini impurity or how much our data is mixed up etc.
3. Generate list of all question which needs to be asked at that node.
4. Partition rows into True rows and False rows based on each question asked.
5. Calculate information gain based on Gini impurity and partition of data from previous step.
6. Update highest information gain based on each question asked.
7 Update question based on information gain (higher information gain).
8. Divide the node on question. Repeat again from step 1 until we get pure node (leaf nodes).

Que 1.21. What are the advantages and disadvantages of decision tree method ?

Answer

Advantages of decision tree method are :

1. Decision trees are able to generate understandable rules.
2. Decision trees perform classification without requiring computation.
3. Decision trees are able to handle both continuous and categorical variables.
4. Decision trees provide a clear indication for the fields that are important for prediction or classification.

Disadvantages of decision tree method are :

1. Decision trees are less appropriate for estimation tasks where the goal is to predict the value of a continuous attribute.
2. Decision trees are prone to errors in classification problems with many class and relatively small number of training examples.
3. Decision tree are computationally expensive to train. At each node, each candidate splitting field must be sorted before its best split can be found.
4. In decision tree algorithms, combinations of fields are used and a search must be made for optimal combining weights. Pruning algorithms can also be expensive since many candidate sub-trees must be formed and compared.

Que 1.22. Write short note on Bayesian belief networks.

Answer

1. Bayesian belief networks specify joint conditional probability distributions.
2. They are also known as belief networks, Bayesian networks, or probabilistic networks.

3. A Belief Network allows class conditional independencies to be defined between subsets of variables.
4. It provides a graphical model of causal relationship on which learning can be performed.
5. We can use a trained Bayesian network for classification.
6. There are two components that define a Bayesian belief network :

a. Directed acyclic graph :

i. Each node in a directed acyclic graph represents a random variable.

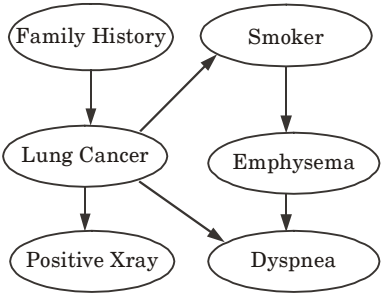
ii. These variable may be discrete or continuous valued.

iii. These variables may correspond to the actual attribute given in the data.

Directed acyclic graph representation : The following diagram shows a directed acyclic graph for six Boolean variables.

i. The arc in the diagram allows representation of causal knowledge.

ii. For example, lung cancer is influenced by a person's family history of lung cancer, as well as whether or not the person is a smoker.



iii. It is worth noting that the variable Positive X-ray is independent of whether the patient has a family history of lung cancer or that the patient is a smoker, given that we know the patient has lung cancer.

b. Conditional probability table :

The conditional probability table for the values of the variable LungCancer (LC) showing each possible combination of the values of its parent nodes, FamilyHistory (FH), and Smoker (S) is as follows :

	FH,S	FH,-S	-FH,S	-FH,S
LC	0.8	0.5	0.7	0.1
-LC	0.2	0.5	0.3	0.9

Que 1.23. Write a short note on support vector machine.

Answer

1. A Support Vector Machine (SVM) is machine learning algorithm that analyzes data for classification and regression analysis.
2. SVM is a supervised learning method that looks at data and sorts it into one of two categories.
3. An SVM outputs a map of the sorted data with the margins between the two as far apart as possible.
4. Applications of SVM :
   i. Text and hypertext classification
   ii. Image classification
   iii. Recognizing handwritten characters
   iv. Biological sciences, including protein classification

Que 1.24. Explain genetic algorithm with flow chart.

Answer

Genetic algorithm (GA) :

1. The genetic algorithm is a method for solving both constrained and unconstrained optimization problems that is based on natural selection.
2. The genetic algorithm repeatedly modifies a population of individual solutions.
3. At each step, the genetic algorithm selects individuals at random from the current population to be parents and uses them to produce the children for the next generation.
4. Over successive generations, the population evolves toward an optimal solution.

Flow chart : The genetic algorithm uses three main types of rules at each step to create the next generation from the current population :

a. Selection rule : Selection rules select the individuals, called parents, that contribute to the population at the next generation.

b. Crossover rule : Crossover rules combine two parents to form children for the next generation.

c. Mutation rule : Mutation rules apply random changes to individual parents to form children.

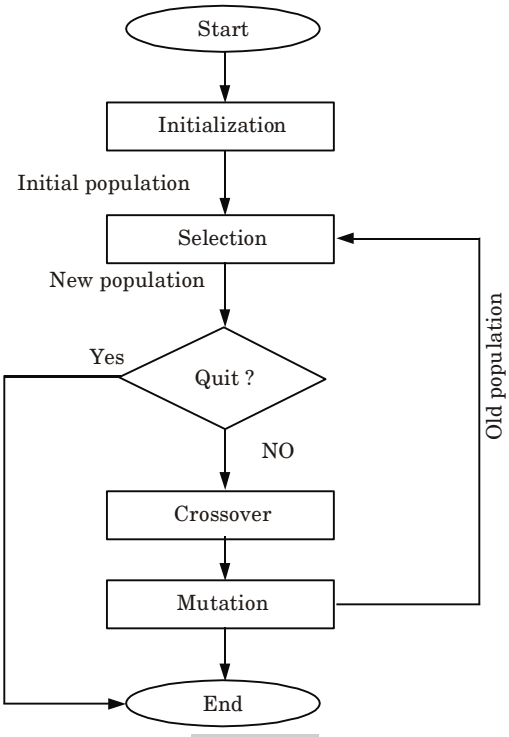


Fig. 1.24.1.

PART-4

Issues in Machine Learning and Data Science Vs. Machine Learning.

Questions-Answers

Long Answer Type and Medium Answer Type Questions

Que 1.25. **Briefly explain the issues related with machine learning.**