UNIT IV  PLANNING AND MACHINE LEARNING

PART - A

1. Why does uncertainty arise ?
   Agents almost never have access to the whole truth about their environment.
   Agents cannot find a categorical answer.
   Uncertainty can also arise because of incompleteness, incorrectness in agents understanding of properties of environment.

2. State the reason why first order logic fails to cope with the mind like medical diagnosis.
   Three reasons
   a. Laziness: it is hard to lift complete set of antecedents of consequence, needed to ensure an exceptionless rule.
   b. Theoretical Ignorance: medical science has no complete theory for the domain.
   Practical ignorance: even if we know all the rules, we may be uncertain about a particular item needed.

3. Define the term utility?
   The term utility is used in the sense of "the quality of being useful.", utility of a state is relative to the agents, whose preferences the utility function is supposed to represent.

4. What is the need for probability theory in uncertainty?
   Probability provides the way of summarizing the uncertainty that comes from our laziness and ignorance. Probability statements do not have quite the same kind of semantics known as evidences.

5. What is the need for utility theory in uncertainty?
   Utility theory says that every state has a degree of usefulness, or utility to in agent, and that the agent will prefer states with higher utility. The use utility theory to represent and reason with preferences.

6. What is called as principle of maximum expected utility?
   The basic idea is that an agent is rational if and only if it chooses the action that yields the highest expected utility, averaged over all the possible outcomes of the action. This is known as MEU.

7. What is called as Decision Theory?
   Decision Theory = Probability Theory + Utility Theory.

8. Define Prior Probability?
   p(a) for the Unconditional or Prior Probability Is That the Proposition A is True. It is important to remember that p(a) can only be used when there is no other information.

9. Define conditional probability?
   Once the agents has obtained some evidence concerning the previously unknown propositions making up the domain conditional or posterior probabilities with the notation p(A/B) is used. This is important that p(A/B) can only be used when all be is known.

10. Define probability distribution:
    Eg. P(weather) = (0.7,0.2,0.08,0.02). This type of notations simplifies many equations.

11. Define an atomic event?
    An atomic event is an assignment of particular values to all variables, in other words, the complete specifications of the state of domain.

12. Define joint probability distribution
    This completely specifies an agent's probability assignments to all propositions in the domain. The joint probability distribution p(x1,x2,------xn) assigns probabilities to all possible atomic events; where X1,X2------Xn 10 =variables.

13. Give the Baye's rule equation
    W.K.T P(A ^ B) = P(A/B) P(B) -------------------------- 1
P(A \land B) = P(B/A) P(A) \quad \text{-------------------------- 2} \\
\text{DIVIDING BY P(A)}; \\
\text{WE GET} \\
P(B/A) = P(A/B) P(B) \quad \text{-------------------- P(A)}

14. What is meant by belief network? 
A belief network is a graph in which the following holds 
\begin{itemize} 
\item A set of random variables 
\item A set of directive links or arrows connects pairs of nodes. 
\item The conditional probability table for each node 
\item The graph has no directed cycles. 
\end{itemize}

15. What are the ways in which one can understand the semantics of a belief network? 
There are two ways to see the network as a representation of the joint probability distribution to view it as an encoding of collection of conditional independence statements.

16. What is the basic task of a probabilistic inference? 
The basic task is to reason in terms of prior probabilities of conjunctions, but for the most part, we will use conditional probabilities as a vehicle for probabilistic inference.

17. What are called as Poly trees? 
The algorithm that works only on singly connected networks known as Poly trees. Here at most one undirected path between any two nodes is present.

18. Define casual support 
E+X is the casual support for X - the evidence variables "above" X that are connected to X through its parent.

19. Define evidential support 
E-X is the evidential support for X - the evidence variables "below" X that are connected to X through its children.

20. What is called as multiple connected graph? 
A multiple connected graph is one in which two nodes are connected by more than one path.

21. What is the purpose of learning? 
The idea behind learning is that percepts should be used not only for acting but also for improving the agent’s ability to act in the future.

22. What are issues in learning element? 
\begin{itemize} 
\item i. Component 
\item ii. Feedback 
\item iii. Representation 
\end{itemize}

23. What are the types of machine learning? 
\begin{itemize} 
\item i. Supervised Learning 
\item ii. Unsupervised Learning 
\item iii. Reinforcement Learning 
\end{itemize}


<table>
<thead>
<tr>
<th>Supervised Learning</th>
<th>Unsupervised Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>It involves learning a function from examples of its inputs And outputs</td>
<td>It involves learning patterns in the input when no specific output values are supplied.</td>
</tr>
<tr>
<td>Example: Applying Brake on the wet road, we can even skip on the road is a result.</td>
<td>Example: Day by day agent will learn about “Good traffic days” and “Bad traffic days” without any advice.</td>
</tr>
</tbody>
</table>
25. Define Reinforcement Learning.
This learning is rather than being told what to do by teacher, a reinforcement learning agent must learn from occasional rewards. Example
If taxi driver does not get a tip at the end of journey, it gives him a indication that his behavior is undesirable.

An algorithm for supervised learning is given as input the correct value of the unknown function for particular inputs and it must try to recover the unknown function.

27. Define Classification Learning.
Learning a discrete valued function is called called classification learning.

Learning a continuous valued function is called regression learning.

29. What is parity and majority function?
Parity Function: It Returns 1 if and only if an even number of inputs are 1.
Majority function: It Returns 1 if more than half of its inputs are 1.

30. What is training set?
The complete set of examples is called the training set.
Example
Restaurant problem
Goal predicate “will wait”

31. Define Information gain.
Information gain from the attribute test is the difference between the original information requirement and the new requirement.
Gain (A) = I(p/(p+n)), n/(p+n)) – Remainder(A)

32. What is test set?
Prediction is good if it turns out to be true, so can assess quality of hypotheses by Checking its predictions against the correct classification once we know it. We do this on a set of examples is known as Test Set.

33. What is over fitting?
Whenever there is a large set of possible hypotheses, one has to be careful not to use the resulting freedom to find meaningless “regularity” in the data. This problem is called over fitting.

34. What is the purpose of cross validation?
It reduces over fitting. It can be applied to any learning algorithm, not just decision tree learning. The basic idea is to estimate how well each hypotheses will predict unseen data.

35. Mention the exercises which broaden the applications of decision trees.
i. Missing data
ii. Multivalued attributes
iii. Continuous and integer valued input attributes
iv. Continuous valued output attributes.

36. What is ensemble learning?
The idea of this learning is to select a whole collection or ensemble, of hypotheses from the hypotheses space and combine their predictions.

An learning algorithm that return hypotheses that are approximately correct is called PAC learning algorithm.

38. Define Decision list.
It is a logical expression of a restricted form, It consists of a series of tests, each of which conjunction of literals. If test succeeds, value is returned. If test fails, processing continues with the next test in the list.

39. What is the purpose of current best hypotheses search?
This search is to maintain a single hypothesis and to adjust it as new examples arrive in order to maintain consistency.
40. Differentiate generalization and specialization.
The extension of the hypotheses must be increased to include it. This is called generalization.
The extension of the hypotheses must be decreased to exclude the example. This is called specialization.

41. Define Boundary set.
Each boundary will not be a point but rather a set of hypotheses called a Boundary set.

42. What are the two boundary sets?
i. G Set : a most general boundary set.
ii. S Set : a most specific boundary set.
43. Show the relationship of an entailment constraint.
Hypothesis ^ Descriptions |= classifications

44. Define EBL.
Explanation based learning, from the prior knowledge (or) information; we can infer a general rule. This kind of generalization process called explanation based learning (or) EBL.

45. What is the entailment constraints satisfied by EBL?
Hypothesis ^ Description |= classification
Background |= Hypothesis

46. Define RBL.
Relevance based Learning; the prior knowledge background concerns the relevance of a set of features to the goal predicate. This knowledge together with the observations, Allows the agent to infer a new, general rule that explains the observations.
Hypothesis ^ Description |= classifications,
Background ^ Description ^ classifications |= Hypothesis.

47. Define knowledge based Inductive learning.
KBIL algorithm finds inductive hypotheses that explain sets of observations with the help of background knowledge.

ILP techniques perform KBIL on knowledge that is expressed in first order logic. ILP methods can learn relational knowledge that is not expressible in attribute based systems.

49. What is the purpose of memorization?
Memorization used in computer science to speed up programs by saving the results of computation. The basic idea of memo function is to accumulate a database of input and output pairs, when the function is called; it first checks the database to see whether it can avoid solving the problem.

50. What is the basic EBL process step?
i. Construct a proof using the available background knowledge.
ii. Construct a generalized proof tree for the variabilized goal using the same inference steps as in the original proof.
iii. Construct a new rule where LHS consists of the leaves of the proof tree and R.H.S is the variabilized goal.
iv. Drop any conditions that are true.

51. Define constructive induction algorithm.
Algorithms that can generate new predicates are called constructive induction algorithms.

52. What are the two main subroutines used for generating literals?
i. NEW – LITERALS
ii. CHOOSE - LITERALS

53. What are the 3 kinds of literals that can be added?
i. Literals using Predicate
ii. Equality and inequality literals
iii. Arithmetic comparisons

54. Define Bayesian Learning.
It calculates the probability of each hypotheses, given the data and makes predictions on that basis, (i.e.) predictions are made by using all the hypotheses, weighted by their probabilities rather than by using just single “best” hypotheses.

55. Define MAP.
Maximum A Posteriori. A very common approximation is to make predictions based on single most probable hypotheses. This is MAP.

56. Define MDL.
The MDL (Maximum Description Length), is a learning method which attempts to minimize the size of the hypotheses and data encodings rather than work with probabilities.

57. What is Maximum – Likelihood hypotheses?
ML – it is reasonable approach when there is no reason to prefer one hypotheses over another a prior.

58. What are the methods for maximum likelihood parameter learning?
i. Write down an expression for the likelihood of the data as a function of the parameter.
ii. Write down the derivative of the log likelihood with respect to each parameter.
iii. Find the parameter values such that the derivatives are zero.

59. Define Naïve Bayes model.
In this model, the “class” variable C is the root and the “attribute” variable XI are the leaves. This model assumes that the attributes are conditionally independent of each other, given the class.

60. Define sum of squared errors.
The difference between the actual value yj and the predicated value (θ1 xj + θ2) so E is the sum of squared errors.

61. Define EM.
Expectation Maximization: the idea of EM is to pretend that we know the parameters of the model and then to infer the probability that each data point belongs to each component. After that we refit the components to the data, where each component is fitted to the entire data set with each point weighted by the probability.

62. What are the 2 steps in mixture model parameters?
i. E – Step
ii. M – Step

63. Define Neural Networks.
It consists of nodes or units connected by directed links. A link propagates the activation. Each link has a numeric weight which determines the strength and sign of the connection.

64. Give Activation function.
\[ n \]
\[ a_i = g(\text{ini}) = g(\Sigma W_{j,i} a_j) \]
\[ j=0 \]

65. What are the two functions in Neural network’s Activation functions?
i. Threshold function
ii. Sigmoid function

66. What are the categories of neural network structures?
i. Acyclic (or) Feed – forward networks
ii. Cyclic (or) Recurrent Networks

67. What is single layer feed forward neural network?
A network with all the inputs connected directly to the outputs is called a single layer neural network or a perceptron networks.

68. What is multilayer feed forward neural networks?
It consists of many hidden units. Each hidden unit act as a perceptron that represents a soft threshold functions in the input space. Output unit act as a soft threshold linear combination of several such functions.

The agent’s policy is fixed and the task is to learn the utilities of states, this could also involve
learning a model of the environment.

70. Define Active Learning.
The agent must learn what to do. An agent must experience as much as possible of its environment in order to learn how to behave in it.

71. Define TD.
Temporal Difference learning: The key of TD is to use the observed transitions to adjust the values of the observed states so that they agree with the constraint equations.
Part - B

1. What are the components of planning system and explain in detail.

Reasoning is the act of deriving a conclusion from certain premises using a given methodology.

- Reasoning is a process of thinking; reasoning is logically arguing; reasoning is drawing inference.
- When a system is required to do something, that it has not been explicitly told how to do, it must reason. It must figure out what it needs to know from what it already knows.
- Many types of Reasoning have long been identified and recognized, but many questions regarding their logical and computational properties still remain controversial.
- The popular methods of Reasoning include abduction, induction, model based, explanation and confirmation. All of them are intimately related to problems of belief revision and theory development, knowledge assimilation, discovery and learning.

List the Machine learning algorithms and explain in detail

A formal language may be viewed as being analogous to a collection of words or a collection of sentences.

**AI – Reasoning:** a collection

In computer science, a formal language is defined by precise mathematical or machine process able formulas.

- A formal language \( L \) is characterized as a set \( F \) of finite-length sequences of elements drawn from a specified finite set \( A \) of symbols.
- Mathematically, it is an unordered pair \( L = \{ A, F \} \)
- If \( A \) is words then the set \( A \) is called alphabet of \( L \), and the elements of \( F \) are called words.
- If \( A \) is sentence then the set \( A \) is called the lexicon or vocabulary of \( F \), and the elements of \( F \) are then called sentences.
- The mathematical theory that treats formal languages in general is known as **formal language theory**.

2. List out the different Methods of Reasoning?

Mostly three kinds of logical reasoning: Deduction, Induction, Abduction.

- **Deduction**

  - Example: "When it rains, the grass gets wet. It rains. Thus, the grass is wet."
  
  This means in determining the conclusion; it is using rule and its precondition to make a conclusion.
  
- Applying a general principle to a special case.
- Using theory to make predictions
- Usage: Inference engines, Theorem provers, Planning.
**Induction**

‡ Example: "The grass has been wet every time it has rained. Thus, when it rains, the grass gets wet."

This means in determining the rule; it is learning the rule after numerous examples of conclusion following the precondition.

‡ Deriving a general principle from special cases
‡ From observations to generalizations to knowledge
‡ Usage: Neural nets, Bayesian nets, Pattern recognition

Get useful study materials from www.rejinpaul.com

**AI - Reasoning**

**Abduction**

‡ Example: "When it rains, the grass gets wet. The grass is wet, it must have rained."

Means determining the precondition; it is using the conclusion and the rule to support that the precondition could explain the conclusion.

‡ Guessing that some general principle can relate a given pattern of cases
‡ Extract hypotheses to form a tentative theory
‡ Usage: Knowledge discovery, Statistical methods, Data mining.

**Analogy**

‡ Example: "An atom, with its nucleus and electrons, is like the solar system, with its sun and planets."

Means analogous; it is illustration of an idea by means of a more familiar idea that is similar to it in some significant features. And thus said to be analogous to it.

‡ finding a common pattern in different cases
‡ usage: Matching labels, Matching sub-graphs, Matching transformations.

Note: Deductive reasoning and Inductive reasoning are the two most commonly used explicit methods of reasoning to reach a conclusion.

3. **Describe Bayes’ Theorem?**

- Bayesian view of probability is related to degree of belief.
- It is a measure of the plausibility of an event given incomplete knowledge.
- Baye's theorem is also known as Bayes' rule or Bayes' law, or called Bayesian reasoning.
- The probability of an event A conditional on another event B ie P(A|B) is generally different from probability of B conditional on A ie P(B|A).
- There is a definite relationship between the two, P(A|B) and P(B|A), and Bayes' theorem is the statement of that relationship.
- Bayes theorem is a way to calculate P(A|B) from a knowledge of P(B|A).
- Bayes' Theorem is a result that allows new information to be used to update the conditional probability of an event.
- **Bayes' Theorem**

  Let $S$ be a sample space.

  Let $A_1, A_2, \ldots, A_n$ be a set of mutually exclusive events from $S$.

  Let $B$ be any event from the same $S$, such that $P(B) > 0$.

  Then Bayes' Theorem describes following two probabilities:

  $$P(A_k \cap B)$$

  $$P(A_k \mid B) = \frac{P(A_k \cap B)}{P(A_1 \cap B) + P(A_2 \cap B) + \cdots + P(A_n \cap B)}$$

  and

  by invoking the fact $P(A_k \cap B) = P(A_k).P(B \mid A_k)$ the probability

  $$P(A_k \mid B) = \frac{P(A_k).P(B \mid A_k)}{P(A_1).P(B \mid A_1) + P(A_2).P(B \mid A_2) + \cdots + P(A_n).P(B \mid A_n)}$$

- **Applying Bayes' Theorem:**

  Bayes' theorem is applied while following conditions exist.

  1. the sample space $S$ is partitioned into a set of mutually exclusive events $\{A_1, A_2, \ldots, A_n\}$.

  2. within $S$, there exists an event $B$, for which $P(B) > 0$.

  3. the goal is to compute a conditional probability of the form: $P(A_k \mid B)$.

  4. you know at least one of the two sets of probabilities described below

     - $P(A_k \cap B)$ for each $A_k$
4. **Discuss in details about Rule Based Systems**

A rule is an expression of the form "if A then B" where
- A is an assertion and B can be either an action or another assertion.
- Example: Trouble shooting of water pumps
  1. If pump failure then the pressure is low
  2. If pump failure then check oil level
  3. If power failure then pump failure
- Rule based system consists of a library of such rules.
- Rules reflect essential relationships within the domain.
- Rules reflect ways to reason about the domain.
- Rules draw conclusions and points to actions, when specific information about the domain comes in. This is called inference.
- The inference is a kind of chain reaction like: If there is a power failure then (see rules 1, 2, 3 mentioned above)
- Rule 3 states that there is a pump failure, and Rule 1 tells that the pressure is low, and Rule 2 gives a (useless) recommendation to check the oil level.

It is very difficult to control such a mixture of inference back and forth in the same session and resolve such uncertainties.

5. **Explain in details about Bayesian Networks and Certainty Factors**

A Bayesian network (or a belief network) is a probabilistic graphical model that represents a set of variables and their probabilistic independencies.

For example, a Bayesian network could represent the probabilistic relationships between diseases and symptoms. Given symptoms, the network can be used to compute the probabilities of the presence of various diseases.

Bayesian Networks are also called: Bayes nets, Bayesian Belief Networks (BBNs) or simply Belief Networks. Causal Probabilistic Networks (CPNs).

A Bayesian network consists of: a set of nodes and a set of directed edges between nodes. The edges reflect cause-effect relations within the domain. The effects are not completely deterministic (e.g. disease -> symptom) the strength of an effect is modeled as a probability.

**More Complicated Bayesian Networks:** The previous network was simple contained three nodes. Let us look at a slightly more complicated one in the context of heart disease.

Given the following facts about heart disease.
- Either smoking or bad diet or both can make heart disease more likely.
- Heart disease can produce either or both of the following two symptoms:
  † high blood pressure
  ‡ an abnormal electrocardiogram
- Here smoking and bad diet are regarded as causes of heart disease.

The heart disease in turn is a cause of high blood pressure and an abnormal electrocardiogram.
6. Describe in details about Fuzzy Logic

We have discussed only binary valued logic and classical set theory like:
A person belongs to a set of all human beings, and if given a specific subset, say all
males, then one can say whether or not the particular person belongs to this set.
This is ok since it is the way human reason. e.g.,
IF person is male AND a parent THEN person is a father. The rules are formed using
operators.
Here, it is intersection operator "AND" which manipulates the sets.
However, not everything can be described using binary valued sets.
The grouping of persons into "male" or "female" is easy, but as "tall" or "not tall" is problematic.
A set of "tall" people is difficult to define, because there is no distinct cut-off point at which tall begins.

Fuzzy logic was suggested by Zadeh as a method for mimicking the ability of human reasoning using a small number of rules and still producing a smooth output via a process of interpolation.

- **Description of Fuzzy Logic**

With fuzzy logic an element could partially belong to a set represented by the set membership. Example, a person of height 1.79 m would belong to both tall and not tall sets with a particular degree of membership.

**Difference between binary logic and fuzzy logic**

<table>
<thead>
<tr>
<th>Grade of truth</th>
<th>Not tall</th>
<th>Tall</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0 m height x</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

A fuzzy logic system is one that has at least one system component that uses fuzzy logic for its internal knowledge representation.

Fuzzy system communicate information using fuzzy sets.

Fuzzy logic is used purely for internal knowledge representation and externally it can be considered as any other system component.