# ARITIFICIAL INTELLIGENT AND MACHINE LEARNING

# UNIT I PROBLEM SOLVING

Introduction to AI - AI Applications - Problem solving agents – search algorithms – uninformed search strategies – Heuristic search strategies – Local search and optimization problems – adversarial search – constraint satisfaction problems (CSP)

# Part A

1. **What is Artificial Intelligence?**

Artificial Intelligence is the study of how to make computers do things which at the moment people do better.

1. **What are the different types of agents?**

**A human agent** has eyes, ears, and other organs for sensors and hands, legs, mouth, and other body parts for actuators.

**A robotic agent** might have cameras and infrared range finders for sensors and various motors for actuators.

**A software agent** receives keystrokes, file contents, and network packets as sensory inputs and acts on the environment by displaying on the screen, writing files, and sending network packets.

**Generic agent** - A general structure of an agent who interacts with the environment.

1. **List down the characteristics of intelligent agent.**

**Internal characteristics are**

**Learning/reasoning:** an agent has the ability to learn from previous experience and to successively adapt its own behaviour to the environment.

**Reactivity:** an agent must be capable of reacting appropriately to influences or information from its environment.

**Autonomy:** an agent must have both control over its actions and internal states. The degree of the agent’s autonomy can be specified. There may need intervention from the user only for important decisions.

**Goal-oriented:** an agent has well-defined goals and gradually influence its environment and so achieve its own goals. **External characteristics are**

**Communication:** an agent often requires an interaction with its environment to fulfil its tasks, such as human, other agents, and arbitrary information sources.

**Cooperation:** cooperation of several agents permits faster and better solutions for complex tasks that exceed the capabilities of a single agent.

**Mobility:** an agent may navigate within electronic communication networks.

**Character:** like human, an agent may demonstrate an external behaviour with many human characters as possible.

1. **What are various applications of AI? or What can AI do today?** 
   * + Robotic vehicles
     + Speech recognition
     + Autonomous planning and scheduling
     + Game playing
     + Spam fighting
     + Logistics planning
     + Robotics
     + Machine Translation

1. **Is AI a science, or is it engineering? Or neither or both? Explain.**

AI is both science and engineering. Observing and experimenting, which are at the core of any science, allows us to study artificial intelligence. From what we learn by observation and experimentation, we are able to engineer new systems that encompass what we learn and that may even be capable of learning themselves.

1. **What are the various agent programs in intelligent systems?**

Simple reflex agents

Model-based reflex agents

Goal-based agents

Utility-based agents

1. **Give example for real world end toy problems. Real world problem examples:** 
   * + - * Airline travel problem.
         * Touring problem.
         * Traveling salesman problem
         * VLSI Layout problem
         * Robot navigation
         * Automatic Assembly
         * Internet searching **Toy problem Examples:**
         * 8 – Queen problem
         * 8 – Puzzle problem
         * Vacuum world problem

1. **How will you measure the problem-solving performance?**

We can evaluate an algorithm’s performance in four ways:

**Completeness**: Is the algorithm guaranteed to find a solution when there is one?

**Optimality**: Does the strategy find the optimal solution?

**Time complexity**: How long does it take to find a solution?

**Space complexity**: How much memory is needed to perform the search?

1. **What is the application of BFS?**

It is simple search strategy, which is complete i.e. it surely gives solution if solution exists. If the depth of search tree is small then BFS is the best choice. It is useful in tree as well as in graph search.

1. **Evaluate performance of problem-solving method based on depth-first search algorithm?**

DFS algorithm performance measurement is done with four ways – 1) Completeness – It is complete (guarantees solution) 2) Optimality – it is not optimal.

* 1. Time complexity – It’s time complexity is O (b).
  2. Space complexity – its space complexity is O (b d+1).

1. **List some of the uninformed search techniques.**

The uninformed search strategies are those that do not take into account the location of the goal. That is these algorithms ignore where they are going until they find a goal and report success. The various uninformed search strategies are

* + - * + Breadth-first search
        + Uniform-cost search
        + Depth-first search
        + Depth-limited search
        + Iterative deepening depth-first search
        + Bidirectional search

1. **Differentiate Blind Search and Heuristic Search.**

|  |  |  |
| --- | --- | --- |
| **Parameters** | **Blind search** | **Heuristic search** |
| Known as | It is also known Uninformed Search | It is also known Informed Search |
| Using  Knowledge | It doesn’t use knowledge for the searching process. | It uses knowledge for the searching process. |
| Performance | It finds solution slow as compared to an informed search. | It finds a solution more quickly. |
| Completion | It is always complete. | It may or may not be complete. |
| Cost Factor | Cost is high. | Cost is low. |
| Time | It consumes moderate time because of slow searching. | It consumes less time because of quick searching. |
| Direction | No suggestion is given regarding the solution in it. | There is a direction given about the solution. |
| Implementation | It is lengthier while implemented. | It is less lengthy while implemented. |
| Computational requirements | Comparatively higher computational requirements. | Computational requirements are lessened. |

# UNIT II PROBABILISTIC REASONING

Acting under uncertainty – Bayesian inference – naïve bayes models. Probabilistic reasoning – Bayesian networks – exact inference in BN – approximate inference in BN – causal networks.

# Part A

1. **Why does uncertainty arise?**

Agents almost never have access to the whole truth about their environment. Uncertainty arises because of both laziness and ignorance. It is inescapable in complex, nondeterministic, or partially observable environments Agents cannot find a categorical answer.

 Uncertainty can also arise because of incompleteness, incorrectness in agents understanding of properties of environment.

1. **Define principle of maximum expected utility (MEU)?**

`The fundamental idea of decision theory 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 called the principle of maximum expected utility (MEU).

1. **Mention the needs of probabilistic reasoning in AI.** 
   * + When there are unpredictable outcomes.
     + When specifications or possibilities of predicates becomes too large to handle.
     + When an unknown error occurs during an experiment.

1. **What does the full joint probability distribution specify?**

The full joint probability distribution specifies the probability of each complete assignment of values to random variables. It is usually too large to create or use in its explicit form, but when it is available it can be used to answer queries simply by adding up entries for the possible worlds corresponding to the query propositions.

1. **State Bayes' Theorem in Artificial Intelligence.**

Bayes' theorem is also known as **Bayes' rule, Bayes' law**, or **Bayesian reasoning**, which determines the probability of an event with uncertain knowledge. It is a way to calculate the value of P(B|A) with the knowledge of P(A|B). Bayes' theorem allows updating the probability prediction of an event by observing new information of the real world.

**Example**: If cancer corresponds to one's age then by using Bayes' theorem, we can determine the probability of cancer more accurately with the help of age.

P(A/B)=[P(A)\*P(B/A)]/P(B)

1. **Given that P(A)=0.3,P(A|B)=0.4 and P(B)=0.5, Compute P(B|A).**

0.4 = (0.3\*P(B/A))/0.5

P(B/A) = 0.66

1. **What is Bayesian Belief Network?**

A Bayesian network is a probabilistic graphical model which represents a set of variables and their conditional dependencies using a directed acyclic graph. It is also called a Bayes network, belief network, decision network, or Bayesian model.

Bayesian networks are probabilistic, because these networks are built from a probability distribution, and also use probability theory for prediction and anomaly detection.

A Bayesian network is a directed graph in which each node is annotated with quantitative probability information. The full specification is as follows:

* + 1. Each node corresponds to a random variable, which may be discrete or continuous.
    2. A set of directed links or arrows connects pairs of nodes. If there is an arrow from node X to node Y , X is said to be a parent of Y. The graph has no directed cycles (and hence is a directed acyclic graph, or DAG).
    3. Each nodeXi has a conditional probability distribution P(Xi |Parents(Xi)) that quantifies the effect of the parents on the node.

# UNIT III SUPERVISED LEARNING

Introduction to machine learning – Linear Regression Models: Least squares, single & multiple variables, Bayesian linear regression, gradient descent, Linear Classification

Models: Discriminant function – Probabilistic discriminative model - Logistic regression, Probabilistic generative model – Naive Bayes, Maximum margin classifier – Support vector machine, Decision Tree, Random forests.

# PART - A

1. **What is Machine Learning?**

Machine learning is a branch of computer science which deals with system programming in order to automatically learn and improve with experience. For example: Robots are programed so that they can perform the task based on data they gather from sensors. It automatically learns programs from data.

1. **What is ‘Overfitting’ in Machine learning?**

In machine learning, when a statistical model describes random error or noise instead of underlying relationship ‘overfitting’ occurs. When a model is excessively complex, overfitting is normally observed, because of having too many parameters with respect to the number of training data types. The model exhibits poor performance which has been overfit.

1. **What are the five popular algorithms of Machine Learning?** 
   * + Decision Trees
     + Neural Networks (back propagation)
     + Probabilistic networks
     + Nearest Neighbor
     + Support vector machines

1. **What are the different Algorithm techniques in Machine Learning?**

The different types of techniques in Machine Learning are

* + - Supervised Learning
    - Unsupervised Learning
    - Semi-supervised Learning
    - Reinforcement Learning
    - Transduction

1. **What is ‘Training set’ and ‘Test set’?**

In various areas of information science like machine learning, a set of data is used to discover the potentially predictive relationship known as ‘Training Set’. Training set is an examples given to the learner, while Test set is used to test the accuracy of the hypotheses generated by the learner, and it is the set of example held back from the learner. Training set are distinct from Test set.

1. **What is the difference between artificial learning and machine learning?**

Designing and developing algorithms according to the behaviours based on empirical data are known as Machine Learning. While artificial intelligence in addition to machine learning, it also covers other aspects like knowledge representation, natural language processing, planning, robotics etc.

1. **What are the advantages of Naive Bayes?**

In Naïve Bayes classifier will converge quicker than discriminative models like logistic regression, so you need less training data. The main advantage is that

it can’t learn interactions between features.

1. **What is the main key difference between supervised and unsupervised machine learning?**

|  |  |
| --- | --- |
| **supervised learning** | **Unsupervised learning** |
| The supervised learning technique needs labelled data to train the model. For example, to solve a classification problem (a supervised learning task), you need to have label data to train the model and to classify the data into your labelled groups. | Unsupervised learning does not need any labelled dataset. This is the main key difference between supervised learning and unsupervised learning. |

1. **What is a Linear Regression?**

In simple terms, linear regression is adopting a linear approach to modeling the relationship between a dependent variable (scalar response) and one or more independent variables (explanatory variables). In case you have one explanatory variable, you call it a simple linear regression. In case you have more than one independent variable, you refer to the process as multiple linear regressions.

1. **What is the principle of least squares?**

Principle of Least Squares" states that the most probable values of a system of unknown quantities upon which observations have been made, are obtained by making the sum of the squares of the errors a minimum.

1. **What Is Bayesian Linear Regression?**

In Bayesian linear regression, the mean of one parameter is characterized by a weighted sum of other variables. This type of conditional modeling aims to determine the prior distribution of the regressors as well as other variables describing the allocation of the regress and eventually permits the out-of-sample forecasting of the regress and conditional on observations of the regression coefficients.

1. **What are types of classification models?** 
   * Logistic Regression
   * Naive Bayes
   * K-Nearest Neighbors
   * Decision Tree
   * Support Vector Machines

1. **What is probabilistic discriminative model?**

Discriminative models are a class of supervised machine learning models which make predictions by estimating conditional probability P(y|x). In order to use a generative model, more unknowns should be solved: one has to estimate probability of each class and probability of observation given class.

1. **What is SVM?**

It is a supervised learning algorithm used both for classification and regression problems. A type of discriminative modelling, support vector machine (SVM) creates a decision boundary to segregate n-dimensional space into classes. The best decision boundary is called a hyperplane created by choosing the extreme points called the support vectors.

1. **What is Decision tree?**

A type of supervised machine learning model where data is continuously split according to certain parameters. It has two main entities–decision nodes and leaves. While leaves are the final outcomes or decisions, nodes are the points where data is split.

1. **What is Random forest?**

It is a flexible and easy-to-use machine learning algorithm that gives great results without even using hyper-parameter tuning. Because of its simplicity and diversity, it is one of the most used algorithms for both classification and regression tasks.

1. **What is Decision Tree Classification?**

A decision tree builds classification (or regression) models as a tree structure, with datasets broken up into ever-smaller subsets while developing the decision tree, literally in a tree-like way with branches and nodes. Decision trees can handle both categorical and numerical data.

1. **What Is Pruning in Decision Trees, and How Is It Done?**

Pruning is a technique in machine learning that reduces the size of decision trees. It reduces the complexity of the final classifier, and hence improves predictive accuracy by the reduction of overfitting.

Pruning can occur in:

* + Top-down fashion. It will traverse nodes and trim subtrees starting at the root
  + Bottom-up fashion. It will begin at the leaf nodes

There is a popular pruning algorithm called reduced error pruning, in which:

* + Starting at the leaves, each node is replaced with its most popular class
  + If the prediction accuracy is not affected, the change is kept
  + There is an advantage of simplicity and speed

# UNIT IV ENSEMBLE TECHNIQUES AND UNSUPERVISED LEARNING

Combining multiple learners: Model combination schemes, Voting, Ensemble Learning - bagging, boosting, stacking, Unsupervised learning: K-means, Instance Based Learning: KNN, Gaussian mixture models and Expectation maximization

# PART - A

1. **What is bagging and boosting in ensemble learning?**

Bagging is a way to decrease the variance in the prediction by generating additional data for training from dataset using combinations with repetitions to produce multi-sets of the original data. Boosting is an iterative technique which adjusts the weight of an observation based on the last classification.

1. **What is stacking in ensemble learning?**

Stacking is one of the most popular ensemble machine learning techniques used to predict multiple nodes to build a new model and improve model performance. Stacking enables us to train multiple models to solve similar problems, and based on their combined output, it builds a new model with improved performance.

1. **Which are the three types of ensemble learning?**

The three main classes of ensemble learning methods are bagging, stacking, and boosting, and it is important to both have a detailed understanding of each method and to consider them on your predictive modeling project.

1. **Why ensemble methods are used?**

There are two main reasons to use an ensemble over a single model, and they are related; they are: Performance: An ensemble can make better predictions and achieve better performance than any single contributing model. Robustness: An ensemble reduces the spread or dispersion of the predictions and model performance.

1. **What is a voting classifier?**

A voting classifier is a machine learning estimator that trains various base models or estimators and predicts on the basis of aggregating the findings of each base estimator. The aggregating criteria can be combined decision of voting for each estimator output

1. **What type of classifiers are used in weighted voting method?**

The performance-weighted-voting model integrates five classifiers including logistic regression, SVM, random forest, XGBoost and neural networks. We first used crossvalidation to get the predicted results for the five classifiers.

1. **What is difference between K means and Gaussian mixture?**

K-Means is a simple and fast clustering method, but it may not truly capture heterogeneity inherent in Cloud workloads. Gaussian Mixture Models can discover complex patterns and group them into cohesive, homogeneous components that are close representatives of real patterns within the data set.

1. **What are Gaussian mixture models How is expectation maximization used in it?**

Expectation maximization provides an iterative solution to maximum likelihood estimation with latent variables. Gaussian mixture models are an approach to density estimation where the parameters of the distributions are fit using the expectation-maximization algorithm.

1. **What is k-means unsupervised learning?**

K-Means clustering is an unsupervised learning algorithm. There is no labeled data for this clustering, unlike in supervised learning. K-Means performs the division of objects into clusters that share similarities and are dissimilar to the objects belonging to another cluster. The term 'K' is a number.

1. **What is the difference between K-means and KNN?**

KNN is a supervised learning algorithm mainly used for classification problems, whereas K-Means (aka K-means clustering) is an unsupervised learning algorithm. K in K-Means refers to the number of clusters, whereas K in KNN is the number of nearest neighbors (based on the chosen distance metric).

1. **What is expectation maximization algorithm used for?**

` The EM algorithm is used to find (local) maximum likelihood parameters of a statistical model in cases where the equations cannot be solved directly. Typically these models involve latent variables in addition to unknown parameters and known data observations.

1. **What is the advantage of Gaussian process?**

Gaussian processes are a powerful algorithm for both regression and classification. Their greatest practical advantage is that they can give a reliable estimate of their own uncertainty.

1. **What are examples of unsupervised learning?**

Some examples of unsupervised learning algorithms include K-Means Clustering, Principal Component Analysis and Hierarchical Clustering.

1. **How do you implement expectation maximization algorithm?**

The two steps of the EM algorithm are:

E-step: perform probabilistic assignments of each data point to some class based on the current hypothesis h for the distributional class parameters;

M-step: update the hypothesis h for the distributional class parameters based on the new data assignments.

1. **What is the principle of maximum likelihood?**

The principle of maximum likelihood is a method of obtaining the optimum values of the parameters that define a model. And while doing so, you increase the likelihood of your model reaching the “true” model.

# UNIT V NEURAL NETWORKS

Perceptron - Multilayer perceptron, activation functions, network training – gradient descent optimization – stochastic gradient descent, error backpropagation, from shallow networks to deep networks –Unit saturation (aka the vanishing gradient problem) – ReLU, hyperparameter tuning, batch normalization, regularization, dropout.

**1. What is perceptron and its types?**

A Perceptron is an Artificial Neuron. It is the simplest possible Neural Network. Neural Networks are the building blocks of Machine Learning.

1. **Which activation function is used in multilayer perceptron?**

Image result for Perceptron - Multilayer perceptron, activation functions

The Sigmoid Activation Function: Activation in Multilayer Perceptron Neural Networks.

1. **What are the activation functions of MLP?**

In MLP and CNN neural network models, ReLU is the default activation function for hidden layers. In RNN neural network models, we use the sigmoid or tanh function for hidden layers. The tanh function has better performance. Only the identity activation function is considered linear.

1. **Does MLP have activation function?**

Multilayer perceptrons (MLP) has been proven to be very successful in many applications including classification. The activation function is the source of the MLP power. Careful selection of the activation function has a huge impact on the network performance.

1. **What is the difference between a perceptron and a MLP?**

The Perceptron was only capable of handling linearly separable data hence the multi-layer perception was introduced to overcome this limitation. An MLP is a neural network capable of handling both linearly separable and non-linearly separable data.

1. **What are the types of activation function?**

Popular types of activation functions and when to use them

* + Binary Step Function
  + Linear Function
  + Sigmoid
  + Tanh
  + ReLU
  + Leaky ReLU
  + Parameterised ReLU
  + Exponential Linear Unit

1. **What is MLP and how does it work?**

A multilayer perceptron (MLP) is a feedforward artificial neural network that generates a set of outputs from a set of inputs. An MLP is characterized by several layers of input nodes connected as a directed graph between the input and output layers. MLP uses backpropogation for training the network.

1. **Why do you require Multilayer Perceptron?**

MLPs are useful in research for their ability to solve problems stochastically, which often allows approximate solutions for extremely complex problems like fitness approximation.

1. **What are the advantages of Multilayer Perceptron?**

**Advantages of Multi-Layer Perceptron:**

A multi-layered perceptron model can be used to solve complex non-linear problems.

It works well with both small and large input data.

It helps us to obtain quick predictions after the training.

It helps to obtain the same accuracy ratio with large as well as small data.

1. **What do you mean by activation function?**

An activation function is a function used in artificial neural networks which outputs a small value for small inputs, and a larger value if its inputs exceed a threshold. If the inputs are large enough, the activation function "fires", otherwise it does nothing.

1. **What are the limitations of perceptron?**

Perceptron networks have several limitations. First, the output values of a perceptron can take on only one of two values (0 or 1) because of the hard-limit transfer function. Second, perceptrons can only classify linearly separable sets of vectors.

1. **How many layers are there in perceptron?**

This is known as a two-layer perceptron. It consists of two layers of neurons. The first layer is known as hidden layer, and the second layer, known as the output layer, consists of a single neuron.

1. **is stochastic gradient descent same as gradient descent?**

Compared to Gradient Descent, Stochastic Gradient Descent is much faster, and more suitable to large-scale datasets. But since the gradient it's not computed for the entire dataset, and only for one random point on each iteration, the updates have a higher variance.

1. **How is stochastic gradient descent used as an optimization technique?**

Stochastic gradient descent is an optimization algorithm often used in machine learning applications to find the model parameters that correspond to the best fit between predicted and actual outputs. It's an inexact but powerful technique. Stochastic gradient descent is widely used in machine learning applications.

1. **Does stochastic gradient descent lead to faster training?**

Gradient Descent is the most common optimization algorithm and the foundation of how we train an ML model. But it can be really slow for large datasets. That's why we use a variant of this algorithm known as Stochastic Gradient Descent to make our model learn a lot faster.

1. **What is stochastic gradient descent and why is it used in the training of neural networks?**

Stochastic Gradient Descent is an optimization algorithm that can be used to train neural network models. The Stochastic Gradient Descent algorithm requires gradients to be calculated for each variable in the model so that new values for the variables can be calculated.

1. **What are the three main types gradient descent algorithm?**

There are three types of gradient descent learning algorithms: batch gradient descent, stochastic gradient descent and mini-batch gradient descent.

1. **What are the disadvantages of stochastic gradient descent?**

SGD is much faster but the convergence path of SGD is noisier than that of original gradient descent. This is because in each step it is not calculating the actual gradient but an approximation. So we see a lot of fluctuations in the cost.

1. **How do you solve the vanishing gradient problem within a deep neural network?**

The vanishing gradient problem is caused by the derivative of the activation function used to create the neural network. The simplest solution to the problem is to replace the activation function of the network. Instead of sigmoid, use an activation function such as ReLU

1. **What is the problem with ReLU?**

Key among the limitations of ReLU is the case where large weight updates can mean that the summed input to the activation function is always negative, regardless of the input to the network. This means that a node with this problem will forever output an activation value of

0.0. This is referred to as a “dying ReLU“

1. **Why is ReLU used in deep learning?**

The ReLU function is another non-linear activation function that has gained popularity in the deep learning domain. ReLU stands for Rectified Linear Unit. The main advantage of using the ReLU function over other activation functions is that it does not activate all the neurons at the same time.

1. **Why is ReLU better than Softmax?**

As per our business requirement, we can choose our required activation function. Generally , we use ReLU in hidden layer to avoid vanishing gradient problem and better computation performance , and Softmax function use in last output layer .