## QUESTION BANK

**Name of the Department : M.E Computer Science and Engineering**

**Subject Code & Name : CP5191- MACHINE LEARNING TECHNIQUES**

**Year & Semester : I & I**

## UNIT I INTRODUCTION

### PART-A

**1 Give precise definition of learning.**

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.

**2. Find T, P and E for checkers learning problem, handwriting recognition learning problem, robot driving learning problem**

Checkers learning problem

- Task T: playing checkers
- Performance measure P: percent of games won against opponents
- Training experience E: playing practice games against itself.

Handwriting recognition learning problem

- Task T: recognizing and classifying handwritten words within images
- Performance measure P: percent of words correctly classified
- Training experience E: a database of handwritten words with given classifications

Robot driving learning problem

- Task T: driving on public four-lane highways using vision sensors
- Performance measure P: average distance travelled before an error
- Training experience E: a sequence of images and steering commands recorded while observing a human driver

**3. Define machine learning**.

Machine learning is a subfield of computer science (CS) and artificial intelligence (AI) that deals with the construction and study of systems that can **learn from data**, rather than follow only explicitly programmed instructions. (i.e. Machine learning is the science of getting computers to act without being explicitly programmed)

**4. List some disciplines and examples of their influence on machine learning**

- Artificial intelligence
- Bayesian methods
- Computational complexity theory
- Control theory
- Information theory
- Philosophy
- Psychology and neurobiology
- Statistics

**5. List some successful applications of machine learning**

- Learning to recognize spoken words
- Learning to drive an autonomous vehicle
- Learning to classify new astronomical structures
- Learning to play world-class backgammon.

**What is the difference between artificial intelligence and machine learning methods?**

Artificial intelligence, and is sometimes known as **symbolic processing** because the computer manipulates symbols that reflect the environment. In contrast, machine learning methods are sometimes called **subsymbolic** because no symbols or symbolic manipulation are involved.

**7. List the types of machine learning.**

- Supervised learning
- Unsupervised learning

- Reinforcement learning
- Evolutionary learning

**8. What is supervised learning?**

A **training set** of examples with the **correct responses** (targets) are provided and, based on this training set, the algorithm **generalizes** to respond correctly to all possible inputs. This is also called learning from exemplars.

**9. What is unsupervised learning?**

**Correct responses are not provided**, instead the algorithm tries to identify similarities between the inputs so that inputs that have something in common are **categorized** together.

The statistical approach to unsupervised learning is known as **density estimation**.

**10. What is reinforcement learning?**

This is somewhere between supervised and unsupervised learning. The algorithm gets told when the answer is wrong, but does not get told how to correct it. It has to explore and try out different possibilities until it works out how to get the answer right. Reinforcement learning sometimes called **learning with critic** because of this monitor that scores the answer, but does not suggest improvements.

**11. What is evolutionary learning?**

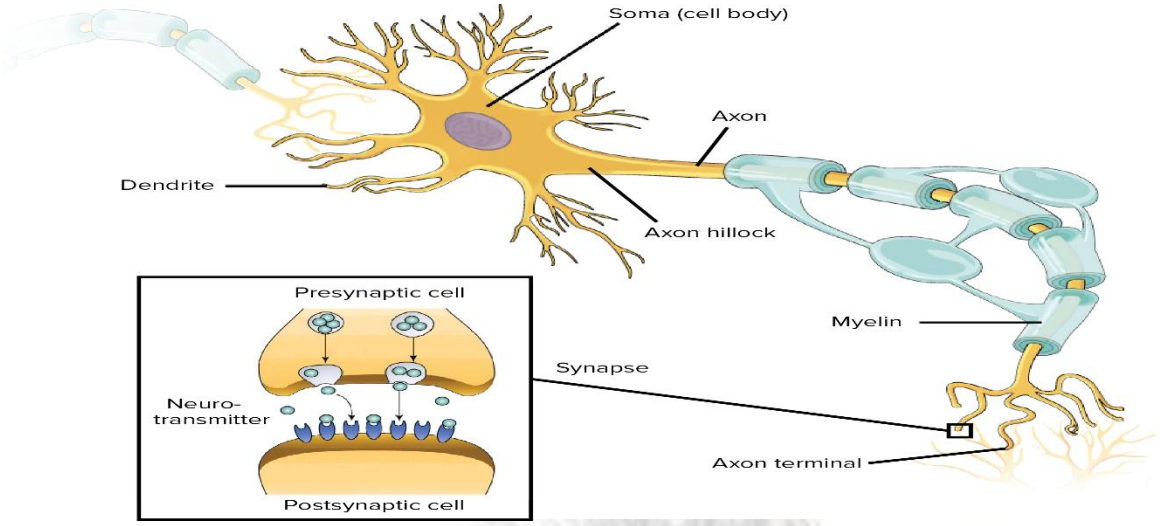
Biological evaluation can be seen as a learning process: biological organisms adapt to improve their survival rates and chance of having offspring in their environment.

**12. What are the basic functions of a neuron?**

- Receive signals (or information).
- Integrate incoming signals (to determine whether or not the information should be passed along).
- Communicate signals to target cells (other neurons or muscles or glands).

These neuronal functions are reflected in the anatomy of the neuron.

**13. Draw the basic structure of neuron**




**14. List the steps in designing a learning system.**

- Choosing the training experience
- Choosing the target function
- Choosing the representation for the target function
- Choosing a function approximation algorithm.

  Estimating training values

  Adjusting the weights
- The final design

**15. State Hebb's rule.**

Hebb's rule says that the changes in the strength of synaptic connections are proportional to the correlation in the firing of the two connecting neurons. So if two neurons consistency fire simultaneously, then any connection between them will change in strength, becoming stronger. However, if the two neurons never fire simultaneously, the connection between them will die away.

**16. How McCulloch and Pitts modelled a neuron?**

McCulloch and Pitts modelled a neuron AS:

- **A set of weighted inputs** $w_i$ that correspond to the synapses

- **An adder** that sums the input signals (equivalent to the membrane of the cell that collects electrical charge)
- **An activation function** (initially a threshold function) that decides whether the neuron fires for the current inputs.

**17. List the limitations of MCP model.**

- Weights and thresholds are analytically determined. Cannot learn
- Very difficult to minimize size of a network
- What about non-discrete and/or non-binary tasks?

**18. What is Perceptron?**

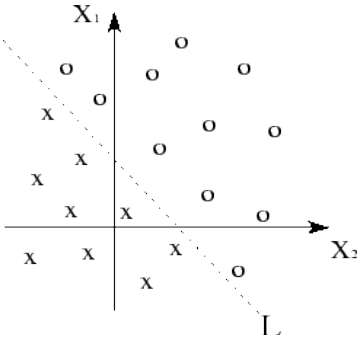
The Perceptron is nothing more than a collection of McCulloch and Pitts neurons together with a set of inputs and some weights to fasten the inputs to the neurons.

**. What is Linear Discriminant?**

Linear discriminant analysis (LDA) is a generalization of Fisher's linear discriminant, a method used in statistics, pattern recognition and machine learning to find a linear combination of features that **characterizes or separates two or more classes of objects or events**. The resulting combination may be used as a linear classifier, or, more commonly, for dimensionality reduction before later classification.

**20. What is linear reparability?**

Consider two-input patterns ($X_1$, $X_2$) being classified into two classes (as shown in figure). Each point with either symbol of $x$ or $o$ represents a pattern with a set of values($X_1$, $X_2$) . Each pattern is classified into one of two classes. Notice that these classes can be separated with a single line $L$. They are known as *linearly separable* patterns. *Linear separability* refers to the fact that classes of patterns with $n$-dimensional vector x = ($x_1$,$x_2$….,$x_n$)can be separated with a single *decision surface*. In the case above, the line $L$ represents the decision surface.




**21. What is linear regression?**

In statistics, linear regression is a linear approach for modeling the relationship between a scalar dependent variabley and one or more explanatory variables (or independent variables) denoted X. The case of one explanatory variable is called simple linear regression. For more than one explanatory variable, the process is called multiple linear regression.

## PART-B

1. Define learning. What are the three features of learning? Explain the three features of learning with the following problem.
   a) Checkers learning problem
   b) Handwriting recognition learning problem
   c) Robot driving learning problem.

2 a) List and explain some successful applications of machine learning?
   b) List some disciplines and examples of their influence on machine learning.

3. Explain the various steps in designing a learning system.

4.Elaborate the Perspectives and issues in machine learning

5. Explain concept learning task with the example of ENJOYSPORT.

6. Explain concept learning as search with the example of ENJOYSPORT.

7. Write FIND-S algorithm. Trace the algorithm with example.

8. Explain version space and List-Then-Eliminate algorithm with an example.

9. Explain CANDIDATE ELIMINATION learning algorithm with an example.

10. List and explain the various remarks on version spaces and candidate elimination.

## UNIT-II LINEAR MODELS

## PART A

**1. What is multilayer perceptron (MLP)?**

A multilayer perceptron (MLP) is a class of feed forwardartificial neural network. The MLP consists of **three or more layers** (an input and an output layer with one or more *hidden layers*) of nonlinearly-activating nodes making it a deep neural network. Since MLPs are fully connected, each node in one layer connects with a certain weight $w_{ij}$to every node in the following layer.

**2. What are the 2 steps of MLP training?**

Training the MLP consists of two parts:

- Working out what the outputs are for the given inputs and the current weights
- Updating the weights according to the error, which is a function of the difference between the outputs and the targets.

**3. What is biases in MLP?**

Just like in the perceptron case, we need to include a bias input to each neuron. We do this in the same way, by having **extra input that is permanently set to -1**, and adjusting the weights to each neuron as part of the training. Thus each neuron in the network has 1 extra input, with fixed value.

**4. What do you mean by going forwards and backwards through the network?**

Training the MLP consists of two parts:

- Working out what the outputs are for the given inputs and the current weights
- Updating the weights according to the error, which is a function of the difference between the outputs and the targets. These are generally known as going forwards and backwards through the network.

**5. What do you mean by back propagation?**

Back propagation is a method used in artificial neural networks to calculate the error contribution of each neuron after a batch of data (e.g. in image recognition, multiple images) is processed. This is used by an enveloping optimization algorithm to adjust the weight of each neuron, completing the learning process for that case.

**6. What is backward propagation of errors?**

Technically it calculates the gradient of the loss function. It is commonly used in the gradient descent optimization algorithm. It is also called backward propagation of errors, because the error is calculated at the output and distributed back through the network layers.

**7. What is activation function?**

The **activation function** of a node defines the output of that node given an input or set of inputs. A standard computer chip circuit can be seen as a digital network of activation functions that can be "ON" (1) or "OFF" (0), depending on input. This is similar to the behavior of thelinear perceptron in neural networks. However, only nonlinear activation functions allow such networks to compute nontrivial problems using only a small number of nodes. In artificial neural networks this function is also called the transfer function.

**8. Why can’t we do it without activating the input signal?**

If we do not apply activation function then the output signal would simply be a simple linear function. A linear function is just a polynomial of one degree. Now, a linear equation is easy to solve but they are limited in their complexity and have less power to learn complex functional mappings from data. A Neural Network without Activation function would simply be a linear regression Model, which has limited power and does not performs good most of the times.

**9. Why do we need Non-Linearities?**

Non-linear functions are those which have degree more than one and they have a curvature when we plot a Non-Linear function. Now we need a Neural Network Model to learn and represent almost anything and any arbitrary complex function which maps inputs to outputs. Neural-Networks are considered Universal Function Approximators. It means that

they can compute and learn any function at all. Almost any process we can think of can be represented as a functional computation in Neural Networks.

Hence it all comes down to this, we need to apply a Activation function f(x) so as to make the network more powerful and add ability to it to learn something complex and complicated form data and represent non-linear complex arbitrary functional mappings between inputs and outputs. Hence using a nonlinear Activation we are able to generate non-linear mappings from inputs to outputs.

**10. List the various activation function.**

- Step function
- Linear function
- Sigmoid function
- Tanh function

**11. Write about Sigmoid Activation function.**

It is an activation function of form f(x) = 1 / 1 + exp (-x). Its Range is between 0 and 1. It is an S - shaped curve. It is easy to understand.

**12. What are the limitations of sigmoid function?**

- Vanishing gradient problem
- Secondly, its output isn't zero centered. It makes the gradient updates go too far in different directions. 0 < output < 1, and it makes optimization harder.
- Sigmoid saturate and kill gradients.
- Sigmoid have slow convergence.

**13. What is Tanh function?**

Hyperbolic Tangent function- Tanh:its mathematical formula is **f(x) = 1 — exp (-2x) / 1 + exp (-2x).** Now its output is zero centered because its range in between -1 to 1 i.e -1 < output < 1. Hence optimization is *easier* in this method hence in practice it is always preferred over Sigmoid function. But still it suffers from Vanishing gradient problem.

**14. What are the two ways to implement BP learning?**

- Sequential mode
- Batch mode

**15. What is sequential mode?**

In this mode of BP learning, adjustments are made to the free parameters of the network on an example-by-example basis. The sequential mode is also suited for pattern classification problem. Sequential mode also referred as on-line mode or stochastic mode.

**16. What is batch mode?**

In this mode of BP learning, adjustments are made to the free parameters of the network on an epoch-by-epoch basis, where each epoch consists of the entire set of training examples. This batch mode is best suited for nonlinear regression.

**17. How many weights are there in MLP with one hidden layer?**

For the MLP with one hidden layer there are (m+1) x n + (n x1) x p weights where m, n, p are the number of nodes in the input, hidden and output layers respectively. The extra +1 come from the bias node, which also have adjustable weights.

**18. What are the various parameters that the hidden unit depends on?**

The best number of hidden units depends in a complex way on many factors, including:

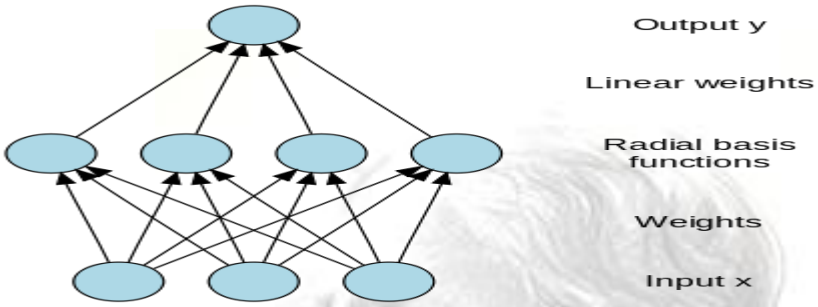
- The number of training patterns
- The numbers of input and output units
- The amount of noise in the training data
- The complexity of the function or classification to be learned
- The type of hidden unit activation function
- The training algorithm

**19. What is a radial basis function network?**

A Radial basis function network is an artificial neural network that uses radial basis functions as activation functions. The output of the network is a linear combination of radial basis functions of the inputs and neuron parameters. Radial basis function networks have

many uses, including function approximation, time series prediction, classification, and system control.

**20. Draw the architecture of a radial basis function network.**




An input vector x is used as input to all radial basis functions, each with different parameters. The output of the network is a linear combination of the outputs from radial basis functions.

**21. What is SVM?**

In machine learning, support vector machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilisticbinarylinear classifier

## PART-B

1. Explain how the XOR problem can be solved by an MLP?
2. Explain the multilayer perceptron algorithm in detail.
3. Explain the following with respect to MLP.
   a) Initializing the weights

b) Different output activation functions

4. Explain the following in detail

a) Sequential and batch training

b) Local minima

5. What are the choices that can be made about the network in order to use it for real problems? Explain in detail.

6. How regression problem can be solved using MLP? Explain with example.

7. Explain classification with MLP in detail.

8. Explain time series prediction problem with MLP in detail.

9. Explain the auto associative network in detail.

10. Explain Radial Basis Function (RBF) network & training RBFN in detail.

11. Explain the following in detail

a) The curse of dimensionality

b) Interpolation and basis function.

12. Explain support vector machine in detail.

## UNIT-III TREE AND PROBABILISTIC MODEL

## PART-A

**1. What is the idea of decision tree?**

The idea of a decision tree is that we break classification down into set of choices about each feature in turn, starting at the **root** (base) of the tree and processing down to the **leaves**, where we receive the classification decision.

**2.List the advantages and disadvantages of decision tree.**

- Are simple to understand and interpret. People are able to understand decision tree models after a brief explanation.
- Have value even with little hard data. Important insights can be generated based on experts describing a situation (its alternatives, probabilities, and costs) and their preferences for outcomes.
- Allow the addition of new possible scenarios.
- Help determine worst, best and expected values for different scenarios.
- Use a white box model. If a given result is provided by a model.
- Can be combined with other decision techniques.

Disadvantages of decision trees:

- For data including categorical variables with different number of levels, information gain in decision trees is biased in favor of those attributes with more levels.
- Calculations can get very complex, particularly if many values are uncertain and/or if many outcomes are linked.

**3. Define (information) Entropy**

Information entropy, which describes the amount of impurity in a set of features. The entropy H of a set of probabilities $p_i$ is

$$H(p) = -\Sigma\, p_i \log_2 p_i$$

where the logarithm is base 2

**4. Write function for computing the entropy.**

```
defcalc_entropy (p):
    if p ! =0:
        return –p * log2 (p)
    else:
        return 0
```

**5. Define information gain.**

Information gain is defined as the entropy of the whole set minus the entropy when a particular feature is chosen. This is defined by

$$Gain(S,F) = Entropy(S) - \sum f \in Values(F)|S_f| / |S| \cdot Entropy(S_f)$$

Where S is the set of examples, F is a possible feature out of the set of all possible ones, and $|S_f|$ is a count of the number of members of S that have value f for feature F.

**6. List some decision tree algorithms.**

There are many specific decision-tree algorithms. Notable ones include:

- ID3 (Iterative Dichotomiser 3)
- C4.5 (successor of ID3)
- CART (Classification And Regression Tree)
- CHAID (CHi-squared Automatic Interaction Detector). Performs multi-level splits when computing classification trees.
- MARS: extends decision trees to handle numerical data better.
- Conditional Inference Trees. Statistics-based approach that uses non-parametric tests as splitting criteria, corrected for multiple testing to avoid overfitting. This approach results in unbiased predictor selection and does not require pruning

**7. Write note on ID3 algorithm.**

- Calculate the entropy of every attribute using the data set

- Split the set S into subsets using the attribute for which the resulting entropy (after splitting) is minimum (or, equivalently, information gain is maximum)
- Make a decision tree node containing that attribute
- Recurse on subsets using remaining attributes.

**8. What is CART?**

Decision trees used in data mining are of two main types:

- Classification tree analysis is when the predicted outcome is the class to which the data belongs.
- Regression tree analysis is when the predicted outcome can be considered a real number (e.g. the price of a house, or a patient's length of stay in a hospital).

The term Classification And Regression Tree (CART) analysis is an umbrella term used to refer to both of the above procedures, first introduced by Breiman et al. Trees used for regression and trees used for classification have some similarities - but also some differences, such as the procedure used to determine where to split.

**9. What is Gini impurity?**

Used by the CART (classification and regression tree) algorithm, Gini impurity is a measure of how often a randomly chosen element from the set would be incorrectly labeled if it was randomly labeled according to the distribution of labels in the subset. Gini impurity can be computed by summing the probability $p_i$ of an item with label I being chosen times the probability $1-p_i$ of a mistake in categorizing that item. It reaches its minimum (zero) when all cases in the node fall into a single target category.

**10. What is the goal of ensemble methods?**

The goal of **ensemble methods** is to combine the predictions of several base estimators built with a given learning algorithm in order to improve generalizability / robustness over a single estimator.

**11. What are the two families of ensemble methods?**

Two families of ensemble methods are usually distinguished:

- In **averaging methods,** the driving principle is to build several estimators independently and then to average their predictions. On average, the combined estimator is usually better than any of the single base estimator because its variance is reduced.

  Examples:Bagging methods, Forests of randomized trees

- By contrast, in**boosting methods**, base estimators are built sequentially and one tries to reduce the bias of the combined estimator. The motivation is to combine several weak models to produce a powerful ensemble.

  Examples:AdaBoost, Gradient Tree Boosting

**12. What is Boosting?**

Boosting is a general ensemble method that attempts to create a strong classifier from a number of weak classifiers.

This is done by building a model from the training data, then creating a second model that attempts to correct the errors from the first model. Models are added until the training set is predicted perfectly or a maximum number of models are added.

**13. What is stumping?**

There is a very extreme form of boosting that is applied to trees. It goes by the descriptive name of stumping. Stumping consists of simply taking the root of the tree and using as the decision maker.

**14. What is bagging?**

Bootstrap aggregating, also called bagging, is a machine learning ensemblemeta-algorithm designed to improve the stability and accuracy of machine learning algorithms used in statistical classification and regression. It also reduces variance and helps to avoid overfitting. Although it is usually applied to decision tree methods, it can be used with any type of method. Bagging is a special case of the model averaging approach.

**15 What is expectation–maximization (EM) algorithm?**

In statistics, an expectation–maximization (EM) algorithm is an iterative method to find maximum likelihood or maximum a posteriori (MAP) estimates of parameters in statistical models, where the model depends on unobserved latent variables. The EM iteration alternates between performing an expectation (E) step, which creates a function for the expectation of the log-likelihood evaluated using the current estimate for the parameters, and a maximization (M) step, which computes parameters maximizing the expected log-likelihood found on the E step. These parameter-estimates are then used to determine the distribution of the latent variables in the next E step.

**16. What is Gaussian mixture model?**

A Gaussian mixture model is a probabilistic model that assumes all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters. One can think of mixture models as generalizing k-means clustering to incorporate information about the covariance structure of the data as well as the centers of the latent Gaussians.

**17. What is kernel smoother?**

A kernel smoother is a statistical technique for estimating a real valued functionby using its noisy observations, when no parametric model for this function is known. The estimated function is smooth, and the level of smoothness is set by a single parameter.

This technique is most appropriate for low-dimensional ($p < 3$) data visualization purposes. Actually, the kernel smoother represents the set of irregular data points as a smooth line or surface.

**18. What is k-NN algorithm?**

In pattern recognition, the k-nearest neighbors algorithm (k-NN) is a non-parametric method used for classification and regression.[1] In both cases, the input consists of the k closest training examples in the feature space. The output depends on whether k-NN is used for classification or regression:

In k-NN classification, the output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common

among its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbor.

In k-NN regression, the output is the property value for the object. This value is the average of the values of its k nearest neighbors.

**19. What is vector quantization?**

Vector quantization (VQ) is a classical quantization technique from signal processing that allows the modeling of probability density functions by the distribution of prototype vectors. It was originally used for data compression. It works by dividing a large set of points (vectors) into groups having approximately the same number of points closest to them. Each group is represented by its centroid point, as in k-means and some other clustering algorithms.

**20. What is SOFM?**

A self-organizing map (SOM) or self-organizing feature map (SOFM) is a type of artificial neural network (ANN) that is trained using unsupervised learning to produce a low-dimensional (typically two-dimensional), discretized representation of the input space of the training samples, called a map, and is therefore a method to do dimensionality reduction. Self-organizing maps differ from other artificial neural networks as they apply competitive learning as opposed to error-correction learning (such as back propagation with gradient descent), and in the sense that they use a neighborhood function to preserve the topological properties of the input space.

**PART-B**

1. Explain ID3 algorithm with example.
2. Explain classification and regression trees with example.
3. Make a decision tree that computes the logical AND function. How does it compare to the perceptron solution?
4. Explain AdaBoost algorithm with example.
5. Explain the different ways to combine classifiers.
6. Explain the following
    a) Turning data into probabilities
    b) Naïve Bayes classifier

7. Explain the important statistical concepts in detail.
8. Explain the Gaussian Mixture models in detail.
9. Explain nearest neighbor methods in detail.
10. Explain the k-means algorithm in detail

## UNIT-IV DIMENSIONALITY REDUCTION

## PART-A

**1. Define dimensionality reduction**

In machine learning and statistics, dimensionality reduction or dimension reduction is the process of reducing the number of random variables under consideration via obtaining a set of principal variables. It can be divided into feature selection and feature extraction.

**2. List the different ways to do dimensionality reduction**

- Feature selection
- Feature derivation
- Clustering

**3. Define feature selection.**

Feature selection means looking through the features that are available and seeing whether or not they are actually useful, i.e correlated to the output variables.

**4. Define feature derivation**

Feature derivation means deriving new features from the old ones, generally by applying transforms to the dataset that simply change the axes 9coordinate system) of the graph by moving and rotating them, which can be written simply as a matrix that we apply to the data.

**5. What is Linear Discriminant Analysis (LDA)?**

Linear Discriminant Analysis (LDA) is most commonly used as dimensionality reduction technique in the pre-processing step for pattern-classification and machine learning

applications. The goal is to project a dataset onto a lower-dimensional space with good class-separability in order avoid overfitting ("curse of dimensionality") and also reduce computational costs.

**6. What is the idea of principal component analysis?**

The idea of principal component analysis is that it is a direction in the data with the largest variation. The algorithm first centres the data by subtracting off the mean, and then chooses the direction with the largest variation and places an axis in that direction, and then looks at the variation that remains and finds another axis that it orthogonal to the first and covers as much of the remaining variation as possible. It then iterates this until it has run out of possible axes.

**7. Define spectral decomposition.**

When$A=S\Lambda S^{-1}$is a real-symmetric (or Hermitian) matrix, itseigenvectors can be chosen orthonormal and hence$S=Q$is orthogonal (or unitary).Thus,$A=Q\Lambda Q^{T}$, which is called thespectral decompositionofA.

**8. Find the spectral decomposition for$A=(3\ 2\ 2\ 3)$**

The characteristic equation for$A$is$\lambda^{2}-6\lambda+5=0$. Thus the eigenvalues of$A$are$\lambda_1=1$ and$\lambda_2=5$. For$\lambda_1=1$, the eigenvector is$v_1=(1-1)$. For$\lambda_2=5$, theeigenvector is$v_2=(11)$. Thus the spectral decomposition of$A$is$A=(\sqrt{2}/2\sqrt{2}/2-\sqrt{2}/2\sqrt{2}/2)(1\ 0\ 0\ 5)(\sqrt{2}/2-\sqrt{2}/2\sqrt{2}/2\sqrt{2}/2)$

**9. What is kernel PCA?**

In the field of multivariate statistics, kernel principal component analysis (kernel PCA) is an extension of principal component analysis (PCA) using techniques of kernel methods. Using a kernel, the originally linear operations of PCA are performed in a reproducing kernel Hilbert space.

**10. What is factor analysis?**

The idea of factor analysis is to ask whether the data that is observed can be explained by a smaller number of uncorrelated factors ot latent variables.

**11. What is genetic algorithm?**

The genetic algorithm is a method for solving both constrained and unconstrained optimization problems that is based on natural selection, the process that drives biological evolution. The genetic algorithm repeatedly modifies a population of individual solutions. 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. Over successive generations, the population "evolves" toward an optimal solution.

**12. What are three main types of rules the genetic algorithm uses at each step to create the next generation from the current population?**

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

- *Selection rules* select the individuals, called *parents* that contribute to the population at the next generation.
- *Crossover rules* combine two parents to form children for the next generation.
- *Mutation rules* apply random changes to individual parents to form children.

**13. Enumerate the difference between genetic algorithm and classical optimization algorithm.**

The genetic algorithm differs from a classical, derivative-based, optimization algorithm in two main ways, as summarized in the following table.

| **Classical Algorithm** | **Genetic Algorithm** |
|---|---|
| Generates a single point at each iteration. The sequence of points approaches an optimal solution. | Generates a population of points at each iteration. The best point in the population approaches an optimal solution. |
| Selects the next point in the sequence by a deterministic computation. | Selects the next population by computation which uses random number generators. |

**14. List the methods of various selection of chromosomes.**

- Fitness proportionate selection (SCX) The individual is selected on the basis of fitness. The probability of an individual to be selected increases with the fitness of the individual greater or less than its competitor's fitness.
- Simplex Crossover (SPX)
- Boltzmann selection
- Tournament selection
- Rank selection
- Steady state selection
- Truncation selection
- Local selection

**15. What is truncation selection?**

Truncation selection is the simplest and arguably least useful selection strategy. Truncation selection simply retains the fittest x% of the population. These fittest individuals are duplicated so that the population size is maintained. For example, we might select the fittest 25% from a population of 100 individuals. In this case we would create four copies of each of the 25 candidates in order to maintain a population of 100 individuals. This is an easy selection strategy to implement but it can result in premature convergence as less fit candidates are ruthlessly culled without being given the opportunity to evolve into something better. Nevertheless, truncation selection can be an effective strategy for certain problems.

**16. What is *fitness-proportionate selection*?**

A better approach to selection is to give every individual a chance of being selected to breed but to make fitter candidates more likely to be chosen than weaker individuals. This is achieved by making an individual's survival probability a function of its fitness score. Such strategies are known as *fitness-proportionate selection.*

**17. What are the three basic tasks to be performed when you want to apply genetic algorithm?**

- Encode possible solutions as strings
- Choose a suitable fitness functions
- Choose suitable genetic operators.

**18. What is crossover?**

The crossover operator is analogous to reproduction and biological crossover. In this more than one parent is selected and one or more off-springs are produced using the genetic material of the parents. Crossover is usually applied in a GA with a high probability

**19. What is Markov decision processes (MDPs)**

Markov decision processes provide a mathematical framework for modeling **decision** making in situations where outcomes are partly random and partly under the control of a **decision** maker. MDPs are useful for studying a wide range of optimization problems solved via dynamic programming and reinforcement learning.

**20. What is partially observable Markov decision process (POMDP)?**

A partially observable Markov decision process (POMDP) is a generalization of a Markov decision process (MDP). A POMDP models an agent decision process in which it is assumed that the system dynamics are determined by an MDP, but the agent cannot directly observe the underlying state. Instead, it must maintain a probability distribution over the set of possible states, based on a set of observations and observation probabilities, and the underlying MDP.

## PART-B

1. Explain the three different ways to do dimensionality reduction in detail.
2. Explain Linear Discriminant analysis in detail.
3. Explain principal component analysis algorithm in detail.
4. Explain Factor analysis in detail.
5. Use the LDA on the Iris dataset. Compare the results with using PCA, which is not supervised and will not therefore be able to find the same space.
6. Explain Levenberg-Marquardt algorithm in detail.
7. Explain Conjugate gradients algorithm in detail.
8. Explain the three basic approaches in search techniques.
9. Explain evolutionary learning in detail.

10. Explain reinforcement learning in detail



## UNIT-V GRAPHICAL MODELS

## PART-A

**1. What is linear congruential generator (LCG)?**

A linear congruential generator (LCG) is an algorithm that yields a sequence of pseudo-randomized numbers calculated with a discontinuous piecewise linear equation. The method represents one of the oldest and best-known pseudorandom number generator algorithms.[1] The theory behind them is relatively easy to understand, and they are easily implemented and fast, especially on computer hardware which can provide modulo arithmetic by storage-bit truncation.

The generator is defined by the recurrence relation:

$$x_{n+1} = (ax_n + c) \bmod m$$

Where a,c and m are parameters that have to be chosen.

**2. What is Mersenne prime?**

In mathematics, a Mersenne prime is a prime number that is one less than a power of two. That is, it is a prime number of the form $M_n = 2^n - 1$ for some integer n.

**3. What is Box-Muller transform?**

The Box–Muller transform, by George Edward Pelham Box and Mervin Edgar Muller is a pseudo-random number sampling method for generating pairs of independent, standard, normally distributed (zero expectation, unit variance) random numbers, given a source of uniformly distributed random numbers.

**4. What are the two forms that the Box–Muller transform is commonly expressed?**

The Box–Muller transform is commonly expressed in two forms. The basic form as given by Box and Muller takes two samples from the uniform distribution on the interval [0, 1] and maps them to two standard, normally distributed samples. The polar form takes two samples from a different interval, [−1, +1], and maps them to two normally distributed samples without the use of sine or cosine functions.

**5. What is chain?**

In probabilistic term, a chain is a sequence of possible states, where the probability of being in state s at time t is a function of the previous state.

**6. What is Markov chain?**

Markov chain is a chain with the Markov property. That is the probability at time t depends only on the state at t-1.

**7. What is graphical model?**

A graphical model or probabilistic graphical model (PGM) is a probabilistic model for which a graph expresses the conditional dependence structure between random variables. They are commonly used in probability theory, statistics—particularly Bayesian statistics—and machine learning.

**8. What is Metropolis–Hastings Algorithm?**

The Metropolis–Hastings algorithm is a Markov chain Monte Carlo (MCMC) method for obtaining a sequence of random samples from a probability distribution for which direct sampling is difficult. This sequence can be used to approximate the distribution (e.g., to generate a histogram), or to compute an integral (such as an expected value). Metropolis–Hastings and other MCMC algorithms are generally used for sampling from multi-dimensional distributions, especially when the number of dimensions is high

**9. What is Gibbs sampling?**

Gibbs sampling is one MCMC technique suitable for the task. The idea in Gibbs sampling is to generate posterior samples by sweeping through each variable (or block of variables) to **sample** from its conditional distribution with the remaining variables fixed to their current values.

**10. What is Transition Probabilities**?

The one-step transition probability is the probability of transitioning from one state to another in a single step. The **Markov** chain is said to be time homogeneous if the transition probabilities from one state to another are independent of time index.

**11. What is ergodic in Markov chain?**

A Markov chain is called an ergodic chain if it is possible to gofrom every state to every state (not necessarily in one move).Ergodic Markov chains are also called irreducible.A Markov chain is called a regular chain if some power of thetransition matrix has only positive elements.

**12. What is Markov chain Monte Carlo?**

Markov chain Monte Carlo (MCMC) methods are a class of algorithms for sampling from a probability distribution based on constructing a Markov chain that has the desired distribution as its equilibrium distribution. The state of the chain after a number of steps is then used as a sample of the desired distribution. The quality of the sample improves as a function of the number of steps.

**13. What is simulated annealing?**

Simulated annealing is a method for solving unconstrained and bound-constrained optimization problems. The method models the physical process of heating a material and then slowly lowering the temperature to decrease defects, thus minimizing the system energy.

**14. What are the two types of graphical models?**

Generally, probabilistic graphical models use a graph-based representation as the foundation for encoding a complete distribution over a multi-dimensional space and a graph that is a compact or factorized representation of a set of independences that hold in the

specific distribution. Two branches of graphical representations of distributions are commonly used, namely, Bayesian networks and Markov random fields. Both families encompass the properties of factorization and independences, but they differ in the set of independences they can encode and the factorization of the distribution that they induce

**15. What is Markov Random Fields**.?

A Markov Random Field (MRF) is a graphical model of a joint probability distribution. It consists of an undirected graph in which the nodes represent random variables

**16. What is hidden Markovmodel?**

Hidden Markov Model (HMM) is a statisticalMarkov model in which the system being modeled is assumed to be a Markov process with unobserved (i.e. *hidden*) states.The hidden Markov model can be represented as the simplest dynamic Bayesian network.

**17. What is HMM Viterbi algorithm?**

The **Viterbi Algorithm** is a dynamic programmingalgorithm for finding the most likely sequence of hidden states – called the **Viterbi path** – that results in a sequence of observed events, especially in the context of Markov information sources and hidden Markov models.

**18. What are the two methods of tracking?**

- Kalman filter
- Particle filter

**19. What is Kalman filter?**

A Kalman filter is an optimal estimator- ie infers parameters of interest from indirect, inaccurate and uncertain observations. It is recursive so that new measurements can be processed as they arrive. (cf batch processing where all data must be present).

**20. What is particle filters?**

Particle filters or Sequential Monte Carlo (SMC) methods are a set of genetic, Monte Carlo algorithms used to solve filtering problems arising in signal processing and Bayesian

statistical inference. The filtering problem consists of estimating the internal states in dynamical systems when partial observations are made, and random perturbations are present in the sensors as well as in the dynamical system.

## PART-B

1. Explain sampling in MCMC.
2. Explain Monte Carlo or Bust in detail.
3. Explain the proposal distribution in detail.
4. Explain Metropolis-Hastings in detail.
5 Explain Bayesian network in detail.
6. Explain Markov random field in detail.
7. Explain HMM-Viterbi algorithm in detail.
8. Explain HMM Baum-Welch (Forward-Backward) algorithm in detail.
9. Explain the particle filter in detail
10. Explain Kalman filter algorithm in detail.