MUTHAYAMMAL ENGINEERING COLLEGE
(An Autonomous Institution)

(Approved by AICTE, New Delhi, Accredited by NAAC & Affiliated to Anna University)
Rasipuram - 637 408, Namakkal Dist., Tamil Nadu

MKC

MUST KNOW CONCEPTS

CSE

2021-2022

Course Code & Course Name : 19CSC29 /Machine Learning Techniques

Year/Sem/Sec :III/V/A

S. No.	Term	Notation (Symbol)	Concept/Definition/Meaning/Units/Equation/ Expression	Units
UNIT - I INTRODUCTION TO SUPERVISED LEARNING				
1	Machine Learning		Machine learning is an application of AI which deals with system programming in order to automatically learn and improve with experience without being explicitly programmed. Eg: Robots	
2	Types of machine learning		Supervised learning, Unsupervised Learning & Reinforcement learning.	
3	Supervised Learning		Learn from trained data and predict output for new input.	
4	Unsupervised Learning		Predict output from hidden pattern without any external trained data.	
5	Reinforcement Learning		Learner is a decision-making agent that takes actions in an environment and receives reward (or penalty) for its actions in trying to solve a problem .	
6	Types of Supervised learning		Classification & Regression	
7	Classification		Classification is Supervised learning technique to categorize into a desired and distinct number of classes. example: Male and Female	
8	Regression		A regression problem is the output variable is a real or continuous value, such as “salary” or “weight”	
9	Example for Classification		pattern recognition, optical character recognition, face recognition, medical diagnosis, speech recognition & Biometrics etc.	
10	Noise		Machine learning techniques often have to deal with noisy data, which may affect the accuracy of the resulting data models. Therefore, effectively dealing with noise is a key aspect in supervised learning to obtain reliable models from data.	
11	Multi class in supervised learning		Multiclass classification is a classification task that consists of more than two classes	
12	Model Selection		Model selection is the process of selecting one final machine learning model from among a collection of candidate machine learning models for a training dataset. Model selection is a process that can be applied both across different types of models (e.g. logistic regression, SVM, KNN, etc.)	

13	Generalization		Generalization refers to your model's ability to adapt properly to new, previously unseen data, drawn from the same distribution as the one used to create the model.	
14	Bayesian machine learning		We can think of machine learning as learning models of data. The Bayesian framework for machine learning states that you start out by enumerating all reasonable models of the data and assigning your prior belief P(M) to each of these models.	
15	Decision tree is used in machine learning		Decision Trees are a non-parametric supervised learning method used for both classification and regression tasks. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features.	
16	Bayesian decision theory		Bayesian decision theory refers to a decision theory which is informed by Bayesian probability. It is a statistical system that tries to quantify the tradeoff between various decisions, making use of probabilities and costs.	
17	Over fitting		Trained set which are trained with lot of data and produce inaccurate output by the noise.	
18	Under fitting		Trained set which have less number of data and not used to generalize a new data.	
19	posterior probability		posterior = prior × likelihood / evidence	
20	Bias		Difference between predicted output and actual value.	
21	Variance		Used to analyse linear relationship between two variables.	
22	Covariance		Used to analyse difference between two attributes.	
23	ill-posed problem		ill-posed problem where the data byitself is not sufficient to find a unique solution.	
24	Generalization		A model trained on the training set predicts the right output for new instances is called generalization	
25	Validation set		validation set and is used to test the generalization ability	
UNIT II PARAMETRIC AND SEMI-PARAMMMETRIC METHODS				
26	Parametric classification		Parametric methods, like Discriminant Analysis Classification, fit a parametric model to the training data and interpolate to classify test data. Nonparametric methods, like classification and regression trees, use other means to determine classifications.	
27	Parametric regression		For example, polynomial regression consists of performing multiple regression with variables. in order to find the polynomial coefficients (parameters). These types of regression are known as parametric regression since they are based on models that require the estimation of a finite number of parameters.	
28	Model complexity		Model complexity can be characterized by many things, and is a bit subjective. In machine learning, model complexity often refers to the number of features or terms included in a given predictive model, as well as whether the chosen model is linear, nonlinear, and so on	
29	Parametric models		parametric models, which are well-defined in the finite-dimensional space, and	
30	Non-parametric models		Non-parametric models, where the parameters can all span an infinite space, a semi parametric model has a component that is finite-dimensional (i.e. it's easy to research and understand), and another that is infinite	

31	model selection		Machine learning model selection is the second step of the machine learning process, following variable selection and data cleansing. Selecting the right machine learning model is a critical step, as a model which does not appropriately fit the data will yield inaccurate results	
32	Parameter estimates		Parameter estimates (also called coefficients) are the change in the response associated with a one-unit change of the predictor, all other predictors being held constant.	
33	Multivariate Regression		Multivariate Regression is a method used to measure the degree at which more than one independent variable (predictors) and more than one dependent variable (responses), are linearly related.	
34	Binary classification		Binary classification is the task of classifying the elements of a set into two groups on the basis of a classification rule.	
35	Clustering		Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group and dissimilar to the data points in other groups	
36	Types of clustering		Hierarchical clustering,K-Means clustering.	
37	K-means algorithm		If k is given, the K-means algorithm can be executed in the following steps: Partition of objects into k non-empty subsets. ... Compute the distances from each point and allot points to the cluster where the distance from the centroid is minimum	
38	Hierarchical Clustering.		Minimum distance clustering is also called as single linkage hierarchical clustering or nearest neighbor clustering	
39	Maximum likelihood estimation		Maximum likelihood estimation is a method that determines values for the parameters of a model.	
40	Bernoulli density		It describes a single trial of a Bernoulli experiment. A closed form of the probability density function of Bernoulli distribution is P (x) = p x (1 − p) 1 − x P(x) = p^{x}(1-p)^{1-x} P(x)=px(1−p)1−x	
41	prior distribution		It is a combination of the prior distribution and the likelihood function, which tells you what information is contained in your observed data (the “new evidence”).	
42	posterior distribution		The posterior distribution summarizes what you know after the data has been observed	
43	Independent variables		Independent variables (also referred to as Features) are the input for a process that is being analyzes.	
44	Dependent variables		Dependent variables are the output of the process	
45	least squares method		The least squares method is a statistical procedure to find the best fit for a set of data points by minimizing the sum of the offsets or residuals of points from the plotted curve	
46	Least squares regression		Least squares regression is used to predict the behavior of dependent variables	
47	Polynomial Regression		Polynomial Regression is a form of linear regression in which the relationship between the independent variable x and dependent variable y is modeled as an nth degree polynomial.	
48	Relative Squared Error		The relative squared error is relative to what it would have been if a simple predictor had been used. More specifically, this simple predictor is just the average of the actual values.	
49	Cross-validation		Cross-validation is a resampling procedure used to evaluate machine learning models on a limited data sample	
50	Regularization		Regularization is the process of adding information in order to solve an well-posed problem or to prevent overfitting.	

UNIT III ARTIFICIAL NEURAL NETWORKS				
51	Artificial neuron		An artificial neuron is a mathematical function conceived as a model of biological neurons, a neural network. ... Usually each input is separately weighted, and the sum is passed through a non-linear function known as an activation function or transfer function.	
52	Neural network learning		An artificial neural network learning algorithm, or neural network, or just neural net. , is a computational learning system that uses a network of functions to understand and translate a data input of one form into a desired output, usually in another form	
53	Perceptron		A Perceptron is an algorithm used for supervised learning of binary classifiers. Binary classifiers decide whether an input, usually represented by a series of vectors, belongs to a specific class	
54	Perceptron Learning Rule		Perceptron Learning Rule states that the algorithm would automatically learn the optimal weight coefficients	
55	Gradient descent		Gradient descent is a first-order iterative optimization algorithm for finding a local minimum of a differentiable function	
56	Delta rule		The Delta rule in machine learning and neural network environments is a specific type of backpropagation that helps to refine connectionist ML/AI networks, making connections between inputs and outputs with layers of artificial neurons.	
57	Multilayer networks		Multilayer networks solve the classification problem for non linear sets by employing hidden layers, whose neurons are not directly connected to the output	
58	Backpropagation algorithm		The Backpropagation algorithm looks for the minimum value of the error function in weight space using a technique called the delta rule or gradient descent.	
59	Gradient descent		Gradient descent is an optimization algorithm used to minimize some function by iteratively moving in the direction of steepest descent as defined by the negative of the gradient. In machine learning, we use gradient descent to update the parameters of our mode	
60	Multilayer networks		Multilayer networks solve the classification problem for non linear sets by employing hidden layers, whose neurons are not directly connected to the output.	
61	Multilayer perceptron		A multilayer perceptron (MLP) is a class of feedforward artificial neural network (ANN). MLP utilizes a supervised learning technique called backpropagation for training. Its multiple layers and non-linear activation distinguish MLP from a linear perceptron. It can distinguish data that is not linearly separable.	
62	Activation function		The activation function also helps the perceptron to learn, when it is part of a multilayer perceptron (MLP). Certain properties of the activation function, especially its non-linear nature, make it possible to train complex neural networks.	
63	Representation on of Neural Networks		The connections between the different neurons are represented by the edge connecting two nodes in the graph representation of the artificial neural network. They are called weights and are typically represented as wij. The weights on a neural network is the particular case of the parameters on any parametric model	
64	Threshold unit		A linear threshold unit is a simple artificial neuron whose output is its thresholded total net input. That is, an LTU with threshold T calculates the weighted sum of its inputs, and then outputs 0 if this sum is less than T, and 1 if the sum is	

			greater than T	
65	Need of Backpropagation		Backpropagation simplifies the network structure by removing weighted links that have a minimal effect on the trained network. It is especially useful for deep neural networks working on error-prone projects, such as image or speech recognition.	
66	Difference between Cost and Loss function		The terms cost and loss functions almost refer to the same meaning.The cost function is calculated as an average of loss functions. The loss function is a value which is calculated at every instance. So, for a single training cycle loss is calculated numerous times, but the cost function is only calculated once	
67	Error-Correction Learning		Error-Correction Learning, used with supervised learning, is the technique of comparing the system output to the desired output value, and using that error to direct the training.	
68	Difference between neuron and Perceptron		is that neuron is (cytology) a cell of the nervous system, which conducts nerve impulses; consisting of an axon and several dendrites neurons are connected by synapses while perceptron is an element, analogous to a neuron, of an artificial neural network consisting of one or more layers of artificial neurons.	
69	Perceptron algorithms		Perceptron algorithms can be categorized into single-layer and multi-layer perceptrons. The single-layer type organizes neurons in a single layer while the multi-layer type arranges neurons in multiple layers	
70	Problem in Neural Network		If you accept most classes of problems can be reduced to functions, this statement implies a neural network	
71	Advantages of Neural Network		Neural Networks have the ability to learn by themselves and produce the output that is not limited to the input provided to them. The input is stored in its own networks instead of a database, hence the loss of data does not affect its working.	
72	Applications of Neural Network		Neural networks can be used to recognize handwritten characters. Image Compression - Neural networks can receive and process vast amounts of information at once, making them useful in image compression	
73	prevent Overfitting in a neural network		• Early Stopping:Early stopping is a form of regularization while training a model with an iterative method, such as gradient descent. • Use Data Augmentation • Use Regularization • Use Dropouts	
74	Types of Neural Network		Feedforward Neural Network – Artificial Neuron Radial basis function Neural Network Kohonen Self Organizing Neural Network Recurrent Neural Network(RNN) – Long Short Term Memory Convolutional Neural Network,Modular Neural Network	
75	Recurrent neural networks		Recurrent neural networks (RNN) are the state of the art algorithm for sequential data and are used by Apple's Siri and and Google's voice search. It is the first algorithm that remembers its input, due to an internal memory, which makes it perfectly suited for machine learning problems that involve sequential data.	
UNIT IV INSTANCE BASED LEARNING				
76	Instance-based		Definition. Instance-based learning refers to a family of techniques for classification and regression, which produce	

	learning		a class label/predication based on the similarity of the query to its nearest neighbor(s) in the training set.	
77	Why instance based learning is called as lazy learning		Instance-based learning includes nearest neighbor, locally weighted regression and case-based reasoning methods. Instance-based methods are sometimes referred to as lazy learning methods because they delay processing until a new instance must be classified.	
78	lazy learner technique		A lazy learner simply stores the training data and only when it sees a test tuple starts generalization to classify the tuple based on its similarity to the stored training tuples	
79	Lazy algorithm		A lazy learning algorithm is simply an algorithm where the algorithm generalizes the data after a query is made. The best example for this is KNN	
80	Why KNN algorithm is used		KNN algorithm is one of the simplest classification algorithm and it is one of the most used learning algorithms.KNN is a non-parametric, lazy learning algorithm	
81	Why KNN is a lazy learner		K-NN is a lazy learner because it doesn't learn a discriminative function from the training data but “memorizes” the training dataset instead.	
82	What does K mean in kNN		'k' in KNN is a parameter that refers to the number of nearest neighbours to include in the majority of the voting process.	
83	Is K means supervised or unsupervised		The 'K' in K-Means Clustering has nothing to do with the 'K' in KNN algorithm. k-Means Clustering is an unsupervised learning algorithm that is used for clustering whereas KNN is a supervised learning algorithm used for classification	
84	What is nearest Neighbour rule		Nearest Neighbor Rule selects the class for x with the assumption that: If x' and x were overlapping (at the same point), they would share the same class	
85	Which is a disadvantage of KNN		The main disadvantage of the KNN algorithm is that it is a lazy learner, i.e. it does not learn anything from the training data and simply uses the training data itself for classification.	
86	What are advantages of KNN		The main advantages of kNN for classification are: Very simple implementation. Robust with regard to the search space; for instance, classes don't have to be linearly separable. Classifier can be updated online at very little cost as new instances with known classes are presented	
87	K Nearest Neighbor algorithm in machine learning		The k-nearest neighbors (KNN) algorithm is a simple, supervised machine learning algorithm that can be used to solve both classification and regression problems. It's easy to implement and understand, but has a major drawback of becoming significantly slows as the size of that data in use grows	
88	Locally weighted regression		Locally weighted regression (LWR) attempts to fit the training data only in a region around the location of a query example. LWR is a type of lazy learning, therefore the processing of training data is often postponed until the	

			target value of a query example needs to be predicted.	
89	weighted kNN		In weighted kNN, the nearest k points are given a weight using a function called as the kernel function	
90	Remarks on Locally weighted regression		• broad range of methods for distance weighting the training examples • range of methods for locally approximating target functions	
91	Radial basis functions		Radial basis functions are means to approximate multivariable (also called multivariate) functions by linear combinations of terms based on a single univariate function (the radial basis function). This is radialised so that in can be used in more than one dimension	
92	Case-based reasoning		Case-based reasoning (CBR) is a paradigm of artificial intelligence and cognitive science that models the reasoning process as primarily memory based. Case-based reasoners solve new problems by retrieving stored 'cases' describing similar prior problem-solving episodes and adapting their solutions to fit new needs	
93	Lazy Learning		A lazy learning algorithm is simply an algorithm where the algorithm generalizes the data after a query is made. The best example for this is KNN.	
94	Eager Learning		In artificial intelligence, eager learning is a learning method in which the system tries to construct a general, input-independent target function during training of the system, as opposed to lazy learning, where generalization beyond the training data is delayed until a query is made to the system	
95	Euclidean distance		The Euclidean distance between two points in either the plane or 3-dimensional space measures the length of a segment connecting the two points. It is the most obvious way of representing distance between two points	
96	Why do we use Euclidean distance		Euclidean Distance gives the distance from each cell in the raster to the closest source	
97	Why Euclidean distance is used in Knn		Usually, the Euclidean distance is used as the distance metric. Then, it assigns the point to the class among its k nearest neighbours (where k is an integer).	
98	Radial basis function		An RBF is a function that changes with distance from a location. For example, suppose the radial basis function is simply the distance from each location, so it forms an inverted cone over each location	
99	What is Gaussian radial basis function		A radial basis function (RBF) is a real-valued function whose value depends only on the distance between the input and some fixed point, either the origin, so that , or some other fixed point , called a center, so that .	

100	Can Knn be used for prediction		KNN algorithm can be used for both classification and regression problems. The KNN algorithm uses 'feature similarity' to predict the values of any new data points. ... The average of the values is taken to be the final prediction	
UNIT V ADVANCED LEARNING				
101	Bayesian network		A *Bayesian network* is a probabilistic graphical model that represents a set of variables and their conditional dependencies via a directed acyclic graph (DAG).	
102	*Directed Acyclic Graph* (*DAG*)		In computer science and mathematics, a DAG is a *graph* that is *directed* and without cycles connecting the other edges.	
103	causal graph		A *causal graph* will depict whatever your assumptions that you're making about the relationship between these variables.	
104	Conditional Independence		Two events A and B are conditionally independent given an event C with P(C)>0 if P(A∩B|C)=P(A|C)P(B|C)(1.8) Recall that from the definition of conditional probability, P(A|B)=P(A∩B)P(B), if P(B)>0.	
105	Diagnostic Inference		Diagnostic or bottom-up inference.	
106	Probabilistic Database		A probabilistic database is an uncertain database in which the possible worlds have associated probabilities.	
107	Hidden Variables		Confounding, in statistics, an extraneous variable in a statistical model that correlates (directly or inversely) with both the dependent variable and the independent variable.	
108	Direct Influence		Direct influence means that we can take specific steps to try to get the thing done.	
109	Multinomial Variable		Multinomial logistic regression is used to predict a nominal dependent variable given one or more independent variables.	
110	Generative Model		A Generative Model is a powerful way of learning any kind of data distribution using unsupervised learning and it has achieved tremendous success in just few years.	
111	Phylogenetic Tree		A phylogenetic tree is a diagram that represents evolutionary relationships among organisms. Phylogenetic trees are hypotheses, not definitive facts.	
112	Hidden Markov Model (HMM)		Hidden Markov models (HMMs) have proven to be one of the most widely used tools for learning probabilistic models of time series data.	
113	Kalman Filter		A Kalman Filter can be applied to take in the GPS data from the car, however GPS devices are not always entirely accurate.	
114	Bayes' ball		*Bayes ball* is an efficient algorithm for computing d-separation by passing simple messages between nodes of the graph.	
115	Junction Trees		The *junction tree* algorithm (also known as 'Clique *Tree*') is a method used in machine learning to extract marginalization in general graphs.	
116	Markov random		A Markov Random Field is a graphical model of a joint probability distribution.	

	field			
117	Maximal Clique		A maximal clique is a clique that cannot be extended by including one more adjacent vertex, meaning it is not a subset of a larger clique.	
118	Factor Graph		A factor graph is a type of probabilistic graphical model.	
119	Sum-Product Algorithm		*Sum-product algorithm*, which operates in a factor graph and at- tempts to compute various marginal functions associated with the global function.	
120	Max-Product Algorithm		*Max-product* is a standard belief propagation *algorithm* on factor graph models.	
121	Decision Node		A decision node is a node in an activity at which the flow branches into several optional flows.	
122	Sensor Fusion		where the data from different sensors are integrated to extract more information for a specific application.	
123	*Random Subspace*		The *random subspace* method for constructing decision forests.	
124	Error-Correcting Output Codes		ECOC is an ensemble method designed for multi-class classification problem.	
125	Bayesian network		A *Bayesian network* is a probabilistic graphical model that represents a set of variables and their conditional dependencies via a directed acyclic graph (DAG).	
GATE QUESTIONS				
126	Multiple Expert		*Multiple Expert* classification methods rely on a large training dataset in order to be properly utilized.	
127	Ensemble		An ensemble is itself a supervised learning algorithm, because it can be trained and then used to make predictions.	
128	Linear Opinion Pools		An important question when eliciting opinions from experts is how to aggregate the reported opinions.	
129	Hamming distance		Hamming distance is a metric for comparing two binary data strings.	
130	Bagging		Bagging is used when the goal is to reduce the variance of a decision tree classifier.	
131	*Boosting*		The term '*Boosting*' refers to a family of algorithms which converts weak learner to strong learners.	
132	AdaBoost		AdaBoost is an ensemble learning method (also known as “meta-learning”) which was initially created to increase the efficiency of binary classifiers.	
133	Decision Stump		A decision stump is a machine learning model consisting of a one-level decision tree.	
134	Mixture of experts		*Mixture of experts* refers to a *machine learning* technique where multiple experts (learners) are used to divide the problem space into homogeneous regions.	
135	Dynamic Classifier Selection		Dynamic Classifier Selection based on Multiple Classifier. Ensembles using Accuracy and Diversity. Measure accuracy and diversity.	
136	Stacked Generaliza		Stacked generalization, a scheme for minimizing the generalization error rate of one or more generalizers.	

	tion			
137	Cascading		Cascading is a multistage method	
138	Spoofing		Spoofing is the act of disguising a communication from an unknown source as being from a known, trusted source.	
139	Multiple kernel learning		*Multiple kernel learning* (MKL) algorithms aim to find the best convex combination of a set of kernels to form the best classifier.	
140	k-armed bandit		In the classical *k-armed bandit* problem, there are k alternative arms, each with a stochastic reward whose probability distribution is initially unknown.	
141	Markov decision process		Markov decision process (MDP) is a discrete-time stochastic control process.	
142	finite-horizon		A stopping rule problem has a finite horizon if there is a known upper bound on. the number of stages at which one may stop.	
143	infinite-horizon		*Infinite horizon* problems are further characterized by the fact that the number of stages N is infinite.	
144	Optimal Policy		An Optimal Policy is a policy where you are always choosing the. action that maximizes the “return”/”utility” of the current state.	
145	Bellman’s equation		The *Bellman Equations*. Step-by-step derivation, explanation, and demystification of the most important equations in reinforcement learning.	
146	Value iteration		Value iteration is a method of computing an optimal MDP policy and its value. Value iteration starts at the "end" and then works backward, refining an estimate of either Q^* or V^*.	
147	Policy Iteration		In Policy Iteration - You randomly select a policy and find value function corresponding to it , then find a new policy based on the previous value function, and so on this will lead to optimal policy	
148	Temporal difference		*Temporal difference* (TD) learning is an approach to learning how to predict a quantity that depends on future values of a given signal.	
149	greedy search		A greedy search algorithm is an. algorithm that uses a heuristic for making locally optimal choices at each stage with the hope of finding a global optimum.	
150	Q-learning		Q-learning is an off policy reinforcement learning algorithm that seeks to find the best action to take given the current state.	

Faculty Team Prepared **Signatures**

1.

2. **HoD**