INSTITUTE OF AERONAUTICAL ENGINEERING
(Autonomous)
Dundigal, Hyderabad - 500 043

COMPUTER SCIENCE AND ENGINEERING
DEFINITION AND TERMINOLGY

Course Title	**FOUNDATIONS OF MACHINE LEARNING**				
Course Code	ACAC03				
Program	B.Tech				
Semester	IV	CSE (AI & ML)			
Course Type	Core				
Regulation	IARE - UG 20				
Course Structure	Theory			Practical	
	Lecture	Tutorials	Credits	Laboratory	Credits
	3	0	3	1	2
Course Coordinator	Dr. Shaik Jakeer Hussain, Associate Professor				

COURSE OBJECTIVES:
The students will try to learn:

I	The underlying mathematical principles from probability, linear algebra and optimization.
II	The knowledge of using machine learning to make predictions in a scientific computing environment.
III	The knowledge on machine learning algorithms correlated with paradigms of supervised and un supervised learning.
IV	The knowledge advanced supervised learning applications in machine learning, such as Neural Networks, Support vector machines etc.

COURSE OUTCOMES:
After successful completion of the course, students should be able to:

CO 1	**Understand** the need for Machine Learning, various learning tasks, and statistical learning framework	Understand
CO 2	**Make use of** different supervised learning algorithms to solve data classification problems.	Apply
CO 3	**Apply the** Ensemble and Probabilistic learning techniques to combine the predictions from two or more models.	Apply

CO 4	**Acquire the knowledge** about different unsupervised learning algorithms for clustering of the data.	Apply
CO 5	**Discuss** the advanced supervised learning techniques to solve the classification problems.	Apply
CO 6	**Apply the algorithms** to a real problem, optimize the models learned, and evaluate their performance efficiency	Apply

DEFINITION AND TERMINOLOGY:

S.No	DEFINITIONS	CO's
	MODULE I	
	INTRODUCTION TO MACHINE LEARNING	
1	**Define Machine Learning ?** Machine learning algorithms build a model based on sample data, known as training data, in order to make predictions or decisions without being explicitly programmed. machine learning algorithms are used in a wide variety of applications, such as email filtering and computer vision, where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks.	CO 1
2	**Define Classification?** Classification is to assign a category to each item. Applications include document classification, text classification, image classification where number of categories are small. Other applications where there are large or unbounded categories are optical character recognition and speech recognition.	CO 1
3	**What are Features?** It is a set of data attributes often represented as a vector associated to an example. Feature extraction from examples is domain specific task done by the experts, and is critical to the successful prediction. If an example has N attributes and all of them can be represented by real numbers, then the feature set X is equal to R .	CO 1
4	**What are Labels?** Values of categories assigned to examples. Labels are discrete for classification and real valued for regression. The set of labels is typically denoted by Y	CO 1
5	**What is a Sample? What are different kinds of Samples?** Sample is a collection of examples together with their labels is called a sample. There are three kinds of samples.Training sample, examples or samples to train a learning algorithm. Validation sample, samples to tune the free parameters of the learning algorithm. Test sample, samples to evaluate the performance of the learning algorithm.	CO 1
6	**What is a Loss Function?**	CO 1

	Loss function is used to measures the difference or loss between predicted and the true label. Set of predictions are denoted by Y not necessarily equal to the set of labels Y.	
7	**What is a Hypothesis set?** Hypothesis set is the set H is a subset of Y, X of functions mapping features to the set of labels.	CO 1
8	**What are the Learning stages of a given sample?** Randomly partition into training, validation, and test sample. Associate features to examples. Fix free learning parameters and pick a hypothesis. Pick the hypothesis with best performance on validation sample. Predict labels of the test examples Evaluated the algorithm using the test labels.	CO 1
9	**What is Supervised Learning** The learner receives a sample for training and validation, and makes prediction for all unseen points. This is common scenarios for classification, regression, and ranking.	CO 1
10	**Define PAC Learning** In computational learning theory, probably approximately correct (PAC) learning is a framework for mathematical analysis of machine learning. In this framework, the learner receives samples and must select a generalization function called the hypothesis.	CO 1
11	**What is Unsupervised Learning?** The learner receives unlabeled examples for training and makes predictions for all unseen points. Difficult to quantitatively evaluate the performance of a learner. Clustering and dimensionality reduction are examples of unsupervised learning	CO 1
12	**What is Active Learning?** The learner adaptively or interactively collects training samples by querying an oracle for new samples. The goal is to achieve comparable performance to the supervised learning with fewer samples.	CO 1
13	**What is Re-inforced Learning?** The learner actively interacts with the environment and receives an immediate reward for each action. The objective is to maximize reward over a course of actions and iterations with the environment.	CO 1

14	**What is Transductive Inference?** The learner receives a labeled training sample along with a set of unlabeled test points, and make predictions for only these test points	CO 1
15	**What is Online Learning?** At each round, the learner receives an unlabeled training example, makes a prediction, receives the true label, and incurs a loss. The objective is to minimize the cumulative loss over all rounds.	CO 1
16	**What is a Support Vector Machine?** Support Vector Machines are a set of related Supervised Learning Methods used for Classification and Regression. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that predicts whether a new example falls into one category or the other	CO 1
17	**What is Cluster Analysis?** Cluster Analysis is the assignment of a set of observations into subsets called clusters so that observations within the same cluster are similar according to some predesignated criterion or criteria, while observations drawn from different clusters are dissimilar.	CO 1
18	**What is a Genetic Algorithm?** A genetic algorithm is a search heuristic that mimics the process of natural selection, and uses methods such as mutation and crossover to generate new genotype in the hope of finding good solutions to a given problem.	CO 1
19	**What is Artificial Intelligence?** Artificial Intelligence is the intelligence exhibited by machines or software. It is also the name of the academic field of study which studies how to create computers and computer software that are capable of intelligent behavior. It is also defined as the study and design of intelligent agents.	CO 1
20	**What is Entropy?** The entropy, H , of a discrete random variable X intuitively is a measure of the amount of uncertainty associated with the value of X when only its distribution is known	CO 1
MODULE II		
SUPERVISED LEARNING ALGORITHMS		
1	**What is Exploratory Data Analysis?** Exploratory data analysis is an approach to analyzing data sets to summarize their main characteristics, often with visual methods. A statistical model can be used or not, but primarily EDA is for seeing what the data can tell us beyond the formal modeling or hypothesis testing task .	CO 2

2	**What is Computational Science?** Computational science also scientific computing or scientific computation is concerned with constructing mathematical models and quantitative analysis techniques and using computers to analyze and solve scientific problems.	CO 1
3	**What is Binary Entropy Function?** The entropy for a random variable with two outcomes is the binary entropy function, usually taken to the logarithmic base 2, thus have the shannon as unit	CO 2
4	**What is Predictive Analysis?** Predictive analytics is a variety of statistical techniques from modeling, machine learning, and data mining that analyze current and historical facts to make predictions about future, or otherwise unknown, events.In business, predictive models exploit patterns found in historical and transactional data to identify risks and opportunities. Models capture relationships among many factors to allow assessment of risk or potential associated with a particular set of conditions, guiding decision making for candidate transactions.	CO 3
5	**What is a Time Series Model?** Time series models are used for predicting or forecasting the future behavior of variables. These models account for the fact that data points taken over time may have an internal structure (such as autocorrelation, trend or seasonal variation) that should be accounted for	CO 2
6	**What is a Classification and Regression Tree?** Classification and regression trees are a non-parametric decision tree learning technique that produces either classification or regression trees, depending on whether the dependent variable is categorical or numeric, respectively.	CO 2
7	**What is a Knot?** Knot is where one local regression model gives way to another and thus is the point of intersection between two splines. In multivariate and adaptive regression splines, basis functions are the tool used for generalizing the search for knots	CO 3
8	**What is Linear Modelling in Classification?** Linear modelling in a classification context consists of regression followed by a transformation to return a categorical output and thereby producing a decision boundary. Really there isn't much to the model which makes diagnosis somewhat simple.	CO 2

9	**Define Multi-class Classification** Multi-class classification is the classification technique that allows us to categorize the test data into multiple class labels present in trained data as a model prediction.	CO 3
10	**Define Multi-Label Classification** Multilabel classification is used when there are two or more classes and the data we want to classify may belong to none of the classes or all of them at the same time, e.g. to classify which traffic signs are contained on an image.	CO 2
11	**Define hypothesis** Hypothesis in Machine Learning is used when in a Supervised Machine Learning, we need to find the function that best maps input to output. This can also be called function approximation because we are approximating a target function that best maps feature to the target.	CO 2
12	**Define Generalization** Generalization is a term used to describe a model's ability to react to new data. That is, after being trained on a training set, a model can digest new data and make accurate predictions. A model's ability to generalize is central to the success of a model.	CO 2
13	**Define Most Specific Hypothesis** A hypothesis, h, is a most specific hypothesis if it covers none of the negative examples and there is no other hypothesis h that covers no negative examples, such that h is strictly more general than h.	CO 3
14	**What is Ada Boosting?** AdaBoost is the first stepping stone in the world of Boosting. AdaBoost is one of the first boosting algorithms to be adapted in solving practices. Adaboost helps you combine multiple weak classifiers into a single strong classifier.	CO 2
15	**What is Bagging ?** Bootstrap Aggregating, also knows as bagging, is a machine learning ensemble meta-algorithm designed to improve the stability and accuracy of machine learning algorithms used in statistical classification and regression. It decreases the variance and helps to avoid overfitting. It is usually applied to decision tree methods.	CO 2
16	**Define Occam's razor?** A scientific and philosophical rule that entities should not be multiplied unnecessarily which is interpreted as requiring that the simplest of competing theories be preferred to the more complex or that explanations of unknown phenomena be sought first in terms of known quantities	CO 2

17	**Define Regression** Simple Linear Regression is a type of Regression algorithms that models the relationship between a dependent variable and a single independent variable. The relationship shown by a Simple Linear Regression model is linear or a sloped straight line, hence it is called Simple Linear Regression.	CO 2
18	**Define Interpolation** Interpolation is a method of deriving a simple function from the given discrete data set such that the function passes through the provided data points. This helps to determine the data points in between the given data ones. This method is always needed to compute the value of a function for an intermediate value of the independent function.	CO 2
19	**Define Extrapolation** In terms of machine learning, extrapolation can be thought of as being trained on a certain range of data and being able to predict on a different range of data. This may be easy with simple patterns, such as simple positive number or negative number	CO 3
20	**Define ill-posed problem** A problem which may have more than one solution, or in which the solutions depend discontinuously upon the initial data	CO 3
MODULE III		
ENSEMBLE AND PROBABILISTIC LEARNING		
1	**Define Inductive Bias** The inductive bias also known as learning bias of a learning algorithm is the set of assumptions that the learner uses to predict outputs given inputs that it has not encountered. In machine learning, one aims to construct algorithms that are able to learn to predict a certain target output.	CO 3
2	**Define Model Selection** Model selection is the task of selecting a statistical model from a set of candidate models, given data. In the simplest cases, a pre existing set of data is considered. However, the task can also involve the design of experiments such that the data collected is well-suited to the problem of model selection.	CO 3
3	**Define Underfitting** Underfitting is a statistical model or a machine learning algorithm is said to have underfitting when it cannot capture the underlying trend of the data. It's just like trying to fit undersized pants, Underfitting destroys the accuracy of our machine learning model.	CO 3

4	**Define Density Estimation?** Density Estimation is a structure to the input space such that certain patterns occur more often than others.	CO 3
5	**Define Overfitting** Overfitting is a modelling error which occurs when a function is too closely fit to a limited set of data points. Overfitting the model generally takes the form of making an overly complex model to explain idiosyncrasies in the data under study. In reality, the data often studied has some degree of error or random noise within it.	CO 3
6	**Define Validation set** In machine learning, a validation set is used to tune the parameters of a classifier. The validation test evaluates the program's capability according to the variation of parameters to see how it might function in successive testing. The validation set is also known as a validation data set, development set or dev set.	CO 3
7	**Define Cross-Validation** Cross validation is a technique that is used for the assessment of how the results of statistical analysis generalize to an independent data set. Cross validation is largely used in settings where the target is prediction and it is necessary to estimate the accuracy of the performance of a predictive model	CO 3
8	**Define Unobservable variables** Observed variables are variables for which you have measurements in your dataset, whereas unobserved or latent variables are variables for which you don't. When your analysis reveals correlations between observed variables, you might look for unobserved variables to explain the correlation, especially in cases where you doubt that there is a direct causal relationship between them. To offer a contrived example, suppose your dataset includes strongly correlated variables.	CO 3
9	**Define Bayes' rule** Bayes' rule allows us to compute the single term P (B or A) in terms of P (A or B), P (B), and P (A). This is very useful in cases where we have a good probability of these three terms and want to determine the fourth one.	CO 3
10	**Define Prior probability** Prior probability, in Bayesian statistical inference, is the probability of an event before new data is collected	CO 3
11	**Define Class Likelihood** A Likelihood function, on the other hand, takes the data set as a given, and represents the likeliness of different parameters for your distribution. X is equal to X 1 ,X 2 and so on is f of x or Theta, where Theta is a parameter. X is equal to x is an observed sample point. Then the function of Theta defined as is your likelihood function.	CO 3

12	**Define Expected Risk** The risk after considering agreed actions that have not yet been implemented Targeted Risk ,The desired optimal level of risk Existing Controls, Controls currently existing in the business Actions Agreed actions to further treat risk	CO 3
13	**Define Discriminant Functions** A discriminant is a function that takes an input vector x and assigns it to one of K classes, denoted by Ck. Discriminating between two classes is easy. We assign a data point to class C1 if y (x) is greater than or equal to 0 else to class C2.	CO 3
14	**Define Decision Regions** The decision regions are separated by surfaces called the decision boundaries. These separating surfaces represent points where there are ties between two or more categories. For a minimum distance classifier, the decision boundaries are the points that are equally distant from two or more of the templates.	CO 3
15	**Define Utility Theory** The main idea of utility theory is really simple: an agent's preferences over possible outcomes can be captured by a function that maps these outcomes to a real number; the higher the number the more that agent likes that outcome. The function is called a utility function.	CO 3
16	**Define Association Rules** Association rules are created by thoroughly analysing data and looking for frequent if or then patterns. Then, depending on the following two parameters, the important relationships are observed, Support indicates how frequently the if or then relationship appears in the database.	CO 3
17	**Define Basket Analysis** Market Basket Analysis is an example of an analytics technique employed by retailers to understand customer purchase behaviors. It is used to determine what items are frequently bought together or placed in the same basket by customers.	CO 3
18	**Define Apriori Algorithm** The Apriori algorithm is an algorithm that attempts to operate on database records, particularly transactional records, or records including certain numbers of fields or items.	CO 3
19	**Define Baye's Estimator** In estimation theory and decision theory a Bayes estimator or a Bayes action is an estimator or decision rule that minimizes the posterior expected value of a loss function .	CO 3

20	**Define Posterior Density** In machine learning, Maximum a Posteriori optimization provides a Bayesian probability framework for fitting model parameters to training data and an alternative and sibling to the perhaps more common Maximum Likelihood Estimation framework. Maximum a posteriori learning selects a single most likely hypothesis given the data.	CO 3
MODULE IV		
UNSUPERVISED LEARNING		
1	**Define Relative Square 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. Thus, the relative squared error takes the total squared error and normalizes it by dividing by the total squared error of the simple predictor.	CO 4
2	**Define Bias or Variance** The Bias variance dilemma or Bias Variance problem is the conflict in trying to simultaneously minimize these two sources of error that prevent supervised learning algorithms from generalizing beyond their training set,The Bias error is an error from erroneous assumptions in the learning algorithm.	CO 4
3	**Define Multi-Variate Data** Multi-Variate data is the data in which analysis are based on more than two variables per observation. Usually, multivariate data is used for explanatory purposes.	CO 4
4	**Define Mean Vector** The mean vector consists of the means of each variable and the variance and covariance matrix consists of the variances of the variables along the main diagonal and the covariances between each pair of variables in the other matrix positions	CO 4
5	**Define Co-Variance matrix** In statistics and probability theory, a square matrix provides the covariance between each pair of components or elements of a given random vector is called a covariance matrix. Any covariance matrix is symmetric and positive semidefinite.	CO 4
6	**Define Sample Mean** In sample, refers to the data that you have, and out of sample to the data you don't have but want to forecast or estimate. The data points used to build the model constitute in sample data whereas all the new data points not belonging to the training sample constitute out of sample data.	CO 4

7	**Define Imputation** It is a technique used for replacing the missing data with some substitute value to retain most of the data or information of the dataset.	CO 4
8	**Define Multi-variate Analysis** Multivariate analysis is a Statistical procedure for analysis of data involving more than one type of measurement or observation. It may also mean solving problems where more than one dependent variable is analyzed simultaneously with other variables.	
9	**Define Linear Discriminant** Linear discriminant functions is a function that is a linear combination of the components of x ,g of x is equal to sum of wtx and w 0 where w is the weight vector and w 0 is the bias	CO 4
10	**Define Naive Bayes Classifier** A classifier is a machine learning model that is used to discriminate different objects based on certain features. A Naive Bayes classifier is a probabilistic machine learning model that is used for classification task. The crux of the classifier is based on the Bayes theorem.	CO 4
11	**Define Nearest Mean Classifier** In machine learning, a nearest mean classifier or nearest prototype classifier is a classification model that assigns to observations the label of the class of training samples whose mean is closest to the observation.	CO 4
12	**Define Template Matching** Template Matching techniques compare portions of images against one another. Sample image may be used to recognize similar objects in source image. If standard deviation of the template image compared to the source image is small enough, template matching may be used.	CO 4
13	**Define Regularized Discriminant Analysis.** This operator performs a regularized discriminant analysis, for nominal labels and numerical attributes. Discriminant analysis is used to determine which variables discriminate between two or more naturally occurring groups, it may have a descriptive or a predictive objective.	CO 4
14	**Define Bayesian Spam Filtering** A Bayesian filter is a filter that learns your spam preferences. When you mark emails as spam, the system will note the characteristics of the email and look for similar characteristics in incoming email, filtering anything that fits the formula directly in to spam for you.	CO 4

15	**Define Multi-Variate Linear Regression** Multivariate linear regression is a natural generalization of the simple linear regression model is a situation including influence of more than one independent variable to the dependent variable, again with a linear relationship mathematically .	CO 4
16	**Define Principal Component Analysis** The central idea of principal component analysis is to reduce the dimensionality of a data set consisting of a large number of interrelated variables, while retaining as much as possible of the variation present in the data set.	CO 4
17	**Define Factor Analysis** It is a theory used in machine learning and related to data mining. The theory behind factor analytic methods is that the information gained about the interdependencies between observed variables can be used later to reduce the set of variables in a dataset.	CO 4
18	**Define Z- Normalization** Z score normalization refers to the process of normalizing every value in a dataset such that the mean of all of the values is 0 and the standard deviation is 1. We use the following formula to perform a z score normalization on every value in a dataset, New value is equal to x minus u divided by sigma	CO 4
19	**Define Tuning Complexity** Time and Space Complexity of a Turing Machine For a Turing machine, the time complexity refers to the measure of the number of times the tape moves when the machine is initialized for some input symbols and the space complexity is the number of cells of the tape written. Time complexity all reasonable functions minus T of n is equal to O of n log n	CO 4
20	**What is Dimensionality Reduction?** By reducing the number of input features, thereby reducing the number of dimensions in the feature space. Dimensionality reduction means reducing your feature set's dimension. Dimensionality reduction brings many advantages to your machine learning data	CO 4

MODULE V		
ADVANCED SUPERVISED LEARNING		
1	**What is Self Organizing Map ?** Self-organizing maps (SOM) are feed-forward networks that use an unsupervised learning approach through a process called self-organization. A Kohonen network consists of two layers of processing units called an input layer and an output layer. There are no hidden units.	CO 5
2	**Define Isomap ?** Isomap is a nonlinear dimensionality reduction method. It is one of several widely used low dimensional embedding methods.Isomap is used for computing a quasiisometric, lowdimensional embedding of a set of high dimensional data points	CO 5
3	**Define Locally Linear Embedding** An eigenvector method for solving the problem of non linear dimensionality reduction. The dimensionality reduction by LLE succeeds in identifying the underlying structure of the manifold.	CO 5
4	**What is Gaussian Mixture Model in Machine Learning?** In machine learning, this is known as Clustering. In real life, many datasets can be modeled by Gaussian Distribution (Univariate or Multivariate). So it is quite natural and intuitive to assume that the clusters come from different Gaussian Distributions.	CO 5
5	**Define k-Means Clustering** K-Means Clustering is an Unsupervised Learning algorithm, which groups the unlabeled dataset into different clusters. Here K defines the number of pre-defined clusters that need to be created in the process, as if K is equal to 2, there will be two clusters, and for K is equal to 3, there will be three clusters.	CO 5
6	**Define Color Quantization** Color Quantization is a method of reducing the number of colors required to represent an image. For example, converting a photograph to GIF format requires the number of colors to be reduced to 256	CO 5
7	**What is Vector Quantization** Vector quantization is used for lossy data compression, lossy data correction, pattern recognition, density estimation and clustering. Lossy data correction, or prediction, is used to recover data missing from some dimensions.	CO 5
8	**Define Re-construction Error** The general definition of the reconstruction error would be the distance between the original data point and its projection onto a lower dimensional subspace	CO 5

9	**Define Expectation Maximization algorithm** The essence of Expectation Maximization algorithm is to use the available observed data of the dataset to estimate the missing data and then using that data to update the values of the parameters.	CO 5
10	**Define Gaussian Mixtures** A Gaussian mixture model is a category of probabilistic model which states that all generated data points are derived from a mixture of a finite Gaussian distributions that has no known parameters.	CO 5
11	**Define Hierarchical Clustering** Hierarchical clustering is separating data into groups based on some measure of similarity, finding a way to measure how they are alike and different, and further narrowing down the data.	CO 5
12	**Define Agglomerative Clustering** The agglomerative clustering is the most common type of hierarchical clustering used to group objects in clusters based on their similarity. The algorithm starts by treating each object as a singleton cluster. Next, pairs of clusters are successively merged until all clusters have been merged into one big cluster containing all objects.	CO 5
13	**Define Divisive Clustering** The Divisive Clustering Algorithm is a top down clustering approach, initially, all the points in the dataset belong to one cluster and split is performed recursively as one moves down the hierarchy. Initially, all points in the dataset belong to one single cluster.	CO 5
14	**Define Single link Clustering** Single linkage clustering is one of several methods of hierarchical clustering. It is based on grouping clusters in bottom up fashion , at each step combining two clusters that contain the closest pair of elements not yet belonging to the same cluster as each other.	CO 5
15	**Define Dendrogram.** A Dendrogram is a diagram representing a tree, in Hierarchical Clustering, it illustrates the arrangement of the clusters produced by the corresponding analysis.	CO 5
16	**Define Multi Dimensonal Scaling?** Multi-dimensional scaling is a tool by which to quantify similarity judgments. Formally, MDS refers to a set of statistical procedures used for exploratory data analysis and dimension reduction	CO 5
17	**Define Subset Selection?** This is a naive approach that essentially tries to find the best model among 2 p models that are trained on all possible subsets of the p variables.	CO 5

18	**Define Structural Risk Minimization ?**	CO 5
	Structural Risk Minimization is an inductive principle of use in machine learning. Commonly in machine learning, a generalized model must be selected from a finite data set, with the problem of overfitting, the model becomes too strongly tailored to the particularities of the training set and poorly generalized to new data.	
19	**Define Least Squares Estimate?**	CO 5
	The Least Square Method is the process of finding the best-fitting curve or line of best fit for a set of data points by reducing the sum of the squares of the offsets residual part of the points from the curve. During the process of finding the relation between two variables, the tre	
20	**Define Bernouli Density ?**	CO 5
	Bernoulli Naive Bayes is one of the variants of the Naive Bayes algorithm in machine learning. It is very useful to be used when the dataset is in a binary distribution where the output label is either present or absent.	

Course Coordinator:
Ms. V Jalaja, Assistant Professor

HOD, CSE (AI & ML)