AMITY UNIVERSITY
KOLKATA

Course Title: Fundamentals of Machine Learning
Course Level: UG
Course Code: CSE313
Credit Units: 4
Course Objectives:

L	T	P/S	SW/FW	TOTAL CREDIT UNITS
2	0	2	2	4

The course aims to introduce the concepts, theories and state-of-the-art algorithms for machine learning. The course deals with theories and practical aspects of machine learning techniques, consists of regression, clustering, classification.

Pre-requisites: Algorithms, Data Structures, Linear algebra, Probability and statistics

Student Learning Outcomes:

At the end of the course, students should be able to

- Understand the concept and application areas of Machine Learning algorithms
- Identify and categorize the problem and apply appropriate Machine Learning algorithms.
- Design model to solve various computing problems
- Evaluate the performance of the model
- Devise appropriate conclusion for Decision making

Course Contents/Syllabus:

	Weightage (%)
Module I: Introduction:	
Concept of Machine Learning, Applications of Machine Learning, Key elements of Machine Learning, Types of learning, Statistical Description of Data and Data Visualization, Introduction of Machine Learning algorithms, Mathematical foundations – Linear algebra, Statistical Learning: Bayesian Method, The Naïve Bayes Classifier. Vectorization, Matrices and Vectors,	**20%**
Module II: Parametric Learning	
Prediction using Linear Regression, Gradient Descent, Linear Regression with one variable, Linear Regression with multiple variables, Feature Scaling/Selection. Dimensionality Reduction Classification using Logistic Regression, Logistic Regression with one variable and with multiple variables. Regularization and its utility	**20%**
Module III: Non linear Learning:	
Introduction, Model Representation, Gradient Descent vs. Perceptron Training, Stochastic Gradient Descent, Multilayer Perceptron, Multiclass Representation, Back propagation Algorithm. Regularization and the bias-variance dilemma; decision theory; Support Vector Machines (SVMs); Boosting and Bagging methods; Nearest neighbors and other instance-based/nonparametric methods;	**20%**

Module IV: Unsupervised Learning	
Unsupervised Learning, Partitioning methods, Hierarchical clustering, Fuzzy clustering, Density-based clustering, Model-based clustering.	**20%**
Module V: Models Evaluation and Ensemble Learning	
Evaluation of learning algorithms performance using training, test and generalization error, cross-validation, Trees, Forest and Ensembles, Learning with imbalance datasets, Outlier detection, large scale	**20%**

Pedagogy for Course Delivery:

The class will be taught using theory and applications. In addition to discussing real life applications, the course instructor will spend considerable time in understanding the concept. The instructor will cover the ways to think innovatively liberally using thinking techniques

Assessment/ Examination Scheme:

Theory L/T (%)	Lab/Practical/Studio (%)	Total
75%	25%	100

Theory Assessment (L&T):

Continuous Assessment/Internal Assessment	End Term Examination

Components (Drop down)	Attendance	Class Test	Assignment	Case Study	
Weightage (%)	5	10	7	8	70

Lab Assessment :

Continuous Assessment/Internal Assessment					End Term Examination
Components (Drop down)	Attendance	Performance	Lab Record	Viva	
Weightage (%)	5	10	10	5	20

For practical Labs for Machine Learning, students may use softwares like MABLAB/Octave or Python. For later exercises, students can create/use their own datasets or utilize datasets from online repositories like UCI Machine Learning Repository (http://archive.ics.uci.edu/ml/).

1. **Perform basic operations on matrices (like addition, subtraction, multiplication) and display specific rows or columns of the matrix.**
2. **Perform other matrix operations like converting matrix data to absolute values, taking the negative of matrix values, additing/removing rows/columns from a matrix, finding the**

maximum or minimum values in a matrix or in a row/column, and finding the sum of some/all elements in a matrix.

3. **Create various type of plots/charts like histograms, plot based on sine/cosine function based on data from a matrix. Further label different axes in a plot and data in a plot.**
4. **Perform vectorized implementation of simple matrix operation like finding the transpose of a matrix, adding, subtracting or multiplying two matrices.**
5. **Implement Linear Regression problem. For example, based on a dataset comprising of existing set of prices and area/size of the houses, predict the estimated price of a given house.**
6. **Based on multiple features/variables perform Linear Regression. For example, based on a number of additional features like number of bedrooms, servant room, number of balconies, number of houses of years a house has been built – predict the price of a house.**
7. **Implement a classification/ logistic regression problem. For example based on different features of students data, classify, whether a student is suitable for a particular activity. Based on the available dataset, a student can also implement another classification problem like checking whether an email is spam or not.**
8. **Use some function for regularization of dataset based on problem 7.**
9. **Use some function for neural networks, like Stochastic Gradient Descent or backpropagation - algorithm to predict the value of a variable based on the dataset of problem 14.**
10. **Implement and evaluate a classification using Naïve Bayes. For example based on different features of students data, classify, whether a student is suitable for a particular activity. Based on the available dataset, a student can also implement another classification problem like checking whether an email is spam or not.**
11. **Implement and evaluate a classification using SVM. For example based on different features of students data, classify, whether a student is suitable for a particular activity. Based on the**

available dataset, a student can also implement another classification problem like checking whether an email is spam or not.

12. **Implement and evaluate a classification using Decision Tree.**
13. **Implement and evaluate a Clustering using K-means algorithm.**
14. **Implement and evaluate a classification for an imbalance dataset**
15. **Implement and evaluate a classification using Random forest**

Text & References:

Text Books:

1. T. Mitchell, Machine Learning, Tata McGraw-Hill, 2013. ISBN: 9781259096952 **Reference Books:**

2. Christopher M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.

3. Kevin Patrick Murphy, Machine Learning: a Probabilistic Perspective, MIT Press (2012).

Online Materials: See http://ai.stanford.edu/people/nilsson/MLBOOK.pdf.
https://jakevdp.github.io/PythonDataScienceHandbook/