BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI

WORK INTEGRATED LEARNING PROGRAMMES

Digital

Part A: Content Design

Course Title	Machine Learning
Course No(s)	ZG565
Credit Units	4
Credit Model	1 - 0.5 - 1.5 1 unit for class room hours, 0.5 unit for Tutorial, 1.5 units for Student preparation. 1 unit = 32 hours
Content Authors	Dr. Sugata Ghosal
Version	1.0
Date	November 11th, 2022

Course Objectives

No	
CO1	Introduce students to the basic concepts and techniques of Machine Learning.
CO2	To gain experience of doing independent study and research in the field of Machine Learning
CO3	To develop skills of using recent machine learning software tools to evaluate learning algorithms and model selection for solving practical problems

Text Book(s)

T1	Tom M. Mitchell, Machine Learning, The McGraw-Hill Companies, Inc. Indian Edition 1997

Reference Book(s) & other resources

R1	Christopher M. Bishop, Pattern Recognition & Machine Learning, Springer, 2006
R2	PANG-NING TAN, MICHAEL STEINBACH, VIPIN KUMAR, Introduction To Data Mining, Pearson, 2nd Edition.
R3	CHRISTOPHER J.C. BURGES: A Tutorial on Support Vector Machines for Pattern Recognition, Kluwer Academic Publishers, Boston, pp. 1–43.

Content Structure

1. Introduction
   - 1.1 Introduction to ML
   - 1.2 Objective of the course
   - 1.3 Taxonomy (types) of Machine Learning
   - 1.4 Design a Learning System
   - 1.5 Challenges in Machine Learning

2. Machine learning Workflow
   - 2.1 Role of Data
   - 2.2 Data Preprocessing, wrangling
   - 2.3 Data skewness removal (sampling)
   - 2.4 Model Training
   - 2.5 Model Testing and performance metrics

3. Linear models for Regression
   - 3.1 Direct Solution Method
   - 3.2 Iterative Method – Gradient Descent (batch/stochastic/mini-batch)
   - 3.3 Linear basis function models
   - 3.4 Bias-variance decomposition

4. Linear models for classification
   - 4.1 Discriminant Functions
   - 4.2 Decision Theory
   - 4.3 Probabilistic Discriminative Classifiers
   - 4.4 Logistic Regression

5. Decision Tree
   - 5.1 Information Theory
   - 5.2 Entropy Based Decision Tree Construction
   - 5.3 Avoiding Overfitting
   - 5.4 Minimum Description Length

5.5 Handling Continuous valued attributes, missing attributes

6. Instance-based Learning
   6.1 k-Nearest Neighbor Learning
   6.2 Locally Weighted Regression (LWR) Learning
   6.3 Radial Basis Functions

7. Support Vector Machine
   7.1 Linearly separable data
   7.2 Non-linearly separable data
   7.3 Kernel Trick (Mercer)
   7.4 Applications to both structured and unstructured data

8. Bayesian Learning
   8.1 MLE Hypothesis
   8.2 MAP Hypothesis
   8.3 Bayes Rule
   8.4 Optimal Bayes Classifier
   8.5 Naïve Bayes Classifier
   8.6 Probabilistic Generative Classifiers
   8.7 Bayesian Linear Regression

9. Ensemble Learning
   9.1 Combining Classifiers
   9.2 Bagging
   9.3 Random Forest
   9.4 Boosting
      9.4.1 ADABoost
      9.4.2 Gradient Boosting
      9.4.3 XGBoost

10. Unsupervised Learning
    10.1 K-means Clustering and variants
    10.2 Review of EM algorithm
    10.3 GMM based Soft Clustering
    10.4 Applications

11. Machine Learning Model Evaluation/Comparison
    11.1 Comparing Machine Learning Models
    11.2 Emerging requirements e.g., bias, fairness, interpretability of ML models

Learning Outcomes:

No	Learning Outcomes
LO1	A strong understanding of the foundations of Machine Learning algorithms
LO2	Able to solve Machine Learning problems using appropriate learning techniques

LO3	Evaluate machine learning solutions to problems
LO4	Identify appropriate tools to implement the solutions to machine learning problems

Part B: Learning Plan

Academic Term	
Course Title	Machine Learning
Course No	ZG 565
Lead Instructor	Dr. Sugata Ghosal

Session No.	Topic Title	Study/HW Resource Reference
1	**Introduction** Objective, What is Machine Learning? Application areas of Machine Learning, Why Machine Learning is important? Design a Learning System, Issues in Machine Learning	T1 – Ch1
2	**Machine learning Workflow** Role of Data, Data Pre-processing, wrangling, Data skewness removal (sampling), Model Training, Model Testing and performance metrics	R2 – Ch2, Ch3 Lecture Notes
3	**Linear models for Regression** Direct Solution Method, Iterative Method – Gradient Descent (batch/stochastic/mini-batch), Linear basis function models	R1 - Ch3
4	**Linear models for Regression (contd)** Bias-variance decomposition **Linear models for classification** Discriminant Functions, Decision Theory, Probabilistic Discriminative Classifiers, Introduction to Logistic Regression	R1 - Ch. 3, 4
5	**Logistic Regression**	R1 - Ch. 4 R2 – Ch. 4

	Logloss Function, Gradient Descent, multi-class classification	
6	**Decision Tree** Information Theory, Entropy Based Decision Tree Construction, Avoiding Overfitting, Minimum Description Length, Handling Continuous valued attributes, missing attributes	T1 – Ch. 3 R2 - Ch. 3
7	**Instance-based Learning** k-Nearest Neighbor Learning, Locally Weighted Regression (LWR) Learning , Radial Basis Functions	T1 – Ch. 8
8	Review of Session 1 to 7	Books, Web references and Slides
9	**Support Vector Machine** Linearly separable data, Non-linearly separable data, Kernel Trick (Mercer), Applications to both structured and unstructured data	R2 - Ch. 4 R3
10	**Bayesian Learning** MLE Hypothesis , MAP Hypothesis, Bayes Rule, Optimal Bayes Classifier	T1 - Ch. 6 R2 – Ch. 4
11	**Bayesian Learning** Naïve Bayes Classifier, Probabilistic Generative Classifiers, Bayesian interpretation of Linear Regression	T1 - Ch. 6 R2 – Ch. 4 R1 – Ch. 4
12	**Ensemble Learning** Combining Classifiers, Bagging, Random Forest, Boosting	R2 – Ch. 4
13	**Ensemble Learning** ADABoost, Gradient Boosting, XGBoost	R2 – Ch. 4 Lecture Notes
14	**Unsupervised Learning** K-Means clustering, Mixture Models for probabilistic clustering, Review of EM algorithm, Applications	T1 – Ch. 6
15	**Machine Learning Model Evaluation/Comparison** Comparing Machine Learning Models, Emerging requirements e.g., bias, fairness, interpretability of ML models	T1 - Ch. 5 Lecture Notes
16	Review of session 9 to 15	Books, Web references and Slides

Detailed Plan for Lab work

Lab No.	Lab Objective	Lab Sheet Access URL	Session Reference
1	End to End Machine Learning		2
2	Linear Regression and Gradient Descent		3, 4
3	Logistic Regression classifier		5
4	Decision Tree and Random Forest		6, 12
5	Naïve Bayes Classification		11

Evaluation Scheme:
Legend: EC = Evaluation Component; AN = After Noon Session; FN = Fore Noon Session

No	Name	Type	Duration	Weight	Day, Date, Session, Time
EC-1	Quiz - Two	Online	~1 hour	10%	
	Assignment-I	Take Home	~2-3 weeks	10%	
	Assignment-II	Take Home	~2-3 weeks	10%	
EC-2	Mid-Semester Test	Closed Book		30%	
EC-3	Comprehensive Exam	Open Book		40%	

Note:
Syllabus for Mid-Semester Test (Open Book): Topics in Session Nos. 1 to 8
Syllabus for Comprehensive Exam (Open Book): All topics (Session Nos. 1 to 16)
Important links and information:
Elearn portal: https://elearn.bits-pilani.ac.in or Canvas
Students are expected to visit the Elearn portal on a regular basis and stay up to date with the latest announcements and deadlines.
Contact sessions: Students should attend the online lectures as per the schedule provided on the Elearn portal.
Evaluation Guidelines:

1. EC-1 consists of either two Assignments or three Quizzes. Students will attempt them through the course pages on the Elearn portal. Announcements will be made on the portal, in a timely manner.
2. For Closed Book tests: No books or reference material of any kind will be permitted.
3. For Open Book exams: Use of books and any printed / written reference material (filed or bound) is permitted. However, loose sheets of paper will not be allowed. Use of calculators is permitted in all exams. Laptops/Mobiles of any kind are not allowed. Exchange of any material is not allowed.
4. If a student is unable to appear for the Regular Test/Exam due to genuine exigencies, the student should follow the procedure to apply for the Make-Up Test/Exam which will be made available on the Elearn portal. The Make-Up Test/Exam will be conducted only at selected exam centres on the dates to be announced later.

It shall be the responsibility of the individual student to be regular in maintaining the self study

schedule as given in the course handout, attend the online lectures, and take all the prescribed evaluation components such as Assignment/Quiz, Mid-Semester Test and Comprehensive Exam according to the evaluation scheme provided in the handout.