GUJARAT TECHNOLOGICAL UNIVERSITY

Bachelor of Engineering (Part Time)
Subject Code: 2971103
Semester – VII
Subject Name: Introduction to Machine Learning

Type of course: Elective

Prerequisite: Basics of Probability and Statistical Theory, and Optimization Concepts

Rationale: The objective of the course is to introduce the students with concepts of machine learning, machine learning algorithms and its various applications.

Teaching and Examination Scheme:

Teaching Scheme			Credits	Examination Marks				Total Marks
				Theory Marks		Practical Marks		
L	T	P	C	ESE (E)	PA (M)	ESE (V)	PA (I)	
3	0	2	4	70	30	30	20	150

Content:

Sr. No.	Content	Total Hrs
1	**Introduction to Machine Learning:** Introduction, Different Types of Learning, Hypothesis Space, Inductive Bias, Evaluation and Cross Validation	03
2	**Basic Machine Learning Algorithms:** Linear Regression, Decision Trees, Learning Decision Trees, K-nearest Neighbour, Collaborative Filtering, Overfitting	06
3	**Dimensionality Reduction:** Feature Selection, Feature Extraction	04
4	**Bayesian Concept of Learning:** Bayesian Learning, Naïve Bayes, Bayesian Network, Exercise on Naïve Bayes	04
5	**Logistic Regression and Support Vector Machine:** Logistic Regression, Introduction to Support Vector Machine, The Dual Formation, Maximum Margin with Noise, Nonlinear SVM and Kernel Function, SVM: Solution to the Dual Problem	06
6	**Basics of Neural Network:** Introduction to neural network, Multilayer Neural Network, Neural Network and Backpropagation Algorithm, Deep Neural Network	08
7	**Computation and Ensemble Learning**: Introduction to Computation Learning, Sample Complexity: Finite Hypothesis Space, VC Dimension, Introduction to Ensembles, Bagging and Boosting	08
8	**Basic Concepts of Clustering:** Introduction to Clustering, K-means Clustering, Agglomerative Hierarchical Clustering	06

GUJARAT TECHNOLOGICAL UNIVERSITY

Bachelor of Engineering (Part Time)
Subject Code: 2971103

Suggested Specification table with Marks (Theory):

Distribution of Theory Marks					
R Level	U Level	A Level	N Level	E Level	C Level
10	20	40	10	10	10

Legends: R: Remembrance; U: Understanding; A: Application, N: Analyze and E: Evaluate C: Create and above Levels (Revised Bloom's Taxonomy)

Reference Books:

1. Machine Learning, Saikat Dull, S. Chjandramouli, Das, Pearson
2. Machine Learning with Python for Everyone, Mark Fenner, Pearson
3. Machine Learning, Anuradha Srinivasaraghavan, Vincy Joseph, Wiley
4. Machine Learning with Python, U Dinesh Kumar Manaranjan Pradhan, Wiley
5. Python Machine Learning, Sebastian Raschka, Vahid Mirjalili, Packt Publishing

Course Outcomes:

Sr. No.	CO statement	Marks % weightage
CO-1	Understand basic concepts of machine learning as well as challenges involved.	15
CO-2	Learn and implement various basic machine learning algorithms.	40
CO-3	Study dimensionality reduction concept and its role in machine learning techniques.	10
CO-4	Realize concepts of advanced machine learning algorithms.	20
CO-5	Comprehend basic concepts of Neural network and its use in machine learning.	15

List of Experiments (Programs can be written in Python or some other language):

1. Generate a synthetic data set using following function, and split it into training, validation, and testing sample points. Use linear regression technique to develop a model, and evaluate on test samples. א is Gaussian noise.
$$y = \frac{x}{2} + sin(x) + \text{א}$$
2. Write a program for Logistic Regression to classify IRIS data for two features (sepal length and width).
3. For the synthetic dataset used in experiment 1, write a program for the concept of decision tree to develop a piecewise linear model and test it as well.
4. Write a program for decision tree to classify IRIS dataset. Consider all four features.

5. Write a program for kNN algorithm for classification of IRIS dataset.
6. Write a program using PCA algorithm for dimensionality reduction in case of Olivetti dataset, and follow it with KNN algorithm for face recognition.
7. Write a program using Bayes algorithm for email classification (spam or non-spam) for the open-sourced data set from the UC Irvine Machine Learning Repository.
8. Write a program using SVM algorithm for Boston house price prediction dataset to predict price of houses from certain features.
9. Write a program using SVM on IRIS dataset and carry out classification.
10. Write a program for artificial neural network for recognition of handwritten digits available in MNIST database. Use Google Tensor Flow library for the said task.

List of e-Learning Resources:

1. **https://www.geeksforgeeks.org/machine-learning/**
2. **https://www.tutorialspoint.com/machine_learning_with_python/index.htm**
3. **https://nptel.ac.in/**
4. **https://www.coursera.org/**