CO327 Machine Learning

Course Coordinator: Prof Anil Singh Parihar

Course Outcomes (Cos)

On completion of the course, students should be able to:

CO1	Explain the principles, types, applications, and ethical implications of machine learning systems.
CO2	Preprocess data, perform exploratory analysis, and engineer features suitable for machine learning models.
CO3	Implement, apply, and differentiate supervised and unsupervised learning algorithms to solve real-world problems.
CO4	Evaluate machine learning models using appropriate metrics and improve them through validation and tuning strategies.
CO5	Develop artificial neural networks and reinforcement learning models for prediction and control tasks.

Program Outcomes (POs/Graduate Attributes)
- PO1: Engineering Knowledge
- PO2: Problem Analysis
- PO3: Design/Development of Solutions
- PO4: Investigation
- PO5: Modern Tool Usage
- PO6: The Engineer and Society
- PO7: Environment and Sustainability
- PO8: Ethics
- PO9: Individual and Teamwork
- PO10: Communication
- PO11: Project Management and Finance
- PO12: Lifelong Learning

Program Specific Outcomes (PSOs)
- PSO1: Design, analyze and develop the engineering problems.
- PSO2: Specify, design, develop, test, and maintain usable systems that behave reliably and efficiently.
- PSO3: Develop systems that perform tasks related to Research, Education and Training, and/or E-governance.

CO-PO Articulation Matrix

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	2	1	1	2	2	1	3	1	2	1	3	2	1	2
CO2	3	3	2	2	3	0	0	1	1	1	1	2	3	3	2
CO3	3	3	3	2	3	0	0	1	2	2	1	2	3	3	3
CO4	3	3	3	3	3	0	0	1	2	2	1	2	3	3	3
CO5	3	3	3	2	3	1	0	2	2	2	2	2	3	3	3
CO6	3	3	3	2	3	2	1	2	3	3	3	3	3	3	3

Syllabus

Prerequisites: Python programming, linear algebra, probability and basic statistics

Course Objectives

To enable students to understand, apply, and evaluate machine learning techniques, design real-world solutions, and gain hands-on experience through projects.

Unit I – Fundamentals of Machine Learning

Definition of machine learning, types of learning (supervised, unsupervised, reinforcement), traditional programming vs machine learning, key components of an ML system, applications of ML in industry, stages of a machine learning pipeline, introduction to ML tools (Python, Jupyter, pandas, scikit-learn, matplotlib), mathematical foundations: vectors, matrices, linear transformations, vector spaces, probability theory: random variables, probability distributions, expectation, variance, conditional probability, Bayes' theorem, statistical learning: empirical risk minimization, generalization, inductive bias, common probability distributions in ML (Bernoulli, Binomial, Normal, Poisson, Uniform), probability density functions (PDF), probability mass functions (PMF), cumulative distribution function (CDF), expectation, variance.

Unit II – Data Handling and Feature Engineering

Data ingestion (CSV, JSON, Excel), handling missing data (deletion, mean/mode imputation), outlier detection, categorical data encoding (label encoding, one-hot), feature scaling (min-max normalization, z-score standardization), creation of new features, feature selection methods (filter, wrapper), exploratory data analysis using statistical plots and heatmaps, population and sample in machine learning datasets, sampling error and its impact on generalization, sample representativeness and selection bias, conceptual introduction to statistical inference, data distribution analysis (normal, skewed, multimodal, and uniform distributions), sampling techniques (random sampling, stratified sampling, bootstrapping, and sampling bias).

Unit III – Supervised Learning: Regression and Classification

Linear regression, least squares, cost function, gradient descent (manual and using libraries), polynomial regression, logistic regression for binary classification, sigmoid activation. Decision trees, Gini index, information gain, overfitting and pruning, ensemble learning (random forest). Bagging, boosting (e.g., AdaBoost), stacking, K-Nearest Neighbors (K-NN).

Unit IV – Model Evaluation and Optimization

Classification metrics (accuracy, precision, recall, F1-score, ROC curve, AUC), confusion matrix, multi-class classification with one-vs-rest, model interpretation
Train/test split, cross-validation (k-fold, stratified), bias-variance tradeoff, hyperparameters vs parameters, grid search, random search, pipeline creation using scikit-learn, evaluating models on imbalanced data, model selection strategies, Regularization techniques: L1 (Lasso), L2 (Ridge), elastic net**,** Validation and learning curves, generalization error bounds (conceptual)**,** Out-of-distribution (OOD) data: definition, impact on generalization, detection techniques, and robustness evaluation.

Unit V – Unsupervised Learning and Dimensionality Reduction

Similarity measures: Euclidean, cosine, Mahalanobis distances, Clustering evaluation: intra-cluster distance, inter-cluster distance, Dunn index. Clustering (k-means, hierarchical, DBSCAN), choosing number of clusters, silhouette score, principal component analysis (PCA), t-SNE, feature compression, customer segmentation, data visualization in reduced dimensions, PCA theory: Eigenvalues, eigenvectors, variance maximization.

Unit VI – Artificial Neural Networks

Structure of an artificial neuron, activation functions (ReLU, sigmoid, tanh), forward propagation, computational graph representation, loss functions (mean squared error, cross-entropy), backpropagation algorithm, chain rule, weight updates, gradient descent variants (SGD, momentum, Adam), learning rate tuning, batch vs mini-batch vs stochastic training, underfitting and overfitting in ANNs, regularization techniques (L2, dropout), Perceptron learning, perceptron convergence theorem, Multi-layer perceptron (MLP), universal approximation theorem (overview).

Unit VII – Explainable ML, Fairness, and Deployment

Explainable ML models, black-box vs interpretable models, SHAP and LIME for model interpretation, bias and fairness in machine learning, algorithmic transparency, real-world ethical concerns, lightweight ML deployment using Flask or Streamlit, reproducibility using Git, GitHub for version control.

Unit VIII – Reinforcement Learning

Reinforcement learning principles, environment, agent, rewards, Q-learning algorithm, ε-greedy exploration, Markov decision processes (basic), Bellman equations, value functions, policy iteration, value iteration, Temporal-Difference Learning, SARSA, Actor-Critic methods (overview), training RL agents in OpenAI Gym (CartPole).

Capstone Project: Project proposal, dataset selection, full ML pipeline implementation, presentation and evaluation of capstone project