CSE-6003

Machine Learning & Deep Learning

CSE-6003 • 4 Credits • 22-Class Lecture Plan

Dr. Mahmudul Hasan
Dept. of CSE
Comilla University

COURSE OVERVIEW

CSE-6003: Machine Learning & Deep Learning

CLO 1

Develop advanced ML models based on statistical theory

CLO 2

Design and optimize deep learning architectures

CLO 3

Analyze explainability, robustness & ethical issues in AI

CLO 4

Apply AI techniques to vision, speech & language problems

PLO Alignment: PLO1 • PLO2 • PLO3 • PLO5

22-CLASS LECTURE PLAN

Course Roadmap & Module Structure

MODULE A Classes 01–04	**Statistical Foundations** Statistical learning theory, bias-variance, regularization, optimization
MODULE B Classes 05–08	**Classical ML Algorithms** Supervised & unsupervised learning, SVMs, ensembles, clustering
MODULE C Classes 09–13	**Deep Learning Core** CNNs, RNNs, LSTMs, Transformers, optimization & training
MODULE D Classes 14–17	**Advanced Architectures** GNNs, GANs, VAEs, diffusion, transfer & representation learning
MODULE E Classes 18–22	**Responsible AI & Applications** XAI, fairness, robustness, CV, NLP, speech & RL applications

CSE-6003

Class 01: Foundations of Statistical Learning Theory

MODULE A — Statistical Foundations

Machine Learning & Deep Learning • 4 Credits

01

Foundations of Statistical Learning Theory

MODULE A — Statistical Learning Theory, Bias-Variance, Generalization

- Introduction to the learning problem: Input space X, output space Y, hypothesis class H, and the goal of finding $h \in H$ minimizing true risk $R(h)$
- Empirical Risk Minimization (ERM): From true risk to empirical risk; when and why ERM succeeds
- PAC Learning framework: Probably Approximately Correct learning, sample complexity, and learnability conditions
- VC Dimension: Shattering, VC dimension of linear classifiers, and its role in generalization bounds
- The Generalization Bound: $R(h) \leq \hat{R}(h) + O(\sqrt{(\text{VC-dim}/m)})$ — interpreting the bound and its implications
- Rademacher Complexity: A tighter, data-dependent measure of hypothesis class richness
- Key Takeaway: The fundamental tradeoff — larger hypothesis classes reduce bias but increase variance and require more data

CSE-6003

Class 02: Bias-Variance Tradeoff & Regularization

MODULE A — Statistical Foundations

Machine Learning & Deep Learning • 4 Credits

02

Bias-Variance Tradeoff & Regularization

MODULE A — Generalization and Regularization Techniques

Bias-Variance Decomposition

- MSE = Bias² + Variance + Noise
- High bias → underfitting (too simple model)
- High variance → overfitting (too complex model)
- The sweet spot: balancing model capacity vs. data size
- Empirical analysis via learning curves
- Double descent phenomenon in modern deep models

Regularization Methods

- L2 / Ridge Regression: weight decay, shrinks weights toward zero
- L1 / Lasso: promotes sparsity, feature selection effect
- Elastic Net: combines L1 + L2 benefits
- Dropout: stochastic regularization for neural networks
- Data augmentation: implicit regularization
- Early stopping: validation-based regularization

CSE-6003

Class 03: Optimization Methods for ML

MODULE A — Statistical Foundations

Machine Learning & Deep Learning • 4 Credits

03

Optimization Methods for Machine Learning

MODULE A — Gradient-Based Methods, Convergence, and Landscapes

- Loss Functions: MSE, Cross-Entropy, Hinge Loss, KL-Divergence — choice of loss and its probabilistic interpretation (MLE/MAP)
- Gradient Descent: Batch GD, convergence guarantees on convex functions, learning rate effects
- Stochastic Gradient Descent (SGD): Mini-batch SGD, noise as regularizer, generalization advantages of SGD
- Momentum Methods: Nesterov momentum, heavy ball method — escaping local minima faster
- Adaptive Methods: AdaGrad, RMSProp, Adam, AdamW — per-parameter learning rates and practical considerations
- Learning Rate Schedules: Step decay, cosine annealing, warm-up strategies, cyclical LR
- Convexity vs. Non-convexity: Why deep learning loss landscapes are non-convex, saddle points, and why flat minima generalize better

CSE-6003

Class 04: Reinforcement Learning Fundamentals

MODULE A — Statistical Foundations

Machine Learning & Deep Learning • 4 Credits

04

Reinforcement Learning Fundamentals

MODULE A — MDP, Policy, Value Functions, Q-Learning

Core RL Framework

- Markov Decision Process: (S, A, P, R, γ) formulation
- Policy π(a|s), value function V^π(s), Q-function Q^π(s,a)
- Bellman equations: recursive definition of value
- Exploration vs. exploitation: ε-greedy, UCB, Thompson Sampling
- Dynamic Programming: Policy/value iteration
- Model-based vs. model-free RL

Key RL Algorithms

- Monte Carlo Methods: Episode-based value estimation
- TD Learning: TD(0), TD(λ), eligibility traces
- Q-Learning: Off-policy, convergence guarantees
- SARSA: On-policy TD control
- Policy Gradient: REINFORCE algorithm
- Actor-Critic: Combining policy and value function

CSE-6003

Class 05: Supervised Learning — Linear & Probabilistic Models

MODULE B — Classical ML Algorithms

Machine Learning & Deep Learning • 4 Credits

05

Supervised Learning — Linear & Probabilistic Models

MODULE B — Regression, Classification, Probabilistic Frameworks

- Linear Regression: OLS, geometric interpretation, closed-form vs. iterative solutions, normal equations
- Logistic Regression: Sigmoid activation, binary cross-entropy, multiclass via softmax, probabilistic interpretation
- Generative vs. Discriminative Models: $P(y|x)$ vs. $P(x,y)$ — when each approach excels
- Naive Bayes Classifier: Conditional independence assumption, Gaussian NB, Bernoulli NB, Multinomial NB
- Linear Discriminant Analysis (LDA): Dimensionality reduction + classification, class separability
- Gaussian Processes: Non-parametric Bayesian approach, kernels, uncertainty quantification
- Practical Tips: Feature scaling, feature engineering, handling class imbalance (SMOTE, class weights)

CSE-6003

Class 06: Support Vector Machines & Kernel Methods

MODULE B — Classical ML Algorithms

Machine Learning & Deep Learning • 4 Credits

06

Support Vector Machines & Kernel Methods

MODULE B — Maximum Margin Classifiers and Kernel Trick

SVM Theory

- Maximum margin classifier: geometric margin and functional margin
- Hard-margin SVM: primal and dual formulation
- Soft-margin SVM: slack variables, C parameter, bias-variance
- KKT conditions and the dual solution
- Support vectors: only a subset of points define the decision boundary
- SVR: Support Vector Regression for continuous outputs

Kernel Methods

- Kernel trick: implicit feature mapping via $k(x,x') = \phi(x)\cdot\phi(x')$
- Common kernels: Linear, Polynomial, RBF/Gaussian, Sigmoid
- Mercer's theorem: conditions for valid kernels
- Kernel PCA: non-linear dimensionality reduction
- Hyperparameter tuning: C, γ via cross-validation
- Scalability: Approximations — Random Fourier Features, Nyström

CSE-6003

Class 07: Ensemble Methods & Decision Trees

MODULE B — Classical ML Algorithms

Machine Learning & Deep Learning • 4 Credits

07

Ensemble Methods & Decision Trees

MODULE B — Tree-Based Models, Bagging, Boosting, Stacking

- Decision Trees: CART algorithm, Gini impurity vs. entropy, information gain, pruning strategies (pre/post)
- Bagging: Bootstrap aggregating, variance reduction, out-of-bag error estimation
- Random Forests: Feature subsampling, decorrelating trees, feature importance via permutation and impurity
- Boosting: AdaBoost — sequential reweighting, exponential loss minimization
- Gradient Boosting Machines (GBM): Gradient descent in function space, shrinkage, subsampling
- XGBoost / LightGBM / CatBoost: Engineering optimizations — column block, histogram binning, leaf-wise growth
- Stacking (Meta-learning): Combining heterogeneous models with a meta-learner; blending strategies

CSE-6003

Class 08: Unsupervised Learning & Dimensionality Reduction

MODULE B — Classical ML Algorithms

Machine Learning & Deep Learning • 4 Credits

08

Unsupervised Learning & Dimensionality Reduction

MODULE B — Clustering, Density Estimation, Manifold Learning

Clustering Algorithms

- K-Means: Lloyd's algorithm, K-Means++ initialization, elbow method
- Hierarchical Clustering: Agglomerative (Ward, complete, average linkage), dendrograms
- DBSCAN: Density-based, handles arbitrary shapes, noise robust
- Gaussian Mixture Models: EM algorithm, soft assignment, model selection (BIC/AIC)
- Spectral Clustering: Graph Laplacian, normalized cuts
- Evaluation: Silhouette score, Davies-Bouldin, adjusted Rand index

Dimensionality Reduction

- PCA: Variance maximization, eigen-decomposition, scree plot, explained variance
- t-SNE: Perplexity, KL divergence minimization, visualization power & pitfalls
- UMAP: Topological data analysis approach, faster than t-SNE, preserves global structure
- Autoencoders: Encoder-decoder for non-linear compression
- ICA: Independent component analysis for source separation
- LDA: Supervised dimensionality reduction

CSE-6003

Class 09: Neural Networks — From Perceptron to Deep Networks

MODULE C — Deep Learning Core

Machine Learning & Deep Learning • 4 Credits

09

Neural Networks — From Perceptron to Deep Networks

MODULE C — Foundations, Backpropagation, Activation Functions

- Historical Context: McCulloch-Pitts neuron → Perceptron → MLP — key milestones in neural network development
- Feedforward Neural Networks: Layers, weights, biases, forward pass computation as function composition
- Activation Functions: Sigmoid (vanishing gradients), Tanh, ReLU (dying ReLU problem), Leaky ReLU, ELU, GELU, Swish
- Backpropagation: Chain rule applied recursively — deriving gradients layer by layer, computational graph perspective
- Universal Approximation Theorem: A single hidden layer can approximate any continuous function (with enough neurons)
- Depth vs. Width: Why deep networks are more expressive per parameter than wide shallow networks
- Initialization Strategies: Xavier/Glorot (tanh), He (ReLU) — preventing vanishing/exploding gradients from the start

CSE-6003

Class 10: Training Deep Networks — Optimization & Stability

MODULE C — Deep Learning Core

Machine Learning & Deep Learning • 4 Credits

10

Training Deep Networks — Optimization & Stability

MODULE C — Batch Normalization, Gradient Flow, Regularization

Optimization & Stability

- Vanishing/Exploding Gradients: Causes, impact on deep networks, gradient clipping
- Batch Normalization: Normalizing layer inputs, learnable γ and β, train vs. inference behavior
- Layer Normalization: For sequence models where batch statistics are unstable
- Residual Connections (ResNets): Skip connections enable gradient flow through 100+ layers
- Dense Connections (DenseNets): All-to-all connections, feature reuse
- Mixed Precision Training: FP16/BF16 for speed + FP32 for stability (loss scaling)

Practical Training Tips

- Hyperparameter tuning: Grid search, random search, Bayesian optimization
- Learning rate finding: LR range test (Smith's method)
- Batch size effects: Large batch → sharp minima; small batch → generalization
- Gradient accumulation: Simulating large batches on limited GPU memory
- Checkpointing: Save best model by validation loss
- Distributed training: Data parallelism, model parallelism, ZeRO optimization

CSE-6003

Class 11: Convolutional Neural Networks (CNNs)

MODULE C — Deep Learning Core

Machine Learning & Deep Learning • 4 Credits

11

Convolutional Neural Networks (CNNs)

MODULE C — Architecture Design, Spatial Features, Computer Vision Backbone

- Convolution Operation: Cross-correlation vs. convolution, kernel/filter, stride, padding (same vs. valid), receptive field
- CNN Architectural Components: Conv → BatchNorm → Activation → Pool; feature map size formulas
- Pooling: Max pooling (translation invariance), average pooling, global average pooling
- Classic Architectures: LeNet-5 → AlexNet → VGG → GoogLeNet/Inception → ResNet → EfficientNet — design philosophy evolution
- Depth-wise Separable Convolutions: MobileNet — drastically reducing parameters while preserving accuracy
- Dilated (Atrous) Convolutions: Expanding receptive field without increasing parameters — key for segmentation
- Applications: Image classification, object detection (YOLO, Faster R-CNN), semantic segmentation (U-Net, DeepLab)

CSE-6003

Class 12: Recurrent Neural Networks & Sequence Models

MODULE C — Deep Learning Core

Machine Learning & Deep Learning • 4 Credits

12

Recurrent Neural Networks & Sequence Models

MODULE C — RNNs, LSTMs, GRUs, Sequence Modeling

RNN Foundations

- Vanilla RNN: Hidden state h_t = f(W_hh h_{t-1} + W_xh x_t), BPTT
- Vanishing gradients in time: Why RNNs forget long-range dependencies
- LSTM: Cell state, input gate, forget gate, output gate — long-range memory
- GRU: Simplified LSTM with reset and update gates, competitive performance
- Bidirectional RNNs: Processing sequences in both directions
- Stacked RNNs: Hierarchical feature extraction from sequences

Sequence-to-Sequence

- Encoder-Decoder Architecture: Compressing input to context vector
- Attention Mechanism (Bahdanau): Dynamic context — learning where to focus
- Beam Search: Better decoding than greedy, balancing exploration
- Applications: Machine translation, text summarization, speech-to-text
- Teacher Forcing: Stabilizing training of seq2seq models
- Limitations of RNNs → motivation for Transformers

CSE-6003

Class 13: Transformers & Attention Mechanisms

MODULE C — Deep Learning Core

Machine Learning & Deep Learning • 4 Credits

13

Transformers & Attention Mechanisms

MODULE C — Self-Attention, Multi-Head Attention, Position Encoding

- The Attention Equation: Attention(Q,K,V) = softmax(QK^T / √d_k)V — intuition, scaling, and complexity
- Multi-Head Attention: Parallel attention heads attending to different representation subspaces
- Positional Encoding: Sinusoidal and learned position embeddings — why position matters in permutation-invariant attention
- Transformer Encoder: Self-attention + Feed-Forward + Add & Norm — pre-norm vs. post-norm variants
- Transformer Decoder: Masked self-attention + cross-attention — causal language modeling
- Scaling Laws: Performance scales predictably with parameters, data, and compute (Kaplan et al.)
- BERT (encoder-only) vs. GPT (decoder-only) vs. T5 (encoder-decoder) — pretraining objectives and use cases

CSE-6003

Class 14: Graph Neural Networks

MODULE D — Advanced Architectures

Machine Learning & Deep Learning • 4 Credits

14

Graph Neural Networks (GNNs)

MODULE D — Message Passing, Graph Tasks, Real-World Applications

GNN Foundations

- Graph Representation: G = (V, E, X) — nodes, edges, node features, adjacency matrix
- Message Passing Framework: h_v^{k} = UPDATE(h_v^{k-1}, AGGREGATE({h_u: u ∈ N(v)}))
- Graph Convolutional Networks (GCN): Spectral graph convolution, normalized adjacency
- GraphSAGE: Inductive learning, neighborhood sampling for large graphs
- Graph Attention Networks (GAT): Attention-weighted neighbor aggregation
- Expressive power: GNN vs. Weisfeiler-Lehman graph isomorphism test

Applications & Advanced Topics

- Node Classification: Semi-supervised with few labeled nodes (Cora, Citeseer)
- Link Prediction: Knowledge graph completion, recommendation systems
- Graph Classification: Molecular property prediction, drug discovery
- Heterogeneous Graphs: Multiple node/edge types (HAN, HGT)
- Temporal Graphs: Dynamic graph learning, temporal GNNs
- Scalability: Cluster-GCN, GraphSAINT for billion-scale graphs

CSE-6003

Class 15: Generative Models — GANs, VAEs & Diffusion

MODULE D — Advanced Architectures

Machine Learning & Deep Learning • 4 Credits

15

Generative Models — GANs, VAEs & Diffusion

MODULE D — Generative Modeling, Latent Spaces, Distribution Learning

- Variational Autoencoders (VAEs): Evidence Lower Bound (ELBO) = Reconstruction loss + KL divergence; reparameterization trick
- Generative Adversarial Networks (GANs): Minimax game — Generator vs. Discriminator; Nash equilibrium as training objective
- GAN Training Challenges: Mode collapse, training instability, non-convergence — Wasserstein GAN (WGAN) as solution
- Progressive GAN / StyleGAN: High-resolution image synthesis, disentangled style control in latent space
- Conditional Generation: cGAN, class-conditional generation, image-to-image translation (Pix2Pix, CycleGAN)
- Diffusion Models: Forward noising process, reverse denoising, score matching — DDPM, DDIM, stable diffusion
- Comparison: VAEs (fast inference, blurry), GANs (sharp, unstable training), Diffusion (highest quality, slow sampling)

CSE-6003

Class 16: Representation Learning & Transfer Learning

MODULE D — Advanced Architectures

Machine Learning & Deep Learning • 4 Credits

16

Representation Learning & Transfer Learning

MODULE D — Self-Supervised, Contrastive, Fine-Tuning Paradigms

Self-Supervised Learning

- The Pretext Task idea: creating labels from data itself
- Contrastive Learning: SimCLR — maximize agreement between views, InfoNCE loss
- BYOL / SimSiam: Self-supervised without negative pairs
- Masked Autoencoders (MAE): Reconstruct masked patches — highly efficient pretraining
- Momentum Contrast (MoCo): Memory bank for large negative set
- DINO / DINOv2: Knowledge distillation for visual representations

Transfer Learning

- Transfer Learning Paradigm: Pretrain on large dataset → fine-tune on target task
- Feature Extraction: Freeze pretrained layers, train new head
- Fine-Tuning Strategies: Full fine-tuning, layer-wise LR decay
- Domain Adaptation: Covariate shift, domain adversarial training (DANN)
- Parameter-Efficient Fine-Tuning: LoRA, adapter layers, prefix tuning
- Foundation Models: CLIP, DALL-E, GPT-4V — vision-language alignment

CSE-6003

Class 17: Deep Reinforcement Learning

MODULE D — Advanced Architectures

Machine Learning & Deep Learning • 4 Credits

17

Deep Reinforcement Learning

MODULE D — DQN, Policy Gradients, Actor-Critic, RLHF

- Deep Q-Network (DQN): Approximating Q-values with deep networks, experience replay, target network stabilization
- Double DQN & Dueling Networks: Correcting overestimation bias, separating state value from advantage
- Policy Gradient Methods: REINFORCE with baseline, variance reduction via baselines
- Proximal Policy Optimization (PPO): Clipped surrogate objective, stable on-policy updates — workhorse of modern RL
- Soft Actor-Critic (SAC): Maximum entropy RL, off-policy efficiency, entropy regularization
- Model-Based Deep RL: World models (MuZero, Dreamer), planning in latent space for sample efficiency
- Reinforcement Learning from Human Feedback (RLHF): Reward model training + PPO fine-tuning — how ChatGPT/Claude are aligned

CSE-6003

Class 18: Explainable AI (XAI)

MODULE E — Responsible AI & Applications

Machine Learning & Deep Learning • 4 Credits

18

Explainable AI (XAI)

MODULE E — Interpretability, Explanation Methods, Trustworthy ML

Intrinsically Interpretable Models

- Linear models: coefficients as feature importance
- Decision trees and rule-based models
- Generalized Additive Models (GAMs): shape functions
- When to prefer interpretable over black-box
- Monotone networks and constrained models
- Sparse models: L1 regularization for interpretability

Post-hoc Explanation Methods

- LIME: Local linear approximation of any model
- SHAP: Shapley values — theoretically grounded, consistent feature attribution
- Grad-CAM: Class Activation Maps for CNN visualizations
- Integrated Gradients: Attribution from baseline to input
- Attention visualization: When and why attention ≠ explanation
- Counterfactual Explanations: Minimal changes to flip prediction

CSE-6003

Class 19: Fairness, Robustness & Ethical AI

MODULE E — Responsible AI & Applications

Machine Learning & Deep Learning • 4 Credits

19

Fairness, Robustness & Ethical AI

MODULE E — Bias, Adversarial Attacks, Privacy, Ethical Frameworks

- Algorithmic Fairness: Demographic parity, equalized odds, individual vs. group fairness — formal mathematical definitions
- Sources of Bias: Historical bias, representation bias, measurement bias, aggregation bias in datasets and models
- Fairness-Accuracy Tradeoff: Impossibility theorems — why all fairness criteria cannot be satisfied simultaneously
- Adversarial Robustness: FGSM and PGD attacks, adversarial training, certified robustness (randomized smoothing)
- Privacy in ML: Differential privacy — ε-DP, gradient perturbation, federated learning; membership inference attacks
- Responsible AI Frameworks: EU AI Act risk categories, NIST AI RMF, model cards, datasheets for datasets
- Case Study Discussion: Algorithmic bias in hiring, criminal justice (COMPAS), healthcare — real-world lessons

CSE-6003

Class 20: Computer Vision Applications

MODULE E — Responsible AI & Applications

Machine Learning & Deep Learning • 4 Credits

20

Computer Vision Applications

MODULE E — Detection, Segmentation, Generation, Vision-Language

Core Vision Tasks

- Image Classification: Top-1/5 accuracy, ImageNet benchmark, ViT vs. CNN performance
- Object Detection: One-stage (YOLO, SSD, FCOS) vs. Two-stage (Faster R-CNN, Cascade RCNN)
- Semantic Segmentation: FCN, U-Net for medical imaging, DeepLab v3+ with atrous convolutions
- Instance Segmentation: Mask R-CNN — joint detection + mask prediction
- Panoptic Segmentation: Unifying semantic + instance segmentation
- 3D Vision: Point clouds, PointNet, NeRF — neural radiance fields for novel view synthesis

Modern Vision Systems

- Vision Transformers (ViT): Patchify → position embed → transformer encoder
- DETR: End-to-end object detection with transformers, bipartite matching loss
- Segment Anything Model (SAM): Promptable segmentation, zero-shot generalization
- Vision-Language Models: CLIP zero-shot classification, Flamingo, LLaVA
- Autonomous Driving: Sensor fusion, bird's-eye view prediction, end-to-end driving
- Medical Imaging: Radiology AI, pathology, FDA-approved AI devices

CSE-6003

Class 21: NLP & Speech Processing Applications

MODULE E — Responsible AI & Applications

Machine Learning & Deep Learning • 4 Credits

21

NLP & Speech Processing Applications

MODULE E — Language Models, Text Understanding, Speech Systems

- Word Representations: Word2Vec (CBOW & Skip-gram), GloVe, FastText — from discrete tokens to continuous vectors
- Contextual Representations: ELMo → BERT → RoBERTa → DeBERTa — how pretraining objectives shape representations
- Large Language Models (LLMs): GPT architecture, autoregressive pretraining, emergent abilities, in-context learning
- Instruction Tuning & RLHF: Fine-tuning LLMs to follow instructions — InstructGPT, ChatGPT, Claude
- Core NLP Tasks: Named Entity Recognition (NER), sentiment analysis, question answering, summarization, machine translation
- Speech Processing: Acoustic model (CTC, RNN-T), language model, end-to-end ASR (Whisper), Text-to-Speech (WaveNet, VITS)
- Multimodal Applications: Visual QA, image captioning, speech-driven face generation, video understanding

CSE-6003

Class 22: Case Studies, Frontiers & Research Directions

MODULE E — Responsible AI & Applications

Machine Learning & Deep Learning • 4 Credits

22

Case Studies, Frontiers & Research Directions

MODULE E — Industry Applications, Open Problems, Research Frontiers

Industry Case Studies

- Healthcare: AlphaFold2 — protein structure prediction, drug discovery pipelines
- Scientific Discovery: GNoME (materials discovery), weather forecasting (GraphCast)
- Code Generation: GitHub Copilot, DeepMind AlphaCode, LLM-based software engineering
- Autonomous Systems: Self-driving (Tesla FSD, Waymo), robotic manipulation
- Creative AI: DALL-E 3, Midjourney, Sora — text-to-image/video generation
- Finance: High-frequency trading, credit scoring, fraud detection at scale

Open Research Frontiers

- Continual Learning: Catastrophic forgetting, lifelong learning in neural networks
- Causality in ML: Beyond correlation — causal representation learning (Schölkopf et al.)
- Neuromorphic Computing: Spiking neural networks, energy-efficient AI hardware
- AI Safety & Alignment: Scalable oversight, interpretability research, constitutional AI
- Emergent Capabilities: Understanding why scale leads to sudden capability jumps
- Post-Transformer Architectures: State Space Models (Mamba), RWKV, xLSTM

COURSE SUMMARY

CSE-6003: 22-Class Lecture Plan

Module A (01–04)

Statistical Theory, Bias-Variance, Optimization, RL Fundamentals

Module B (05–08)

Supervised, SVMs, Ensembles, Unsupervised & Dimensionality Reduction

Module C (09–13)

Neural Networks, CNNs, RNNs, Transformers, Training Stability

Module D (14–17)

GNNs, Generative Models, Representation Learning, Deep RL

Module E (18–22)

XAI, Fairness & Ethics, Computer Vision, NLP, Speech, Case Studies