Lahore University of Management Sciences
CS 535/EE 514 Machine Learning
Fall 2023-24

Course description
Machine learning (ML) techniques allow computers to adapt to data and solve new problems related to previously encountered problems more efficiently. Such methods enable machines to perform practical exploratory and predictive tasks without being explicitly programmed. ML finds applications in speech recognition and synthesis, machine translation, object recognition, chatbots, question-answering, natural language understanding, anomaly detection, medical diagnosis and prognosis, autonomous vehicles and robots, time series forecasting, and much more. This introductory course covers the theoretical foundations and practical applications of ML and the design, implementation, and analysis of various ML algorithms. Students will learn to compare across and choose the most appropriate algorithms for multiple problem types and be able to design and implement their solutions. Students will be prepared for industry and academia and for pursuing advanced courses.

Course distribution	
Elective	This is an elective course.
Open for Student Category	Juniors, seniors, and graduates.
Close for Student Category	Please see the prerequisites below.

Course prerequisites

- Undergrads (Seniors/Juniors) must have passed:
  - An Ugrad/Grad course in Probability (MATH230 (Probability) OR DISC203 (Probability & Statistics) OR CS501 (Applied Probability)) OR ECON230 (Statistics and Data Analysis))
  - And a programming course (CS200/EE201 (Intro. to Programming)
  - And a course on Linear Algebra (MATH120 (LA with Diff. Equations))
- Grads are strongly advised to brush up their programming skills and take CS501 (Applied Probability), may be in parallel with ML
- All students must possess strong programming skills and proficiency in algorithm implementation in JAVA/C/Python/MATLAB

Course Offering Details				
Credit Hours	3 hours			
Lecture(s)	Nbr of lec(s) per week	2	Duration	75 minutes
Recitation/Lab (per week)	Nbr of lec(s) per week		Duration	
Tutorial (per week)	Nbr of lec(s) per week	1 (optional)	Duration	50 minutes

Instructor	Agha Ali Raza
Room No.	SBASSE 9-G49A
Office Hours	TBA
Email	agha.ali.raza@lums.edu.pk
Telephone	3336
Secretary/TA	Zohaib Khan (24100074@lums.edu.pk), Nida Tanveer (24100091@lums.edu.pk), Saad Hasan Iqbal (24020320@lums.edu.pk), Syeda Mah Noor Asad (24100243@lums.edu.pk), Alina Faisal (24100314@lums.edu.pk), Muawiz Feroze Khan (24100052@lums.edu.pk), Fatima Ali (24100048@lums.edu.pk), Rafey Rana (24100103@lums.edu.pk), Zain Ali Khokhar (24100130@lums.edu.pk), Muhammad Haad Zahid (24100134@lums.edu.pk), Mughees Ur Rehman (24100086@lums.edu.pk), Syed Kabir Ahmad (24100249@lums.edu.pk)
TA Office Hours	TBA
Course URL (if any)	From last offering: https://www.c-salt.org/courses/machine-learning-f2021

Course Teaching Methodology (Please mention the following details in plain text)
● **Lectures:** In-person. ● **TA Sessions:** TAs will conduct asynchronous and synchronous sessions (in-person and online) to cover tutorials related to assignments. ● **Exams:** Exams will be conducted in person in pre scheduled sessions. ● **Quizzes:** Quizzes will be conducted during announced class timings. ● **Class discussions:** There will be a slack channel for all discussions (general, assignments, quizzes, etc.)

PROGRAM EDUCATIONAL OBJECTIVES (PEOs)	
PEO-01	Demonstrate excellence in the profession through in-depth knowledge and skills in the field of Computing.
PEO-02	Engage in continuous professional development and exhibit a quest for learning.
PEO-03	Show professional integrity and commitment to societal responsibilities.

Course Objectives
The goal of this course is to get the students excited about Machine Learning and to enable them to: • Develop a firm grip on the theory behind statistical learning • Understand and rigorously go through the phases of the design, implementation, and evaluation of fundamental ML algorithms • Choose the appropriate algorithm for each problem type and be able to compare the strengths and weaknesses of the algorithms • Appreciate the end-to-end organic integration of ML in its application areas, from data sources, annotation pipelines, and choice of algorithms to societal biases, explainability of models, and potential to impact and even disrupt existing processes

COURSE LEARNING OUTCOMES (CLOs)	
	By the end of the course, students should be able to:
CLO1:	• Develop an appreciation for what is involved in learning models from data, and integrating ML in existing real-world processes
CLO2:	• Thoroughly understand the ML pipeline from design and data gathering to meaningful and relevant evaluation
CLO3:	• Learn a wide variety of learning algorithms, and formulate and implement solutions to machine learning problems
CLO4:	• Apply algorithms to real-world problems, optimize the trained models and report on the expected performance

CLO	**CLO Statement**	**Bloom's Cognitive Level**	**PLOs/Graduate Attributes (Seoul Accord)**
CLO1	Develop an appreciation for what is involved in learning models from data, and integrating ML in existing real-world processes	C2, C3	PLO2
CLO2	Thoroughly understand the ML pipeline from design and data gathering to meaningful and relevant evaluation	C3, C4, C5	PLO2, PLO3, PLO4
CLO3	Learn a wide variety of learning algorithms, and formulate and implement solutions to machine learning problems	C2, C5, C6	PLO4, PLO5
CLO4	Apply algorithms to real-world problems, optimize the trained models and report on the expected performance	C5, C6	PLO5

Grading Breakup and Policy

Assessment	**Weight (%)**	**Related CLOs**	**ACM Recommended Disposition**
Programming assignment(s)	25%	CLO2, CLO3, CLO4	D3, D4, D7, D9, D10
Quizzes	25%	CLO1, CLO2	D4, D7, D9, D10
Project	20%	CLO1 – CLO4	D1, D3, D4, D5, D6, D7, D8, D9, D11
Reading assignment(s)/homework(s)/Implementation of Research Paper(s)/viva	15%	CLO1 – CLO4	D3, D4, D7, D9, D10
Final examination + viva	15%	CLO1 – CLO4	D4, D7, D9, D10

Examination detail	
Midterm Exam	Yes/No: No Duration: Exam Specifications:

Final Exam	Yes/No: Duration: Exam Specifications:	Yes 2.5 – 3 hours In-person exam

SSE Council on Equity and Belonging
In addition to LUMS resources, SSE's **Council on Belonging and Equity** is committed to devising ways to provide a safe, inclusive and respectful learning, living, and working environment for students, faculty and staff. To seek counsel related to any issues, please feel free to approach either a member of the council or email at cbe.sse@lums.edu.pk.
Mental Health Support at LUMS
For matters relating to counselling, kindly email student.counselling@lums.edu.pk, or visit https://osa.lums.edu.pk/content/student-counselling-office for more information. You are welcome to write to me or speak to me if you find that your mental health is impacting your ability to participate in the course. However, should you choose not to do so, please contact the Counseling Unit and speak to a counsellor or speak to the OSA team and ask them to write to me so that any necessary accommodations can be made.
Harassment Policy
SSE, LUMS and particularly this class, is a harassment free zone. Harassment of any kind is unacceptable, whether it be sexual harassment, online harassment, bullying, coercion, stalking, verbal or physical abuse of any kind. Harassment is a very broad term; it includes both direct and indirect behavior, it may be physical or psychological in nature, it may be perpetrated online or offline, on campus and off campus. It may be one offense, or it may comprise of several incidents which together amount to sexual harassment. It may include overt requests for sexual favors but can also constitute verbal or written communication of a loaded nature. Further details of what may constitute harassment may be found in the LUMS Sexual Harassment Policy, which is available as part of the university code of conduct. LUMS has a Sexual Harassment Policy and a Sexual Harassment Inquiry Committee (SHIC). Any member of the LUMS community can file a formal or informal complaint with the SHIC. If you are unsure about the process of filing a complaint, wish to discuss your options or have any questions, concerns, or complaints, please write to the Office of Accessibility and Inclusion (OAI, oai@lums.edu.pk) and SHIC (shic@lums.edu.pk) —both of them exist to help and support you and they will do their best to assist you in whatever way they can. You can find more details regarding the LUMS sexual harassment policy here. **To file a complaint, please write to harassment@lums.edu.pk.**
Rights and Code of Conduct for Online Teaching
A misuse of online modes of communication is unacceptable. TAs and faculty will seek consent before the recording of live online lectures or tutorials. Please ensure if you do not wish to be recorded during a session to inform the faculty member in a timely manner. Please also ensure that you prioritize formal means of communication (email, LMS) over informal means to communicate with course staff.

Course overview

W	Topics	Recommended Readings	Related CLOs	ACM Comp Knowledge Landscape
1.	**Course overview** ● What is ML? Traditional CS vs. ML, history of ML, AI vs. ML ● Classification and Regression with examples. Training and Testing. ● Rules vs. Patterns, Deterministic vs. Probabilistic, Certainty vs. Uncertainty. ● Learning: Supervised, unsupervised, semi-supervised ● Labeled data sources: Expert annotators, crowd ● Example ML application areas: Speech and Language Technologies ● Challenges and Opportunities of ML: ○ Explainability ○ Fairness and Societal Biases ○ ML for Social Good, ML for Development (ML4D), Speech and Language Technologies for Development (SLT4D)	● Murphy chapter 1 ● Alpaydin, chapter 1	CLO1, CLO2	
2.	**Supervised Learning** ● Features, Labels, Training, Testing, Classification, Regression. ● Formalizing the supervised learning setup ● Feature spaces and feature vectors ○ Sparse and dense feature vectors, one-hot vectors ○ Bag-of-word features	● Murphy: 1.1, 1.2, 1.4.2, 1.4.3, 1.4.9 **Recommended topics:** ● Goals of Cross Validation:	CLO1, CLO2, CLO3	

	● Label spaces ○ Label spaces for classification (binary and multiclass) and regression ● Hypothesis spaces ○ The No Free Lunch theorem ○ Choosing the hypothesis class H and hypothesis $h \in H$ ○ Various Algorithms for traversing hypothesis classes: ■ Pick h randomly ■ Try every h ■ Just output the label of the training data (memorizer) ● Evaluating hypotheses: Loss functions and goals of optimization ○ Zero-One ○ Squared ○ Absolute ● Loss reduction and Generalization in Learning ○ Memorizers ○ Smoothing and Priors ○ Tradeoff between Bias and Variance ● Sampling from the distribution $P(X, Y)$ ○ Representative datasets ○ Training, validation, and testing ● How to split the dataset D? ○ Time series data ○ Independent and Identically Distributed (IID) ● The weak law of large numbers ($\epsilon_{TE} \rightarrow \epsilon \ as\ \lvert D_{TE} \rvert \rightarrow +\infty$) ● How to prevent overfitting to test data? Do's and Don'ts ● Validation sets (dev sets) and Cross Validation	Model selection, training, and performance estimation ● Types of Cross Validation and Pros and Cons ● Exhaustive ● Leave-p-out ● Leave-one-out ● Non-Exhaustive ● k-fold ● Holdout ● Repeated random subsampling ● Nested ● $k * l$ fold ● k-fold with validation and test sets ● Bootstrapping ● Stratified cross validation ● Time series cross validation (forward chaining – Rolling origin)		
3.	**The K-Nearest Neighbor Classifier** (An instance-based, lazy, discriminative, non-linear, non-parametric classifier) ● KNN – The Basics ○ Nearest neighbor classification rule ○ KNN formal definition ○ KNN decision boundaries and Voronoi Tessellations ○ Properties of KNN: Non-parametric, used for classification and regression, instance-based, lazy ● KNN Similarity/Distance measures and constraints ○ Minkowski Distances (Manhattan, Euclidean and Chebyshev) ● The KNN algorithm and implementation ○ KNN regression and classification with examples ○ Space and Time complexity ○ Bias/variance tradeoff as ○ $K \rightarrow 1$ and $K \rightarrow n$ ○ Tuning the hyperparameter: K ○ KNN: The good, the bad and the ugly ● KNN error bounds as $n \rightarrow \infty$ ○ Bayes Error ○ 1-NN error as $n \rightarrow \infty$ ● KNN Enhancements ○ Parzen Windows and Kernels ○ K-D trees ○ Inverted lists ○ Locality sensitive hashing ● The Curse of Dimensionality ○ Demonstration and examples ○ Challenges and opportunities	● Murphy: 1.1, 1.2, 1.4.2, 1.4.3, 1.4.9 ● Videos of different values of k ● Video describing nearest neighbors ● A nice explanation of nearest neighbors	CLO3, CLO4	

	○ Lower dimensional subspaces and manifolds in higher dimensional ambient space ***Tutorial Topics:*** ● Review of KNNs ● Review of Voronoi Tessellations ● Review (and Expansion) of Error Bounds ● Review (and Expansion) of the Curse of Dimensionality			
4.	**Evaluation metrics** ● The Confusion matrix (contingency tables) – binary and multi-label ○ True and False – Positives and Negatives ○ Type I and type II errors ● Performance Metrics ○ Accuracy, Sensitivity (recall, TPR), Specificity (TNR), Precision (Positive Predictive Value), Negative predictive value ○ False acceptance rate, False rejection rate ○ Examples – pros and cons ● The need for a combined measure ○ Types of averages: AM, GM, HM ○ F-β-measure, F-1-measure ● Multiclass Classification ○ Any-of (multi-label) classification ○ One-of (multinomial) classification- ○ Micro and Macro averaging ● Gold labels and annotation of data ○ Inter-annotator agreements ○ Cohen's Kappa and Krippendorff's alpha ● Evaluation of Classifiers, thresholds, comparing classifiers, imbalanced classes ○ Receiver operating Characteristic (ROC) and Precision-Recall (P-R) Curves ○ ROC *Area Under the Curve* (AUC) ○ Equal Error Rate (EER) and Biometric Systems ***Tutorial Topics:*** ● Review of Cohen's Kappa ● Review of ROC and Precision-Recall Curves ● Assignment 1	● SLP3: 4.7-4.9	CLO2, CLO3	
5.	**Linear Regression** ● Motivation for linearity ● Revision ○ Lines, planes, hyperplanes, and vectors ■ Lines and planes: Normal form and slope-intercept form ■ Decision boundaries with perpendicular weight vectors ■ Distance between a hyperplane and a point ○ The Dot Product ○ The geometric interpretation of absorbing the bias term ○ Visualizing n dimensions ● Linear Regression ○ Intuition ○ Derivation and implementation ■ Linear regression with one variable ■ Cost function: Square Loss, Mean Square Loss (MSE), ■ $\frac{1}{2}$ Mean Square Loss ■ Motivating gradient descent ● The Gradient Descent Algorithm ○ Gradients ○ The step size, ○ α ○ Convex and non-convex cost functions – Global and local optima ○ The batch gradient descent algorithm ○ Types of Gradient Descent: Batch, Minibatch, and Stochastic ● Multivariate Linear Regression ○ Multivariate Gradient Descent ○ Vectorizing the notation	● ESLII Ch3 ● Murphy 7-7.5.1, 7.5.4	CLO2, CLO3, CLO4	

	● Practical Issues of linear regression ○ Feature Scaling, local minima, ravines, saddle points, tracking progress in GD ○ Hyperparameters: Learning rate ● Polynomial regression ***Tutorial Topics:*** ● Review of Hyperplanes ● Review of Gradient Descent ● Demo on Polynomial Regression ● Review (and Expansion) of Standardization for Linear Regression			
6.	**Linear Regression: Bias and Variance** ● How to recognize high variance/high bias scenarios? ○ Underfitting and Overfitting ○ How to reduce bias and variance? ■ Cross validation ■ Feature selection ● Manual Feature Selection ○ Scatter Diagrams and Plots ○ Eyeballing those Correlations! ● Regularization ○ Motivation – The fitting problem ○ L2 Regularization or Ridge Regression ○ L1 Regularization or Lasso Regression ■ Automatic Feature selection ● Comparison of Ridge and Lasso regression ● Elastic Net Regression – Intuition **Logistic Regression** (A linear, discriminative, parametric classifier) ● Intuition and derivation ○ Regression for classification ○ “Squishing” between 0 and 1 using a non-linear activation function: The sigmoid ● A simple sentiment classifier ● Visualizing the logistic regression decision boundary ● Hyperplanes, linear and non-linear decision boundaries ● Cost function: Derivation of the cross-entropy loss function (log loss) ● Learning algorithm: Batch, Stochastic and Mini-batch Gradient Descent ● Multiclass (multinomial) classification: One-vs-all (one-vs-rest), One-vs-one ● The SoftMax activation function and multivariate log loss ***Tutorial Topics:*** ● Review of Bias and Variance: detection and techniques to deal with them ● Review (and Expansion) of Feature Selection ● Review (and Expansion) of Sigmoid and Softmax: derivatives, visualizations, dealing with overflow ● Review of using Binary Classifiers to setup Multiclass Classification ● Assignment 2	● Ben Taskar’s under- and overfitting ● MLaPP: 1.4.7 ● Andrew Ng’s lecture – ML debugging ● Ben Taskar’s Notes on Bias Variance ● Notes by Scott Foreman-Roe ● ELSII Chapter 2.9 ● Murphy: 6.2.2 ● SLP3 Ch5, ESLII Ch4 ● Murphy 8, 13.3, 13.5.3 ● Ben Taskar’s notes ● TM chapter: Naive Bayes and Logistic Regression ● Nice blogpost on Gradient Descent, Adagrad, Newton’s method	CLO2, CLO3, CLO4	
7.	**The Perceptron** (A discriminative, linear, parametric classifier) ● The McCulloch-Pitts Neuron and its limitations ● The Perceptron and its limitations ○ The Heaviside step function ○ Boolean functions: AND, OR and XOR! ○ One perceptron, two perceptrons, … ● Linear separability in low and high dimensional spaces ● From the step function to other activation functions ● The perceptron learning algorithm and its geometric interpretation ● Proof of convergence ○ Relation between margin and rate of convergence **Maximum Margin Classifiers: Support Vector Machines (SVMs)** (A discriminative, linear/non-linear classifier) ● Intuition and motivation ○ The perceptron and the optimal separating hyperplane ● Hard Margin Linear Support Vector Machines: Derivation	● The Perceptron Wiki page ● Murphy 8.5.4 ● Murphy: 14.5 - 14.5.2.2 ● Ben Taskar’s Notes on SVMs ● Kernel Cookbook by David Duvenaud ● Laurent El Ghaoui’s lectures on duality ● “Idiot’s guide to SVM”	CLO2, CLO3, CLO4	

	○ Constrained optimization and Lagrange Multipliers ● Soft Margin Linear Support Vector Machines: Derivation ● Hinge-loss ● Kernels and Kernel SVMs ***Tutorial Topics:*** ● Decision Boundaries and Hyperplanes of SVMs ● Review (and Expansion) of Soft Margin Classifiers and their Hyperparameters ● Review of Margins, Hinge Loss, and optimizing SVMs ● Review of Kernels: Linear, Polynomial, RBFs ● Supplementary: the Kernel Trick and Quadratic Programming			
8.	**Neural Networks** ● The Neuron and linear decision boundaries. Can we do better? ○ Review Logistic regression and gradient descent ● Changing the representation of the data ○ Kernels ○ Neural networks ● Non-linear activation ● Multi-layer Perceptron ● Neural Networks ● Deep Learning ● Intuition: How NNs work? ○ Non-linear and complex decision boundaries ○ Universal approximation and regression ● Layers: Depth vs. width ● Gradient descent for NNs ○ Overfitting and the importance of Stochastic Gradient Descent (SGD) ● Formal Notation for Logistic Regression and NNs ○ Vectorizing LR and NNs ○ Forward propagation (LR, NN) ■ Simple, vectored with a single instance ■ Vectored with m instances ● Activation functions, gradients, and pros and cons ○ Sigmoid ○ Tanh ○ ReLU and Leaky ReLU ● Backward propagation (LR, NN) ○ Simple, vectored with a single instance and m instances ○ Training the NN ***Tutorial Topics:*** ● Review of the Forward Pass and Backward Pass (in equations) ● Using Decision Boundaries to infer the architecture of a NN ● Review of strategies to deal with Overfitting and Underfitting ● Assignment 3 ● Supplementary: programming demo on using MLPs for character-level Language Modeling	● SLP Ch7, ESLII Ch11	CLO2, CLO3, CLO4	
9.	**Sequence Models** ● Notion of Sequence Modeling tasks ○ Text and Speech (One-Hot) ○ Time-Series ○ Videos as sequences of images ● Types of Sequence Modeling problems and specifications ○ 1-to-Many - Image Captioning ○ Many-to-1 - Video Classification, Sentiment Classification ○ Many-to-Many (Seq2Seq) - Machine Translation, Summarization etc. ● Why do Feedforward NNs perform poorly at Sequence Modeling tasks? ● Introducing RNNs ○ Idea of the Hidden State ○ The overall architecture ○ "Unrolling" a unit ○ Intuition of (truncated) BPTT ● Challenges of training RNNs, and how to deal with them ○ Exploding/Vanishing Gradients	Andrej Karpathy's The Unreasonable Effectiveness of RNNs Jay Alammar's Visualizing A Neural Machine Translation Model SLP3	CLO2, CLO3, CLO4	

	○ Choice of Activation functions ○ Gradient Clipping ○ Changing the architecture (intro to LSTMs) ● Introducing LSTMs ○ Changes to the architecture ○ Idea of storing the "memory" in a cell ● Exploring Machine Translation ○ Setup as a Seq2Seq problem ○ Embeddings ○ Using the Encoder-Decoder framework ○ Information bottleneck: passing on only one hidden state ○ Improvements ■ Passing all the hidden states ■ Intuition of the Attention Mechanism ***Tutorial Topics:*** ● Review of the representations of text: Label Encoding, One-Hot Encoding, Embeddings ● Programming demo on Embeddings in Python ● Review of RNNs and LSTMs ● Review of the Attention Mechanism in Machine Translation			
10.	**Attention Mechanism and Transformers** ● The Attention Mechanism in Machine Translation ● Self Attention ○ Dot Product Attention ○ The idea of contextualized word embeddings ● Introducing the Transformer and its contributions ○ Highly Parallelizable ○ Contextualized Embeddings vs. Plain Embeddings ○ Long-Term Dependencies ● Transformer Architecture in a nutshell ○ Role of an Encoder ○ Role of a Decoder and Masked Attention ○ Query, Keys, Values ● The Transformer in equations ○ Positional Embeddings ○ Projections to QKV ○ Self Attention as Dot Product Attention ○ Role of Feedforward layers ○ Multi Headed Self Attention: the role and its equations ○ (ignore Layer Norms for brevity) ● Case Studies: BERT, T5, and GPT-3/LLMs ***Tutorial Topics:*** ● Review (and Expansion) of Self-Attention ● Differences between the Encoder and Decoder: masking, role, architecture ● Demo on coding out GPT in Python	Jay Alammar's The Illustrated Transformer Andrej Karpathy's Let's Build GPT SLP3		
11.	**Decision (Classification/Regression) Trees** (Discriminative, non-linear, parametric/non-parametric) ● Clustering using K-D Trees and nearest neighbor methods ○ Why do we still need nearest neighbors? ● Revision of Trees ○ Graph: Nodes, edges, directed/undirected, path, cyclic/acyclic, ○ Tree: A rooted directed acyclic graph (Rooted DAG) ■ Parent, children, siblings, root, leaves, degree, height, arity, various relationships ■ Forests ● From K-D trees to decision trees ● Decision tree examples: classification and regression ○ Categorical and real-valued attributes ● Bias and variance ○ Tree height ● Growing tree automatically – Decision tree training ○ The ID3 and CART algorithms	● Decision Tree wiki page ● Ben Taskar's old notes ● Murphy: 16.2 ● ESLII 8.7, ESLII Ch10, 15, 16	CLO3, CLO4	

	○ Splits and purity/impurity ■ Gini and Entropy ■ Information Gain and Information Gain Ratio ■ Multiclass ■ Regression: Variance ○ Real-valued attributes ● Strengths and weakness of CARTs ○ Automatic feature selection ○ Generalizability of splitting on attributes with a large set of values ○ Missing data ○ Axis-aligned splits ○ Over fitting ● Side notes on Huffman coding ● A deep dive into Entropy ○ Quantifying expected surprise ■ Events in isolation ■ Probability distributions ○ Information, surprise, and uncertainty ○ Evaluating Language Models ■ Comparing distributions using cross entropy ● A discussion on Parametric/Non-parametric models **Ensemble methods: Bagging and Random Forests** ● Decomposition of Generalization Error ○ Bias/Variance/Noise ○ Detecting high bias and high variance regimes ● Variance reduction ○ The weak law of large numbers ○ M datasets sampled from P ○ Bootstrapping ■ Bootstrapped Aggregation (Bagging) ○ Are we even drawing from the same P ○ Summary and advantages of bagging ● Random Forests ○ Algorithm ○ Examples and benefits ■ Out-of-box performance ■ The hyper parameters ■ m and k ■ No need of normalization and feature-scaling ■ Resilience to the curse of dimensionality ■ Feature selection ■ Missing data and clustering ○ Variants ***Tutorial Topics:*** ● Review of Gini/Entropy and Decision Trees ● Demo on Random Forests in Python: exploring Feature Importances for Feature Selection ● Standardization in the context of Decision Trees ● Decision Trees vs. Neural Networks			
12.	**Ensemble methods: Boosting, Gradient Boosted Trees, and AdaBoost** ● Bias Reduction ○ Intuition ○ Vectors ○ Gradient Descent in function space ○ Generic Boosting (AnyBoost) ■ Algorithm and geometric interpretation ● Gradient Boosted Regression Trees ○ Algorithm ○ Detailed walk-through ● AdaBoost ○ Setting ○ Odds Ratio and log-odds ○ Step size proportional to error reduction	● ESLII 8.7, ESLII Ch10, 15, 16 ● SLP Ch4 ● Murphy 2.2, 3.1-3.4 ● ESLII 6.6.3	CLO1, CLO2, CLO3, CLO4	

	○ Instance weights proportional to (mis)classification and the say of the classifier ○ Algorithm and detailed walk-through ○ Properties and Summary **Bayes Theorem** ● Review of probability, joint and conditional probability, and derivation of the Bayes Theorem ● Maximum a posteriori (MAP) and Maximum Likelihood Estimation (MLE) ○ Posterior, likelihood, prior and evidence ○ Classification using MAP and MLE ● Example problems and solutions using Bayes Theorem ○ Binary and multiclass ○ Monty Hall problem, medical testing, Language Modeling ● Generative and Discriminative classifiers ○ Solving SPAM vs. Not-SPAM			
13.	**The Naïve Bayes Classifier** (A linear, generative, parametric classifier) ● Derivation and implementation ○ Classification using the Bayes Theorem ○ Learning by example: The SPAM vs. Not-SPAM problem ○ The "zeros" and how to get rid of them! ● Independence, mutual exclusion, and conditional independence ○ The challenge of "how much of the context to use?" - Ngrams ○ Naïve Assumptions: Conditional Independence and Bag-of-Words ● Data sparsity and Out-of-vocabulary (OOV) items ○ Laplace Add-1 smoothing ● Another example: Sentiment analysis ● Text generation using Naïve Bayes ○ Infinite monkeys on typewriters ○ The Shannon visualization method for Ngrams ○ Approximating Shakespeare and the Wall Street Journal ● Real-valued features: Gaussian Naïve Bayes ● Probability vs. likelihood ● A worked example ● Naïve Bayes decision boundary ● Under assumptions and general case ● Naïve Bayes: Strengths and weaknesses ***Tutorial Topics:*** ● Probability vs. Likelihood (and other terminologies) ● Review of Naive Bayes ● Review (and Expansion) of add-k smoothing ● Demo on using Bigrams for Language Modeling	● Ben Taskar's notes on Naïve Bayes ● TM chapter on Naive Bayes (ch 1-3) ● Xiaojin Zhu's notes on Multinomial Naïve Bayes ● Mannings' description of Multinomial Naive Bayes	CLO2, CLO3, CLO4	
14.	**Unsupervised Learning** ● The unsupervised learning setup ● Use cases of unsupervised learning ○ Clustering ○ Anomaly detection ○ Feature selection and dimensionality reduction ● Types of clustering ○ Monothetic and Polythetic ○ Hard and Soft ○ Flat and Hierarchical ● Clustering ○ K-D Trees ■ Monothetic, hard boundary, hierarchical, divisive (top-down) ■ Motivation ■ Algorithm ○ Vector Quantization ■ Motivation and method ■ Codebook and distance metric ■ Euclidean and Mahalanobis distances ○ K-means ■ Polythetic, hard boundary, flat ■ Lloyd/Forgy method	● ESLII 8.7, ESLII Ch10, 15, 16	CLO3, CLO4	

        - Expectation Maximization (EM) and K-means
        - The K-means objective
        - Optimal Number of Clusters
        - Categorical data and K-modes
        - Vector Quantization using K-means
    - Evaluating Clustering
        - Extrinsic and Intrinsic evaluation
    - Gaussian Mixture Models
        - Polythetic, soft boundary, flat, probabilistic
        - K-means vs. GMMs
        - EM for GMMs
        - Mixture models in 1-dimension and n-dimensions
        - Likelihoods, cluster assignments, and cluster update rules
        - The covariance matrix
        - How many Gaussians?
    - Hierarchical clustering
        - Recursive K-means
            - Polythetic, hard boundary, hierarchical, top-down
        - Agglomerative Clustering
            - Polythetic, hard boundary, hierarchical, bottom-up
            - Examples
            - Distances: Single link, complete link, average link, centroids, Ward's method

- Dimensionality Reduction
    - Feature selection vs. feature reduction
    - Motivation
        - Visualization
        - Redundant and correlated features
        - Real vs. apparent dimensionality
        - The curse of dimensionality
    - Principal Component Analysis (PCA)
        - The dimension of greatest variability
        - A side-note on Matrices, Linear Transformations, the determinant, Eigenvalues and Eigenvectors
        - The Eigenvectors of the Covariance Matrix
        - How many dimensions?
        - Strengths and weaknesses
    - Linear Discriminant Analysis (LDA)
        - Supervised setup
        - Discrimination vs. spread
- Anomaly Detection
    - Anomalies
    - How do we define anomalies?
    - Why unsupervised?
    - Examples and challenges
    - Detection
        - One Class Classification
        - Density estimation
        - One feature and multiple features
        - Algorithm
        - Example
        - Evaluation
        - Unsupervised vs. supervised

Big Challenges and Opportunities in AI and ML

- The case for Explainable AI
- The case for Fair AI
- Societal biases
- Imbalanced classification
- Machine Learning for Development (ML4D)

Tutorial Topics:

- Review of Clustering
- Review of PCA

	● Programming demo of PCA In Python as a preprocessing step			
	Other topics – to be covered if we have time			
15.	**Bayes: Advanced topics** (supplementary) ● Hypothesis spaces ● Frequentist viewpoint ○ Intuition, derivation, pros and cons, extreme data ● Bayesian viewpoint ○ Intuition, derivation, MAP, pros and cons ○ Conjugate priors: The Beta and Dirichlet distributions ● Comparison of MAP and MLE ○ Laplace smoothing ● The Bayes Optimal Classifier	● SLP Ch4 ● Murphy 2.2, 3.1-3.4 ● ESLII 6.6.3	CLO2, CLO3, CLO4	
16.	**Graphical Sequence Processing Models** ● Hidden Markov Models (HMMs) ● Maximum Entropy Markov Models (MEMMs) ● Undirected Graphical Models (Markov Random Fields) ● Conditional Random Fields (CRFs) ● Directed Graphical Models (Bayes Nets)	● SLP A, ESLII Ch17	CLO3, CLO4	

Textbook(s)/Supplementary Readings

Text Books

- Machine Learning, Tom Mitchell, McGraw Hill, 1997 – TM
- The Elements of Statistical Learning: Data mining, Inference, and Prediction, Hastie, Trevor, Robert Tibshirani, and Jerome Friedman, Springer Science & Business Media, 2009 – ESLII

Reference Books

- Speech and Language Processing by Jurafsky and Martin, Ed 3 (online draft) – SLP
- Machine Learning: A Probabilistic Perspective, Murphy, Kevin P. MIT press, 2012 – Murphy.
- Pattern Recognition and Machine Learning, Christopher M. Bishop, Springer, 2006 – Bishop.
- Introduction to Machine Learning, Ethem Alpaydin, Ed 2, MIT Press, 2010 – Alpaydin.
- Deep Learning, Ian Goodfellow and Yoshua Bengio and Aaron Courville, 2016 – Goodfellow

Course policies

Use of electronic devices (e.g., mobile phones and laptops) in the class is strictly forbidden. A violation could result in deduction of marks and other strict penalties

Late arrival: You may not be allowed in the class 10 minutes after the start time

Plagiarism: All work MUST be done independently. In certain assignments students will be allowed to have discussions with peers, in which case they must mention the name and roll number of the student with whom the discussion took place and the nature of the discussion. Even in those assignments, all implementations need to be done independently. Any plagiarism or cheating of work from others or the internet will be immediately referred to the DC. If you are confused about what constitutes plagiarism, it is YOUR responsibility to consult with the instructor or the TA in a timely manner. No "after the fact" negotiations will be possible.

- Submitting someone else's assignment as your own "by mistake" would count as plagiarism. If this indeed happens accidentally, please let us know immediately (within minutes) along with an explanation and do not wait until we find it out on our own. In the latter case, it would be considered plagiarism.

Quizzes: Quizzes will be unannounced. We will be following an n-x (x=2) policy for the quizzes. <u>There is no makeup for a missed quiz</u>. If you have missed up to x quizzes, you will be covered only using the n-x policy (even if you have an approved petition with the OSA). If you have missed more than x quizzes, then you would be awarded the average marks (across all the quizzes that you attempted) for each missed quiz, provided that your case has been approved by the Office of Student Affairs.

Non-uniform weightage: All subcomponents (e.g., quizzes, assignments) may not carry the same weight. These weights may not be announced prior to the submission of the components and will be determined by the course instructor based on factors including (but not limited to) the length, difficulty level, amount of help available, etc. for each subcomponent.

Programming: Strong programming skills are expected for this course. Please keep in mind that this is a programming intensive course, and you will be spending a lot of time designing and coding up your solutions.

Assignments: There is negative marking for skipped assignments and there is no n-x policy for assignments. Assignments are a basic building block of this course, and it will be ensured that students, who pass the course, have significant hands-on experience.

- You will be awarded 0 marks or investigated for plagiarism for submitting incorrect/corrupted files and/or older assignments. We will not accept resubmissions in these cases even if the system date shows that the file was not modified after the deadline.
- You are allowed 5 grace days for the entire semester. No late submission of assignments is allowed after your grace days have expired. We do not have any deduction policy for late submissions in addition to the grace days. All grace days must be utilized before the start of the dead week and any remaining grace days will expire as soon as the dead week begins.
- Please do not wait until the last moment to submit assignments and other components. Any requests to accommodate late submissions due to last minute issues (submission of partial or incorrect files, assignment server down-time, internet and power failures, personal problems, etc.) would not be accommodated.

Appendix A
Bloom's Taxonomy

BLOOM's TAXONOMY*	
1 - Remember	Recall facts and basic concepts
2 - Understand	Explain ideas or concepts
3 - Apply	Use information in new situations
4 - Analyze	Draw connection among ideas
5 - Evaluate	Justify a stand or decision
6 - Create	Produce new or original work

https://cft.vanderbilt.edu/guides-sub-pages/blooms-taxonomy/

Appendix B

ACM Dispositions Table - I

ACM Dispositions			
Element	**Elaboration**	**Element**	**Elaboration**
D1 Adaptable:	Flexible; agile, adjust in response to change	D7 Professional:	Professionalism, discretion, ethical, astute
D2 Collaborative:	Team player; willing to work with others	D8 Purpose-driven:	Goal driven, achieve goals, business acumen
D3 Inventive:	Exploratory; Look beyond simple solutions	D9 Responsible:	Use judgment, discretion, act appropriately
D4 Meticulous:	Attentive to detail; thoroughness, accurate	D10 Responsive:	Respectful; react quickly and positively
D5 Passionate:	Conviction, strong commitment, compelling	D11 Self-directed:	Self-motivated, determination, independent
D6 Proactive:	With initiative, self-starter, independent		

ACM Dispositions Table - II

Class Assessments and Proposed Dispositions												
Assessment Type	**D1 Adaptable**	**D2 Collaborative**	**D3 Inventive**	**D4 Meticulous**	**D5 Passionate**	**D6 Proactive**	**D7 Professional**	**D8 Purpose-driven**	**D9 Responsible**	**D10 Responsive**	**D11 Self-directed**	**Included**
Quiz				✓			✓		✓			Yes
Assignment-Individual			✓	✓			✓		✓			Yes
Assignment-Group		✓	✓	✓			✓		✓	✓		Yes
Project-Individual	✓		✓	✓	✓	✓	✓	✓	✓		✓	Yes
Project-Group	✓	✓	✓	✓	✓	✓	✓	✓	✓			Yes
Presentation-Individual				✓			✓		✓	✓	✓	Yes
Presentation-Group		✓		✓			✓		✓	✓		Yes
Labs-Individual			✓	✓			✓		✓			Yes
Labs- Group		✓	✓	✓			✓		✓	✓		Yes
Exams				✓			✓		✓			Yes
Included	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	

Appendix C
ACM Computing Knowledge Landscape Table

ACM Computing Knowledge Landscape (CK)			
1. **Users and Organizations**	CK1.1: Social Issues and Professional Practice CK1.2: Security Policy and Management CK1.3: IS Management and Leadership CK1.4: Enterprise Architecture CK1.5: Project Management CK1.6: User Experience Design	**4.** **Software Development**	CK4.1: Software Quality, Verification and Validation CK4.2: Software Process CK4.3: Software Modeling and Analysis CK4.4: Software Design CK4.5: Platform-Based Development
2. **Systems Modeling**	CK2.1: Security Issues and Principles CK2.2: Systems Analysis & Design CK2.3: Requirements Analysis and Specification CK2.4: Data and Information Management	**5.** **Software Fundamentals**	CK5.1: Graphics and Visualization CK5.2: Operating Systems CK5.3: Data Structures, Algorithms and Complexity CK5.4: Programming Languages CK5.5: Programming Fundamentals CK5.6: Computing Systems Fundamentals
3. **Systems Architecture and Infrastructure**	CK3.1: Virtual Systems and Services CK3.2: Intelligent Systems (AI) CK3.3: Internet of Things CK3.4: Parallel and Distributed Computing CK3.5: Computer Networks	**6.** **Hardware**	CK6.1: Architecture and Organization CK6.2: Digital Design CK6.3: Circuits and Electronics CK6.4: Signal Processing