CSE-4125: Introduction to Machine Learning
4th Year 1st Semester, 2024 (Held in March-July, 2025)

Reference Books and Materials

1) Introduction to Statistical Learning by James, Witten, Hastie and Tibshirani, 2023 (Available on the web) [ISR]
   https://drive.google.com/file/d/1I7NNIsMnm1NYfros-t1-itjXdJ3fD27x/view?usp=drive_link
2) Prof. Andrew Ng's course materials (will be provided as needed) [PAN]
3) Stanford University's CS229 MS Course Notes (will be provided as needed). [SUCS229]
4) An Introduction to Machine Learning by Miroslav Kubat, 2017 (Available on the web) [IML]
   https://drive.google.com/file/d/1rQ3HtH9PJxvXE8-n8VBBWeCtae863uOv/view?usp=drive_link
5) Pattern Recognition and Machine Learning by Bishop, 2006 (PRML)
   https://drive.google.com/file/d/1y6cDHvOaNzplrjL6xUFKu2dFn2fO97tN/view?usp=drive_link

Note: The following topics may be changed slightly as we go through the course.

Week	Heading	Topics	Reading Materials	Required Math/Stat Concepts	Advanced Topics (Optional)
1	Overview and Linear Regression	Human intelligence and civilization. Digital computer and artificial intelligence. From human learning to Machine Learning (ML). Settings of ML: supervised, unsupervised etc. Evolution of ML from an algorithmic perspective. Linear regression: assumption, evaluation, refinement phases. Gradient descent.	**PAN:** Lecture Notes 01 **ISR:** Pp. 1 - 6, 9 - 11, 26 - 29 **My class notes**	Continuity, derivative, partial derivative, gradient, convexity of a function.	Probabilistic derivation of linear regression.
2	Linear and Polynomial Regression	Closed form solution, pseudocode, learning rate. Validation set, pseudocode with learning rate tuning. Multivariate linear regression. Polynomial regression.	**PAN:** Lecture Notes 02, 04 **ISR:** Pp. 15 - 17, 59 - 63, 71 - 75 **My class notes**	Basics of matrix algebra.	Formal proof of convexity of cost function. More closed form solutions. More model selection methods.
3	Polynomial	Recap of polynomial regression. Overfitting and underfitting,	**PAN:** Lecture Notes 07		Statistical derivation of GE =

	regression and bias-variance.	learning curves. Bias and variance, generalization error, irreducible error. Cross validation as model selection. Regularization (ridge and lasso) for regression.	**ISR:** Pp. 17 - 18, 33 - 36, 87 - 94, 197 - 205, 237 - 242 Machine Learning Yearning book**:** Pp. 1 - 19 **My class notes**		squared bias plus variance. More regularization methods.
4	Classification: perceptron. Error metrics.	Problem formulation, hyperplane, decision boundary. Lifting data points by adding a constant feature. Perceptron hypothesis, loss function, update rule, prediction, pseudocode, limitations. Error metrics: confusion matrix, accuracy, precision, recall, F1-score, ROC.	**ISR:** Pp. 29 - 33, 37 - 37, 368 - 370 **PRML:** Pp. 179-183, 192-196, **IML:** Pp. 69 - 73, 211 - 219 Sec 4.7 of Chapter 4 of Jurafsky's book **My class notes**	Points divided by a hyperplane. Point-normal form of a line/plane. Dot product of two vectors.	0-1 (instead of +1 -1) labeling convention with perceptron and the change in update rule. Relation of perceptron with gradient descent. Formal proof of perceptron's convergence.
5	Logistic Regression, maximum likelihood estimation.	From perceptron to logistic regression, hypothesis, loss function, optimization, prediction. Multiclass classification: one-vs-all, one-vs-one. Multinomial logistic regression. Maximum Likelihood Estimation (MLE) framework. MLE for logistic regression and linear regression.	**ISR:** 129 - 140 Chapter 5 of Jurafsky's book **PAN:** Lec 6 **My class notes**	Distance between a point and a hyperplane. Probability basics: joint probability, conditional probability, Bayes theorem, Bernoulli distribution, Gaussian distribution.	Non-linearity and non-convexity (using Hessian). Deep understanding of loss functions of one-vs-all, multinomial with binary cross-entropy loss (ylog(h(x)) + (1-y)log(1-h(x))), and multinomial with categorical cross-entropy loss (ylog(h(x))).
6	Naive bayes. Generative vs discriminative models. Neural network.	Maximum APosteriori (MAP) as regularization. Naive bayes: hypothesis, learning, prediction. Comparison between generative and discriminative models. Neural network: motivation, hypothesis, forward propagation.	**PRML:** Pp. 43 - 44, 196 - 202, 203 - 206, 225 - 231 Chapter 4 of Jurafsky's book **PAN:** Lec 08 **My class notes**	Chain rule of probability.	Detailed derivation of ridge regression and lasso regression using MAP. Gaussian Discriminative Analysis (GDA), generalized linear models
7	Neural network.	Neural network: cost function, training, backpropagation, benefits, limitations, characteristics, activation functions.	**PAN:** Lec 09 **PRML:** Pp. 241 - 245 Andrew Ng 229 Notes	Chain rule of partial derivatives. Composition of convex	Proof of non-convexity of cost function of NN. Fully vectorized implementation of NN. Variance

			My class notes	and of linear functions.	reduction techniques of NN. Details of activation functions.
8		Incourse Exam Syllabus: Up to (i.e., including) forward propagation of NNs.			
9	SVM and kernel methods	Optimal margin classifiers: functional and geometric margins, SVM optimization objective, kernel trick for SVM, SVM with soft margin.	**PAN:** Lec 12 Andrew Ng CS229 Notes MIT Notes 2006 **My class notes**	Linear algebra topics from 4th and 5th weeks.	SVM for multiclass classification, for regression, details of dual representation, of Representer theorem, of Lagrangian and Lagrange multipliers.
10	Machine learning engineering	For real-life machine learning projects: empirical bias-variance analysis, learning curve analysis, optimizing and satisfying metrics, error analysis, comparing with human-level performance, correcting mislabeled examples, data mismatch problem, transfer learning, multi-task learning, end-to-end learning.	Machine Learning Yearning book**:** Sections 1 - 43, 47 - 49 Course # 3 of Andrew Ng's Deep Learning Specialization course. **PAN:** Lec 10, 11	Machine learning basics.	More real-life challenges in ML system design and deployment: class imbalance, feature selection algorithms, feature engineering, data preprocessing: missing values imputation, noisy features and labels, data scaling.
11	Clustering, Anomaly detection	Unsupervised learning. K-means: motivation, algorithm, convergence, choosing K, initialization. Outlier detection motivation, density-based methods: univariate gaussian and multivariate gaussian, use of labeled cross-validation set.	**PAN:** Lec 13 and 15. Handout of Stanford's course. **ISR:** Pp. 514 - 519 (k-means)	Maximum likelihood estimation basics.	Hierarchical clustering, DBScan, kernel density estimation
12	EM algorithm, Gaussian mixture models, dimensionality reduction	Gaussian Discriminant Analysis (GDA) (including linear and quadratic discriminant analysis). From GDA to Gaussian Mixture Model (GMM). Expectation-Maximization (EM) algorithm for GMM. Compare and contrast k-means clustering and GMM clustering. GMM for anomaly detection. PCA: motivation, formulation, reconstruction, choosing k.	**SUCS229:** Notes (Pages 35 - 40, 148 - 151) **SUCS229:** Notes (Pages 165 - 170) **PAN:** Lec 14 **ISR:** Pp. 252 - 256 (PCA) **[Optional] ISR:** Sections 4.4.1, 4.4.2, 4.4.3 (GDA), Pp. 495 - 508 (PCA).	Gaussian Discriminant Analysis basics. Basics of change of basis of a vector space, eigen vectors.	Details of Jensen's inequality, general EM algorithm, convergence of EM algorithm. Factor analysis. Independent component analysis.

13	k-NN, Decision Tree, Bagging, Random Forest, Boosting	Non-parametric methods for supervised learning: k-nearest neighbor method: training and prediction. From k-nn to decision tree (for both regression and classification): motivation, training, prediction, node impurity measures for classification, pruning principle, benefits and drawbacks. Bagging and random forest for variance reduction, boosting for bias reduction.	**ISR:** Pp. 39 - 42 (k-nn), Pp. 329 - 348 (trees and tree ensembles)	Basics of Information theory.	Variants of random forest and boosting. Ensembling other ML algorithms. Details of bagging and random forest (My paper on RF).
	[OPTIONAL] Theory of ML		Handout, My paper on bias-variance (Sections 2 and 4)		