CSXXX	Machine Learning	3L:1T: 0P	4 credits	Pre-Reqs:

Learning outcomes of course (i.e., statements on students' understanding and skills at the end of the course the student shall have):

Essential (<=6):

1. Understanding popular ML algorithms with their associated mathematical foundations for appreciating these algorithms.
2. Capability to implement basic algorithms using basic machine learning libraries mostly in python. Gain hands-on experience in applying ML to problems encountered in various domains. In addition, obtain exposure to high-level ML libraries or frameworks such as TensorFlow, PyTorch.
3. Make aware of the role of data in the future of computing, and also in solving real-world problems using machine learning algorithms.
4. Help connect real-world problems to appropriate ML algorithm(s) for solving them. Enable formulating real world problems as machine learning tasks.
5. Appreciate the mathematical background behind popular ML algorithms.
6. Ensure awareness about importance of core CS principles such as algorithmic thinking and systems design in ML

Desirable/Advanced (<= 3):

1. Have sound mathematical understanding of popular ML algorithms
2. Preparedness to use state of the art machine learning algorithms in formulating and solving new problems.
3. Capability to train (or solve optimization problems) ML models with applications in real-world use cases.

Detailed contents for essential learning outcomes:

Module (appx duration of 3 weeks or 9-12 Hrs)	**Topics**	**Pedagogy / teaching suggestions**	**Nature of lab / assignment / practice**
Introduction to ML	(i) Motivation and role of machine learning in computer science and problem solving (ii) Representation (features), linear transformations, Appreciate	(i) Connect machine learning to the broader theme of Computer Science (ii) Expose broad canvas of machine learning; brief history and importance	(i) Experiments/notebooks/code that refresh Python, programming frameworks used for the course (ii) Experiments/Code that allows students to appreciate mathematics and data

	linear transformations and matrix vector operations in the context of data and representation. (iii) Problem formulations (classification and regression). (iv) Appreciate the probability distributions in the context of data, Prior probabilities and Bayes Rule. (v) Introduce paradigms of Learning (primarily supervised and unsupervised. Also a brief overview of others)	(iii)Role of data, Connection to the knowledge /experience in learning. (iv)Show successful examples of machine learning in Industry/working (v) Motivate students by Showing how ML and Data driven solutions could help in our day to day problems around (interdisciplinary such as agriculture, healthcare, education, living etc.) (vi)Refresh the basic mathematical notions that students may know (vectors, matrices, probabilities, etc.) with examples in ML (vii) Make students aware of relevant topics like "what is learnable"? And "what are the disadvantages of data driven solutions".	manipulation. Appreciate (a) Features, Representation of the data/real-world phenomena (b) mathematical operations or transformations that manipulate the data (c) plot/visualise the data distributions (say in 2D) (d) eigen values, eigen vectors, rank of matrices. (iii)Lab/Experiments that appreciate the problem of Classification and problem of Regression (iv)Lab/Experiments that appreciates the notions related to "Training" and "Testing" by considering algorithms like decision trees, nearest neighbour as black boxes.
Fundamentals of ML	(i) PCA and Dimensionality Reduction, (ii) Nearest Neighbours and KNN. (iii) Linear Regression (iv) Decision Tree Classifiers (iv) Notion of Generalization and concern of Overfitting (v) Notion of Training, Validation and Testing; Connect to generalisation and overfitting.	(i)Focus on mathematical and algorithmic precise description of the content. (ii)Insights into these algorithms, why? When? What are the limitations? Why multiple algorithms exists for a specific problem (iii)Insights into the notion of generalization. Challenges for generalization. Assumptions to make (iv)Practical insights and tops on avoiding overfitting.	(i) Dimensionality Reduction using PCA and its applications in (a) removing irrelevant features (b) compression /compaction (c) efficient ML pipeline (ii)Experiment related to Nearest neighbour classifier, (a) visualize the decision boundaries (b) appreciate the role of hyperparameter K. Role of validation data in choice of hyper parameters (iii) Decision Tree as a classifier and see the overfitting with "deep" trees. How the overfitting can be controlled by seeing validation performance during the training.
Selected Algorithms	(i) Ensembling and RF (ii) Linear SVM, (iii) K Means, (iv) Logistic Regression (v) Naive Bayes	(i) Make students appreciate the role of optimization in machine learning. Challenges in optimization and why we are sometimes happy with sub-optimal solutions. How assumptions make the algorithms simple/tractable. (ii) Make students appreciate the role of uncertainty in data and machine learning problems/solutions. Give probabilistic insights into Loss functions (em MSE, cross entropy)	(i)Experiments related to K-Means, by varying in "K", "initialization". How the "analysis of the algorithm" can be seen in the lab (eg. change of objective across iterations). Try multiple datasets. Appreciate that "unsupervised discovery" makes sense in the problem under consideration. (ii)An experiment that demonstrates how SVM can yield

		(iii)Introduce iterative algorithms, convergence, role of initialization etc. in a class of ML solutions. (iv)Connect the geometric view of Margin (eg, linear SVM) and Probabilistic View of Margin (Logistic Regression) and the need of Generalization	a solution better than a simple linear separating solution. Appreciate the role of support vectors. Appreciate how SVMs extend to problems even if data is not linearly separable. (iii)An experiment that makes students appreciate the utility of naive Bayes classifier in practice (say designing a text classifier).
Neural Network Learning	(i) Role of Loss Functions and Optimization, (ii) Gradient Descent and Perceptron/Delta Learning, (iii) MLP, (iv) Backpropagation (v) MLP for Classification and Regression, (vi) Regularisation, Early Stopping (vii) Introduction to Deep Learning (viii) CNNs	(i)Appreciate (a) the neuron model (b) the neural network and its utility in modelling and solving the problem. Connect to the biological motivations and parallelism. (ii)Expose the simple elegant optimization scheme of gradient descent with associated mathematical rigour and insights. (iii)Expose the practical issues in extending GD to multiple layers and how the backpropagation algorithm efficiently computes the gradients. (iv)Expose the practical challenges in training a neural network (such as non-convexity, initialization, size of data, number of parameters) and how they are taken care of in the practical implementations of today. (v)Appreciate the need for empirical skills in training neural networks.	(i)Experiment that exposes the GD and BP in simple neural networks. Show the learning process (graphs) and performances. (ii)Experiment that use a modern library and implementation of a deep neural network, expose computational graphs, expose the generalized way of appreciating BP as a learning algorithm in Deep Neural Networks (iii)Experiment that uses a popular CNN architecture for practical application (say image classification). (iv)Experiments that strengthen the empirical skills in training with (a) initializations (b) update strategies (c) regularisation (d) multi fold validation on a small/medium size deep neural network that can be trained in 5 minutes.

Detailed Contents for Desirable Learning Outcomes (optional, <= 3 modules):

Module	**Topics**	**Pedagogy teaching suggestions**	**Nature of lab / assignment / practice**
Key Concepts from ML	Kernels (with SVM), Bayesian Methods, Generative Methods, HMM, EM, PAC learning	Focus on mathematical and analytical skills. Expose the intuition behind these algorithms. Introduce	Python notebooks that demonstrate the use of these algorithms on public datasets

		analysis of machine learning using a PAC model.	
Deep Learning Architectures	Popular CNN Architectures, RNNs, GANS and Generative Models,	Introduce popular architectures, models, and the use of it in various settings.	1. Use of popular architectures for pretrained features and transfer learning 2. Use of RNNs in learning “language models” in large text corpus (charRNN) 3. Capability and practical challenged in working with GANS
Training Todays Neural Networks	Advances in Backpropagation and Optimization for Neural Networks Adversarial Learning	Appreciate the challenges in large non-convex optimization and how many of today's design choices have helped.	1. See how (i) initialization (ii) momentum (iii) update rules have helped in getting better minima/soln. 2. Experience how regularisation helps in avoiding overfitting and getting better solutions

Notes on Exercises/Labs/Homeworks

1. Use iPython/Jupyter notebooks to hand out assignments; such notebooks allow embedding instructions/videos/etc, making it easy for instructors, students as well as TAs
2. It may be good to have both theory and programming components in the assignment/homework component, to allow students to appreciate and learn both aspects of machine learning
3. Based on target audience, have a healthy mix of using an off-the-shelf machine learning library (e.g. using sklearn’s decision tree function) and writing an algorithm from scratch (e.g. coding up decision tree from scratch)
4. Consider having a Kaggle-style hackathon in the course, where students get used to competitive machine learning

Notes on Sequencing Lectures:

1. Consider your target audience (undergrad vs grad; 2nd-year vs 3rd/4th-year students;etc) in deciding the prerequisites that the students may fulfil, and also the topics that you may want to cover.
2. It may be worthwhile visiting publicly available course outlines to verify and confirm that no important topics are missed out from both fundamental and contemporary perspectives (it is important to keep both these considerations in mind – to raise the spark of fundamental curiosity and also to provide utilitarian value)..

3. Sequencing can be done considering the target audience in mind – one option may be to start with simple algorithms and gradually move towards mathematically involved ones.
4. Sequencing can also be modified based on whether there are any other machine learning or AI or related courses in the curriculum/elective list.

Suggested text books

T1 Marc Peter Deisenroth, A. Aldo Faisal, Cheng Soon Ong, **Mathematics for Machine Learning,** Cambridge University Press (23 April 2020)

T2 Tom M. Mitchell- **Machine Learning** - McGraw Hill Education, *International Edition*

T3 Aurélien Géron **Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow,** O'Reilly Media, Inc. 2nd Edition

Reference Books:

R1 Ian Goodfellow, Yoshoua Bengio, and Aaron Courville **Deep Learning** MIT Press Ltd, *Illustrated edition*

R2 Christopher M. Bishop **Pattern Recognition and Machine Learning** - Springer, *2nd edition*

R3 Trevor Hastie, Robert Tibshirani, and Jerome Friedman - **The Elements of Statistical Learning: Data Mining, Inference, and Prediction** - Springer, *2nd edition*

Prepared by:

1. Mitesh Khapra *IIT Madras*
2. Vineeth N Balasubramanian *IIT Hyderabad*
3. Preethi Jyothi *IIT Bombay*
4. Pabitra Mitra *IIT Kharagpur*
5. Chetan Arora *IIT Delhi*
6. P J Narayanan *IIIT Hyderabad*
7. C. V. Jawahar *IIIT Hyderabad*