Fusemachines Research and Training Center

Syllabus

Machine Learning

Microdegree in Artificial Intelligence Program

Version	Significant Changes (Marked with a Symbol)	Modified by	Modification Date
1.0	First Version of the course	Fuse.ai Content Team	

Contents

Introduction

The syllabus of this course is updated regularly to keep up with the current state-of-the-art techniques and practices employed in academia as well as the industries.

Syllabus Aims

The Syllabus aims to:

- Provide a worthwhile learning experience for all learners and enable them to acquire sufficient knowledge and skills to get started in the domain of AI in both academic research and industrial application
- Facilitate and Standardise Course Content Development and Delivery

Course Objectives

After Completion of the course students will be able to

- Select and implement appropriate libraries, framework, techniques for different problems
- Explain the math and code behind some algorithms and build and make changes to the code
- Assess the performance, evaluate and compare different models to design and deploy an end-to-end solution
- Run experiment to change some details in code to improve the algorithm

Guided Sessions

Session	**Units Covered**	**Teaching Guide (Instructions for Instructors)**	
1	Course Orientation and M1		
2	Linear Regression,		
3			
4	Logistic Regression, Metrics		
5	Overfitting and Regularisation		
6			
7			
8			

9			
10			
11			
12	Mid-Examination		
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23	Final Project Presentation		
24	Examination		

References

Code	Reference	Link
PRML	Pattern Recognition and Machine Learning, by Christopher Bishop	Pattern Recognition and Machine Learning
ESL	The Elements of Statistical Learning: Data Mining, Inference, and Prediction by Hastie, Tibshirani and Friedman	Elements of Statistical Learning: data mining, inference, and prediction. 2nd Edition.

Pre-requisites

Assessment

The Assessment will be based on the cognitive domain of Bloom’s Taxonomy to classify learning objectives into different levels of complexity and specificity, viz. Remember, Understand, Apply, Analyze, Evaluate, Create

Assessment Objectives	**Categories**	**Objective**	**Action Words**
AO1	**Remember**	Recall Facts and basic concepts	Define, list, memorize, repeat, state, recognize
AO2	**Understand**	Explain Ideas or Concepts	Classify, describe, discuss, explain, identify, locate, recognize, report, select, translate, interpret, exemplify,
AO3	**Apply**	Use Information in a new situation	Execute, implement, solve, use, demonstrate, interpret, operate, schedule, sketch
AO4	**Analyze**	Draw connections among ideas	Differentiate, organize, relate, compare, contrast, distinguish, examine, experiment, question, test
AO5	**Evaluate**	Justify a stand or decision	Check, Appraise, argue, defend, judge, select, support, value, critque, weigh
AO6	**Create**	Produce new or original Work	Design, assemble, construct, conjecture, develop, generate, plan, produce, formulate, investigate

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

Weight Distribution on

Components	**AO1**	**AO2**	**AO3**	**AO4**	**AO5**	**AO6**	**Weights**
Quizzes	20%	20%	20%	20%	10%	-	
Programming Assignment	10%	20%	20%	20%	10%	10%	
Projects	-	-	10%	20%	20%	50%	
Classroom Assessment							

Prerequisites

Distinct Features Used in the Syllabus

Bold Outcomes refers to **Must Have** learning outcomes (Bare Minimum Criteria),

Normal Text refers to Should Have learning outcomes,

Green refers to Good to Have or Higher Level Learning Outcomes for high achieving students.

Cyan Highlight refers to Implementation Learning Outcomes

Red Outcomes refers to kept in the course but optional Outcomes completely ignorable.

Chapter Number #.#.#, Chapter Name	Learning Objectives follows Students should be able to: A. Answer this question Can Include additional contents that are available in platform or for teaching	Resources for reference Also include guide for teaching the particular content

Course Contents

Module 1. Introduction to the Course

1.1. Introduction to the Course

Students will be able to:

1.1.1. Introduction to the Course	A. Discuss what is meant by learning and why do we want machines to learn ? B. Describe what a machine learning problem and solution look like. (eg: sales prediction, disease classification, movie rating and recommendation)	- Why do we need machines to learn? - What is machine learning and how is it being applied in the real world - Different types of examples - Road map of the course and What students will be able to do after completing this course, and what kind of jobs they can apply -
1.1.2. Course Logistics	A. Restate the prerequisites required for the course B. Show awareness about the expectations from the student and the overall time commitment C. Recall and follow the grading and honor code policy	

1.2. Introduction to Machine Learning

Students will be able to:

1.2.1. Introduction to the Machine Learning	A. Define a well posed learning problem by choosing an appropriate experience, tasks and performance measure as defined by Tom Mitchel B. Discuss types of Reasoning: deductive, abductive and inductive. ML is based on inductive reasoning (specific to generality) vs Deductive Reasonin C. Describe and differentiate between prediction, inference, and description D. Describe model representation and the process of learning (Learning = Representation + Evaluation + Optimization) E. Describe and recognize the type of Machine Learning a. Supervised Learning b. Unsupervised Learning c. Reinforcement Learning	- Tom Mitchel Definition of ML with emphasis on each term with examples - - Analogy of a kid learning to recognize object given examples, or ability to group objects with similarity or learn to do things by failing. -
Possible Assignment	Different Sources of Data, Data Engineering (convert platform data to analyzable data frames) Data Wrangling, Collecting data from various sources, And Data Visualisation	

Module 2. Supervised Machine Learning

2.1. Introduction to Supervised ML

Students will be able to:

2.1.1. Introduction to Supervised ML	A. Define supervised Machine Learning with examples B. Differentiate between Regression and Classification problems C. Different types of models a. Linear and Non-linear Models b. Probabilistic and non-probabilistic	- -

2.2. Linear Regression

Students will be able to:

2.2.1. Least Squares Estimation	A. Define linear regression problem: for one or more variables. B. Discuss the key assumptions of linear regression C. Discuss similarity of regression with function approximation D. Deduce least square regression model and explain the E. shortcoming of the model F. Normal Equations	
2.2.2. Gradient Descent	A. Explain why use Sum of Squared Error vs absolute error B. Gradient Descent a. Stochastic and Batch Gradient Descent C. Implement Gradient Descent to estimate parameters in Linear regression	
2.2.3. Regularised Regression and other Methods	A. Explain the concept of Overfitting and Underfitting B. Describe and use the following Regularisation techniques C. L1/LASSO, L2 / Ridge Regression, Elastic Net a. Basics implementation in sklearn b. D. Other Methods: [Just Introduction] a. FISTA b. ADMM c. Conjugate Gradient	
Assignment	Implement Linear Regression from Scratch as well as Sklearn	

2.3. Logistic Regression

Students will be able to:

2.3.1. Logistic Regression	A. Discuss how logistic regression differs from linear	https://florianhartl.com/logistic-

	regression with examples of why and where linear regression won't work and how logistic regression is more suitable. B. Geometrical Interpretation ~~C. Recall concept of overfitting and underfitting and apply L1 and/or L2 Regularisation:~~	regression-geometric-intuition.html http://www.cs.cmu.edu/~tom/mlbook/NBayesLogReg.pdf
2.3.2. Probabilistic view of Logistic Regression [Optional]	A. Define Logistic Function (Sigmoid) and Logit B. Probabilistic view point of Logistic Regression model a. Likelihood function b. how to find parameters(gradient Descent) c. Implementation from Scratch and ScikitLearn C. Recall concept of overfitting and underfitting and apply L1 and/or L2 Regularisation:	https://medium.com/@premvardhankumar/a-deep-understanding-of-logistic-regression-with-geometric-probabilistic-and-loss-minimization-2ced042bdcc7 1. CS229 Lectures(Andrew NG) - http://cs229.stanford.edu/notes/cs229-notes1.pdf
Assignment	Fake Advertisement Dataset - Whether a user clicked on the data or not. Data- https://www.kaggle.com/sazid28/advertising.csv, Or using Amazon Fine food Review dataset (needs preprocessing, preprocessing step would be skipped) Data -https://www.kaggle.com/snap/amazon-fine-food-reviews	

2.4. Neural Network

Students will be able to:

2.4.1. Introduction to Neural Network	A. Describe basics of neural network and correlate it with biological neurons (this may also be redundant because of deep learning content) B. Explain the feature learning introduced by neural nets with examples C. Examine and recognise the problems where the use of Neural network is appropriate	***Neural net in nutshell:*** a. Chapter with Neural nets
2.4.2. Perceptron and PLA	A. Describe Rosenblatt's perceptron as a linear unit with a step activation function B. Explain how Perceptron Learning Algorithm helps to learn the parameter for the perceptron C. Explain why perceptron fails in XOR Problem and other limitations of the perceptron D. Implement Perceptron from scratch [Assignment]	a. Perceptron github link b. Perceptron from scratch link c. Perceptron implementation link d. Perceptron limitation slide link
2.4.3. MLP	A. Explain how neural networks can represent the nonlinear functions by stacking the layers of	a. MLP deeplearning link b. MLP implementation

	artificial neurons with non-linear activation functions B. Understand MLP with code implementation (if possible) C. Construct a feed forward network D. Introduce vanilla RNN as a different neural network than feed forward network [Among Neural Network Zoo]	link c. Feed forward network link
2.4.4. Back propagation	A. Sketch and explain the concept of a computational graph to explain and implement the process of forward propagation B. Implement Forward Propagation from Scratch C. Describe how errors propagate back through the computation graph using the chain rule. (auto differentiation) D. Sketch and estimate the error propagated to each node during back propagation on simple computation graph E. Implement Back Propagation from Scratch	***Medium blog:*** a. Gradient Descent link b. Backpropagation link c. Feed - forward from scratch d. Forward backprop implementation link ***Hackernoon blog:*** Gradient Descent link ***Others:*** Backprop link feedforward implementation link
2.4.5. Stochastic Gradient Descent	A. explain concept of gradient descent to reduce the error function by changing the parameters (moving down the loss landscape) B. Derive the Gradient Descent rule for Linear Regression C. Define the terminologies associated with the learning the network: a. epoch, batch size, learning rate. b. stochastic vs mini-batch vs batch gradient descent c. Local vs global minimum D. Train a linear unit with Gradient Descent	
Assignment		

2.5. Decision Trees

Students will be able to:

2.5.1. **Introduction to Decision Tree**	A. Describe working of decision trees B. Explain the terminologies of root, internal and leaf nodes C. Ability to handle different forms of data (binary, nominal, ordinal, and continuous attributes) D. Criteria for node splitting a. Recall concept of Entropy b. Information Gain (Gain Ratio) c. Gini Index d. Likelihood-Ratio Chi–Squared Statistics E. Select appropriate parameters as stopping criteria	References: https://www-users.cs.umn.edu/~kumar001/dmbook/ch4.pdf http://www.ise.bgu.ac.il/faculty/liorr/hbchap9.pdf

	for learning in decision tree a. Impurity Threshold b. Depth c. Minimum samples in leaf or min sample in node ,.... F. Scikit learn explanation for different parameters	
2.5.2. Pruning Methods	A. Explain the importance of pruning in decision tree construction B. Explain different pruning methods a. Cost–Complexity Pruning b. Reduced Error Pruning c. Minimum Error Pruning d. Error–based Pruning (EBP)	See above
2.5.3. Decision Tree Inducers	A. Algorithm for decision tree inducer B. Brief explanation of decision tree inducers a. ID3 b. C4.5 c. CART C. Discuss Issues with decision tree induction	References:See above
Assignment	Predicting whether a person will die or survive using the titanic dataset, Write a report on analysis playing with parameters on the decision tree.	Dataset: https://www.kaggle.com/c/titanic/data

2.6. Instance Based Learning

Students will be able to:

2.6.1. Instance-based learning	A. Differentiate between model-based and Instance-based learning (a.k.a memory-based learning) and discuss their use cases B. Discuss characteristic advantages and disadvantage of Instance based learning C. state different instance based algorithms, (KNN, RBF, Kernel based Methods)	https://cogsys.uni-bamberg.de/teaching/ss05/ml/slides/cogsysII-8.pdf
2.6.2. Nearest Neighbors algorithm	A. Introduction with some use cases B. Discuss about different variants of Nearest Neighbour Algorithms a. K-Nearest Neighbor b. Approximate Nearest Neighbor C. Distances D. Curse of dimensionality E. Use different methods to choosing value of k such as Elbow method F. Implement KNN from Scratch	Thoughtful Machine Learning with Python- Matthew Kir's book Wiki for other resources: https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm
Assignment	Geospatial data ?	

2.7. Support Vector Machines (SVM)

Students will be able to:

2.7.1. Introduction	A. Describe Margin, Support Vectors, B. Explain Soft and Hard Margin C. Implement SVM for linear separation (separable Case) D. Optimal Hyperplanes	https://monkeylearn.com/blog/introduction-to-support-vector-machines-svm/
2.7.2. Kernels	1. Introduce Kernel 2. Kernel Tricks 3. Properties of Kernels 4. Various Kernels a. Polynomial kernels b. Gaussian RBF Kernel	http://cs229.stanford.edu/notes2019fall/cs229-notes3.pdf
2.7.3. Support Vector Machines	1. The Non-Separable Case 2. Explain SVM as the non-linear classifier 3. Primal and Dual Solutions (Lagrangian Solution) 4. The Karush-Kuhn-Tucker Conditions ~~5. Support Vector machine Algorithm~~	
2.7.4. Extensions to the SVM	A. Regression with SVM B. Differentiate Binary Classification and Multi-class Classification C. Approach for Multiclass Classification a. One-vs-all b. One-v-one	
Assignment	Use libsvm, sklearn to work on Spiral dataset, show both linear and nonlinear separation	

2.8. Ensemble Methods

Students will be able to:

2.8.1. Introduction to Ensemble Methods	A. Describe what is ensembling B. Understand Bias-variance decomposition C. Explain strong (stable) and weak(unstable) classifiers D. Discuss what are the main challenges for developing ensemble methods	
2.8.2. Bagging	A. Understand Bootstrapping B. Parallelization during model training	Ensemble methods: bagging, boosting and stacking
2.8.3. Boosting	A. Understand boosting and how it differs from bagging. B. Describe Adaptive and Gradient Boosting C. Sequential during model training	The Elements of Statistical Learning Stochastic Gradient Boosting
2.8.4. Stacking	A. Understand the concept of heterogeneous weak learners. B. Describe stacking C. Differentiate bagging, boosting and stacking.	
2.8.5. Popular Ensemble Methods	A. Describe and use **Random Forest** B. XGBOOST a. Explain xgboost as a software library b. Features(model feature, system feature,	https://medium.com/@williamkoehrsen/random-forest-simple-explanation-377895a60d2d

	algorithm features)	https://machinelearningmastery.com/gentle-introduction-xgboost-applied-machine-learning/
Assignment	XGboost eg: Sales prediction ?	

2.9. Probabilistic Models

Students will be able to:

2.9.1. Bayes Theorem	A. Differentiate Probabilistic and non-probabilistic models B. Solve simple problems on conditional Probability C. Derive Bayes Theorem by applying chain rule on conditional probability D. Describe each term (Posterior, Prior, Likelihood, and Evidence) in the Bayes rule and their significance E. Recall Bayes theorem to Solve probability problems (eg. probability quite accurate lab test to detect a very rare disease) F. Use bayes theorem for inference in parameter estimation G. Introduce Gaussian Process	***Bayes Theorem*** https://brilliant.org/wiki/bayes-theorem/ ***Derivation:*** https://www.hackerearth.com/practice/machine-learning/prerequisites-of-machine-learning/bayes-rules-conditional-probability-chain-rule/tutorial/ ***Prior probability*** https://en.wikipedia.org/wiki/Prior_probability ***Posterior Probability*** https://en.wikipedia.org/wiki/Posterior_probability ***Likelihood*** https://en.wikipedia.org/wiki/Likelihood_function ***Marginal Likelihood*** https://en.wikipedia.org/wiki/Marginal_likelihood ***Bayesian Inference*** https://towardsdatascience.com/probability-concepts-explained-bayesian-inference-for-parameter-estimation-90e8930e5348
2.9.2. **Naive Bayes**	A. Optimal Bayes Algorithm B. Assumptions of Naive Bayes C. Pros and Cons of Naive Bayes D. Describe a. Gaussian Naive Bayes b. Multinomial Naive Bayes c. Bernoulli Naive Bayes	***Optimal Bayes*** ***https://www.edx.org/course/machine-learning*** ***Naive Bayes*** https://www.machinelearningplus.com/predictive-modeling/how-naive-

		bayes-algorithm-works-with-example-and-full-code/ https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/ ***Assumptions*** https://towardsdatascience.com/machine-learning-part-16-naive-bayes-classifier-in-python-c9d3fa496fa4 ***Types of NB*** https://scikit-learn.org/stable/modules/naive_bayes.html
2.9.3. Probabilistic Models	A. Maximum Likelihood Estimation a. Define independent and identically distributed (iid) assumption and its consequence b. MLE approach B. Bayesian Parameter Estimation a. The Full Bayesian Approach b. Maximum a-posteriori (MAP) approximation	http://www.cs.toronto.edu/~rgrosse/csc321/probabilistic_models.pdf
2.9.4. Bayesian Networks	A. Brief introduction and explanation about the Bayesian networks B. Inference and learning C. Dynamic Bayesian Networks	https://www.coursera.org/learn/probabilistic-graphical-models
2.9.5. Markov Networks	A. Brief introduction and explanation about the Markov networks B. Inference	https://www.coursera.org/learn/probabilistic-graphical-models
Assignment	PGM	

2.10. Module Summary

Students should be able to :

2.10.1. Module Summary	A. Compare and contrast different algorithms covered this module B. discuss advantage and disadvantages (and/or challenges) of different algorithms C. Discuss best use cases for each of the algorithms covered D. Integrate the concepts covered in this module with other topics covered in the course	

Module 3. Unsupervised Machine Learning

3.1. Introduction to Module

Students will be able to:

3.1.1. Introduction to unsupervised learning	A. Describe and exemplify unsupervised learning problem (eg, Netflix prize, Google News Aggregator) B. Describe how unsupervised methods are used as pre-processing step	

3.2. Dimensionality Reduction

Students will be able to:

3.2.1. Singular Value Decomposition (SVD)	A. Describe SVD a. SVD approximates any matrix X as sum of rank one uv^T matrices b. Rank r approximation of matrix X as stated by Eckart-Young Theorem c. SVD interpretation as correlation among columns of X and Rows of X B. Discuss about various application of SVD a. Dimensionality Reduction b. Image Reconstruction (Noise Removal, Deblurring,) c. Describe how SVD helps to find similarity(correlation) between variables(e.g. Recommender system)	
3.2.2. Principal Component Analysis (PCA)	A. Explain curse of dimension and the importance of dimensionality reduction B. Describe eigenvalue decomposition and its statistical properties C. Steps for principal components analysis. D. Discuss real life applications of PCA (eg. image compression)	1. https://towardsdatascience.com/a-one-stop-shop-for-principal-component-analysis-5582fb7e0a9c 2. https://setosa.io/ev/principal-component-analysis/ 3. https://www.youtube.com/watch?v=FgakZw6K1QQ 4. https://www.youtube.com/watch?v=FYKcBpzszmw
3.2.3. Random Projection	A. Discuss the computational complexities of PCA, SVD. B. State Johnson-Lindenstrauss Lemma C. Explain Random Projection	Random projection in dimensionality reduction: Applications to image and text data
3.2.4. Autoencoder	A. Neural network based dimensionality reduction a. Encoder b. Decoder B. Comparison with PCA	Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion

3.2.5. Probabilistic PCA	A. Limitation of PCA and motivation behind probabilistic PCA B. Latent variable models C. Probabilistic PCA	
3.2.6. Manifold learning	A. Limitations of PCA in datas with non-linear relationship B. Describe how Kernel PCA differs from linear PCA and discuss its application C. Describe Manifold learning D. Application and Limitations of Manifold learning	1. https://jakevdp.github.io/PythonDataScienceHandbook/05.10-manifold-learning.html 2. https://towardsdatascience.com/manifold-learning-the-theory-behind-it-c34299748fec 3. https://www.youtube.com/watch?v=j8080l9Pvic 4. https://en.wikipedia.org/wiki/Nonlinear_dimensionality_reduction
3.2.7. TSNE	A. Compare PCA and TSNE B. Describe how TSNE works and its use cases C. Describe data visualization with TSNE D. Limitation of TSNE	1. https://distill.pub/2016/misread-tsne/ 2. https://towardsdatascience.com/an-introduction-to-t-sne-with-python-example-5a3a293108d1 3. https://en.wikipedia.org/wiki/T-distributed_stochastic_neighbor_embedding 4. https://www.youtube.com/watch?v=NEaUSP4YerM 5. https://www.youtube.com/watch?v=PTNfRk0v6YM
3.2.8. Matrix Factorisation	A. Matrix Factorization(LU) B. Probabilistic Matrix factorisation C. Relation between matrix factorisation and Ridge regression D. How SVD can be used to get Word Embeddings. (Truncated SVD) E. Discussion on Topic modeling and LDA F. Non-negative Matrix Factorization(NMF) G. Application of Matrix Factorisation	1. How does Netflix recommend movies? Matrix Factorization 2. https://github.com/opokualbert/Topic-Modeling-with-Non-Negative-Matrix-Factorization-NMF-
Assignment	PCA and TSNE in Higher dimensional dataset, (protein dataset)	

3.3. Clustering Algorithms

Students should be able to :

3.3.1. Introduction to Clustering	A. Describe Clustering B. Classify different types of Clustering algorithms a. Partitioning-based b. Hierarchical-based c. Density-based d. Grid-based e. Model-based C. Hard and Soft Clustering D. Application of Clustering	https://www.analyticsvidhya.com/blog/2016/11/an-introduction-to-clustering-and-different-methods-of-clustering/ https://towardsdatascience.com/the-5-clustering-algorithms-data-scientists-need-to-know-a36d136ef68
3.3.2. Partitioning-based Clustering	A. Describe **K-means** clustering algorithm B. Explain the importance of choosing initial centroids C. Elbow method for finding suitable cluster method. D. Limitation of K means(problem when cluster differs in size, density and non-globular shapers) E. Bisecting K means: An extension of basic k means F. State Different types of partitioning based methods	https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e644b48a
3.3.3. Hierarchical-based Clustering	A. Understand dendrogram as a visualization tool to see how data contributes to an individual cluster. B. Two types of clustering a. Agglomerative b. Divisive C. Hierarchical algorithm a. Start with example b. Compute proximity matrix I. Defining inter-cluster similarity (Few approach: MIN, MAX, Group Average, Distance between centroid etc) c. Form first level cluster d. For intermediate proximity matrix for cluster e. Form second level cluster D. How hierarchical clustering helps to select the optimal number of clusters.	https://towardsdatascience.com/understanding-the-concept-of-hierarchical-clustering-technique-c6e8243758ec https://www.analyticsvidhya.com/blog/2019/05/beginners-guide-hierarchical-clustering/
3.3.4. Density-Based Clustering	A. Explain importance of density based clustering with respect to other types of clustering B. Terminologies a. Core Point b. Border Point c. Nose Point C. Describe Original Query-based DBScan algorithm D. Advantages and disadvantages of dbscan	https://www.geeksforgeeks.org/dbscan-clustering-in-ml-density-based-clustering/ https://blog.dominodatalab.com/topology-and-density-based-clustering/
3.3.5. Cluster Validation and Usecases	A. Numerical measure of cluster validity a. External Index(Entropy and Purity) b. Internal Index(Sum of Squared Error) i. Cluster Cohesion ii. Cluster Separation (Dunn Index) B. Use cases of different Clustering Techniques	BOOK: Introduction to the data mining by PANG-NANG TAN

Assignment	Iris ?	

3.4. Gaussian Mixture Models

Students will be able to:

3.4.1. Expectation Maximization	A. Recall Maximum likelihood B. Describe Expectation maximization algorithm C. Implement EM algorithm	***PRML link*** ***(good to follow this book)***
3.4.2. Gaussian Mixture Models	A. Explain how GMM solves problems in hard clustering algorithms such as K-means B. Generate data from mixture of models C. Definition and simple derivation D. Implement Maximum likelihood EM for the GMM E. Applications of GMM in real world problems	***Blog:*** GMM explained link ***Wiki:*** Expectation maximization link ***GMM paper link:*** GMMs paper link ***Youtube link:*** GMM explanation GMM with mathematical concept
Assignment		

3.5. Module Summary

Students should be able to :

3.5.1. Module Summary	A. Compare and contrast different algorithms covered this module B. discuss advantage and disadvantages (and/or challenges) of different algorithms C. Discuss best use cases for each of the algorithms covered D. Integrate the concepts covered in this module with other topics covered in the course	

Module 4. Feature and Data Engineering

Learn about the advantages and limitations of each technique and their underlying assumption, when to use them and implement them.

4.1. Introduction to Feature and Data Engineering

Students will be able to:

4.1.1. Data Engineering	A. Introduction to Data Engineering B. Differentiate between data engineers and data scientist C. Describe different roles of data engineers D. Different tools used by data engineering for a. Database Management: MySQL, Postgres b. Data Processing Tools: Spark, Hive c. Scheduling Tools: Airflow, oozie, cron	References https://towardsdatascience.com/who-is-a-data-engineer-how-to-become-a-data-engineer-1167ddc12811 https://www.dataquest.io/blog/what-is-a-data-engineer/ https://www.springboard.com/blog/data-engineer-vs-data-scientist/
4.1.2. Feature Engineering	A. Explain the importance of Feature engineering to improve ML model, despite feature learning in Deep learning B. Discuss different sources of data and data availability C. Describe different types of variables: Numerical (Continuous/Discrete), Categorical (Nominal/Ordinal), Dates, Text, Mixed D. Describe different variable Characteristics: Missing data, Cardinality, Frequency of labels, distributions, outliers, feature magnitude, which characteristics are compatible while which will affect the ML model	https://elitedatascience.com/feature-engineering

4.2. Missing Data Imputation

Students will be able to:

4.2.1. Missing Data Mechanism	A. Discuss types of Missing Data (Ignorable and Non-Ignorable) and different methods to handle these errors a. MAR b. MCAR c. MNAR	
4.2.2. **General Missing Data Imputation**	A. Deletion a. Deleting rows b. Pairwise deletion c. Deleting columns B. Categorical a. Missing Indicator (Binary Variable) or NA C. Continuous a. Mean, Median, Mode Imputation b. Random sample Imputation c. Arbitrary Value imputation d. Imputation using Regression e. MICE Multivariate Imputation using Chained Equations) f. Others	http://www.stat.columbia.edu/~gelman/arm/missing.pdf Book: Python-Feature-Engineering-Cookbook-Over-70-recipes-Galli.pdf
4.2.3. Time-Series data Imputation	A. Data without Trend and Seasonality (Mean, Median, mode, Random Sample Imputation) B. Data with Trend and without Seasonality (Linear	

	Interpolation) C. Data with Trend & with Seasonality (Seasonal adjustment + interpolation)	

4.3. Feature Transformation

Students will be able to:

4.3.1. **Categorical Variable encoding**	A. Describe what is data transformation and its importance B. Describe and use a. One hot encoding b. Ordinal encoding c. count/frequency encoding d. Mean/Target Encoding e. Weight of Evidence f. Rare Label Encoding g. Feature Hashing and Binary Encoding	
4.3.2. Mathematical Transformations	A. Describe the methods their, use cases, advantages, and disadvantages of different mathematical transformation methods a. Logarithmic b. Reciprocal c. Exponential d. Power e. Yeo-Johnson f. Box-Cox	
4.3.3. Feature Selection	A. Discuss the importance of feature selection B. Feature Selection Algorithms a. Filter Methods i. Pearson Correlation ii. Chi squared test b. Wrapper Methods i. Recursive Feature Elimination c. Embedded Methods i. Lasso, Ridge and Elastic net regression.	https://machinelearningmastery.com/an-introduction-to-feature-selection/ https://towardsdatascience.com/the-5-feature-selection-algorithms-every-data-scientist-need-to-know-3a6b566efd2 https://www.analyticsvidhya.com/blog/2016/12/introduction-to-feature-selection-methods-with-an-example-or-how-to-select-the-right-variables/
4.3.4. Concept hierarchy Generation	A. Describe concept hierarchy generation. (What is it and how it can be useful in machine learning) B. Concept hierarchy generation for numeric data a. Binning b. Histogram Analysis c. Clustering Analysis d. Entropy-based discretization e. Segmentation by natural partitioning. C. Concept hierarchy generation for categorical data a. Specification of a set of attributes, but not of their partial ordering.	what is Concept Hierarchy? How Concept Hierarchy is generated for Numerical and categorical data? https://www.youtube.com/watch?v=JkO8hj3tkpo

	b. Specification of only a partial set of attributes.	
4.3.5. Feature Scaling	A. Implement following feature scaling using sklearn as well as using basic python/numpy a. Standardization b. MinMax Scaling c. Robust Scaling d. Maximum Absolute Scaling e. Mean Normalisation f. Scaling to unit length -Vector Norm	

4.4. Data Sampling, Reduction and Discretization Techniques

Students will be able to:

4.4.1. Sampling Techniques	A. Power and effect Size, B. Determine the sample size for a particular problem C. Random sampling D. Stratified random sampling	• Coe, Robert. "It's the effect size, stupid: What effect size is and why it is important." (2002). • Groves, Robert M., et al. *Survey methodology*. Vol. 561. John Wiley & Sons, 2011. • Ellis, Paul D. *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. Cambridge University Press, 2010. • Heeringa, Steven G., Brady T. West, and Patricia A. Berglund. *Applied survey data analysis*. CRC press, 2010.
4.4.2. Data Reduction	A. Dimensionality Reduction B. Attribute Subset Selection C. Numerosity Reduction D. Data Cube Aggregation	
4.4.3. Discretization	A. Equal-width, Equal-frequency, fixed Interval B. Discretization with decision trees C. Discretization with clustering	

4.5. Outlier analysis and handling

Students will be able to:

4.5.1. Outliers	A. Describe what is an outlier B. Discuss causes for presence of outliers in dataset C. Describe different Types of outliers: Univariate and Multivariate outliers D. Justify the impact of outliers in ML models	https://www.analyticsvidhya.com/blog/2015/02/outliers-detection-treatment-dataset/?source=post_page---------------------------
4.5.2. Outlier Detection	A. Using visualization tools (box plots, scatterplots,..) B. Using mathematical functions (z-score, IQR score) C. Isolation Forests	https://towardsdatascience.com/ways-to-detect-and-remove-the-outliers-404d16608dba

	D. Extreme Value analysis	
4.5.3. Outlier Handling	A. Capping B. Trimming C. Transforming and binning D. Imputing the outliers E. Treating outliers separately if there are a lot of them	https://www.analyticsvidhya.com/blog/2015/02/outliers-detection-treatment-dataset/?source=post_page----- https://medium.com/@swethalakshmanan14/outlier-detection-and-treatment-a-beginners-guide-c44af0699754

4.6. Features Creation

Students will be able to:

4.6.1. Features from existing Data	A. Use Polynomial features B. Create/Extract year/month/day/week/ time elapsed from Date-Time data in different format and work around with different time zones C. Mixed Data (number/String/mixed)	
4.6.2. **Features from Images**	A. Image Features B. Intro to OpenCV, Scikit-image C. Use sklearn feature extraction	
4.6.3. Features from Text	A. CountVectorizer B. TfidfVectorizer C. For an example of using TF-IDF in a classification problem use Naive Bayes Classification.	

4.7. Model Selection and Evaluation

Students will be able to:

4.7.1. No Free Lunch (NFL) Theorem	A. State No Free Lunch Theorem	
4.7.2. Information and Model Selection	A. Describe Entropy and KL divergence B. Information Criteria a. AIC b. BIC	
4.7.3. Cross-Validation	A. Recall Concept of Overfitting B. Train test split C. Exhaustive cross-validation a. Leave p-out (case where p=1) D. Non-exhaustive cross-validation a. K-Fold CV b. Hold Out Methods c. Monte Carlo cross-validation E. Nested Cross Validation [Eg: Nested 10-fold CV] F. Stratified Cross Validation G. How CV is different in Time Series Data	
4.7.4. Hyper-parameter Search	A. Grid Search	

	B. Random grid Search C. Genetic Search	
4.7.5. Metrics and Scoring	A. Regression : RMSE, MSE, MAE, R2 score B. Classification: a. Accuracy b. Confusion Matrix: Precision,Recall,F1 Score, Specificity c. Precision/Recall tradeoff d. PR curve e. ROC-AUC	
4.7.6. Similarity Metrics	A. Applications of Similarity Metrics B. Popular Similarity Metrics a. Euclidean distance b. Manhattan distance c. Minkowski distance d. Cosine similarity e. Jaccard similarity	https://dataaspirant.com/2015/04/11/five-most-popular-similarity-measures-implementation-in-python/
4.7.7. Feature and Metric Learning	A. Metric Learning B. Sk-learn metric learn The last two references can be used for reading purposes. Please give a slight introduction of metric learning and then you can follow this link: http://contrib.scikit-learn.org/metric-learn/introduction.html	http://contrib.scikit-learn.org/metric-learn/introduction.html http://researchers.lille.inria.fr/abellet/talks/metric_learning_tutorial_CIL.pdf http://people.bu.edu/bkulis/pubs/ftml_metric_learning.pdf

4.8. Machine Learning Pipeline

Students will be able to:

4.8.1. ML Pipeline	A. Discuss the Importance of Pipelining in Machine Learning Application B. Describe various stages of ML Pipeline in Detail C. Build end-to-end sklearn pipelines for classification and regression problems D. Discuss about different steps in ETL (Extract Transform Load)	
4.8.2. Automating ML Pipeline	A. discuss the importance of automating ML pipelines compared to manual ML pipelines B. Discuss on different Technologies for automating ML pipeline a. Workflow creation technologies like AzureML Schedule,Cron, Oozie, MLFlow b. ML Workflow for streaming data (eg in Kafka, pub/sub) C. Schedule automated workflow using cron and apache airflow (eg to trigger and monitor ETL, trigger automated training process)	

4.9. Handling Imbalanced Dataset

Students will be able to:

4.9.1. Introduction to Imbalance Dataset	A. Discuss challenges of Imbalanced Dataset B. How to Handle It a. Accuracy Paradox b. Resampling c. Switching performance metric d. Switching algorithm	https://machinelearningmastery.com/tactics-to-combat-imbalanced-classes-in-your-machine-learning-dataset/
4.9.2. Undersampling	A. Describe undersampling techniques B. Undersampling vs oversampling C. Random undersampling	https://eprint.iacr.org/2018/476.pdf
4.9.3. Oversampling	A. Describe oversampling B. SMOTE(Synthetic Minority Oversampling Technique) oversampling technique C. SMOTE + ENN (Synthetic Minority Oversampling Technique with Edited Nearest Neighbor)	https://eprint.iacr.org/2018/476.pdf
4.9.4. Algorithm level methods of handling imbalance	A. Threshold method B. One-class learning C. Cost-sensitive learning	Handling imbalanced datasets: A review
4.9.5. Evaluation Metrics	A. Evaluation Metrics for Imbalanced Dataset a. Recall Confusion Matrix b. ROC-AUC c. Cost/Weighted evaluation metrics	https://machinelearningmastery.com/tour-of-evaluation-metrics-for-imbalanced-classification/
4.9.6. Imbalance-learn	A. Example usage and implementation with “imbalanced-learn” toolbox	API DOCS: https://imbalanced-learn.readthedocs.io/en/stable/api.html

4.10. Module Summary

Students should be able to :

4.10.1. Module Summary	A. Compare and contrast different algorithms covered this module B. discuss advantage and disadvantages (and/or challenges) of different algorithms C. Discuss best use cases for each of the algorithms covered D. Integrate the concepts covered in this module with other topics covered in the course	

Module 5. Machine Learning Application [Optional]

5.1. Recommender System

Students will be able to:

5.1.1. Introduction to RS	A. Introduction to recommender system B. Application and importance of recommendation system. C. Taxonomy of Recommender System	Recommender_systems_handbook.pdf RecsysSummerSchool-XavierAmatriain.pdf
5.1.2.	A. Populating utility matrix a. Implicit and Explicit Data	

	b. Rating c. Inferencing B. Non-Personalized Summary Statistics: Examples, Best seller, Most popular, trending C. Eg: Classification algorithms, Decision Tree	
5.1.3. Content-Based Filtering	A. Content-Based Filtering B. Item Profiles: The profile consists of some characteristics of the item that are easily discovered. C. Feature discovery and comparing similarity. Examples, Jaccard Distance, Cosine Similarity	
5.1.4. Collaborative Filtering	A. User-User Collaborative Filtering B. Item-Item Collaborative Filtering C. Matrix factorization Techniques + reducing the dimensionality of the user-product preference space D. Clustering user and items (Nearest Neighbor)	
5.1.5. Evaluation metrics	A. Strengths and Weaknesses of different recommendation system algorithms B. Accuracy of predictions and rank– Usefulness of recommendations C. Correctness D. Computational performance E. decision-support, and other factors such as diversity, product coverage, and serendipity	https://www.analyticsvidhya.com/blog/2018/06/comprehensive-guide-recommendation-engine-python/
5.1.6. Factorization machines and Deep Recommenders	A. Factorization machines (FM) B. Field-aware Factorization Machines (FFM) C. Introduction to Deep RS: some of many ways such as Wide-and-Deep, DCF	https://www.csie.ntu.edu.tw/~b97053/paper/Rendle2010FM.pdf
5.1.7. Recommender System in Practise	A. Concept of LIBRA B. Documenting the analysis of Deep Neural Networks for YouTube Recommendations the selected solution, and the justification for that solution. C. Discuss how Hybrid and Ensemble solutions are used in practice D. Discuss how RS could be formulated in an Active Learning Framework	Deep Neural Networks for YouTube Recommendations

5.2. Time Series Forecasting

Students will be able to:

5.2.1. Introduction to Time Series.	A. Definition of Time series B. Time series data format C. Time Series common Terminologies like a. Moving|Rolling b. Lagged c. Stationarity d. Autocovariance e. Seasonality f. White Noise etc D. Why does the Time Series have to be stationary?	https://medium.com/analytics-vidhya/starting-off-with-time-series-56056c4f4b78
5.2.2. Time Series Data Analysis	A. Time Series Analysis Vs Forecasting	https://medium.com/analytics-

and preprocessing	B. Extracting Validation Set from dataSet without distorting Time Series pattern. C. Some Preprocessing Steps D. Analysis of ACF and PACF Plots E. Mode Decomposition (EMD)	vidhya/preprocessing-for-time-series-forecasting-3a331dbfb9c2
5.2.3. Auto-regressive And Moving Average (Implement)	A. Explanation on AR models. B. Explanation on MA Models C. Differencing (How Stationarity)	References: https://otexts.com/fpp2/ https://machinelearningmastery.com/time-series-forecasting/ https://machinelearningmastery.com/arima-for-time-series-forecasting-with-python/ https://towardsdatascience.com/unboxing-arima-models-1dc09d2746f8 https://medium.com/datadriveninvestor/holt-winters-exponential-smoothing-for-time-series-forecasting-5905fad390e5
5.2.4. ARMA,ARIMA ,ARIMAX,SARIMA	A. Explain different models like ARMA , ARIMA , ARIMAX B. Introduce Seasonality with SARIM Models C. Accuracy Metrics For time Series Forecasting	
5.2.5. Exponential Smoothing and Holt Winters´ Exponential Smoothing	A. Explanation on Exponential Smoothing B. Holt-Winters´ Exponential Smoothing.	
Assignment	Time Series Forecasting with ARIMA from scratch	Dataset link():https://raw.githubusercontent.com/jbrownlee/Datasets/master/daily-min-temperatures.csv

5.3. Outlier/Anomaly Detection

Students will be able to:

5.3.1. Intro to Anomaly Detection	A. Anomalies and Outliers B. Causes of Anomaly C. types of anomalies I. Point Anomaly II. Contextual anomaly II. Collective anomalies D. Application of anomaly Detection	References: http://courses.washington.edu/css581/lecture_slides/18_anomaly_detection.pdf http://cucis.ece.northwestern.edu/projects/DMS/publications/

		AnomalyDetection.pdf http://www.jmlr.org/papers/volume6/steinwart05a/steinwart05a.pdf
5.3.2. Anomaly Detection Techniques	A. Statistical(outlier detection , grubbs test, likelihood approach) B. Proximity based C. Density based(outlier score) D. Clustering based E. Neural Network Based F. Rule Based G. Isolation forest H. one-class SVM	
5.3.3. Assignment	Use kNN to detect anomaly or possible fraud from “Credit Card Fraud Detection” dataset available on kaggle, Fault in Machine Vibration	Dataset: https://www.kaggle.com/mlg-ulb/creditcardfraud

Module 6. Reinforcement Learning

6.1. Introduction to the Module

Students will be able to:

6.1.1. Introduction to Reinforcement Learning	A. Discuss the Importance of RL and the type of problem to be solved using Reinforcement Learning. B. Comparison of RL with supervised and unsupervised learning C. Understand the flow of the contents in the module (overview) and their learning outcomes	

6.2. Fundamentals of RL

Students will be able to:

6.2.1. **Basic RL Concept.**	A. Recall Distributions and sampling B. Components in the RL a. Agent b. State c. Action d. Reward e. Environment C. Block Diagram of RL stating the components and their interaction	
6.2.2. K-armed Bandit Problem	A. Solve k-armed bandit problem	

	B. Understand the Exploration vs Exploitation tradeoff	
6.2.3. Markov Decision Process (MDP)	A. Understand the Markov property B. Understand the additional elements of MDP a. Transition Probability b. Reward Probability c. Discount factor C. Understand the relation between rewards and returns D. Bellman Equation E. Understand: a. Policy functions b. Value functions c. Action value functions F. Use a 4x4 gridworld as an example to calculate the state values solving the Bellman equations	
6.2.4. Dynamic Programming	A. Understand Policy Evaluation and Policy Improvement B. Understand about Policy Iteration for solving MDP C. Understand about Value Iteration for solving MDP D. Identify the problem with DP and how asynchronous DP can address the issue.	

6.3. Sample-based Learning Methods

Students will be able to:

6.3.1. Monte Carlo Methods	A. Understand about the Monte Carlo Methods and Model free B. Discuss on types of Value approximation in Monte Carlo (Prediction) a. First visit b. Every visit C. Use policy evaluation and policy improvement to define Generalized Policy Iteration	
6.3.2. Temporal Difference Learning	A. Understand how Temporal Difference (TD) combines Monte Carlo (MC) method and Dynamic Programming (DP) B. Understand about TD controls: a. SARSA: On-policy TD Control b. Q-learning: Off-policy TD Control C. Differentiate between Q-learning and SARSA	

6.4. Value-based Learning Method

6.5. Policy-based Learning Method

6.6. Module Summary

Students should be able to :

6.6.1. Module Summary	A. Review different algorithms covered in this module	

	B. discuss advantage and disadvantages (and/or challenges) of different algorithms C. Integrate the concepts covered in this module with other topics covered in the course	