Data 622 – Big Data

L01 Introduction to Learning. Machine Learning

### Subject Matter of the Course

* Machine Learning: Birds Eye View
* Supervised Learning: Classifiers Strategies for Supervised Learning
* Analytical Foundation of Classifiers
  + Density Estimation:MLE
  + Optimization:Loss Function
  + Convergence: GD, SGD
* Comparative Analysis of Classifiers
* Sources of Error Estimation
* Sources of: Bias vs Variance Error Reduction Strategies
  + Resampling
  + Methods Penalization Methods
  + Cross Validation
* Ensemble Methods
  + Combining Like Classifiers
  + Combining Unlike Classifiers
* Brief Introduction to Big Data
* Opportunities of Parallelism

### Recommended Reading

------------R1----------Reading Material for Lecture 01----------------------------------------------  
Lecture 01 is an introductory session to review Machine Learning.  
Chapter 1 -- ISLR -- http://www-bcf.usc.edu/~gareth/ISL/ISLR%20Sixth%20Printing.pdf  
There is a more comprehensive incarnate of this book by almost the same authors  
http://statweb.stanford.edu/~tibs/ElemStatLearn  
This book, "Elements of Statistical Learning" requires more determination and more quantitative  
You are encouraged to read introductory chapters of other acclaimed ML books  
-- [ALP] Ethem Alpaydin (technical) -- https://github.com/kerasking/book-1/blob/master/ML MIT.Introduction to ML.2Ed.pdf  
-- [KMP] Kevin Patrick Murphy (mathematical)  
(https://github.com/kerasking/book-1/blob/master/ML%20Machine%20Learning-A%20Probabilistic%20Perspective.pdf)  
-- [PRML] Christopher M. Bishop (technical) -- 1st Chapter 1 hour to 2 hours and relevant for 1st and 2nd Lecture  
(https://github.com/kerasking/book-1/blob/master/ML%20Pattern%20Recognition%20and%20Machine%20Learning.pdf)  
-- http://www.cs.cmu.edu/~tom/mlbook/ 3 chapters relevant for the entire course  
-- https://www.cs.cmu.edu/~tom/pubs/MachineLearning.pdf -- approximate 0.5 to 1 hour  
References cited in Lecture 01  
Tom Mitchell -- chapter 1

-----------END of Reading for L01

### Course Focus

**We will develop required expertise to:**

* Prepare dataset,
* Load Clean (missing values, range, outlier, unbalanced)
* Exploratory Analysis (discriptive statistics)
* Feature selection/scale/transform run one or more classifiers training phase test/validation phase
* Write a report on summary findings.

### Lecture Notes:Introduction to Learning

Motivation, Introduction to learning.

[L01.pdf](https://github.com/samriti0202/DATA622/blob/main/L01%20Introduction%20to%20Learning.%20Machine%20Learning/L01%20(3).pdf)

### **Review Questions**

Each Lecture module will also have an item (Review Questions) comprised of Survey, short quiz or q&a -- all of it not for grade items. The rationale is to motivate learners to review the material and validate their understanding of the material.

What are the major kinds of M/L computing tasks?

Which of those do we want to cover in this course?

What exactly are we trying to achieve?

How is M/L different from others forms of computational problem solving?

Why is M/L necessary?

### [The Discipline of Machine Learning](https://github.com/samriti0202/DATA622/blob/main/L01%20Introduction%20to%20Learning.%20Machine%20Learning/MachineLearning.pdf)

Over the past 50 years the study of Machine Learning has grown from the efforts of a handful of computer engineers exploring whether computers could learn to play games, and a field of Statistics that largely ignored computational considerations, to a broad discipline that has produced fundamental statistical-computational theories of learning processes, has designed learning algorithms that are routinely used in commercial systems for speech recognition, computer vision, and a variety of other tasks, and has spun off an industry in data mining to discover hidden regularities in the growing volumes of online data. This document provides a brief and personal view of the discipline that has emerged as Machine Learning, the fundamental questions it addresses, its relationship to other sciences and society, and where it might be headed.

Tom M. Mitchell July 2006 CMU-ML-06-108

### [Difference between Machine Learning, Data Science, AI, Deep Learning, and Statistics](https://www.datasciencecentral.com/profiles/blogs/difference-between-machine-learning-data-science-ai-deep-learning)

In this article, I clarify the various roles of the data scientist, and how data science compares and overlaps with related fields such as machine learning, deep learning, AI, statistics, IoT, operations research, and applied mathematics. As data science is a broad discipline, I start by describing the different types of data scientists that one may encounter in any business setting: you might even discover that you are a data scientist yourself, without knowing it. As in any scientific discipline, data scientists may borrow techniques from related disciplines, though we have developed our own arsenal, especially techniques and algorithms to handle very large unstructured data sets in automated ways, even without human interactions, to perform transactions in real-time or to make predictions.

[Vincent Granville](https://www.datasciencecentral.com/profile/VincentGranville)

### [What is Machine Learning?](https://mlplatform.nl/what-is-machine-learning/)

* What is Machine Learning?
* The History of Machine Learning
* ELIZA
* Neural Networks
* Statistical AI
* Big Data
* Current Status

### [A Short History of Machine Learning -- Every Manager Should Read](https://www.forbes.com/sites/bernardmarr/2016/02/19/a-short-history-of-machine-learning-every-manager-should-read/#16bd6eb415e7)

### [The real prerequisite for machine learning isn’t math, it’s data analysis](https://www.r-bloggers.com/the-real-prerequisite-for-machine-learning-isnt-math-its-data-analysis/)

### [A Few Useful Things to Know About Machine Learning](https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf)

### [Top 10 Machine Learning Algorithms](https://www.datasciencecentral.com/profiles/blogs/top-10-machine-learning-algorithms)

### [Data Mining From A to Z](https://www.sas.com/content/dam/SAS/en_us/doc/whitepaper1/data-mining-from-a-z-104937.pdf)

How to Discover Insights and Drive Better Opportunities

* + Contents Introduction .......................................................................1
  + The SAS® Analytical Life Cycle: Combining Data, Discovery and Deployment............................................2
  + What Can Data Mining Help You Discover? ..........................................................3
  + A Closer Look at the Role of Data Mining in the Discovery Process .................................................5
  + Step 1: Turn a Business Question Into an Analytical Hypothesis ......................................................5
  + Step 2: Prepare the Data for Data Mining .......................5
  + Step 3: Explore the Data ......................................................5
  + Step 4: Model the Data ........................................................6
  + SAS® Data Mining Solutions ...........................................6
  + Using SAS® Enterprise Miner™ for Data Mining and Machine Learning..................................6
  + Using SAS® Factory Miner for an Automated Approach to Data Mining..............................8
  + Scaling Your Discovery Process to Handle Big Data and Complex Problems ......................................8
  + Integration Eases Model Deployment, Monitoring and Management...........................................10
  + Conclusion........................................................................11
  + Learn More.......................................................................11