

Introduction to Data Science

Course Description

This course is split into 2 parts:

1. Mathematical and statistical foundation
2. Derivation of how each algorithm works
 - a. How each algorithm works, and what are important to the algorithm
 - b. The advantages and limits of each algorithm

This course covers the mathematical basics and concepts (and derivation) of what goes behind each algorithm and provide a sound foundation and understanding of what each model or algorithm does. The notions of calculus, mathematical theories and assumptions are explained. These notions then form the bases and build up to each algorithm that will be taught in this course.

Learning Objectives

By the end of this course, you will be able to:

- Use techniques and theorems from calculus to solve basic problems, such as optimization and functional approximation
- Use the concepts taught to read, appreciate and understand foundational Machine Learning literature
- Use the concepts taught to apply to and solve real world problems using data science (or mathematics/statistics)

Pre-Requisite Knowledge

No pre-requisite knowledge is required, everything will be covered in class. Even if you have no knowledge in calculus, it will be re-taught in class.

Class Timings

Every Saturday, 1030 – 1330 (ish)

39 Robinson Road, Level 9, Potala

Recommended Text and Readings

Any good advanced calculus textbook will be useful as supplementary reading

- Understanding Machine Learning: From Theory to Algorithms; Shai Shalev-Shwartz and Shai Ben-David
- Elementary Linear Algebra, Howard Anton
- Numerical Analysis, Timothy Sauer
- A First Course in Differential Equations, Brooks/Cole
- Time Series Analysis, Jan Grandell
- Time Series Analysis and Its Applications, Robert H. Shumway and David S. Stoffer

Lesson Plans (and how does everything tie in together)

<i>Lesson No.</i>	<i>Specific Learning Objective for the Lesson</i>	<i>Topics Covered</i>	<i>Which Topic does it Relate to</i>
1	Intro to Basic Python	<ul style="list-style-type: none"> • Installation of python and jupyter notebook • Installation of python packages • Basic syntax • Basic string manipulations 	<ul style="list-style-type: none"> • Used in coding
2	Intro to Pandas and Visualization	<ul style="list-style-type: none"> • Basic data manipulation and wrangling and transformation • Basic creation of functions • Basic visualization of data 	<ul style="list-style-type: none"> • Used in coding
3	Intro to Python Class Files	<ul style="list-style-type: none"> • Creation of python class files • Basic idea of code modularization with class files 	<ul style="list-style-type: none"> • Used in coding
4	Intro to SQL and Version Control	<ul style="list-style-type: none"> • SQL syntax • Join, union, window functions, having and with clauses • Introduction of Git and setup (SSH) 	<ul style="list-style-type: none"> • Used in data extraction and basic transformation • Git is used in storage of files and tracking of progress
5	Intro to Statistics and Probability Theory	<ul style="list-style-type: none"> • Intro to distributions • Intro to probability theory and assumptions • Intro to statistical testing • Intro to MLE • Intro to MGF 	<ul style="list-style-type: none"> • Used in ML algorithms • ML algorithms have underlying distributions like Normal or exponential distribution • Getting the parameters of ML algos, MLE is often used
6	Intro to Linear Regression*	<ul style="list-style-type: none"> • Intro to univariate • Intro to the loss function • Intro to getting parameters of regression using MLE • Intro to R-square, and adjusted R-square • Intro to variables significance • Intro to multiple regression 	<ul style="list-style-type: none"> • Loss function is the objective function for ML algos • Linear regression is also a machine learning algo
7	Intro to Linear Algebra*	<ul style="list-style-type: none"> • Intro to basic properties of matrices and vector spaces • Intro to solving system of linear equations with LA 	<ul style="list-style-type: none"> • LA can also be used to solve linear regression

		<ul style="list-style-type: none"> • Intro to eigen values/vectors • Intro to orthogonal • Intro to PCA • Intro to matrix factorization 	<ul style="list-style-type: none"> • LA is often used in dimension reduction • LA links to ML algo like SVM and recommendation systems
8	Intro to Time Series Analysis*	<ul style="list-style-type: none"> • Intro to stationary and non stationary series • Intro to stationary tests • Intro to ARIMA • Intro to exponential smoothing • Intro to state space models (BSTS) • Intro to distribution forecasting • Intro to probabilistic forecasting 	<ul style="list-style-type: none"> • Not everything needs to be solved by ML
9	Intro to Optimization and Numerical Methods*	<ul style="list-style-type: none"> • Intro to different optimization methods • Intro to splines • Intro gradient descent 	<ul style="list-style-type: none"> • Solving for loss function or the intuition of the loss function
10	Intro to Classical Machine Learning Algorithms*	<ul style="list-style-type: none"> • Intro to supervised and unsupervised learning • Intro to unsupervised learning (clustering) • Intro to supervised learning <ul style="list-style-type: none"> ○ Linear regression ○ Logistics regression ○ Generalized linear model ○ Regularization ○ Support Vector Machine (SVM) ○ Linear Discriminant Analysis ○ Nearest Neighbour ○ Decision Trees ○ Bagging ○ Boosting ○ Gradient Boosting 	<ul style="list-style-type: none"> • What is learnt above will be applied to here
11	Intro to Neural Networks*	<ul style="list-style-type: none"> • Intro to NN • Intro to cNN, rNN and dNN (Deep learning) 	<ul style="list-style-type: none"> • Gradient descent • Everything else above • Differential equations

*Code will be shown appropriately