

Lesson 1: Business Concepts

Session 1: you will be taught analytics life cycle and how analytics is used in real time with case studies and example

- Analytics landscape and components
- Analytics frame-work-CRISP-DM,
- Real life analytics project examples

Lesson 2: Foundations of probability and statistics

Session 1: This session introduces you with how statistics is used in business with basic statistical concepts like levels of data and measures of central tendencies

- Measures of central tendencies on ungrouped data (mean, median, mode)
- Measures of Variability on Ungrouped data (Range, IR, Variance Standard Deviation, Z scores, coefficient of variance)
- Measures of shape (Skewness and the Relationship of the Mean, Median, and Mode, Coefficient of Skewness Kurtosis, Box-and-Whisker Plots, Histograms)
- Introduction to probability

Session 2: You will get into the deeper aspects of various distributions. You will understand the parameters that define the probability distributions and differences between discrete and continuous distributions.

- Discrete probability distributions: Bernoulli, Binomial, Geometric, Poisson and properties of each.
- Continuous probability distributions: Normal distribution; t-distribution , Exponential Distribution

Session 3: You will also start making statistical inferences about populations from samples.

- Sampling
- Estimating the Population Mean Using the Z-Statistic and T- statistic
- Hypothesis testing
- confidence interval

Session 4: Till this point you will have received the complete picture how to understand the data, attributes, distributions, sample versus population, and procedure for statistical testing, etc. While you continue the analysis of a variable, you will extend that understanding to analyse the relationship between variables.

- chi-square test
- ,t-test, z-test, F-test
- one- way -ANOVA , two -way -Anova

Assignment on statistics

Lesson 3: Introduction to Programming

Session 1: In this course you will learn how to program in R and how to use R for effective data analysis. You will learn how to install and configure software necessary for a statistical programming environment and describe generic programming language concepts as they are implemented in a high-level statistical language. The course covers practical issues in statistical computing which includes programming in R, reading data into R, accessing R packages, writing R functions, debugging, profiling R code, and organizing and commenting R code. Topics in statistical data analysis will provide working examples.

Control structures and functions. : `str()`, `class()`, `length()`, `nrow()`, `ncol()`, `seq()`, `cbind()`, `rbind()`, `merge()`,

Data manipulation techniques : The various steps involved in Data Cleaning, functions used in Data Inspection, tackling the problems faced during Data Cleaning, uses of the functions like `grepl()`, `grep()`, `sub()`, Coerce the data, uses of the `apply()` functions.

Session 2: We will continue with R concepts and execute statistical concepts using R

- Pre-processing Techniques: Binning, Filling missing values, Standardization & Normalization, type conversions, train-test data split, ROC1
- other R concepts
- Exploratory Data analysis

Session 3: Preparing Data as an input for machine learning algorithms

Case study

Assignment on R understandings

Session 4: you will be executing all machine algorithms in R

Lesson 5: Regression concepts

Predictive analytics search for patterns found in historical and transactional data to understand a business problem and predict future events. In many business problems, we try to deal with data on

Several variables, sometime more than the number of observations. Regression models help us understand the relationships among these variables and how the relationships can be exploited to make decisions. Primary objective of this module is to understand how regression and causal forecasting models can be used to analyse real-life business problems such as prediction, classification and discrete choice problems. The focus will be case-based practical problem-solving using predictive analytics techniques to interpret model outputs.

Session 1: Linear Regression

- Simple linear regression
- Coefficient of determination,
- Significance Tests,
- Residual Analysis,
- Confidence and Prediction intervals
- Multiple linear regression:
- Coefficient of determination,
- Interpretation of regression coefficients,
- Categorical variables in regression
- Heteroscedasticity, Multi-collinearity outliers,
- R-square and goodness of fit
- Hypothesis testing of Regression Model
- Transformation of variables
- Polynomial Regression

Case Study

Session 2: Logistic Regression

Logistic regression is a method for classifying data into discrete outcomes. For example, we might use logistic regression to classify an email as spam or not spam. In this module, we introduce the notion of classification, the cost function for logistic regression, and the application of logistic regression to multi-class classification.

- Logistic function,
- Estimation of probability using logistic regression,
- Model Evaluation
- Confusion Matrix

Case study

Session 3: Time series data

The focus is on analysing and understanding Time Series with financial markets as the case study.

- Trend analysis
- Cyclical and Seasonal analysis
- Smoothing; Moving averages;
- Auto-correlation;
- ARIMA; ARIMAX
- Applications of Time Series in financial markets

Case study

Lesson 6: Machine learning

Session 1: Clustering

-
- What is Clustering?
 - Clustering Examples in Business Verticals
 - Solution strategies for Clustering
 - Finding pattern and Fixed Pattern Approach
 - Limitations of Fixed Pattern Approach
 - Machine Learning Approaches for Clustering
 - Iterative based K-Means & K-Medoid Approaches
 - Hierarchical Agglomerative Approaches
 - Density based DB-SCAN Approach
 - Evaluation Metrics for Clustering
 - Cohesion, Coupling Metrics
 - Correlation Metric

Case study

Session 2 : Naive Bayes

- Conditional probability.
- Conditional independence.
- Bayes rule and examples.
- Naive Bayes algorithm.
- Space and Time complexity: train and test time.

- Laplace/Additive Smoothing.
- Under fitting and over fitting.
- Feature importance and interpretability

Case study

Session 3: KNN

- Intuitive idea of KNN classification
- KNN learning
- Limitations of KNN
- KNN Regression
- Applying KNN and parameter tuning
- Pros and Cons of the Model

Case study

Session 4: Decision Trees

- Geometric Intuition:
 - Axis parallel hyper planes.
 - Nested if-else conditions.
 - Sample Decision tree.
 - Building a decision Tree:
 - Entropy, Information Gain
 - Gini Impurity (CART)
 - Depth of a tree: Geometric and programming intuition.
 - Categorical features with many levels.
 - Regression using Decision Trees.
 - Bias-Variance trade-off.
 - Limitations

Case study

Session 5: Support vector machines (SVM)

- Geometric intuition.
- Mathematical derivation.
- Loss function (Hinge Loss) based interpretation.
- Support vectors.
- Linear SVM.
- Non-linear svm and kernel function
- Primal and Dual.
- Kernelization.
- RBF-Kernel.
- Polynomial kernel.
- Domain specific Kernels.
- Train and run time complexities.
- Bias-variance trade-off: Under fitting and Over fitting
- Nu-SVM: control errors and support vectors.
- SVM Regression.

Case study

Session 6: Neural networks

- History of Neural networks and Deep Learning.
- Perceptron's
- Self-organizing maps
- Auto encoders

- Back propagation and typical feed forward algorithm
- Sigmoidal Activation functions.
- Mathematical formulation.
- Back propagation and chain rule of differentiation
- Vanishing Gradient problem.
- Bias-Variance Trade-off.
- Determining the number of levels.
- Decision surfaces.

Case study

Session 7: Ensemble Methods

- Understanding Weak Learners
- Approaches for Ensemble learning: Boosting, Bagging and Randomization
- Bagging Idea in depth and why it works?
- Bootstrapped Aggregation (Bagging)
- Random Forest and their construction.
- Bias-Variance trade-off
- Gradient Boosting and XGBoost
- Algorithm.
- Loss function and advantages.
- XGBoost code samples
- AdaBoost: geometric intuition.
- Cascading models
- Stacking models.

Case Study

Session 8: Association Rules

- Apriority Model
- Intuitive Idea
- Apriority Model
- Applying the Algorithm and tuning
- Pros and Cons of the Model
- Recommender systems
- user - user
- item -item
- content based

Case Study

Session 9: Feature engineering

- Dimensionality Reduction
- PCA and EDA
- Principal Component Analysis.
- Why learn it.
- Geometric intuition.
- Mathematical objective function.
- Alternative formulation of PCA: Distance minimization
- Eigen values and Eigen vectors.
- PCA for dimensionality reduction and visualization.
- Visualize MNIST dataset.
- Limitations of PCA

Case Study

Session 10 And Session 11: Business case analysis

- The objective of this session is to provide an application and end-to-end view of solving a Data Science problem and defend your analysis.

We provide a business case in advance in which you will be required to apply all the data pre-processing steps and prepare the input for ML algorithms learnt thus far. The lab is designed such that everyone participates in a discussion, design the solution approach for the given business case and defend the analysis approach

- Hands on
- Revision

Lesson 8: Data Visualization

- Tableau
- Real Time Dashboards

Lesson 9 : Python Programming

- I python or Jupiter setup
- control structures
- Functions
- Packages
- pandas, numpy, scipy, matplotlib, seaborn, sklearn
- space and time complexity

Lesson 10 : Text Mining with Natural language processing Techniques

In this course, we will introduce a variety of basic principles, techniques and modern advances in text mining. Topics to be covered include (the schedules are tentative and subject to change, please keep track of it on the course website):

1. Introduction: We will highlight the basic organization and major topics of this course, and go over some logistic issues and course requirements.
2. Natural language processing: We will briefly discuss the basic techniques in natural language processing, including tokenization, part-of-speech tagging, chunking, syntax parsing and named entity recognition. Public NLP toolkits will be introduced for you to understand and practice with those techniques.
3. Document representation: We will discuss how to represent the unstructured text documents with appropriate format and structure to support later automated text mining algorithms.
4. Text categorization: It refers to the task of assigning a text document to one or more classes or categories. We will discuss several basic supervised text categorization algorithms, including Naive Bayes, k Nearest Neighbour (kNN) and Logistic Regression. (If time allows, we will also cover Support Vector Machines and Decision Trees.)
5. Text clustering: It refers to the task of identifying the clustering structure of a corpus of text documents and assigning documents to the identified cluster(s). We will discuss two typical types of clustering algorithms, i.e., connectivity-based clustering (a.k.a., hierarchical clustering) and centroid-based clustering (e.g., k-means clustering).
6. Topic modelling: Topic models are a suite of algorithms that uncover the hidden thematic structure in document collections. We will introduce the general idea of topic modelling, two basic topic models, i.e., Probabilistic Latent Semantic Indexing (pLSI) and Latent Dirichlet Allocation (LDA), and their variants for different application scenarios, including classification; imagine annotation, collaborative filtering, and hierarchical topical structure modelling.

7. Document summarization: It refers to the process of reducing a text document to a summary that retains the most important points of the original document. Extraction based summarization methods will be covered.
8. Social media and network analysis: We will discuss the unique characteristic of social network: inter-connectivity, and introduce Google's winning algorithm PageRank. Based on this, we will discuss social influence analysis and social media analysis
9. Sentiment analysis: It refers to the task of extracting subjective information in source materials. We will discuss several interesting problems in sentiment analysis, including sentiment polarity prediction, review mining, and aspect identification,
10. Project presentation

Lesson 11: Artificial Intelligence and Decision Science

Deep learning techniques are so powerful because they learn the best way to represent the problem while learning how to solve the problem.

This is called representation learning.

Representation learning is perhaps the biggest differentiation between deep learning models and classical machine learning algorithm.

It is the power of representation learning that is spurring such great creativity in the way the techniques are being used. For example: Combinations of deep learning models are being used to both identify objects in photographs and then generate textual descriptions of those objects, a complex multi-media problem that was previously thought to require large artificial intelligence systems.

Deep learning is hot; it is delivering results and now is the time to get involved. But where do you start?

Artificial Neural Networks

- How do Neural Networks work
- Gradient Descent
- Stochastic Gradient Descent
- Building an ANN
- Multilayer perceptron

Deep Learning

Introduction to deep learning

Why is Deep Learning taking off?

Gradient Descent

Logistic Regression Gradient Descent

Logistic Regression with a Neural Network mindset

Deep learning Framework:

Tensorflow

Keras

Theano or torch

Convolution Neural networks

ConvNet Architecture with layers

Kernel/Filter & Stride & Padding

pooling

Dilated Convolutions AlexNet & Inception Modules

Residual Networks

DenseNets

Case Study: Image Object Detection & Localization

In this session, you should design a model to detect multiple objects in images. This is a multi-tasks problem, the first one is localization and second is classification.

Recurrent Neural networks

- Building RNN
- LSTM
- Sequence learning
- GRU

Case Study: Image Caption

In this session, you should design a model that can be given an image, then generates suitable caption which can describe the image.

Lesson 12: Deployment

GPU

Amazon cloud

Microsoft Azure