

Data Science Program

Week 1: Basics of Python Programming

1. Session 1: Python Basics

- About Python
- Python Output/print function
- Python Data Types
- Python Variables
- Python comments
- Python Keywords and Identifiers
- Python User Input
- Python Type conversion
- Python Literals

2. Session 2: Python Operators + if-else + Loops

- Start of the session
- Python Operators
- Python if-else
- Python Modules
- Python While Loop
- Python for loop

3. Session 3: Python Strings

- Introduction
- Solving Loop problems
- Break, continue, pass statement in loops
- Strings
- String indexing
- String slicing
- Edit and delete a string
- Operations on String
- Common String functions

Week 2: Python Data Types

1. Session 4: Python Lists

- Introduction
- Array vs List
- How lists are stored in a memory
- Characteristics of Python List
- Code Example of Lists
 - i. *Create and access a list*
 - ii. *append(), extend(), insert()*
 - iii. *Edit items in a list*
 - iv. *Deleting items from a list*
 - v. *Arithmetic, membership and loop operations on a List*
 - vi. *Various List functions*
 - vii. *List comprehension*
 - viii. *2 Ways to traverse a list*
 - ix. *Zip() function*
 - x. *Python List can store any kind of objects*
- Disadvantages of Python list

2. Session 5: Tuples + Set + Dictionary

- **Tuple**
 - i. *Create and access a tuple*
 - ii. *Can we edit and add items to a tuple?*
 - iii. *Deletion*
 - iv. *Operations on tuple*
 - v. *Tuple functions*
 - vi. *List vs tuple*
 - vii. *Tuple unpacking*
 - viii. *Zip () on tuple*
- **Set**
 - i. *Create and access a set*
 - ii. *Can we edit and add items to a set?*
 - iii. *Deletion*
 - iv. *Operations on set*
 - v. *set functions*
 - vi. *Frozen set (immutable set)*
 - vii. *Set comprehension*
- **Dictionary**
 - i. *Create dictionary*
 - ii. *Accessing items*
 - iii. *Add, remove, edit key-value pairs*
 - iv. *Operations on dictionary*

- v. Dictionary functions
- vi. Dictionary comprehension
- vii. Zip() on dictionary
- viii. Nested comprehension

3. Session 6: Python Functions

- Create function
- Arguments and parameters
- args and kwargs
- How to access documentation of a function
- How functions are executed in a memory
- Variable scope
- Nested functions with examples
- Functions are first class citizens
- Deletion of function
- Returning of function
- Advantages of functions
- Lambda functions
- Higher order functions
- map(), filter(), reduce()

Week 3: Object Oriented Programming (OOP)

1. Session 7: OOP Part1

- What is OOP?
- What are classes and Objects?
- Methods vs Functions
- Magic/Dunder methods
- What is the true benefit of constructor?
- Concept of 'self'
- Create Fraction Class
- __str__, __add__, __sub__, __mul__, __truediv__

2. Session 8: OOP Part2

- How objects access attributes
- Attribute creation from outside of the class
- Reference Variables
- Encapsulation
- Collection of objects
- Static variables and methods

3. Session 9: OOP Part3

- Class Relationship
- Inheritance and Inheritance class
- Constructor example
- Method Overriding
- Super keyword
- Super constructor
- Types of Inheritance (Single, Multilevel, Hierarchical, Multiple)
- Hybrid Inheritance
- Polymorphism
- Method Overriding and Method Overloading
- Operator Overloading

4. Session on Abstraction

- What is Abstraction?
- Abstract class

Week 4: Advanced Python:

1. Session 10: File Handling + Serialization & Deserialization

- How File I/O is done
- Writing to a new text file
- What is open()?
- append()
- Writing many lines
- Saving a file
- Reading a file -> read() and readline()
- Using context manager -> with()
- Reading big file in chunks
- Seek and tell
- Working with Binary file
- Serialization and Deserialization
- JSON module -> dump() and load()
- Serialization and Deserialization of tuple, nested dictionary and custom object
- Pickling
- Pickle vs JSON

2. Session 11: Exception Handling

- Syntax Error with Examples
- Exception with Examples
- Why we need to handle Exception?

- Exception Handling (Try-Except-Else-Finally)
- Handling Specific Error
- Raise Exception
- Create custom Exception

3. Session 12: Decorators and Namespaces

- Namespaces
- Scope and LEGB rule
- Hands-on local, enclosing, global and built-in scope
- Decorators with Examples

4. Session on Iterators

- What are iterators
- What are iterables
- How for loop works in Python?
- Making your own for loop
- Create your own range function
- Practical example to use iterator

5. Session on Generator

- What is a generator?
- Why to use Generator?
- Yield vs Return
- Generator Expression
- Practical Examples
- Benefits of generator

Week 5: Numpy

1. Session 13: Numpy Fundamentals

- Numpy Theory
- Numpy array
- Matrix in numpy
- Numpy array attributes
- Array operations
- Scalar and Vector operations
- Numpy array functions
 - Dot product*
 - Log, exp, mean, median, std, prod, min, max, trigo, variance, ceil, floor, slicing, iteration*
 - Reshaping*

iv. Stacking and splitting

2. Session 14: Advanced Numpy

- Numpy array vs Python List
- Advanced, Fancy and Boolean Indexing
- Broadcasting
- Mathematical operations in numpy
- Sigmoid in numpy
- Mean Squared Error in numpy
- Working with missing values
- Plotting graphs

3. Session 15: Numpy Tricks

- Various numpy functions like sort, append, concatenate, percentile, flip, Set functions, etc.

4. Session on Web Development using Flask

- What is Flask library
- Why to use Flask?
- Building login system and name entity recognition with API

Week 6: Pandas

1. Session 16: Pandas Series

- What is Pandas?
- Introduction to Pandas Series
- Series Methods
- Series Math Methods
- Series with Python functionalities
- Boolean Indexing on Series
- Plotting graphs on series

2. Session 17: Pandas DataFrame

- Introduction Pandas DataFrame
- Creating DataFrame and read_csv()
- DataFrame attributes and methods
- Dataframe Math Methods
- Selecting cols and rows from dataframe
- Filtering a Dataframe
- Adding new columns
- Dataframe function – astype()

3. Session 18: Important DataFrame Methods

- Various DataFrame Methods

- Sort, index, reset_index, isnull, dropna, fillna, drop_duplicates, value_counts, apply, etc.

4. Session on API Development using Flask

- What is API?
- Building API using Flask
- Hands-on project

5. Session on Numpy Interview Question

Week 7: Advanced Pandas

1. Session 19: GroupBy Object

- What is GroupBy?
- Applying builtin aggregation functions on groupby objects
- GroupBy Attributes and Methods
- Hands-on on IPL dataset

2. Session 20: Merging, Joining, Concatenating

- Pandas concat method
- Merge and join methods
- Practical implementations

3. Session on Streamlit

- Introduction to Streamlit
- Features of Streamlit
- Benefits of Streamlit
- Flask vs Streamlit
- Mini-project on Indian Startup Funding Dataset using Streamlit – Part 1

4. Session on Pandas Case Study (Indian Startup Funding)

- Data Analysis on Indian Startup Funding Dataset and display results on the Dashboard made by Streamlit – Part 2

5. Session on Git:

- What is Git?
- What is VCS/SCM?
- Why Git/VCS is needed?
- Types of VCS
- Advantages
- How Git works?
- Installing git
- Creating and Cloning repo
- add, commit, add ., gitignore
- seeing commits (log -> oneline)
- Creating versions of a software

6. Session on Git and GitHub:

- Nonlinear Development (Branching)
- Merging branches
- Undoing changes
- Working with a remote repo

Week 8: Advanced Pandas Continued

1) Session 21: MultiIndex Series and DataFrames

- About Multiindex objects
- Why to use Multiindex objects
- Stacking and unstacking
- Multiindex DataFrames
- Transpose Dataframes
- Swaplevel
- Long vs wide data
- Pandas-melt

2) Session 22: Vectorized String Operations | Datetime in Pandas

- Pivot table
- Agg functions
- Vectorized String operations
- Common functions
- Pandas Datetime

Week 9: Data Visualization

1) Session 23: Plotting Using Matplotlib

- Get started with Matplotlib
- Plotting simple functions, labels, legends, multiple plots
- About scatter plots
- Bar chart
- Histogram
- Pie chart
- Changing styles of plots

2) Session 24: Advanced Matplotlib

- Colored Scatterplot
- Plot size, annotations
- Subplots
- 3D plots

- Contour plots
- Heatmaps
- Pandas plot()

3) Session on Plotly (Express)

- About Plotly
- Disadvantages
- Introduction about Plotly Go, Plotly Express, Dash
- Hands-on Plotly

4) Session on Plotly Graph Objects (go)

5) Session on Plotly Dash

- Basic Introduction about Dash

6) Making a COVID-19 dashboard using Plotly and Dash

7) Deploying a Dash app on Heroku

Week 10: Data Visualization Continued

1) Session 25: Plotting Using Seaborn- part 1

- Why seaborn?
- Seaborn roadmap
- Main Classification plotting
- Relational plots
- Distribution plots
- KDE plot
- Matrix plot

2) Session 26: Plotting Using Seaborn- Part 2

- Categorical Plots
- Stripplot
- Figure level function — catplot
- Swarmplot
- Categorical Distribution Plots
- Boxplot
- Violinplot
- Categorical Estimate Plot — for central tendency
- Barplot
- Pointplot
- Countplot
- Faceting
- Doubt - error bar issue
- Regression Plots

- Regplot
- Lmplot
- Residual Plot
- FacetGrid
- Pairplot and Pairgrid
- JointGrid and Jointplot
- Utility Function in Seaborn - load_dataset
- Blog idea

Week 11: Data Analysis Process – Part I

1) Session 27: Data Gathering | Data Analysis

- Data Analysis Process
- Import Data from various sources (CSV, excel, JSON, text, SQL)
- Export data in different file formats
- Gather Data through API or Web Scrapping

2) Session 28: Data Assessing and Cleaning

- Data assessing
- Types of unclean data
- Write summary of data
- Types of Assessment
- Manual and Automatic Assessment
- Data Quality Dimension
- Data Cleaning

3) Session on ETL using AWS RDS

- Introduction about Extraction, transform and Load pipeline
- Fetch data from AWS
- Apply transformation on the data
- Upload transformed data into AWS RDS

4) Session on Advanced Web Scrapping using Selenium

- Introduction to Selenium and Chromedriver
- Automated Web scrapping on Smartprix website

Week 12: Data Analysis Process – Part 2

1) Session on Data Cleaning Case Study – Smartphone dataset

- Quality issues
- Tidiness issues
- Data Cleaning

2) Session 29: Exploratory Data Analysis (EDA)

- Introduction to EDA
- Why EDA?
- Steps for EDA
- Univariate Analysis
- Bivariate Analysis
- Feature Engineering

3) Session on Data Cleaning – Part 2

- Data Cleaning on Smartphone Dataset – Continued

Week 13: SQL Basics

1) Session 30: Database Fundamentals

- Introduction to Data and Database
- CRUD operations
- Properties of database
- Types of Database
- DBMS
- Keys
- Cardinality of Relationship
- Drawbacks of Database

2) Session 31: SQL DDL Commands

- Xampp Software
- Types of SQL commands
- DDL commands

Week 14: SQL Continued – Part 1

1) Session 32: SQL DML commands

- MySQL workbench
- INSERT
- SELECT
- UPDATE

- DELETE
- Functions in SQL

2) Session 33: SQL Grouping and Sorting

- Sorting Data
- ORDER BY
- GROUP BY
- GROUP BY on multiple columns
- HAVING clause
- Practice on IPL Dataset

Week 15: SQL Continued – Part 2

1) Session 34: SQL Joins

- Introduction to SQL joins
- Types of Joins (Cross, inner, left, right, full outer)
- SQL hands-on on joins
- SET operations
- SELF join
- Query execution order
- Practice questions

2) Session 35: Subqueries in SQL

- What is Subquery
- Types of Subqueries
- Independent and Correlated subquery

Week 16: Advanced SQL

1) Session 36: Window Functions in SQL

- What are Window functions?
- OVER(), RANK(), DENSE_RANK(), ROW_NUMBER(), FIRST_VALUE(), LAST_VALUE()
- Concept of Frames
- LAG(), LEAD()

2) Session 37: Windows Functions Part 2

- Ranking
- Cumulative sum and average
- Running average
- Percent of total

3) Session 37: Windows Functions Part 3

- Percent Change
- Quantiles/Percentiles
- Segmentation
- Cumulative Distribution
- Partition by multiple columns

Week 17: Descriptive Statistics

1) Session 38: Descriptive Statistics Part 1

- What is Statistics?
- Types of Statistics
- Population vs Sample
- Types of Data
- Measures of central tendency
- Measure of Dispersion
- Coefficient of variation
- Graphs for Univariate Analysis
- Frequency Distribution table
- Graphs for bivariate Analysis
- Categorical – Categorical Analysis
- Numerical – Numerical Analysis
- Categorical – Numerical Analysis

2) Session on Datetime in SQL

- Remaining topics of EDA using SQL (numerical – categorical, missing values, ppi, price_bracket, one hot encoding)
- Temporal Data types
- Creating and Populating Temporal Tables
- DATETIME Functions
- Datetime Formatting
- Type Conversation
- DATETIME Arithmetic
- TIMESTAMP VS DATETIME
- Case Study – Flights

Week 18: Descriptive Statistics continued

1) Session 39: Descriptive Statistics part 2

- Quantiles and Percentiles
- Five Number Summary
- Boxplots

- Scatterplots
- Covariance
- Correlation
- Correlation vs Causation
- Visualizing multiple variables

2) Session 40: Probability Distribution Functions (PDF, CDF, PMF)

- Random Variables
- Probability Distributions
- Probability Distribution Functions and its types
- Probability Mass Function (PMF)
- Cumulative Distribution Function (CDF) of PMF
- Probability Density Function (PDF)
- Density Estimation
- Parametric and Non-parametric Density Estimation
- Kernel Density Estimate (KDE)
- Cumulative Distribution Function (CDF) of PDF.

Week 19: Probability Distributions

1) Session 41: Normal Distribution

- How to use PDF in Data Science?
- 2D density plots
- Normal Distribution (importance, equation, parameter, intuition)
- Standard Normal Variate (importance, z-table, empirical rule)
- Properties of Normal Distribution
- Skewness
- CDF of Normal Distribution
- Use of Normal Distribution in Data Science

2) Session 42: Non-Gaussian Probability Distributions

- Kurtosis
- Excess Kurtosis and Types of kurtosis
- QQ plot
- Uniform Distribution
- Log-normal distribution
- Pareto Distribution
- Transformations
 - Mathematical Transformation*
 - Function Transformer*
 - Log Transform*
 - Reciprocal Transform / Square or sqrt Transform*

- v. *Power Transformer*
- vi. *Box-Cox Transform*
- vii. *Yeo-Johnson Transformation*

Week 20: Inferential Statistics

1) Session 43: Central Limit Theorem

- Bernouli Distribution
- Binomial Distribution
 - i. *PDF formula*
 - ii. *Graph of PDF*
 - iii. *Examples*
 - iv. *Criteria*
 - v. *Application in Data Science*
- Sampling Distribution
- Intuition of Central Limit Theorem (CLT)
- CLT in code
- Assumptions of making samples

2) Session on Central Limit Theorem Proof

3) Session 44: Confidence Intervals

- Population vs Sample
- Parameter vs Estimate
- Point Estimate
- Confidence Interval
 - i. *Ways to calculate CI*
 - ii. *Applications of CI*
 - iii. *Assumptions of z-procedure*
 - iv. *Formula and Intuition of z-procedure*
 - v. *Interpreting CI*
 - vi. *T-procedure and t-distribution*
 - vii. *Confidence Intervals in code*

Week 21: Hypothesis Testing

1) Session 45: Hypothesis Testing (Part 1)

- Key idea of hypothesis testing
- Null and alternate hypothesis
- Steps in Hypothesis testing
- Performing z-test
- Rejection region and Significance level
- Type-1 error and Type-2 Error

- One tailed vs. two tailed test
- Applications of Hypothesis Testing
- Hypothesis Testing in Machine Learning

2) Session 46: Hypothesis Testing (Part 2) | p-value and t-tests

- What is p-value?
- Interpreting p-value
- P-value in the context of z-test
- T-test
- Types of t-test
 - Single sample t-Test*
 - Independent 2-sample t-Test*
 - Paired 2 sample t-Test*
 - Code examples of all of the above*

3) Session on Chi-square test

- Chi – square distribution (Definition and Properties)
- Chi-square test
- Goodness of fit test (Steps, Assumptions, Examples)
- Test for Independence (Steps, Assumptions, Examples)
- Applications in machine learning

4) Session on ANOVA

- Introduction
- F-distribution
- One-way ANOVA
 - Steps*
 - Geometric Intuition*
 - Assumptions*
 - Python Example*
- Post – Hoc test
- Why t-test is not used for more than 3 categories?
- Applications in Machine Learning

Week 22: Linear Algebra

1) Session on Tensors | Linear Algebra part 1(a)

- What are tensors?
- 0D, 1D and 2D Tensors
- Nd tensors
- Rank, axes and shape
- Example of 1D, 2D, 3D, 4D, 5D tensors

2) Session on Vectors | Linear Algebra part 1(b)

- What is Linear Algebra?
- What are Vectors?
- Vector example in ML
- Row and Column vector
- Distance from Origin
- Euclidean Distance
- Scalar Addition/Subtraction (Shifting)
- Scalar Multiplication/Division [Scaling]
- Vector Addition/Subtraction
- Dot product
- Angle between 2 vectors
- Equation of a Hyperplane

3) Linear Algebra Part 2 | Matrices (computation)

- What are matrices?
- Types of Matrices
- Matrix Equality
- Scalar Operation
- Matrix Addition, Subtraction, multiplication
- Transpose of a Matrix
- Determinant
- Minor
- Cofactor
- Adjoint
- Inverse of Matrix
- Solving a system of Linear Equations

4) Linear Algebra Part 3 | Matrices (Intuition)

- Basis vector
- Linear Transformations
- Linear Transformation in 3D
- Matrix Multiplication as Composition
- Test of Commutative Law
- Determinant and Inverse
- Transformation for non-square matrix?
- Why only square matrix has inverse?
- Why inverse is possible for non-singular matrices only?

Week 23: Linear Regression

1) Session 48: Introduction to Machine Learning

- About Machine Learning (History and Definition)

- Types of ML
 - i. Supervised Machine Learning*
 - ii. Unsupervised Machine Learning*
 - iii. Semi supervised Machine Learning*
 - iv. Reinforcement Learning*
- Batch/Offline Machine Learning
- Disadvantages of Batch learning
- Online Machine Learning
 - i. Importance
 - ii. When to use and how to use
 - iii. Learning Rate
 - iv. Out of core learning
 - v. Disadvantages
- Batch vs Online learning
- Instance based learning
- model-based learning
- Instance vs model-based learning
- Challenges in ML
 - i. Data collection
 - ii. Insufficient/Labelled data
 - iii. Non-representative data
 - iv. Poor quality data
 - v. Irrelevant features
 - vi. Overfitting and Underfitting
 - vii. Offline learning
 - viii. Cost
- Machine Learning Development Life-cycle
- Different Job roles in Data Science
- Framing a ML problem | How to plan a Data Science project

2) Session 49: Simple Linear regression

- Introduction and Types of Linear Regression
- Simple Linear Regression
- Intuition of simple linear regression
- Code example
- How to find m and b?
- Simple Linear Regression model code from scratch
- Regression Metrics
 - i. MAE*
 - ii. MSE*

iii. *RMSE*

iv. *R² score*

v. *Adjusted R² score*

3) Session 50: Multiple Linear Regression

- Introduction to Multiple Linear Regression (MLR)
- Code of MLR
- Mathematical Formulation of MLR
- Error function of MLR
- Minimizing error
- Error function continued
- Code from scratch

4) Session on Optimization the Big Picture

- Mathematical Functions
- Multivariable Functions
- Parameters in a Function
- ML models as Mathematical Function
- Parametric Vs Non-Parametric ML models
- Linear Regression as a Parametric ML model
- Loss Function
- How to select a good Loss Function?
- Calculating Parameters from a Loss Function
- Convex And Non-Convex Loss Functions
- Gradient Descent
- Gradient Descent with multiple Parameters
- Problems faced in Optimization
- Other optimization techniques

Week 24: Gradient Descent

1) Session 51: Gradient descent from scratch

- What is Gradient Descent?
- Intuition
- Mathematical Formulation
- Code from scratch
- Visualization 1
- Effect of Learning Rate
- Adding m into the equation
- Effect of Loss function
- Effect of Data

2) Session 52 (part 1): Batch Gradient Descent

- Types of Gradient Descent
- Mathematical formulation
- Code from scratch

3) Session 52 (part 2): Stochastic Gradient Descent

- Problems with Batch GD
- Stochastic GD
- Code from scratch
- Time comparison
- Visualization
- When to use stochastic GD
- Learning schedules
- Sklearn documentation

4) Session 52 (part 3): Mini-batch Gradient Descent

- **Introduction**
- **Code**
- **Visualization**

Week 25: Regression Analysis

1) Session on Regression Analysis (Part 1)

- What is Regression Analysis?
- Why Regression Analysis is required?
- What's the Statistic connection with Regression analysis?
- Inference vs Prediction
- Statsmodel Linear Regression
- TSS, RSS, and ESS
- Degree of freedom
- F-statistic and Prob(F-statistic)

2) Session on Regression Analysis (Part 2)

- F -test for overall significance
- R-squared (Goodness of fit)
- Adjusted R-squared
- T – Statistic
- Confidence Intervals for Coefficients

3) Session on Polynomial Regression

- Why we need Polynomial Regression?
- Formulation of Polynomial Regression
- Polynomial Regression in python

4) Session on Assumptions of Linear Regression

- Assumptions of Linear Regression
 - i. *Linearity*
 - ii. *Normality of Residuals*
 - iii. *Homoscedasticity*
 - iv. *No Autocorrelation*
 - v. *No or little Multicollinearity*
- What happen when these assumptions failed?
- How to check each of these assumptions?
- What to do when an assumption fails?
- Standard Error

5) Session 53: Multicollinearity

- What is multicollinearity?
- When is Multicollinearity bad?
- Multicollinearity (Mathematically)
- Perfect and non-perfect Multicollinearity
- Types of multicollinearity
- How to detect Multicollinearity
- Correlation
- VIF (Variance Inflation Factor)
- Condition Number
- How to remove Multicollinearity

Week 26: Feature Selection

1) Session 54: feature Selection Part 1

- What is Feature Selection?
- Why to do Feature Selection?
- Types of Feature Selection
- Filter based Feature Selection
 - i. *Duplicate Features*
 - ii. *Variance Threshold*
 - iii. *Correlation*
 - iv. *ANOVA*
 - v. *Chi-Square*
 - vi. *Advantages and Disadvantages*

2) Session 55: Feature Selection Part 2

- Wrapper method
- Types of wrapper method
- Exhaustive Feature Selection/Best Subset Selection

- Sequential Backward Selection/Elimination
- Sequential Forward Selection
- Advantages and Disadvantages

3) Session on Feature Selection part 3

- Embedded Methods
 - i. *Linear Regression*
 - ii. *Tree based models*
 - iii. *Regularized Models*
- Recursive Feature Elimination
- Advantages and Disadvantages

Week 27: Regularization

1) Session on Regularization Part 1 | Bias-Variance Tradeoff

- Why we need to study Bias and Variance
- Expected Value and Variance
- Bias and Variance Mathematically

2) Session on Regularization Part 1 | What is Regularization

- Bias Variance Decomposition
- Diagram
- Analogy
- Code Example
- What is Regularization?
- When to use Regularization?

3) Ridge Regression Part 1

- Types of Regularization
- Geometric Intuition
- Sklearn Implementation

4) Ridge Regression Part 2

- Ridge Regression for 2D data
- Ridge Regression for nD data
- Code from scratch

5) Ridge Regression Part 3

- **Ridge regression using Gradient Descent**

6) Ridge Regression Part 4

- 5 Key Understandings
 - i. *How do the coefficients get affected?*
 - ii. *Higher values are impacted more*
 - iii. *Bias variance tradeoff*
 - iv. *Contour plot*

v. Why is it called Ridge?

7) Lasso Regression

- Intuition
- Code example
- Lasso regression key points

8) Session on Why Lasso Regression creates Sparsity?

9) ElasticNet Regression

- Intuition
- Code example

Week 28: K Nearest Neighbors

1) Session on K nearest Neighbors Part 1

- KNN intuition
- Code Example
- How to select K?
- Decision Surface
- Overfitting and Underfitting in KNN
- Limitations of KNN

2) Session on coding K nearest Neighbors from scratch

3) Session on How to draw Decision Boundary for Classification problems

4) Session on Advanced KNN Part 2

- KNN Regressor
- Hyperparameters
- Weighted KNN
- Types of Distances (Euclidean and Manhattan)
- Space and Time Complexity
- KD-Tree

5) Classification Metrics Part 1

- Accuracy
- Accuracy of multi-classification problems
- How much accuracy is good?
- Problem with accuracy
- Confusion matrix
- Type 1 error
- Type 2 error
- Confusion matrix of multi-classification problems
- When accuracy is misleading

6) Classification Metrics Part 2

- Precision

- Recall
- F1 score
- Multi class Precision and Recall
- Multi class F1 score

Week 29: PCA

1) Session on Curse of Dimensionality

2) PCA part 1

- Introduction
- Geometric Intuition of PCA
- Why is Variance important?

3) PCA part 2

- Mathematical Problem formulation
- What is covariance and covariance matrix?
- Matrices as Linear Transformation
- EigenVectors and Eigenvalues
- Step by step solution of PCA
- How to transform points?
- PCA step-by-step code in python

4) PCA Part 3

- Practical example on MNIST dataset
- PCA demo with sklearn
- Visualization
- What is explained variance
- Find optimum number of Principal components
- When PCA does not work?

5) Session on Eigen vectors and Eigen Values

- What are Matrices?
- What are Eigen Vectors and Eigen Values?
- Intuition – Axis of rotation
- How to calculate Eigen Vectors and Eigen Values
- Properties
- Eigen Vectors in PCA

6) Session on Eigen Decomposition and PCA variants

- Types of PCA variants
- What are some special matrices?
 - i. *Diagonal Matrix*
 - ii. *Orthogonal Matrix*
 - iii. *Symmetric Matrix*

- Matrix as Linear Transformation Visualization tool
- Matrix Composition
- Matrix Decomposition
- Eigen decomposition
- Eigen decomposition of Symmetric Matrix (Spectral Decomposition)
- Advantages of Eigen decomposition
- Kernel PCA
- Code example of Kernel PCA

7) **Session on eigen Singular Value Decomposition (SVD)**

- Intuition of Non-Square Matrix
- Rectangular Diagonal Matrix
- What is SVD
- Applications of SVD
- SVD - The intuition of the mathematical equation
- Relationship with Eigen Decomposition
- Geometric Intuition of SVD
- How to calculate SVD
- SVD in PCA

Week 30: Model Evaluation & Selection

1. **ROC Curve in Machine Learning**

- ROC AUC Curve and its requirements
- Confusion matrix
- True Positive Rate (TPR)
- False Positive Rate (FPR)
- Different cases of TPR & FPR

2. **Session on Cross Validation**

- Why do we need Cross Validation?
- Hold-out approach
- Problem with Hold-out approach
- Why is the Hold-out approach used?
- Cross Validation
 - Leave One Out Cross Validation (LOOCV)
 - Advantages
 - Disadvantages
 - When to use
 - K-Fold Cross Validation

1. Advantages
 2. Disadvantages
 3. When to use
- iii. Stratified K-Fold CV
- 3. Session on Data Leakage**
 - a. What is it and what is the problem
 - b. Ways in which Data Leakage can occur
 - c. How to detect
 - d. How to remove Data Leakage
 - e. Validation set
- 4. Session on Hyperparameter Tuning**
 - a. Parameter vs Hyperparameter
 - b. Why the word "hyper" in the term
 - c. Requirements
 - i. Grid Search CV
 - ii. Randomized Search CV
 - d. Can this be improved?

Week 31: Naive Bayse

- 1. Crash course on Probability Part 1**
 - a. 5 important terms in Probability
 - i. Random Experiment
 - ii. Trials
 - iii. Outcome
 - iv. Sample Space
 - v. Event
 - b. Some examples of these terms
 - c. Types of events
 - d. What is probability
 - e. Empirical vs Theoretical probability
 - f. Random variable
 - g. Probability distribution of random variable
 - h. Mean of 2 random variable
 - i. Variance of Random variable
- 2. Crash course on Probability Part 2**
 - a. Venn diagrams
 - b. Contingency table
 - c. Joint probability
 - d. Marginal probability

- e. Conditional probability
- f. Intuition of Conditional Probability
- g. Independent vs Dependent vs Mutually Exclusive Events
- h. Bayes Theorem

3. Session 1 on Naive Bayes

- a. Intuition
- b. Mathematical formulation
- c. How Naive Bayes handles numerical data
- d. What if data is not Gaussian
- e. Naive Bayes on Textual data

4. Session 2 on Naive Bayes

- a. What is underflow in computing
- b. Log Probabilities
- c. Laplace Additive Smoothing
- d. Bias Variance Trade off
- e. Types
 - i. Gaussian Naive Bayes
 - ii. Categorical Naive Bayes
 - iii. Multinomial Naive Bayes

5. Session 3 on Naive Bayes

- a. Probability Distribution related to Naive Bayes
 - i. Bernoulli Distribution
 - ii. Binomial Distribution
 - iii. Categorical distribution / Multinoulli distribution
 - iv. Multinomial Distribution
 - v. Why do we need these distributions?
- b. Categorical Naive Bayes
- c. Bernoulli Naive Bayes
- d. Multinomial Naive Bayes
- e. Out of Core Naive Bayes

6. End to End Project | Email Spam Classifier

Week 32 : Logistics Regression

1. Session 1 on Logistic Regression

- a. Introduction
- b. Some Basic Geometry
- c. Classification Problem

- i. Basic Algorithm
 - ii. Updation in Basic Algorithm
- d. Sigmoid Function
- e. Maximum Likelihood
- f. Log Loss
- g. Gradient Descent
- h. summary

2. Session on Multiclass Classification using Logistic Regression

- a. What is Multiclass Classification
- b. How Logistic Regression handles Multiclass Classification Problems.
- c. One vs Rest (OVR) Approach
 - i. Intuition
 - ii. Code
- d. SoftMax Logistic Regression Approach
 - i. SoftMax Function
 - ii. Code
- e. When to use what?

3. Session on Maximum Likelihood Estimation

- a. Recap
- b. Some Examples
 - i. Example 1 - Coin Toss
 - ii. Example 2 - Drawing balls from bag
 - iii. Example 3 - Normal Distribution
- c. Probability Vs Likelihood
- d. Maximum Likelihood Estimation
- e. MLE for Normal Distribution
- f. MLE in Machine Learning
- g. MLE in Logistic Regression
- h. Some Important Questions

4. Session 3 on Logistic Regression

- a. Maximum Likelihood in Logistic Regression
- b. FAQ on MAE (Maximum Likelihood Estimation)
 - i. Is MLE a general concept applicable to all ML algorithms?
 - ii. How is MLE related to the concept of loss functions?
 - iii. Why does the loss function exist, why don't we maximize likelihood?
 - iv. Why study about maximum likelihood at all?
- c. An interesting task for you
- d. Assumptions of Logistics Regression

- e. Odds and Log(Odds)
- f. Another interpretation of Logistic Regression
- g. Polynomial Features
- h. Regularization in Logistic Regression

5. Logistic Regression Hyperparameters

Week 33 : Support Vector Machines (SVM)

1. SVM Part 1 – Hard Margin SVM

- a. Introduction
- b. Maximum Margin Classifier
- c. Support Vectors
- d. Mathematical Formulation
- e. How to solve this?
- f. Prediction
- g. Coding Example
- h. Problems with Hard Margin SVM

2. SVM Part 2 | Soft Margin SVM

- a. Problems with Hard Margin SVM
- b. Slack Variable
- c. Soft Margin SVM
- d. Introduction of C
- e. Bias-Variance Trade Off
- f. Code Example
- g. Relation with Logistic Regression

3. Session on Constrained Optimization Problem

- a. Problem with SVC
- b. Kernel's Intuition
- c. Coding examples of Kernel
- d. Types of Kernels
- e. Why is it called Trick?
- f. Mathematics of SVM

4. Session on SVM Dual Problem

- a. SVM in n Dimensions
- b. Constrained Optimization Problems
- c. Karush Kuhn Tucker Conditions
- d. Concept of Duality
- e. SVM Dual Problem

- f. Dual Problem Derivation
- g. Observations

5. Session on Maths Behind SVM Kernels

- a. SVM Dual Formulation
- b. The Similarity Perspective
- c. Kernel SVM
- d. Polynomial Kernel
 - i. Trick
 - ii. What about the other Polynomial terms
- e. RBF Kernel
 - i. Local Decision Boundary
 - ii. Effect of Gamma
- f. Relationship Between RBF and Polynomial K...
- g. Custom Kernels
 - i. String kernel
 - ii. Chi-square kernel
 - iii. Intersection kernel
 - iv. Hellinger's kernel
 - v. Radial basis function network (RBFN) kernel
 - vi. Spectral kernel

Extra Sessions – Feature Engineering

1. Session on Handling Missing Values Part - 1

- a. Feature Engineering
 - i. Feature Transformation
 - ii. Feature Construction
 - iii. Feature Extraction
 - iv. Feature Selection
- b. Types of Missing Values
 - i. Missing Completely at random
 - ii. Missing at Random
 - iii. Missing Not at Random
- c. Techniques for Handling Missing Values
 - i. Removing Missing Values
 - ii. Imputation
- d. Complete Case Analysis

2. Session 2 on Handling Missing Data

- a. Univariate Imputation – Numerical Data
 - i. Mean Imputation

- ii. Median Imputation
 - b. Univariate Imputation – Arbitrary Value & End Distribution Value
 - c. Univariate Imputation – Categorical Data
 - i. Mode Imputation
 - ii. Missing Category Imputation
 - d. Univariate Imputation – Random (Numerical + Categorical)
 - e. Missing Indicator
- 3. Session 3 on Handling Missing Data – Multivariate Imputation**
- a. KNN Imputer
 - i. Steps in KNN Imputation
 - ii. Advantages and Disadvantages in KNN Imputation
 - b. Iterative Imputer
 - i. MICE
 - ii. When To Use
 - iii. Advantages and Disadvantages
 - iv. Demonstration of MICE algorithm
 - v. Steps involved in Iterative Imputer
 - vi. Important parameter in Iterative Imputer

Week 34 : Decision Trees

- 1. Session 1 on Decision Tree**
- a. Introduction
 - b. Intuition behind DT
 - c. Terminology in Decision Tree
 - d. The CART Algorithm – Classification
 - e. Splitting Categorical Features
 - f. Splitting Numerical Features
 - g. Understanding Gini Impurity?
 - h. Geometric Intuition of DT
- 2. Session 2 on Decision Tree**
- a. CART for Regression
 - b. Geometric Intuition of CART
 - c. How Prediction is Done
 - d. Advantages & Disadvantages of DT
 - e. Project Discussion – Real Estate
- 3. Session 3 on Decision Tree**
- a. Feature Importance

- b. The Problem of Overfitting
- c. Why Overfitting happens
- d. Unnecessary nodes
- e. Pruning & its types
 - i. Pre-pruning
 - ii. Post Pruning
- f. Cost Complexity Pruning

Week 35 : Ensemble Methods – Introduction

1. Introduction to Ensemble Learning

- a. Intuition
- b. Types of Ensemble Learning
- c. Why it works?
- d. Benefits of Ensemble
- e. When to use Ensemble

2. Bagging Part 1 – Introduction

- a. Core Idea
- b. Why use Bagging?
- c. When to use Bagging?
- d. Code Demo

3. Bagging Part 2 – Classifier

- a. Intuition through Demo app
- b. Code Demo

4. Bagging Part 3 – Regressor

- a. Core Idea
- b. Intuition through demo web app
- c. Code Demo

5. Random Forest : Session 1

- a. Introduction to Random Forest
- b. Bagging
- c. Random Forest Intuition
- d. Why Random Forest Works?
- e. Bagging vs. Random Forest
- f. Feature Importance

6. Random Forest : Session 2

- a. Why Ensemble Techniques work?
- b. Random Forest Hyperparameters

- c. OOB Score
- d. Extremely Randomized Trees
- e. Advantages and Disadvantages of Random Forest

Week 36 : Gradient Boosting

1. Gradient Boosting : Session 1

- a. Boosting
- b. What is Gradient Boosting
- c. How
- d. What
- e. Why

2. Gradient Boosting : Session 2

- a. How Gradient Boosting works?
- b. Intuition of Gradient Boosting
- c. Function Space vs. Parameter Space
- d. Direction of Loss Minimization
- e. How to update the function
- f. Iterate
- g. Another perspective of Gradient Boosting
- h. Difference between Gradient Boosting and Gradient Descent

3. Gradient Boosting : Session 3 (Classification - 1)

- a. Classification vs. Regression
- b. Prediction

4. Gradient Boosting for Classification - 2 | Geometric Intuition

- a. Geometric Intuition

5. Gradient Boosting for Classification - 3 | Math Formulation

- a. Step 0 : Loss Function
- b. Step 1 : Minimise Loss Function to get $F_0(x)$
- c. Step 2 :
 - i. Pseudo Residuals
 - ii. Training Regression Tree
 - iii. Compute Lambda for all leaf nodes
 - iv. Update the Model
- d. Step 3 : Final Model
- e. Log(odds) vs Probability

