Data Science

Week 1 : Basics of Python Programming

1. Lesson 1: Python Basics

- Short info about DSMP
- 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. Lesson 2: Python Operators + if-else + Loops

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

3. Lesson 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

4. Lesson on Time complexity

- Start of the Lesson (00:00:00 – 00:11:22)

- PPT presentation on Time Complexity (Efficiency in Programming and Orders of Growth)

Week 2: Python Data Types

1. Lesson 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. Lesson 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. Lesson 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()

4. Paid Lesson on Career QnA
5. Array Interview Questions
6. Week 2 Interview Questions

Week 3: Object Oriented Programming (OOP)

1. Lesson 7: OOP Part1

- What is OOP?

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

2. Lesson 8: OOP Part2

- Revision of last session by solving problems
- How objects access attributes
- Attribute creation from outside of the class
- Reference Variables
- Mutability of Object
- Encapsulation
- Collection of objects
- Static variables and methods

3. Lesson 9: OOP Part3

- Class Relationship
- Aggregation and aggregation class diagram
- Inheritance and Inheritance class diagram
- Constructor example
- Method Overriding
- Super keyword
- Super constructor
- Practice questions on Inheritance
- Types of Inheritance (Single, Multilevel, Hierarchical,Multiple)
- Hybrid Inheritance
- Code example and diamond problem
- Polymorphism
- Method Overriding and Method Overloading
- Operator Overloading

4. Lesson on Abstraction

- What is Abstraction?
- Bank Example Hierarchy
- Abstract class
- Coding abstract class (BankApp Class)

5. Lesson on OOP Project

Week 4: Advanced Python:

1. Lesson 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. Lesson 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. Lesson 12: Decorators and Namespaces

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

4. Lesson 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. Lesson on Generator

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

6. Lesson on Resume Building

7. Lesson on GUI Development using Python

- GUI development using tkinter

8. Week 4 Interview Questions

Week 5: Numpy

1. Lesson 13: Numpy Fundamentals

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

2. Lesson 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. Lesson 15: Numpy Tricks

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

4. Lesson 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. Lesson 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. Lesson 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. Lesson 18: Important DataFrame Methods

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

4. Lesson on API Development using Flask

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

5. Lesson on Numpy Interview Question

Week 7: Advanced Pandas

1. Lesson 19: GroupBy Object

- What is GroupBy?
- Applying builtin aggregation fuctions on groupby objects

- GroupBy Attributes and Methods
- Hands-on on IPL dataset

2. Lesson 20: Merging, Joining, Concatenating

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

3. Lesson 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. Lesson 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. Lesson 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. Lesson on Git and GitHub:

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

Week 8: Advanced Pandas Continued

1) Lesson 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) Lesson 22: Vectorized String Operations | Datetime in Pandas

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

3) Lesson on Pandas Case Study – time Series analysis

4) Lesson on Pandas Case Study – Working with textual data

Week 9: Data Visualization

1) Lesson 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) Lesson 24: Advanced Matplotlib

- Colored Scatterplot
- Plot size, annotations
- Subplots
- 3D plots
- Contour plots
- Heatmaps
- Pandas plot()

3) Lesson on Plotly (Express)

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

- Hands-on Plotly

4) Lesson on Plotly Graph Objects (go)

5) Lesson on Plotly Dash

- Basic Introduction about Dash

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

7) Deploying a Dash app on Heroku

8) Lesson on Project using Plotly

- Project using Indian Census Data with Geospatial indexing Dataset

Week 10: Data Visualization Continued

1) Lesson 25: Plotting Using Seaborn- part 1

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

2) Lesson 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 Seabom - load_dataset
- Blog idea

3) Lesson on Open-Source Software – Part 1

4) Lesson on Open-Source Software – Part 1

Week 11: Data Analysis Process - Part1

1) Lesson 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 Scraping

2) Lesson 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) Lesson 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) Lesson on Advanced Web Scraping using Selenium

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

Week 12: Data Analysis Process – Part 2

1) Lesson on Data Cleaning Case Study – Smartphone dataset

- Quality issues
- Tidiness issues
- Data Cleaning

2) Lesson 29: Exploratory Data Analysis (EDA)

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

3) Lesson on Data Cleaning – Part 2

- Data Cleaning on Smartphone Dataset – Continued

4) Lesson on EDA Case Study – Smartphone Dataset

Week 13: SQL Basics

1) Lesson 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) Lesson 31: SQL DDL Commands

- Xammp Software
- Types of SQL commands
- DDL commands

3) Lesson on Tableau – Olympics Dataset (Part 1)

- Download and Install Tableau
- Getting Started with Tableau Desktop
- Overview of Olympic datasets
- Create Dashboards using Tableau

Week 14: SQL Continued – Part 1

1) Lesson 32: SQL DML commands

- MySQL workbench
- INSERT
- SELECT
- UPDATE
- DELETE
- Functions in SQL

2) Lesson 33: SQL Grouping and Sorting

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

3) Lesson on Tableau – Part 2

- Tableau Basics
    i. *Importing Data*
    ii. *Measures and Dimensions*
    iii. *Sheet, dashboard, story*
    iv. *Levels of Granularity*
    v. *Different types of charts*
    vi. *Datetime*
- Hierarchical level of granularity
- Common filters
- Calculated fields and Table Calculations
- Working with Geographical columns
- Dashboard and Interactive filters
- Blending and Dual axis chart
- Connecting to a remote database

Week 15: SQL Continued - Part 2

1) Lesson 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) Lesson on SQL Case Study 1 – Zomato Dataset

- Understanding Dataset through diagram
- Solving SQL Questions on Zomato Dataset

3) Lesson 35: Subqueries in SQL

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

4) Lesson on Making a Flight Dashboard using Python and SQL

- How to Connect MYSQL through Python
- Run SQL queries with Python
- Creating a dynamic dashboard with Streamlit on Flights dataset

5) Lesson on SQL Interview Questions – Part 1

- Database Server Vs Database Client
- Database Engines
- Components of DBMS
- What is Collation?
- COUNT(*) vs COUNT(col)
- Dealing with NULL values
- DELETE Vs TRUNCATE
- Anti joins
- Non-equi joins
- Natural joins
- All and Any operators
- Removing Duplicate Rows
- Metadata Queries

Week 16: Advanced SQL

1) Lesson 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) Lesson 37: Windows Functions Part 2

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

3) Lesson 37: Windows Functions Part 3

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

4) Lesson on Data Cleaning using SQL | Laptop Dataset

- Basic level Data Cleaning and Data exploration using SQL
- Why to use SQL for Data Cleaning

- String Data types
- Wildcards
- String Functions
- Data Cleaning

5) Lesson on EDA using SQL | Laptop Dataset

- EDA on numerical and categorical columns
- Plotting
- Categorical – Categorical Analysis
- Numerical – Numerical Analysis

Week 17: Descriptive Statistics

1) Lesson 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) Lesson 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) Lesson 39: Descriptive Statistics part 2

- Quantiles and Percentiles
- Five Number Summary
- Boxplots
- Scatterplots
- Covariance
- Correlation
- Correlation vs Causation
- Visualizing multiple variables

2) Lesson 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.

3) Lesson on SQL Datetime Case Study on Flights Dataset

4) Lesson on Database Design | SQL Data Types | Database Normalization

- Different SQL Datatypes (Numeric, Text, Datetime, Misc)
- Database Normalization
- ER Diagram

Week 19: Probability Distributions

1) Lesson 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) Lesson 42: Non-Gaussian Probability Distributions

- Kurtosis
- Excess Kurtosis and Types of kurtosis
- QQ plot

- Uniform Distribution
- Log-normal distribution
- Pareto Distribution
- Transformations
    i. *Mathematical Transformation*
    ii. *Function Transformer*
    iii. *Log Transform*
    iv. *Reciprocal Transform / Square or sqrt Transform*
    v. *Power Transformer*
    vi. *Box-Cox Transform*
    vii. *Yeo-Johnson Transformation*

3) Lesson on views and User Defined Functions in SQL
- What are views?
- Types of views
- User Defined Functions (Syntax, Examples, Benefits)

4) Lesson on Transactions and Stored Procedures
- Stored Procedures
- Benefits of using stored procedures
- Transactions (Commit, rollback, savepoint)
- ACID properties of a Transaction

Week 20: Inferential Statistics

1) Lesson 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
- Case study
- Assumptions of making samples

2) Lesson on Central Limit Theorem Proof

3) Lesson 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) Lesson 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) Lesson 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
    i. *Single sample t-Test*
    ii. *Independent 2-sample t-Test*
    iii. *Paired 2 sample t-Test*
    iv. *Code examples of all of the above*

3) Lesson 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) Lesson on ANOVA

- Introduction

- F-distribution
- One-way ANOVA
    i. *Steps*
    ii. *Geometric Intuition*
    iii. *Assumptions*
    iv. *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) Lesson 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) Lesson 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) **Lesson 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) Lesson 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) Lesson 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) Lesson 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

5) Lesson on Differential Calculus

- What is differentiation?
- Derivative of a constant
- Cheatsheet
- Power Rule
- Sum Rule
- Product Rule
- Quotient Rule
- Chain Rule
- Partial Differentiation
- Higher Order Derivatives
- Matrix Differentiation

Week 24: Gradient Descent

1) Lesson 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) Lesson 52 (part 1): Batch Gradient Descent

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

3) Lesson 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) Lesson 52 (part 3): Mini-batch Gradient Descent
- **Introduction**
- **Code**
- **Visualization**

5) Doubt Clearance session on Linear Regression

Week 25: Regression Analysis

1) Lesson 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) Lesson 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) Lesson on Polynomial Regression
- Why we need Polynomial Regression?
- Formulation of Polynomial Regression
- Polynomial Regression in python

4) Lesson 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) Lesson 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) Lesson 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) Lesson 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) Lesson 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) Lesson on Regularization Part 1 | Bias-Variance Tradeoff

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

2) Lesson 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) Lesson on Why Lasso Regression creates Sparsity?

9) ElasticNet Regression

- Intuition

- Code example
10) **Doubt Clearance session on regularization**

Week 28: K Nearest Neighbors

1) **Lesson 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) **Lesson on coding K nearest Neighbors from scratch**

3) **Lesson on How to draw Decision Boundary for Classification problems**

4) **Lesson 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) **Lesson 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) Lesson 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) Lesson 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) **Lesson 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**
   a. ROC AUC Curve and it's requirements
   b. Confusion matrix
   c. True Positive Rate (TPR)
   d. False Positive Rate (FPR)
   e. Different cases of TPR & FPR

2. **Lesson on Cross Validation**
   a. Why do we need Cross Validation?
   b. Hold-out approach
   c. Problem with Hold-out approach
   d. Why is the Hold-out approach used?
   e. Cross Validation
      i. Leave One Out Cross Validation (LOOCV)
         1. Advantages
         2. Disadvantages
         3. When to use
      ii. K-Fold Cross Validation
         1. Advantages
         2. Disadvantages
         3. When to use
      iii. Stratified K-Fold CV

3. **Lesson 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. Lesson 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. Lesson 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 Bayse on Textual data

4. Lesson 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. Lesson 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. Lesson 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. Lesson 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?

    f. Tasks

3. Lesson 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. Lesson 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. **Lesson 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 if SVM

4. **Lesson 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. **Lesson 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 Lessons - Feature Engineering

1. Lesson 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. Lesson 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. Lesson 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. Lesson 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. Lesson 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. Lesson 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

4. Lesson on Decision Tree Visualization

      a.   dtreeviz Demo - Coding

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 : Lesson 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 : Lesson 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 : Lesson 1
   a.   Boosting
   b.   What is Gradient Boosting
   c.   How
   d.   What
   e.   Why

2. Gradient Boosting : Lesson 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 : Lesson 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

Capstone Project:

1. Lesson 1 on Capstone Project | Data Gathering
   a.   Project overview in details
   b.   Gather data for the project
   c.   Details of the data

2. Lesson 2 on Capstone Project | Data Cleaning

    a. Merging House and Flats Data

    b. Basic Level Data Cleaning

3. Lesson 3 on Capstone Project | Feature Engineering

    a. Feature Engineering on Columns:

        i. additionalRoom

        ii. areaWithType

        iii. agePossession

        iv. furnishDetails

        v. features : luxury Score

4. Lesson 4 on Capstone Project | EDA

    a. Univariate Analysis

    b. PandasProfiling

    c. Multivariate Analysis

5. Lesson 5 on Capstone Project | Outlier Detection and Removal

    a. Outlier Detection And removal

6. Lesson 6 on Capstone Project | Missing Value Imputation

    a. Outlier Detection and Removal on area and bedroom

    b. Missing Value Imputation

7. Lesson 7 on Capstone Project | Feature Selection

    a. Feature Selection

        i. Correlation Technique

        ii. Random Forest Feature Importance

        iii. Gradient Boosting Feature Importance

        iv. Permutation Importance

        v. LASSO

        vi. Recursive Feature Elimination

        vii. Linear Regression with Weights

        viii. SHAP (Explainable AI)

    b. Linear Regression - Base Model

        i. One-Hot Encoding

        ii. Transformation

        iii. Pipeline for Linear Regression

    c. SVR

8. Lesson 8 on Capstone Project | Model Selection & Productionalization

    a. Price Prediction Pipeline

        i. Encoding Selection

            1. Ordinal Encoding

            2. OHE

3. OHE with PCA
           4. Target Encoding
       ii.   Model Selection
   b. Price Prediction Web Interface -Streamlit
9. Lesson 9 on Capstone Project | Building the Analytics Module
   a. geo map
   b. word cloud amenities
   c. scatterplot -> area vs price
   d. pie chart bhk filter by sector
   e. side by side boxplot bedroom price
   f. distplot of price of flat and house
10. Lesson 10 on Capstone Project | Building the Recommender System
   a. Recommender System using TopFacilities
   b. Recommender System using Price Details
   c. Recommender System using LocationAdvantages
11. Lesson 11 on Capstone Project | Building the Recommender System Part 2
   a. Evaluating Recommendation Results
   b. Web Interface for Recommendation (Streamlit)
12. Lesson 12 on Capstone Project | Building the Insights Module
13. Lesson 13 on Capstone Project | Deploying the application on AWS

Extreme Gradient Boosting(XGBoost)

1. **Introduction to XGBoost**
   ● Introduction
   ● Features
       ○ Performance
       ○ Speed
       ○ Flexibility
2. **XGBoost for Regression**
   a. Regression Problem Statement
   b. Step-by-Step Mathematical Calculation
3. **XGBoost for Classification**
   a. Classification Problem Statement
   b. Step-by-Step Mathematical Calculation
4. **The Complete Maths of XGBoost**
   a. Prerequisite & Disclaimer
   b. Boosting as an Additive Model

c. XGBoost Loss Function
   d. Deriving Objective Function
   e. Problem With Objective Function and Solution
      i. The Taylor series
      ii. Applying Taylor Series
      iii. Simplification
   f. Output Value for Regression
   g. Output Value for Classification
   h. Derivation of Similarity Score
   i. Final Calculation of Similarity Score

KMeans Clustering

1. Lesson 1 on KMeans Clustering
   a. Plan of Attack (Getting Started with Clustering)
   b. Types of ML Learning
   c. Applications of Clustering
   d. Geometric Intuition of K-Means
   e. Elbow Method for Deciding Number of Clusters
      i. Code Example
      ii. Limitation of Elbow Method
   f. Assumptions of KMeans
   g. Limitations of K Means

2. Lesson 2 on KMeans Clustering
   a. Recap of Last class
   b. Assignment Solution
   c. Silhouette Score
   d. Kmeans Hyperparameters
      i. Number of Clusters(k)
      ii. Initialization Method (K Means++)
      iii. Number of Initialization Runs (n_init)
      iv. Maximum Number of Iterations (max_iter)
      v. Tolerance (tol)
      vi. Algorithm (auto, full, ..)
      vii. Random State
   e. K Means ++

3. Lesson 3 on KMeans Clustering
   a. K-Means Mathematical Formulation (Loyd's Algorithm)
   b. K-Means Time and Space Complexity
   c. Mini Batch Kmeans

      d. Types of Clustering
           i.   Partitional Clustering
           ii.   Hierarchical Clustering
           iii.   Density Based Clustering
           iv.   Distribution/Model-based Clustering

4. K-Means Clustering Algorithms from Scratch in Python

      a. Algorithms implementation from Scratch in Python

Other Clustering Algorithms

1. Lesson on DBSCAN

      a. Why DBSCAN?
      b. What is Density Based Clustering
      c. MinPts & Epsilon
      d. Core Points, Border Points & Noise Points
      e. Density Connected Points
      f. DBSCAN Algorithm
      g. Code
      h. Limitations
      i. Visualization

2. Lesson on Hierarchical Clustering

      a. Need of Other Clustering Methods
      b. Introduction
      c. Algorithm
      d. Types of Agglomerative Clustering
           i.   Min (Single-link)
           ii.   Max (Complete Link)
           iii.   Average
           iv.   Ward
      e. How to find the ideal number of clusters
      f. Hyperparameter
      g. Code Example
      h. Benefits/Limitations