Machine Learning Introduction - (2-3h)
1. What is Machine Learning?
2. Types of Machine Learning
   a. Supervised
   b. Unsupervised
   c. Reinforcement
3. The 7 Steps of Machine Learning
   a. Data Gathering
   b. Data Preparation
      i. Removing Anomalies
      ii. Visualization of Data
      iii. Splitting in Training and Testing Data
   c. Choosing Model
   d. Training the Model
   e. Evaluation
   f. Parameter Tuning
   g. Prediction or Inference

Numerical Python - (3h)
1. Numpy (2h)
   a. What is Numpy used for?
   b. n-dimensional numpy Array and their properties
   c. n-dimensional special Arrays e.g. zeroes, ones, range, random and linspace
   d. Reshaping of arrays
   e. Elementwise operations
   f. Indexing, Slicing and Iterating
   g. Flattening arrays
   h. Transpose of array
   i. Stacking arrays
   j. Splitting arrays
   k. Boolean Masking
2. Scipy (1h)
   a. What is Scipy used for?
   b. Differentiation
   c. Integration
   d. Optimization
   e. Basic Statistics

Data Visualization in Python- (3h)
1. Matplotlib (1-1.5h)
   a. Graph Life Cycle
   b. Basic Plots eg. Scatter plot, Histograms
2. Seaborn (1h)
3. Ggplot(1h)

Pandas- (2h)
1. Dataframes in Pandas
2. Reading data in Pandas using read_csv, read_excel, read_json, read_pickle and read_html
3. Data Cleaning with Pandas
4. Data Manipulation with Pandas

Mathematical Preliminaries for Machine Learning- (9h)
1. Probability (2h)
    a. Need and Significance of Probability in ML
    b. Making sense of Probability
    c. Basics of Probability
    d. Bayes Theorem

2. Statistics with Scipy (2h)
    a. Need and Significance of Statistics in ML
    b. Significance of Mean, Median, Mode, Variance, Standard Deviations, and Covariance with respect to important mathematical models as Gaussian & Binomial.
    c. Null Hypothesis, Alternate Hypothesis, Z-Scores, P-values and Chi Square Test

3. Calculus with Scipy (3h)
    a. Need and Significance of Calculus in ML
    b. Derivatives
    c. Maxima and Minima
    d. Optimizing a function using Gradient Descent

4. Linear Algebra with Numpy (2h)
    a. Need and Significance of Linear Algebra in ML
    b. Making sense of a Matrix
    c. Matrix Addition, Subtraction, Inversion and Multiplication
    d. Eigen values and Eigen Vectors

Linear Models for Supervised ML in Depth- (4h)
1. Linear Regression (Derivation of Cost and Manually Fitting a Small Dataset) (2.5h)
2. Logistic Regression (Sigmoid Function and Translating a Regression model to Classification Model) (2h)

Validation Techniques - (1h)
1. K-fold
2. LOOV

Multivariate Models, their problems and possible solutions- (3h)
1. Polynomial Regression
2. Bias Variance Dichotomy
3. Regularization
4. Normalization
5. Curse of Dimensionality
6. Underfitting and Overfitting
7. Hyperparameters

Machine Learning with Scikit Learn- (11.5)
1. Supervised Learning- (4h)
    a. Regression (1h)
        i. Linear Regression
        ii. Ridge Regression
    b. Classification (3h)
        i. Naive Bayes
        ii. Decision Trees
        iii. Random Forest
        iv. K Nearest Neighbors
        v. Support Vector Machine

2. Unsupervised Learning (2h)
    a. K-Means Clustering
    b. Hierarchical Clustering

3. Dimensionality Reduction (1.5h)
    a. PCA
    b. LDA
    c. t-SNE

4. Anomaly Detection (1h)
    a. Outlier Detection using Local Outlier Factor

What Next? (*What content students should refer to after learning this much*)(0.5h)

Total hours - 45

Prerequisites - Python Fundamentals, Practical, Abit of *interest* in mathematics.

Project to clear ML concepts.

Revisiting the 7 Steps of Machine Learning for House Price Prediction for Ghaziabad- (4-5h)
> **Decide What to Predict ?**

1. Data Gathering -
   Collect House Prices from 99acres.com for Ghaziabad using web-scraping and collect all possible features
   i. plotarea(with metric)
   ii. Property Age
   iii. New Booking/Resale
   iv. Facing Road Width (with metric)
   v. Cost (with unit ie. crore, lacs, thousands)

2. Data Preparation -
   i. Remove Missing Values columns.
   ii. Convert plotarea and Facing Road width to common metric
   iii. Convert NewBooking/Resale to Integer values 0/1
   iv. Convert Cost to common unit
   v. Visualize the data and try to see whether it fits a proper mathematical model eg. Gaussian, Binomial.
   vi. split data into 3:1 or 4:1.

3. Choose the model
   a. Justify that what kind of problem it is, ie. Supervised, Unsupervised etc.
   b. Justify what kind of subproblem it is ie. Classification, Regression.
   c. Try choosing a Model.

4. Train the model.

5. Score the model and try cross validating it.

6. Retune the parameters and retrain the model.

7. Repeat 3 to 6 till satisfied.

8. Make Real Inferences to see how close the result is to the actual, apart from dataset.