Syllabus : Numerical Analysis

Introduction to Numerical Analysis:

Overview of numerical methods and their importance in data science. Errors in numerical computations: sources and types of errors.

Root Finding and Optimization:

Bisection method, Newton-Raphson method for finding roots of equations. Optimization techniques: gradient-based methods (e.g., gradient descent) for optimization problems.

Interpolation and Curve Fitting:

Polynomial interpolation (Lagrange interpolation, Newton's divided differences). Least squares fitting for curve fitting and regression.

Numerical Differentiation and Integration:

Approximating derivatives using finite differences, Numerical integration techniques: Trapezoidal rule, Simpson's rule, and Romberg integration.

Linear Algebra Computations:

Numerical methods for solving linear systems: Gaussian elimination, LU decomposition, iterative methods (e.g., Jacobi, Gauss-Seidel) Eigenvalue computation methods: Power method, QR algorithm.

Numerical Solutions to Differential Equations:

Euler's method for initial value problems, Runge-Kutta methods for ordinary differential equations. Finite difference methods for partial differential equations.

Numerical Linear Algebra:

Matrix factorizations and their applications in numerical computations, Singular Value Decomposition (SVD) and its relevance in data analysis.

Error Analysis and Stability:

Understanding numerical stability and conditioning of problems, Error analysis in numerical methods and their impact on data analysis.

Applications in Data Science:

Implementation and applications of numerical methods in machine learning algorithms (e.g., optimization techniques in training models), Handling large datasets using efficient numerical algorithms.

Essential topics of Matrices

Introduction to Matrices:

Definition and representation of matrices.

Basic matrix operations: addition, subtraction, scalar multiplication.

Matrix dimensions, transpose, and properties.

Matrix Operations:

Properties of matrix multiplication.

Matrix-vector multiplication.

Matrix-matrix multiplication.

Relationship with composition of linear transformations.

Inverse matrices and conditions for invertibility.

Special Matrices:

Identity matrix, zero matrix, diagonal matrix.

Symmetric, skew-symmetric, and orthogonal matrices.

Positive definite matrices,

Rank and Systems of Linear Equations:

Rank of matrices, Row reduction, normal and echelon forms, Linear transformations

Representing systems of linear equations using matrices.

Gaussian elimination and matrix inversion methods for solving systems of equations.

Inverse of Non-singular matrices by Gauss-Jordan method.

Solving system of Homogeneous and Non-Homogeneous equations by Gauss elimination method

Matrix Factorizations and Decompositions:

LU decomposition.

QR decomposition.

Singular Value Decomposition (SVD).

Eigenvalue decomposition and its applications.

Matrix Operations and Applications:

Matrix inversion and solving systems of linear equations.

Determinants and their significance.

Matrix norms and their applications.

Linear dependence and independence

Eigenvalues and Eigenvectors:

Definition, properties, and computation of eigenvalues and eigenvectors.

Linear Transformation and Orthogonal Transformation, Diagonalization of a matrix, Cayley-Hamilton Theorem, finding inverse and power of a matrix by Cayley-Hamilton Theorem, Quadratic forms and Nature of the Quadratic Forms, Reduction of Quadratic form to canonical forms by Orthogonal Transformation.

Principal Component Analysis (PCA) and its relation to eigenvalues/eigenvectors.

Applications in Data Science:

Linear regression and its matrix formulation.

Matrix methods in machine learning algorithms (e.g., gradient descent, regularization).

Applications of matrix decompositions in data analysis, recommendation systems, and dimensionality reduction.

Sparse Matrices:

Definition and representation of sparse matrices.

Efficient storage and computation techniques for sparse matrices.

Advanced Topics (depending on the depth of the course and available time):

Non-negative Matrix Factorization (NMF).

Applications of matrices in graph theory, network analysis, and image processing.

Additional matrix factorizations or specialized techniques.

Differential Calculus

Chapter 1:

Limit, continuity and differentiability of function of single variable, Cauchy’s definition, Heine’s definition, Uniform continuity, Borel’s theorem, boundedness theorem, Bolzano’s theorem, Intermediate value theorem, extreme value theorem, Darboux's intermediate value theorem for derivatives, Chain rule, indeterminate forms.

Chapter 2:

Rolle’s theorem, Lagrange and Cauchy Mean value theorems, mean value theorems of higher order, Taylor's theorem with various forms of remainders, Successive differentiation, Leibnitz theorem, Maclaurin’s and Taylor’s series, Limit and Continuity of functions of two variables, Differentiation of function of two variables, Necessary and sufficient condition for differentiability of functions two variables.

Chapter 3:

Partial differentiation, Euler’s theorem on homogeneous function, Schwarz’s and young theorem, Taylor's theorem for functions of two variables with examples, Maxima and minima for functions of two variables, Lagrange multiplier method, Jacobians, Inverse function theorem and implicit function theorem.

Chapter 4:

Tangents and normals, Asymptotes, Curvature, Envelops and evolutes, Tests for concavity and convexity, Points of inflexion, Multiple points, Parametric representation of curves and tracing of parametric curves, Tracing of curves in Cartesian and Polar forms.

Integral Calculus

Chapter 1:

The Definite Integral, Properties of the Integral and the Average Value, The Fundamental Theorem and Its Consequences, Riemann integral, Integrability of continuous and monotonic functions, Fundamental theorem of integral calculus, Mean value theorems of integral calculus, Differentiation under the sign of Integration.

Chapter 2:

Evaluation of Double Integrals (Cartesian and polar coordinates), change of order of integration (only Cartesian form), Evaluation of Triple Integrals: Change of variables (Cartesian to polar) for double and (Cartesian to Spherical and Cylindrical polar coordinates) for triple integrals. Applications: Areas (by double integrals) and volumes (by double integrals and triple integrals)

Chapter 3:

Areas and Volumes by Slices, Length of a Plane Curve, Area of a Surface of Revolution, Probability and Calculus, Masses and Moments, Force, Work, and Energy

Chapter 4:

Vector Differentiation, Gradient, Divergence and Curl, Normal on a surface, Directional Derivative, Vector Integration, Theorems of Gauss, Green, Stokes and related problems.

Syllabus :Basic Statistics for Data Science

Introduction:

Data, data types, theory of Statistics, applications and real-life case studies, Data Science and its application.

Survey Sampling:

Concept of statistical population, Attributes and variables (discrete and Continuous), Sampling vs. complete enumeration: sampling units and frame, Precision and efficiency of estimators, Simple Random sampling with and without replacement, Use of random number tables in selection of simple random sample.

Statistical Methods:

Different types of scales – nominal, ordinal, ratio and interval, Primary data – designing a questionnaire and schedule, collection of primary data, checking their consistency, Secondary data; scrutiny of data for internal consistency and detection of errors of recording, Ideas of cross validation, Presentation of data; classification, tabulation, diagrammatic & graphical representation of grouped data, Frequency distributions, cumulative frequency distributions and their graphical representations, histogram, frequency polygon and ogives, Stem and Leaf Plot, Box Plot, Measure of central tendency and dispersion, merits and demerits of these measures, Moments and factorial moments, Shephard's correction for moments, Skewness and Kurtosis and their measures, Measures based on quartiles.

Correlation & Regression:

Bivariate data, Method of least squares for curve fitting, Correlation and regression, rank correlation (Spearman's and Kendall's measure), Intra-class correlation, correlation ratio, Partial and Multiple Correlation & Multiple Regression for Tri-variate data.

Probability Theory and Random Variable

Experiment & Probability:

Random experiment, trial, sample point and sample space, events, operations of events, concepts of equally likely, mutually exclusive and exhaustive events. Definition of probability: Classical, relative frequency and axiomatic approaches. Discrete probability space, properties of probability under set theoretic approach, Independence of events, Conditional probability, total and compound probability theorems, Bayes theorem and its applications.

Random Variable & Expectation

Random variables – discrete and continuous, probability mass function (PMF) and probability density function (PDF), Cumulative distribution function (CDF), Joint distribution of two random variables, marginal and conditional distributions, Independence of random variables. Expectation of a random variable (RV) and its properties, expectation of sum of random variables and product of independent random variables, conditional expectation.

Moments & Law of Large number:

Moments, moment generating function (MGF) & their properties, continuity theorem for MGF (without proof), Chebyshev's inequality, Weak law of large numbers and Central Limit Theorem for a sequence of independently and identically distributed random variables and their applications.

Probability Distribution:

Distribution, Univariate distributions: Binomial, Poisson, Hypergeometric, Geometric and Negative Binomial, Uniform (discrete & continuous), Normal, Exponential, Gamma, Beta distributions, Cauchy, Laplace, Pareto, Weibull, log normal Distributions, Normal and Poisson distributions as limiting case of binomial distribution, Distributions of function of random variables: Distribution of sum, product and quotient of two Variable. Reproductive property of standard distributions, χ 2 (chi-square), t and F distributions (Central cases only) and their limiting forms. Bivariate normal distribution and its properties.

Estimation and Testing of Hypothesis

Point Estimation:

Point estimation. Characteristics of a good estimator: Unbiasedness, consistency, sufficiency and efficiency, Method of maximum likelihood and properties of maximum likelihood estimators (without proof), Method of minimum Chi-square, Method of Least squares and method of moments for estimation of parameters.

Interval Estimation:

Concept of Interval Estimation, Pivotal quantity and its use, Confidence Interval, concept of best confidence intervals, Different types of Confidence Interval

Testing of Hypothesis:

Statistical Hypothesis (simple and composite), Testing of hypothesis, Type I and Type II errors, significance level, p-values, power of a test

General Statistical Test: Z test, t-test, Chi-Square test, F-test etc.

ANOVA and Design of Experiment

Analysis of variance:

Analysis of Variance, One way classification, Assumptions regarding model, Two-way classification with equal number of observations per cell, Duncan's multiple comparison test, Analysis of covariance.

Design of Experiment:

Principles of Design of experiments: Randomization, Replication and local control, Choice of size and type of a plot using uniformity trials, CRD, Randomized block design, Concept and definition of efficiency of design, Comparison of efficiency between CRD and RBD.

Latin Square Design:

Latin square Design, Lay-out, ANOVA table, Comparison of efficiencies between LSD and RBD; LSD and CRD, Missing plot technique: estimation of missing plots by minimizing error sum of squares in RBD and LSD with one or two missing observations.

Factorial Design:

Factorial Experiments: general description of factorial experiments; 2^2, 2^3 and 2^n factorial experiments arranged in RBD and LSD. Definition of main effects and interactions in 2^2 and 2^3 factorial experiments, Preparation of ANOVA by Yates procedure, Estimates and tests for main and interaction effects (Analysis without confounding).

Fundamental of Regression Analysis

Overview of Modelling:

Concept of Modelling, Different types of models, live case studies, Understanding the basics of regression analysis, Difference between correlation, regression and causation

Simple Linear Regression:

Understanding the concept of a linear relationship between two variables, graphical representation of relationship, Types of variable and models (Univariate & Multivariate), Assumptions of Linear Regression, Formulation of the simple linear regression model, estimating coefficients (slope and intercept) using Ordinary Least Squares (OLS) method and maximum likelihood estimation.

Multiple Linear Regression:

Introduction to multivariable model, Multicollinearity, Heteroscedasticity and Autocorrelation, Assumption and graphical testing, Coefficient estimation using OLS and MLE, Interpretation of coefficients, adjusted R-squared, and model evaluation, model diagnostics in multiple linear regression, Diagnostic tools: residual analysis, leverage, and influential points.

Variable Selection and Model Building:

Feature selection methods (e.g., forward selection, backward elimination, stepwise regression), Advanced Topics in Linear Regression.

Regularization Techniques:

Introduction to regularization methods (Ridge, Lasso) to handle multicollinearity and overfitting, Implementation of regularization in linear regression models, Practical Applications and Case Studies.

Advanced Regression Analysis

Models for Cross Sectional Data:

Cross sectional and Longitudinal data, Linear probability model (LPM), Logit Model and Probit Model, Ordinal Regression Model, Multinomial Regression Model.

Polynomial Regression Model:

Basic Principle of Polynomial function, Polynomial Models in One Variable, Piecewise polynomial fitting, Nonparametric Regression, Kernel Regression, Locally Weighted Regression (Loess).

Non-Linear Regression:

Linear and Nonlinear Regression Models, Nonlinear Regression Models, Origins of Nonlinear Models, Non-linear Least Squares, Transformation to a Linear Model, Parameter Estimation in a Nonlinear System, Linearization, Other Parameter Estimation Methods, Statistical Inference in Nonlinear Regression, Examples of Nonlinear Regression Models.

Logistic Regression Models:

Models with a Binary Response Variable, Estimating the Parameters in a Logistic Regression Model, Interpretation of the Parameters in a Logistic Regression Model, Statistical Inference on Model Parameters, Diagnostic Checking in Logistic Regression, Other Models for Binary Response Data.

Poisson Regression:

The Generalized Linear Model, Link Functions and Linear Predictors, Parameter Estimation and Inference in the GLM, Prediction and Estimation with the GLM, Residual Analysis in the GLM.

Syllabus : Computational Thinking

Introduction to Computational Thinking:

Understanding the core concepts of computational thinking: decomposition, pattern recognition, abstraction, and algorithmic thinking, Overview of how computational thinking applies to problem-solving in various disciplines.

Analytical Thinking and Problem Decomposition:

Understanding problem decomposition: breaking down complex problems into smaller, manageable parts, Techniques for identifying patterns and recognizing recurring themes within problems.

Algorithm Design:

Introduction to flow chart and various structures, Introduction to algorithms: definition, characteristics, and examples, developing step-by-step procedures and strategies to solve problems.

Logical Reasoning and Logical Operations:

Introduction to logical operations (AND, OR, NOT) and truth tables, Using logic gates and Boolean algebra for problem-solving.

Control Structure and Modular Programming:

Conditional statements, looping statements and control statements in loop, functional programming and its use.

Data Structures and Memory use:

Different types of data structure and how their memory allocation

Syllabus : Basic R Programming

Introduction of R Programming:

Introduction and history of R programming, Basic fundamentals, installation and use of software, R environment, options and function, different windows and support system, use of R as a calculator, Syntax, Identifier, Keyword, Basic Operators, Introduction to R Studio and different platforms (VS code, Jupyter Notebook etc.)

Data Types & Structure:

Data Types: Numeric, Integer, Complex, Double Character, String, Logical (Boolean) etc. Data Structure: Vector, Array, Matrix, Data Frame, List and Factor, Tables etc. Operations on different data structures: indexing, subsetting, and manipulation.

Data Import/Export in R:

Input data manually, taking input from user, Reading and writing data from different file formats (CSV, Excel, text files).

Data Cleaning and Basic Functions:

Handling missing data: identifying, removing, or imputing missing values, Data reshaping, merging, and transforming, Calculating basic Statistics results (Central Tendency and Dispersions).

String Handling:

Creation of String, length, Whitespace characters, Concatenation, Subsetting and Manipulation, Case Conversion of strings

Control Structures:

Conditional Statements: if, if-else Statements, if-else if-else Statements etc. Looping Structures: for loop, while loop, Loop Control Statements: Break, Next, Vectorized Operations vs. Loops

Functions and Base Visualization:

Function type, Function argument, Writing user defined function, utility and operations of data visualization (dot chart, bar chart, frequency chart etc.)

Advanced R Programming

R Package and Automatic Learning tool:

Introduction to R packages, use and functions (dplyr, ggplot2, tidyr, lubridate, readxl etc.) Introduction about Swirl Packages and its use.

Functional Programming and Advanced Control Structure

Using higher-order functions, anonymous functions, and functional idioms in R, Using the apply family functions (apply, lapply, sapply, tapply)

Debugging and Profiling:

Techniques for debugging R code effectively, Profiling code for performance optimization,

Performance Optimization:

Efficient programming techniques and best practices for faster code execution, Identifying and addressing performance bottlenecks.

Object-Oriented Programming in R:

S4 and reference classes: understanding and implementing object-oriented programming paradigms in R, Creating and using classes and methods in R.

Parallel and High-Performance Computing in R:

Parallel computing with R: using parallel, snow, and other packages for parallel processing, Leveraging multicore architectures for improved performance.

Advanced Functionalities and Techniques:

Working with big data: handling large datasets in R (using packages like data.table), Advanced statistical analysis: exploring additional statistical models and techniques.

Featuring R

Add-on Skills

- Introduction to R markdown
- Version Control using GitHub
- R Shiny

Capstone Project:

- Web Scrapper
- Text Mining
- Sentiment Analysis
- Writing and Documenting R packages

Data Visualization using R

Introduction to Data Visualization:

Importance and principles of data visualization, Exploring types of visualizations and their applications

Basic Plotting in Base R:

Creating simple plots using base R functions like plot(), hist(), barplot(), etc. Customizing plot appearance: labels, titles, axes, colors.

ggplot2 Fundamentals:

Understanding the grammar of graphics, Exploring ggplot2 syntax for creating versatile and customizable visualizations.

Advanced ggplot2:

Customizing visualizations with facets, themes, and annotations, Using additional geoms and scales for specialized plots.

Interactive Visualizations:

Creating interactive plots using packages like plotly, ggplotly, and Shiny. Incorporating interactivity for exploration and presentation.

Specialized Plots and Visualization Techniques:

Creating specialized plots: heatmaps, treemaps, network graphs, etc. Visualizing geospatial data with maps and geographical information systems (GIS) using packages like ggmap and leaflet.

Data Storytelling and Dashboard Creation:

Principles of effective data storytelling through visualization. Building dashboards with R using Shiny or other dashboarding packages.

Visualization Best Practices:

Techniques for effectively communicating insights through visualizations. Considerations for color choices, accessibility, and design aesthetics.

Visualizing Time Series and Temporal Data:

Techniques for visualizing time-dependent data: time series plots, calendars, etc. Handling and displaying temporal data effectively.

Big Data Visualization

Visualization strategies for large datasets using packages like ggplot2 extensions, data.table, and others. Handling challenges related to visualizing big data.

Case Studies and Real-World Applications:

Analysing and replicating visualizations from real-world datasets. Creating comprehensive visualizations for various domains (finance, healthcare, social sciences, etc.).

Project Work and Presentations:

Implementing comprehensive data visualization projects using R. Presenting and communicating findings through visualizations.

Syllabus: Basic Python Programming

Unit 1 - Introduction to Python:

Installing python 3.x, Getting started with IDLE, Using Jupyter notebook, Python key features, Python syntax, Getting started with python programming, keywords, Identifiers, Variables, Comments, Operators, Expressions, Statement, Type conversion, IO (input () and print () functions).

Unit 2 - Python Data Types:

Data types in python, Number (Integer along with Boolean, Floating point and Complex), Sequence (String, List and Tuple), Introduction to string, Slicing, operations, Built-in functions. Introduction to list, Slicing, operations, Built-in functions, Nested List, List Comprehension. Introduction to Tuple, Slicing, operations, Built-in functions, Nested Tuple. Sets, None and Mapping (Dictionary). Introduction to Dictionaries, operations, traversal, Built-in functions, Dictionary Manipulation.

Unit 3 - Control Flow in Python:

Introduction, Selection or conditional statements (if, if..elif..else), indentation, Repetition or looping statements(for loop, while loop), Break and Continue Statements, Nested loops.

Unit 4 - Functions and Modules in Python:

Introduction to functions, Functions, Types of functions (In-built and User Defined), return statement, Types of arguments, Concepts of scope of variable (local & global variable), recursion, Lambda Functions, Python Standard Library. Modules, Built-in Modules (e.g. math, random, statistics), Importing a module, importing a function.

Unit 5 - Exception Handling and File Handling in Python:

Introduction to Exceptions, Syntax Errors and Runtime Errors, Exceptions, Built-in Exceptions, Raising Exceptions(raise), Handling Exceptions, need for exception handling, Handling Process, Catching exceptions, try…except…else Clause, Finally Clause Introduction to Files, types of Files, Opening and Closing a text file, Writing to and reading from a text file, Creating and Traversing a text File, CSV files.

Advanced Python Programming

Unit 1 - Regular Expressions in Python

Introduction, Concept of regular expression, Various types of regular expressions, Python re package, Matching vs Searching, re.search(), re.findall(), re.match() Methods, Substituting Strings with the sub() method, Character Classes, The Wildcard Character, Reg-ex symbols, Case-Insensitive Matching.

Unit 2 - Object Oriented Programming in Python

Overview of OOP, Class Definition, Creating Objects, Static and instance variables, Constructor concept, understanding self and __init__ (), Built-in Class Attributes, Inheritance, super keyword, Overloading, Method Overriding, Abstraction, Encapsulation, Data Hiding, Decorators.

Unit 3 - Database connectivity using Python

Overview of Database, installing MySQL-connector-python, Connecting to database, Creation of cursor instance, Creation of a database and tables, Insert and Update queries, commit(), Query execution, Extract data from result (fetchall(), fetchmany(), fetchone() methods), Close the connection.

Unit 4 - Python Numpy Library

Installation, Introduction to Numpy, Numpy array, Numpy array vs python List, Creation of Numpy arrays, Array attributes, indexing and slicing, Operations on Numpy arrays (Arithmetic and Statistical operations), Loading Numpy arrays from File, Saving to files

Unit 5 - Python Pandas Library

Installation, Introduction to Pandas, Series data structure in pandas, Creation of series, Attributes, Accessing the elements, Methods and Mathematical Operations on series, Data Frame data structure in Pandas, Creation of a data frame, operations on rows and columns, Accessing elements (Indexing, Slicing and Filtering), Merging/Joining Dataframes, Attributes, Importing and exporting data from and to CSV file. Pandas statistical functions, Aggregation and handling missing data.

Data Visualization

Unit 1 - Data Visualization with matplotlib

Introduction to Data Visualization, Matplotlib overview, installation, matplotlib objects, Plotting (pyplot module), Figure class, axes class, Multiplots, subplots, grids, plot customization, Bar plot, Histogram, Pie chart, scatter plot, Box plot, Violin plot, 3D plotting.

Unit 2 - Data Visualization with Pandas and Seaborn

The Pandas Plot function (Pandas Visualizations), plot for series and data frame, kind argument in plot() function, different types of plots.

Seaborn installation, Overview, seaborn datasets, Bar plot, pairplot(),Plotting statistical distributions, Categorical Scatter Plots(stripplot(), swarmplot()), Box plots, Violin plots, plot grids.

Syllabus :SQL for Data Science

A SQL syllabus for data science would typically cover a range of topics from fundamental SQL commands to more advanced techniques that are relevant for data science applications.

1. Introduction to SQL and Relational Databases

Overview of SQL and its role in data science , Understanding relational databases and their components , Introduction to popular database management systems (DBMS) like MSSQL, MySQL, PostgreSQL, or SQLite

2. Basic SQL Commands

SQL syntax and structure , SELECT statement and basic querying , Filtering and sorting data using WHERE and ORDER BY clauses , Working with DISTINCT, LIMIT, and OFFSET

3. Joins and Relationships

Understanding relationships in databases (one-to-one, one-to-many, many-to-many) , INNER JOIN, LEFT JOIN, RIGHT JOIN, and FULL JOIN , Handling NULL values in joins

4. Aggregations and Grouping

Aggregate functions: COUNT, SUM, AVG, MIN, MAX , GROUP BY clause for grouping data , HAVING clause for filtering grouped data

5. Sub queries and Nested Queries:

Introduction to sub queries and their use cases , Correlated sub queries , Sub queries in different parts of a SQL statement

6. Data Modification Commands

INSERT, UPDATE, DELETE statements ,Maintaining data integrity, Transactions and their importance

7. Indexing and Optimization:

Importance of indexing for performance , Creating and using indexes , Query optimization techniques

8. Views and Stored Procedures:

Creating and using views ,Introduction to stored procedures and functions , Benefits of using views and stored procedures

9. Working with Time and Dates

Handling date and time data types , Extracting information from date and time

Performing date arithmetic

10. Advanced Topics in SQL

Window functions (ROW_NUMBER, RANK, DENSE_RANK) , Common Table Expressions (CTEs) , Analytic functions and their applications

11. Case Studies and Real-world Applications

Applying SQL to real-world data science problems , Analysing datasets with complex queries

Solving data-related challenges using SQL

12. SQL and Python/R Integration

Basics of integrating SQL with Python or R , Using SQL in Jupyter notebooks or other data science environments