**Curriculum Plan (ODD SEM 2024): B.Sc. (H) Mathematics III Year (Semester V)**

**DSE-3(i): MATHEMATICAL DATA SCIENCE**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Dr. Tajender Kumar**Assistant ProfessorDepartment of MathematicsKalindi College (University of Delhi)Delhi- 110008Mobile: +91 7417837644**E- mail**: tajenderkumar@kalindi.du.ac.in  |  | **Marks Distribution**  | **Theory** |  90 Marks |
| **Practical** |  40 Marks  |
| **Internal Assessment** | Assignment 30 Marks |
|  |
|  |
| **Classes Assigned** | **Lectures** | 3 per week (Theory) |
| **Practical** | 2 per week |
| **References** |  | 1. Mertz, David. (2021). Cleaning Data for Effective Data Science, Packt Publishing.2. Ozdemir, Sinan. (2016). Principles of Data Science, Packt Publishing.3. Phillips, Jeff M. (2021). Mathematical Foundations for Data Analysis, Springer.(https://mathfordata.github.io/). |
|  | **Week** | **Topics** |  |
|  | **Beginning/1st week with 3 days** **(**01-03,05-10AUG) | Types of Data: nominal, ordinal, interval, and ratio; Steps involved in data science case-study: question, procurement, exploration, modeling, and presentation; |  |
|  | **2nd week (**12-17 AUG) | Structured and unstructured data: streams, frames, series, survey results, scale and source of data – fixed, variable, high velocity, exact and implied/inferred; Overview of problems with data – dirty and missing data in tabular formats – CSV, data frames in R/Pandas. [2]: Chapter 2, Chapter 3, and relevant material for different presentation styles from Chapter 9.[1]: Chapter 1 (up to page 28). |  |
|  | **3rd week (**19-24 AUG) | Anomaly detection, assessing data quality, rectification and creation methods, data hygiene,  |  |
|  | **4th week (**26-31 AUG) | Meta-data for inline data-description-markups such as XML and JSON; Overview of other data-source formats – SQL, pdf, Yaml, HDF5, and Vaex. [1]: Relevant material from Chapters 4, 5, and 6.[1]: Chapter 1 (pages 29- 44, and 58-60). |  |
|  | **5th week (**02-07 SEP) | Model driven data in Rn, Log-likelihoods and MLE, Chebyshev, and Chernoff-Hoeffding inequalities with examples, Importance sampling. [3]: Chapter 1 (pages 12-13), and Chapter 2 (Section 2.2, 2.3 [2.3.1 to 2.3.3], and 2.4). |  |
|  | **6th week (**09-14 SEP) | Norms in Vector Spaces– Euclidean, and metric choices; Types of distances: Manhattan, Hamming, Mahalanobis, Cosine and angular distances, KL divergence;  |  |
|  | **7th week (**16-21 SEP) | Distances applied to sets– Jaccard, and edit distances; Modeling text with distances.[3]: Chapter 3 (Section 3.3), and Chapter 4 (Sections 4.1 to 4.4).  |  |
|  | **8th week (**23-28 SEP) | Linear Regression: Simple, multiple explanatory variables, polynomial, |  |
|  | **9th week (**30 SEP-05 OCT) | Cross-validation, regularized, Lasso, and matching pursuit; Gradient descent.[3]: Chapter 5, and Chapter 6 (Sections 6.1 to 6.3). |  |
|  | **10th week**. (07-12 0CT) | Problem of dimensionality, Principal component analysis, Singular value decomposition (SVD), |  |
|  | **11th week (**14-19 0CT) | Best k-rank approximation of a matrix, Eigenvector and eigenvalues relation to SVD, Multidimensional scaling, Linear discriminant analysis.[3]: Chapter 7 (Sections 7.1 to 7.7).  |  |
|  | **12th week (**21-26 OCT) | Clustering: Voronoi diagrams, Delaunay triangulation, Gonzalez’s algorithm for k-center clustering, |  |
|  | **13th week (**04-09 NOV) | Lloyd’s algorithm for k-means clustering, Mixture of Gaussians, Hierarchical clustering, Density-based clustering and outliers, Mean shift clustering.[3]: Chapter 8. |  |
|  | **14th week (**11-16 NOV) | Classification: Linear classifiers, Perceptron algorithm, Kernels, |  |
|  | **15th week** (18-23 NOV) | Support vector machines, and k-nearest neighbors (k-NN) classifiers.[3]: Chapter 9 (Sections 9.1 to 9.5). |  |
|  | **16th week only with 2 Days** (25-27 NOV) | Revision |  |
| Dispersal of classes, preparation leave and practical examination begin- 28 November, 2024. |