Code No.: **MDS 504**

Course Title:**Mathematics for Data Science** Full Mark: 75

Nature: **Theory** (Compulsory) Credit: 3

***Course Description*:**

The course will cover basic topics in linear algebra to understand high-dimensional vector spaces, matrices and graphs as popular mathematical structures with which to model data (e.g., as models for term-document corpora, high-dimensional regression problems, ranking/classification of web data, adjacency properties of social network data, etc.); and geometric approaches to eigendecompositions, least-squares, principal components analysis, etc. The course requires to solve problems using programming R.

***Learning Objectives*:**

After successful completion of this course the student will be able to

- Understand basic linear algebra techniques which are useful in data science.
- Explain \why different methods do and don't work.
- Understand tools that are used to diagnose problems, to develop new methods, etc.
- Understand how some discrete probability and optimization are used with matrices and graphs, two very common ways to model data.
- Understand the connections between the discrete probability ideas and very related linear algebra ideas.
- Acquire a basic understanding and intuition of why various methods work, so that the student can use them in practical applications data science.
- Use programming R to solve problems of this course.

***Course Contents*:**

Unit 1: Introduction, Motivation, and Overview **[9 Hrs.]**

Linear algebra and machine learning,
Representing data as flat tables versus matrices and graphs;
Different ways probability/randomness/noise interacts with data;
Probability and matrices/graphs in data science versus other areas;
Quantification of the inference step.

Unit 2: Introduction to Matrices and Vectors **[15Hrs.]**

Vectors, Basic properties of R^n;
Norms and balls;
Vector addition and scalar multiplication.
Vector spaces and subspaces;
Matrices,
Operations on matrices, including matrix multiplication;
Functions, linear functions, and linear transformations; Matrices as transformations.
Dot products, angles, and perpendicularity;
Linear combinations, span, and linear independence; Bases, orthonormal bases, and projections.
Applications in the theory of probabilityand data science

Unit 3: Spectral Theorems **[14Hrs.]**

Eigenvectors and Eigenvalues:
Quadratic forms and matrices
Symmetric bi-linear functions;
Connections with conic sections;
Definiteness, indefiniteness, and quadratic forms as a sum/difference of squares;
EigenValue Decomposition (EVD)
Singular Value Decomposition (SVD)
Properties of the SVD
Orthogonal subspaces;
Uses of the spectral decomposition
Applications in data science

Unit 4: System of Linear Equations **[10 Hrs.]**

Solving system of linear equations:
Geometry of linear equations;
Gaussian elimination;
Row exchanges;
Networks and incidence matrices;
The four fundamental subspaces.
Basis transformations;
Orthogonal bases;
Gram-Schmidt Orthogonalization;
Numerical issues.
Applications in data science

References:

1. Nick Fieller (2015).*Basics of Matrix Algebra for Statistics with R*, Chapman and Hall/CRC.
2. Shayle R. Searle& André I. Khuri (2017).*Matrix Algebra Useful for Statistics*, John Wiley & Sons, Inc..
3. Michael W. Mahoney (2018).*Linear Algebra for Data*, University of California Berkeley.
4. Deisenroth, M. P. ,Faisal, A. A. and Ong, C. S. (2019). *Mathematics for Machine Learning*, Cambridge University Press.
5. Jason Brownlee (2018).*Basics of Linear Algebra for Machine Learning*, https://www.mobt3ath.com/uplode/book/book-33342.pdf .

***