Ph. D in

COMPUTER SCIENCE
&
ENGINEERING

Department of Computer Science & Engineering

Dibrugarh University Institute of Engineering andTechnology,
Dibrugarh University
Dibrugarh,

Assam-786004

COURSE I

Course Contents

Course Code	Course Name	L-T-P-Credits
Course-I	**Research Methodology**	**3-1-0-4**

Unit I: Research Methodology –an introduction

Meaning and Objectives of Research, Motivation in Research, Types of Research, Significance of Research, Research Methods verses Methodology, Criteria of Good Research, Problems Encountered by Researchers in India, Research Problem, Features of Good Research Design.

Unit II: Methods of Data Collection

Primary Data, Observation Method, Interview Method, Collection of Data through Questionnaires, Collection of Data through Schedules, Difference between Questionnaires and Schedules, Other Methods of Data Collection, Collection of Secondary Data, Selection of Appropriate Method for Data Collection, Review of previous work and literature.

Unit III: Interpretation and Report Writing

Meaning and Technique of Interpretation, Precaution in Interpretation, Different Steps in Writing Report, Types of Reports, Oral Presentation, Precaution for Writing Research Reports, Reference & Bibliography, Citation index of Publication.

Unit IV: Role of Computer in Research

Computer Application
Software Application

COURSE II

Course-II

Course Code	Course Name	L-T-P-Credits
PhD-CSE-202	**GRAPH THEORY**	**3-1-0-4**

Course Contents

INTRODUCTION:

This course describes graph theory terminologies and problems, and the use of algorithms in both mathematical theory of graphs and their applications. The course will explain in detail the basic theory of various types of graphs. This course introduces some of the algorithms that solve all or part of the problems in graph theory.

COURSE OBJECTIVE:

1. To understand and apply the basic concepts of graph theory.
2. Application of graph theory-based tools to solve practical problems.

LEARNING OUTCOME:

1. Students can apply the concepts and principles of graph theory in practical situations.

DETAILED SYLLABUS:

Unit I: Graph
Marks: 20

Graph : Incidence and degree; Handshaking Lemma; Isomorphism; Subgraphs and Union of graphs; Connectedness; Walks, Paths and Circuits; Components and Connectedness; Walks, Walks, Paths and Circuits; Components and Connectedness algorithms; Shortest Path Algorithms, Eulerian graph, Fleury's algorithm and Chinese postman problem; Hamiltonian graph - necessary and sufficient conditions; Traveling salesman; Bipartite graph. Directed graphs : Binary relations; Directed graphs and connectedness; directed trees; Aborecence; Polish method; Tournaments. Counting of labeled trees : Cayley's theorem; Counting methods; Polya theory.

Unit II: Tree
Marks: 20

Tree : Properties of trees; Pedant vertices in a tree; Center of a tree; Rooted binary trees; Spanning trees - Spanning tree algorithms; Fundamental circuits; Spanning trees of a weighted graph; cut-sets and cut-vertices; Fundamental cut-sets; Connectivity and separativity; network flow; max-flow min-cut theorem.

Unit III: Planner Graph

Marks: 10

Planner Graph: Combinatorial and geometric dual; Kuratowski's graph; detection of planarity; Thickness and crossings.

Unit IV: Matrix
Marks: 10
Matrix representations of graph: Incidence; Adjacency; matrices and their properties. Colourings: Chromatic number : Chromatic polynomial; The six and five colour theorems; The four colour problem.

REFERENCE BOOKS:

1. Deo, N.: Graph Theory with Applications to Engineering and Computer Science.
2. Harary : Graph Theory, PHI (EEE)

Course Code	Course Name	L-T-P-Credits
PhD-CSE-201	**DATA PROCESSING AND VISUALIZATION**	**3-1-0-4**

Objective: The objective of Data Science is to enable society, companies and citizens to understand and use the ever-increasing amount of collected data in ways that make it possible to detect potential problems or improvements to the current state of affairs. Data Science should also empower humans to estimate and understand the potential result of different actions.

After having completed the course the student should be able to:

1. Understand what is meant by Data Science as a concept: where and when Data Science is needed, what types of problems Data Science can solve and what main methodologies and tools of Data Science are.
2. Understand the meaning of various data-based measurements and visualizations commonly used in society, and know how to read and interpret them.

Module 1:

- Introduction to "Python" programming language and tools
- Import and export of data from text files, data bases and other sources
- Data visualization in R, in 2D and 3D
- Map visualizations

Module 2:

- Displaying and working with images in R
- Introduction to other useful data preprocessing and visualization packages
- Linear regression, BLUE, RMSE, shrinkage methods (Lasso, ridge regression)

Module 3

- Linear classification (logistic regression, LDA)
- Principal Components Analysis (PCA) for identifying linear correlations between variables
- Robust PCA and low-rank matrix completion for outliers and missing data,

Module 4:

- K-means clustering
- Nonlinear or nonparametric methods (k-NN, kernel methods, etc.)
- Preparation of data for machine learning, introduction to "caret" machine learning package
- Basic notions of Explainable Artificial Intelligence (XAI)

Books

1. Data Mining – Concepts and Techniques – Jiawei Han & Micheline Kamber, 3rd Edition Elsevier.
2. Data Mining Introductory and Advanced topics – Margaret H Dunham, PEA.

Course Code	Course Name	L-T-P-Credits
PhD-CSE-203	**MATHEMATICS FOR MACHINE LEARNING**	**3-1-0-4**

Course Objective:

CO1: At the end of this course student will have an intuitive understanding of vectors and matrices that will help you bridge the gap into linear algebra problems, and how to apply these concepts to machine learning.

CO2: At the end of this course student will have an intuitive understanding "rise over run" formulation of a slope, the formal definition of the gradient of a function,

CO3: At the end of this course student will have an intuitive understanding of how to calculate vectors that point up hill on multidimensional surfaces.

CO4: At the end of this course student will have an intuitive understanding of how they can use calculus to build approximations to functions, as well as quantifying how accurate those approximations are.

Module 1: Linear Algebra for Machine Learning: (12 Hours)

1.1 **Introduction: (3 Hours)** Solving data science challenges with mathematics, Motivations for linear algebra, getting a handle on vectors, Operations with vectors, Modulus & inner product, Cosine & dot product, Projection, Changing Basis, Basis, vector space, and linear independence, Application of changing basis.

1.2. **Objects that operate on Vectors: (3 Hours**) Matrices, vectors, and solving simultaneous equation problems, How matrices transform space, Types of matrix transformation, Composition or combination of matrix transformations, Solving the apples and bananas problem: Gaussian elimination, Going from Gaussian elimination to finding the inverse matrix, Determinants and inverses.

1.3 **Linear Mapping: (3 Hours)** Introduction: Einstein summation convention and the symmetry of the dot product, Matrices changing basis, Doing a transformation in a changed basis, Orthogonal matrices, The Gram–Schmidt process.

1.4 **Eigenvalues and Eigenvectors: Application to Data Problems (3 Hours)** What are eigenvalues and eigenvectors?, Special eigen-cases, Calculating eigenvectors, Changing to the eigenbasis, Eigenbasis example, Introduction to PageRank

Module 2: Multivariate Calculus (10 Hours)

2.1 **What is calculus (1 Hours)** Introduction to Multivariate Calculus, Functions, Definition of a derivative, Differentiation examples & special cases, Product rule, Chain rule

2.2 **Multivariate calculus (2 Hours)** Variables, constants & context, differentiate with respect to anything, The Jacobian, Jacobian applied, The Sandpit, The Hessian.

2.3 **Multivariate chain rule and its applications (2 Hours)** Multivariate chain rule, More multivariate chain rule, Simple neural networks, More simple neural networks.

2.5 **Taylor series and linearization (2 Hours)** Building approximate functions, Power series, Power series derivation, Power series details, Examples, Linearization, Multivariate Taylor

2.6. **Introduction to Optimization: (2 Hours)** Gradient Descent, Constrained optimization, Newton Rapson in One dimension, Langrage multiplier

2.7 **Regression (2 Hours)**: Simple linear regression, General nonlinear least squares, Doing least squares regression analysis in practice

Books:

1. *Mathematical Methods in the Physical Sciences* by Mary L Boas, John Wiley and Sons, 3rd Ed, 2006. Linear Algebra is in Chapter 3. Most Engineering maths textbooks are also useful, although they often teach 'vectors and matrices' rather than viewing linear algebra as an integrated whole.

2. *Introduction to Linear Algebra* by Gilbert Strang, Wellesley-Cambridge Press, 5th Ed, 2016. http://math.mit.edu/~gs/linearalgebra/

Web Resources:

1. https://www.khanacademy.org/math/linear-algebra
2. https://www.khanacademy.org/math/differential-calculus
3. http://www.3blue1brown.com
4. https://en.wikipedia.org/wiki/Linear_algebra

COURSE III

COURSE III

Course Code	Course Name	L-T-P-Credits
PhD-CSE-301	**NATURAL LANGUAGE PROCESSING**	**3-1-0-4**

Course Contents

INTRODUCTION:

Natural Language Processing (NLP) is a branch of artificial intelligence that uses natural language to handle interactions between humans and computers. The ultimate goal of NLP is to read, understand, and decode human language in a valuable way. Most NLP technologies rely on machine learning to extract meaning from human language.

COURSE OBJECTIVE:

3. To learn the basics of Natural Language Processing.
4. To apply the techniques of Natural Language Processing.
5. To understand the role of sentence semantics and pragmatics.

LEARNING OUTCOME:

2. Students will be able to understand Natural Language Processing.
3. Analyze semantics and discourse differentiation from an NLP perspective.
4. Application of a probabilistic model of defining language and techniques.

DETAILED SYLLABUS:

Unit I: Words
Marks: 20

Words – Morphology and Finite State transducers – Computational Phonology and Pronunciation Modelling – Probabilistic models of pronunciation and spelling – Ngram Models of syntax – Hidden markov models and Speech recognition – Word classes and Part of Speech Tagging.

Unit II: Context Free Grammars
Marks: 20

Context free Grammars for English – Parsing with Context free Grammar – Features and unification – Lexicalized and Probabilistic Parsing -Language and Complexity. Semantics: Representing meaning – Semantic analysis – Lexical semantics – Word sense disambiguation and Information retrieval.

Unit III: Pragmatics
Marks: 20

Pragmatics: Discourse – Dialog and Conversational agents – Natural language generation, Statistical alignment and Machine translation: Text alignment – word alignment – statistical machine translation.

REFERENCE BOOKS:

1. Daniel and Martin J. H., "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition", Prentice Hall, 2009.

2. Manning C. D. and Schutze H., "Foundations of Statistical Natural Language processing", First Edition, MIT Press, 1999 Allen J., "Natural Language Understanding", Second Edition, Pearson Education, 2003.

Course Code	Course Name	L-T-P-Credits
PhD-CSE-302	**ADVANCED COMPUTER NETWORK**	**3-1-0-4**

Course Objective:

The course introduces main concepts of networking; application areas; classification; reference models; transmission environment; technologies; routing algorithms; IP, UDP and TCP protocols; reliable data transferring methods; application protocols; network security; management systems; perspectives of communication networks.

Course Objectives: As a result of successfully completing this course, students will:

1. Become familiar with layered communication architectures (OSI and TCP/IP).
2. Understand the client/server model and key application layer protocols.
3. Learn sockets programming and how to implement client/server programs.
4. Understand the concepts of reliable data transfer and how TCP implements these concepts.
5. Know the principles of congestion control and trade-offs in fairness and efficiency.

Module 1:

Data communication Components: Representation of data and its flow Networks, Various Connection Topology, Protocols and Standards, OSI model, Transmission Media, LAN: Wired LAN, Wireless LANs, Connecting LAN and Virtual LAN, Techniques for Bandwidth utilization:

Module 2:

Data Link Layer and Medium Access Sub Layer: Error Detection and Error Correction - Fundamentals, Block coding, Hamming Distance, CRC;

Module 3:

Flow Control and Error control protocols - Stop and Wait, Go back – N ARQ, Selective Repeat ARQ, Sliding

Window, Piggybacking, Random Access, Multiple access protocols -Pure ALOHA, Slotted ALOHA, CSMA/CD,CDMA/CA

Module 4:

Network Layer: Switching, Logical addressing – IPV4, IPV6; Address mapping – ARP, RARP, BOOTP and DHCP–Delivery, Forwarding and Unicast Routing protocols.

Books:

1. Data Communication and Networking, 4th Edition, Behrouz A. Forouzan, McGraw- Hill.
2. Data and Computer Communication, 8th Edition, William Stallings, Pearson Prentice Hall India. Suggested reference books
3. Computer Networks, 8th Edition, Andrew S. Tanenbaum, Pearson New International Edition.

Course Code	Course Name	L-T-P-Credits
PhD-CSE-303	**GRAPH REPRESENTATION LEARNING**	**3-1-0-4**

Course Outcome:

1. At the end of the course students should be able to distinguish between static graph analysis and feature base graph analysis.
2. At the end of the course students should be able to create graph data using pytorch geometric library, networkX etc.
3. At the end of the course students should be able to design an efficient graph representation learning model.

Module 1: Graph Terminology and Representation (8 Hours)

Graph definition, Storing Graph Information, Graph Degree, and Laplacian of Graph, Definition of learning in Graph Representation Learning, Drawback of existing graph learning models, Practice using Tensor, and Torch Geometric for defining a Graph.

Module 2: From Convolutional Neural Network to Graph Neural Network (6 Hours)

Review of Convolution operation, Graph Convolution, Message Passing Framework

Module 3: Introduction to Different Graph Embedding Methods (10 Hours)

Graph Embedding Problem statement, DeepWalk Algorithm, Practice with RandomWalk using karetclub library, Node2Vec Algorithm, Practice Node2Vec using Karateclub, Pytorch Geometric, GNN Motivation, Simplifying Graph Convolution Network. Practice for Graph Convolution Network using Pytorch Geometric, Graph Attention Network.

Module 4: Induction and Transudative Graph Embedding (10 Hours)

Review of Popular GNN Embedding Methods, Transudative and Inductive Embedding Methods, GraphSAGE

References:

1. Lingfei Wu, Peng Cui, Jian Pei, Liang Zhao (2022). Graph Neural Networks: Foundations, Frontiers, and Applications , Springer Singapore.
2. Ma, Y., & Tang, J. (2021). Deep Learning on Graphs. Cambridge: Cambridge University Press. doi:10.1017/9781108924184

Web Resource:

1. https://antoniolonga.github.io/Pytorch_geometric_tutorials/index.html

COURSE IV

Course Code	**Course Name**	**L-T-P-Credits**
Course-IV	**Technical Writing**	**2-0-0-2**