

Rajiv Gandhi Proudyogiki Vishwavidyalaya Bhopal

M.Tech Computer Science and Engineering (Data Science)

Second Semester Syllabus

MTCD 201: STATISTICAL FOUNDATIONS FOR DATA SCIENCE

UNIT - I Greatest Common Divisors and Prime Factorization: Greatest common divisors, The Euclidean algorithm, The fundamental theorem of arithmetic, Factorization of integers and the Fermat numbers Congruence's: Introduction to congruence's, Linear congruences, The Chinese remainder theorem, Systems of linear congruences

UNIT - II Simple Linear Regression and Correlation: Introduction to Linear Regression, The Simple Linear Regression Model, Least Squares and the Fitted Model, Properties of the Least Squares Estimators, Inferences Concerning the Regression Coefficients, Prediction, Simple Linear Regression Case Study Random Variables and Probability Distributions: Concept of a Random Variable, Discrete Probability Distributions, Continuous Probability Distributions, Statistical Independence. Discrete Probability Distributions: Binomial Distribution, Poisson distribution.

UNIT - III Continuous Probability Distributions: Normal Distribution, Areas under the Normal Curve, Applications of the Normal Distribution, Normal Approximation to the Binomial, Fundamental Sampling Distributions: Random Sampling, Sampling Distributions, Sampling, Distribution of Means and the Central Limit Theorem, Sampling Distribution of S^2 , t -Distribution, F Distribution.

UNIT - IV Estimation & Tests of Hypotheses: Introduction, Statistical Inference, Classical Methods of Estimation. Estimating the Mean, Standard Error of a Point Estimate, Prediction Intervals, Tolerance Limits, Estimating the Variance, Estimating a Proportion for single mean, Difference between Two Means, between Two Proportions for Two Samples and Maximum Likelihood Estimation.

UNIT - V Stochastic Processes and Markov Chains: Introduction to Stochastic processes-Markov process. Transition Probability, Transition Probability Matrix, First order and Higher order Markov process, n step transition probabilities, Markov chain, Steady state condition, Markov analysis.

TEXT BOOKS:

1. Kenneth H. Rosen, Elementary number theory & its applications, sixth edition, Addison Wesley, ISBN 978 0-321-50031-1 R22 M.Tech Data Science JNTUH
2. Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keying Ye, Probability & Statistics for Engineers & Scientists, 9th Ed. Pearson Publishers.
3. S. D. Sharma, Operations Research, Kedarnath and Ramnath Publishers, Meerut, Delhi

REFERENCE BOOKS:

1. S C Gupta and V K Kapoor, Fundamentals of Mathematical statistics, Khanna publications
2. T.T. Soong, Fundamentals of Probability And Statistics For Engineers, John Wiley & Sons Ltd, 2004.
3. Sheldon M Ross, Probability and statistics for Engineers and scientists, Academic Press.

MTCD 202: DEEP LEARNING

Unit-1 History of Deep Learning, Deep Learning Success Stories, review of Neuron model, activation functions, Perceptron Learning, Multilayer Perceptrons (MLPs), Feed forward Neural Networks, Back propagation, weight initialization methods, Batch Normalization, Representation Learning, GPU implementation, Decomposition – PCA and SVD.

Unit-2 Deep Feedforward Neural Networks, Gradient Descent (GD), Momentum Based GD, Nesterov Accelerated GD, Stochastic GD, AdaGrad, Adam, RMSProp, Autoencoder, Regularization in auto-encoders, Denoising auto-encoders, Sparse autoencoders, Contractive auto-encoders, Variational auto-encoder, Auto-encoders relationship with PCA, Dataset augmentation.

Unit-3 Introduction to Convolutional neural Networks (CNN) and its architectures, CNN terminologies: ReLu activation function, Stride, padding, pooling, convolutions operations, Convolutional kernels, types of layers: Convolutional, pooling, fully connected, Visualizing CNN, CNN examples: LeNet, AlexNet, ZF-Net, VGGNet, GoogLeNet, ResNet, RCNN etc. Deep Dream, Deep Art. Regularization: Dropout, drop Connect, unit pruning, stochastic pooling, artificial data, injecting noise in input, early stopping, Limit Number of parameters, Weight decay etc.

Unit-4 Introduction to Deep Recurrent Neural Networks and its architectures, Backpropagation Through Time (BPTT), Vanishing and Exploding Gradients, Truncated BPTT, Gated Recurrent Units (GRUs), Long Short Term Memory (LSTM), Solving the vanishing gradient problem with LSTMs, Encoding and decoding in RNN network, Attention Mechanism, Attention over images, Hierarchical Attention, Directed Graphical Models.

Unit-5 Introduction to Deep Generative Models, Restricted Boltzmann Machines (RBMs), Gibbs Sampling for training RBMs, Deep belief networks, Markov Networks, Markov Chains, Autoregressive Models: NADE, MADE, PixelRNN, Generative Adversarial Networks (GANs), Applications of Deep Learning in Object detection, speech/ image recognition, video analysis, NLP, medical science etc.

Recommended Books:

1. Ian Goodfellow, Yoshua Bengio and Aaron Courville; Deep Learning, MIT Press, 2017.
2. Chris Bishop; Pattern Recognition and Machine Learning, Springer publication, 2006
3. Aurelien Geon, "Hands-On Machine Learning with Scikit-Learn and Tensorflow: Concepts, Tools, and Techniques to Build Intelligent Systems", First Edition, O'Reilly publication, 2017.
4. Francois Chollet, "Deep Learning with Python", First Edition, Manning Publications, 2018.
5. Andreas Muller, "Introduction to Machine Learning with Python: A Guide for Data Scientists", First edition, O'Reilly Edition, 2016.

MTCD 203: COMPUTATIONAL INTELLIGENCE

Pre-Requisite: Computer Programming

Course Outcomes: After completing the course student should be able to:

1. Describe in-depth about theories, methods, and algorithms in computation Intelligence.
2. Compare and contrast traditional algorithms with nature inspired algorithms.
3. Examine the nature of a problem at hand and determine whether a computation intelligent technique/algorithm can solve it efficiently enough.
4. Design and implement Computation Intelligence algorithms and approaches for solving real-life problems.

Course Contents:

UNIT 1 Introduction to Computational Intelligence (CI): Basics of CI, History of CI, Adaptation, Learning, Self Organization and Evolution, CI and Soft Computing, CI Techniques; Applications of CI; Decision Trees: Introduction, Training Decision Trees, Evaluation, Splitting Criteria, Decision Tree Induction Algorithms.

UNIT 2 Evolutionary Computation: Genetic Algorithms: Basic Genetics, Concepts, Working Principle, Creation of Offsprings, Encoding, Fitness Function, Selection Functions, Genetic Operators-Reproduction, Crossover, Mutation; Genetic Modeling, Benefits; Problem Solving; Introduction to Genetic Programming, Evolutionary Programming, and Evolutionary Strategies.

UNIT 3 Fuzzy System: Fuzzy Sets: Formal Definitions, Membership Functions, Fuzzy Operators, Fuzzy Set Characteristics, Fuzzy Relations and Composition, Fuzziness and Probability; Fuzzy Logic and Reasoning: Fuzzy Logic, Fuzzy Rules and Inferencing; Fuzzy Controllers: Components of Fuzzy Controllers, Types, Defuzzification.

UNIT 4 Rough Set Theory: Introduction, Fundamental Concepts, Knowledge Representation, Set Approximations and Accuracy, Vagueness and Uncertainty in Rough Sets, Rough Membership Function, Attributes Dependency and Reduction, Application Domain, Hidden Markov Model (HMM), Graphical Models, Variable Elimination, Belief Propagation, Markov Decision Processes.

UNIT 5 Swarm Intelligence: Introduction to Swarm Intelligence, Swarm Intelligence Techniques: Ant Colony Optimization(ACO): Overview, ACO Algorithm; Particle Swarm Optimization(PSO): Basics, Social Network Structures, PSO Parameters and Algorithm; Application Domain of ACO and PSO; Bee Colony Optimization etc.; Hybrid CI Techniques and applications; CI Tools.

Recommended Books:

1. Russell C. Eberhart and Yuhui Shi, Computational Intelligence: Concepts to Implementations, Morgan Kaufmann Publishers, 2007.
2. Andries P. Engelbrecht, Computational Intelligence: An Introduction, Wiley Publishing, 2007.
3. David E. Goldberg, Genetic Algorithm in Search Optimization and Machine Learning, Pearson Education, 2009.

4. Jagdish Chand Bansal, Pramod Kumar Singh, Nikhil R. Pal, Evolutionary and Swarm Intelligence Algorithms, Springer Publishing, 2019.
5. S. Rajeskar, G.A. VijayalakshmiPai, "Neural Networks, Fuzzy Logic, Genetic Algorithms Synthesis and Applications", PHI, 2003.

Research Journals:

1. IEEE Transactions on Evolutionary Computation
2. IEEE Transactions on Systems, Man and Cybernetics
3. IEEE Transaction on Neural Networks and Learning Systems
4. IEEE Transaction on Fuzzy Systems
5. IEEE Transactions on Pattern Analysis and Machine Intelligence
6. ACM Transactions on Intelligent Systems and Technology
7. ACM Genetic and Evolutionary Computation Conference (GECCO)
8. ACM Journal of Machine Learning Research

MTCD 204: RECOMMENDER SYSTEMS

UNIT-1 Introduction and basic taxonomy of recommender systems (RSs), Traditional and non-personalized RSs. Introduction of Information Retrieval, Retrieval Models, Search and Filtering Techniques: Relevance Feedback, User Profiles, Recommender system functions, Matrix operations, covariance matrices, Understanding ratings, Issues with recommender system.

UNIT-2 Content-based Filtering: High level architecture of content-based systems, Advantages and drawbacks of content based filtering, Item profiles, Discovering features of documents, preprocessing and feature extraction, Obtaining item features from tags, Methods for learning user profiles, Similarity based retrieval, Classification algorithms.

UNIT-3 Collaborative Filtering (CF): Mathematical foundations Mathematical optimization in CF RSs. Baseline predictor through least squares. Regularization, over fitting. User-based recommendation, Item-based recommendation, Model based approaches, Matrix factorization. Recommender systems in personalized web search, knowledge-based recommender system, Social tagging recommender systems, Trust-centric recommendations, Group recommender systems.

UNIT-4 Hybrid Approaches: opportunities for hybridization, Monolithic hybridization design: Feature combination, Feature augmentation, Parallelized hybridization design: Weighted, Switching, Mixed, Pipelined hybridization design: Cascade, Meta-level, Limitations of hybridization strategies. Evaluation of Recommender Systems: Performance Evaluation of RSs Experimental settings. Working with RSs data sets. Evaluation on historical datasets, Offline evaluations. Evaluation metrics: Rating prediction and accuracy. Other metrics (fairness, coverage, diversity, novelty, serendipity).

UNIT-5 General properties of evaluation research, User behavior understanding in RSs Foundations of behavioral science. User choice and decisions models. Digital nudging and user choice engineering principles. Applications of RSs for content media, social media and communities Music and video RSs. Datasets. Group recommender systems. Social recommendations. Recommending friends: link prediction models. Similarities and differences of RSs with task assignment in mobile crowd sensing. Social network diffusion awareness in RSs.

Books and References:

1. Jannach D., Zanker M. and FelFering A., Recommender Systems: An Introduction, Cambridge University Press (2011), 1st ed.
2. Charu C. Aggarwal, Recommender Systems: The Textbook, Springer (2016), 1st ed.
3. Ricci F., Rokach L., Shapira D., Kantor B.P., Recommender Systems Handbook, Springer(2011), 1st ed.
4. Manouselis N., Drachsler H., Verbert K., Duval E., Recommender Systems For Learning, Springer (2013), 1st ed.
5. J. Leskovec, A. Rajaraman and J. Ullman, Mining of massive datasets, 2nd Ed., Cambridge, 2012. (Chapter 9).
6. M. Chiang, Networking Life, Cambridge, 2010. (Chapter 4).

MTCD 205 (A) EXPERT SYSTEM

Course Objectives: After completing the course, student should be able to:

- Understand the basics of Expert Systems and its terminology.
- Learn about programming languages and tools for expert systems
- Apply the different knowledge representation techniques as per requirements.
- Design and implement real world problems using expert system.

Course Contents:

UNIT 1 : Basics of Expert Systems : Introduction to expert system, need and history of expert systems, components of expert system, application of expert systems, types of expert system, advantages and limitation of an expert system, Organization, Characteristics, Prospector and Features of an expert system.

UNIT 2: Expert System Tools: Knowledge Representation and reasoning in expert systems, Expert systems tools, Programming languages for expert systems, Expert system shells, System building aids, Support facilities, Stages in development of expert system tools, knowledge engineering languages and tools.

UNIT 3: Building of an Expert Systems: Tasks in building expert systems, general stages in the development of an expert system, types of errors in the development stages, Choosing a tool for building expert systems, Acquiring the knowledge from the experts, Expert system development life cycle, Handling of uncertainties. Truth Maintenance Systems.

UNIT 4: Difficulties in Development of Expert Systems: Difficulties such as lack of resources, inherent limitations of expert systems, Common pitfalls in planning of expert systems, Pitfalls in choosing the domain of expert systems.

UNIT 5: Expert Systems in Market Place: Where is expert systems work being done, High performance expert systems used in research, business, XCON, Case Studies.

Books and References:

1. Stuart Russel and Peter Norvig, 'Artificial Intelligence A Modern Approach', Second Edition, Pearson Education, 2003 / PHI.
2. Donald A. Waterman, 'A Guide to Expert Systems', Pearson Education.
3. Foundation Artificial Intelligence & Expert Systems by VS Janakiraman K, Sarukesi P Gopalakrishnan Macmillan series in computer science.
4. Janakiraman, K. Sarukesi, 'Foundations of Artificial Intelligence and Expert Systems', Macmillan Series in Computer Science.
5. W. Patterson, 'Introduction to Artificial Intelligence and Expert Systems', Prentice Hall of India, 2003.

MTCD 205(B): RESEARCH METHODOLOGY AND IPR

UNIT-1 Meaning of Research, Types of Research, Research Process, Problem definition, Objectives of Research, Research Questions, Research design, Approaches to Research, Quantitative vs. Qualitative Approach, Understanding Theory, Building and Validating Theoretical Models, Exploratory vs. Confirmatory Research, Experimental vs Theoretical Research, Importance of reasoning in research.

UNIT-2 Problem Formulation, Understanding Modeling & Simulation, Conducting Literature Review, Referencing, Information Sources, Information Retrieval, Role of libraries in Information Retrieval, Tools for identifying literatures, Indexing and abstracting services, Citation indexes. Literature Review: Effective literature review approaches, literature analysis, avoiding plagiarism, ethics in research, data collection, analysis, interpretation.

UNIT-3 Experimental Research: Cause effect relationship, Development of Hypothesis, Measurement Systems Analysis, Error Propagation, Validity of experiments, Statistical Design of Experiments, Field Experiments, Data/Variable Types & Classification, Data collection, Numerical and Graphical Data Analysis: Sampling, Observation, Surveys, Inferential Statistics, and Interpretation of Results.

UNIT-4 IPR: Introduction and significance of intellectual property rights, types of Intellectual Property Rights, copyright and its significance, introduction to patents and its filing, introduction to patent drafting, best practices in national and international patent filing, copyrightable work examples. Patents-copyrights-Trademarks-Industrial design geographical indication. Ethics of ResearchScientific Misconduct- Forms of Scientific Misconduct. Plagiarism, Unscientific practices in thesis work, Ethics in science.

UNIT-5 Patents and its basics, patentable items, designs, process of filing patent at national and international level, process of patenting and development, technological research and patents, innovation, patent and copyright international intellectual property, procedure for grants of patents, need of specifications, types of patent applications, provisional and complete specification, patent specifications and its contents, trade and copyright.

Books and References:

1. Bordens, K. S. and Abbott, B. B., "Research Design and Methods – A Process Approach", 8th Edition, McGraw-Hill, 2011
2. C. R. Kothari, "Research Methodology – Methods and Techniques", 2nd Edition, New Age International Publishers
3. Michael P. Marder, "Research Methods for Science", Cambridge University Press, 2011
4. T. Ramappa, "Intellectual Property Rights Under WTO", S. Chand, 2008
5. Robert P. Merges, Peter S. Menell, Mark A. Lemley, "Intellectual Property in New Technological Age". Aspen Law & Business;
- 6 edition July 2012 6. Stuart Melville, Wayne Goddard, Research Methodology: An Introduction for Science and Engineering Students, Juta& Co Ltd.
7. Ranjit Kumar, Research Methodology: A Step by Step Guide for Beginners, Pearson.
8. T. Ramappa, Intellectual Property Rights under WTO: Tasks before India, S. Chand.

MTCD 205 (C): NATURAL LANGUAGE PROCESSING THEORY

UNIT-1 Introduction to NLP: Different Data Models such as Boolean Model, Vector model, Probabilistic Model, comparison of classical models. Introduction to alternative algebraic models such as Latent Semantic Indexing etc.

UNIT-2 Probabilistic language modeling and its applications. The role of language models. Simple Ngram models. Estimating parameters and smoothing. Evaluating language models. Markov models. Estimating the probability of a word, and smoothing. Generative models of language.

UNIT-3 Part of Speech Tagging and Sequence Labeling: Lexical syntax. Hidden Markov Models Forward and Viterbi algorithms and EM training.

UNIT-4 Syntactic parsing: Grammar formalisms and treebanks. Efficient parsing for context-free grammars (CFGs). Statistical parsing and probabilistic CFGs (PCFGs).

UNIT-5 Semantic Analysis: Lexical semantics and word-sense disambiguation. Compositional semantics. Semantic Role Labeling and Semantic Parsing.

BOOKS RECOMMENDED:

1. Daniel Jurafsky & James H. Martin, Speech and Language Processing, Pearson publication, 2018.
2. Manning and Schütze "Foundations of Statistical Natural Language Processing", MIT Press, 2009

REFERENCE BOOKS:

1. Dipanjan Sarkar, Text Analytics with Python (Apress/Springer, 2016)
2. Handbook of Natural Language Processing, Second Edition—Nitin Indurkha, Fred J. Damerau, Fred J. Damerau (ISBN 13: 978-1420085921)
3. Natural Language Processing with Python by Steven Bird, Ewan Klein, Edward Loper (ISBN 13: 978-0596516499)

MTCD 205 (D) BUSINESS STATISTICS & ANALYTICS

Unit I :

Descriptive Statistics Meaning, Scope, types, functions and limitations of statistics, Measures of Central tendency – Mean, Median, Mode, Quartiles, Measures of Dispersion – Range, Inter quartile range, Mean deviation, Standard deviation, Variance, Coefficient of Variation, Skewness and Kurtosis.

Unit II:

Time Series & Index Number Time series analysis: Concept, Additive and Multiplicative models, Components of time series, Trend analysis: Least Square method - Linear and Non-Linear equations, Applications in business decision-making. Index Numbers:- Meaning , Types of index numbers, uses of index numbers, Construction of Price, Quantity and Volume indices:- Fixed base and Chain base methods.

Unit III:

Correlation & Regression Analysis Correlation Analysis: Rank Method & Karl Pearson's Coefficient of Correlation and Properties of Correlation. Regression Analysis: Fitting of a Regression Line and Interpretation of Results, Properties of Regression Coefficients and Relationship between Regression and Correlation.

Unit IV:

Probability Thoery & Distribution Probability: Theory of Probability, Addition and Multiplication Law, Baye's Theorem Probability Theoretical Distributions: Concept and application of Binomial; Poisson and Normal distributions.

Unit V

Hypothesis Testing& Business Analytics Hypothesis Testing: Null and Alternative Hypotheses; Type I and Type II errors; Testing of Hypothesis: Large Sample Tests, Small Sample test, (t, F, Z Test and Chi Square Test) Concept of Business Analytics- Meaning types and application of Business Analytics, Use of Spread Sheet to anlayze data-Descriptive analytics and Predictive analytics.

REFERENCE BOOKS:

1. G C Beri – Business Statistics, 3rd ed, TATA McGrawHill.
2. Chandrasekaran & Umaparvathi-Statistics for Managers, 1st edition, PHI Learning.
3. Davis , Pecar – Business Statistics using Excel, Oxford
4. Ken Black – Business Statistics, 5th ed., Wiley India
5. Levin and Rubin – statistics for Management, 7th ed., Pearson
6. Lind, Marchal, Wathen – Staistical techniques in business and economics, 13th ed, McGrawHill
7. Newbold, Carlson, Thorne – Statistics for Business and Economics, 6th ed., Pearson
8. S. C.Gupta – Fundamentals of Statistics, Himalaya Publishing
9. Walpole – Probability and Statistics for Scientists and Engineers, 8th ed., Pearson