

M.Sc Applied Statistics and Data Analytics AHEAD
Syllabus
2021 Admissions onwards

SEMESTER-I

21MAT505A Linear Algebra 300 3

Unit-I

Vector Spaces: Vector spaces - Sub spaces - Linear independence - Basis – Dimension.

Unit-II

Inner Product Spaces: Inner products - Orthogonality - Orthogonal basis - Gram Schmidt Process - Least Square Principle.

Unit-III

Linear Transformations: Positive definite matrices - Matrix norm and condition number - QR- Decomposition - Linear transformation - Relation between matrices and linear transformations - Trace and Transpose, Determinants, Symmetric and Skew Symmetric Matrices.

Unit-IV

Eigen values and Eigen vectors: Problems in Eigen Values and Eigen Vectors, Diagonalization, Orthogonal Diagonalization, Quadratic Forms, Diagonalizing Quadratic Forms, Diagonalisation and its applications.

Unit-V

Decompositions : LU and QR.

Text Books

Howard Anton and Chris Rorres, “Elementary Linear Algebra”, Tenth Edition, John Wiley & Sons, 2010.

Reference Books:

1. Nabil Nassif, Jocelyne Erhel, Bernard Philippe, Introduction to Computational Linear Algebra, CRC press, 2015.
2. Gilbert Strang, “Linear Algebra and Its Applications”, Fourth Edition, Cengage, 2006.
3. Kenneth Hoffmann and Ray Kunze, Linear Algebra, Second Edition, Prentice Hall, 1971.
4. I. N. Herstein, ‘Topics in Algebra’, Second Edition, John Wiley and Sons, 2000.

21MAT507A Probability Theory and Estimation

30 0 3

Course outcomes

- CO1: Understand the basics of probability, random variables and distribution functions.
- CO2: Gain knowledge about standard statistical distributions and their properties
- CO3: Know the importance of two dimensional random variables and correlation studies
- CO4: To gain knowledge point estimation and properties
- CO5: To gain knowledge about sampling distributions interval estimations.

Unit-I

Review of probability concepts - conditional probability- Bayes theorem.
Random Variable and Distributions: Introduction to random variable – discrete and continuous random variables and its distribution functions- mathematical expectations.

Unit-II

Standard distributions - Binomial, Poisson, Uniform, exponential, Normal. Mean, variance and applications of these distributions-Central limit theorem.

Unit-III

Joint, marginal and conditional probability distributions for discrete and continuous cases, stochastic independence, expectation of two dimensional random variables

Unit-IV

Point estimation, properties, methods of estimating a point estimator, minimum risk estimators
Sampling distributions of mean, variance, difference of means and ratio of variances.

Unit-V

Interval estimation- Confidence interval for one mean, difference of two means, single variance, ratio of two variances.

TEXT BOOKS /REFERENCE BOOKS:

1. Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers and Keying Ye, Probability and Statistics for Engineers and Scientists, 8th Edition, Pearson Education Asia, 2007.
2. Douglas C. Montgomery and George C. Runger, Applied Statistics and Probability for Engineers , John Wiley and Sons Inc., 2005.
3. Ravichandran, J: Probability and Statistics for engineers, First Reprint Edition, Wiley India, 2012.
4. Amir D. Aczel and Jayavel Sounderpandian. Complete Business Statistics, Sixth Edition, Tata McGraw-Hill Publishing Company, New Delhi. 2006.

21CSA506A Data Structures and Algorithms

3 0 2 4

Unit I

Abstraction - Abstract data types; Data Representation; Elementary data types; Basic concepts of data Structures; Mathematical preliminaries - big-Oh notation; efficiency of algorithms; notion

of time and space complexity; performance measures for data structures.
ADT array - Computations on arrays - sorting and searching algorithms.

Unit-II

ADT Stack, Queue, list - array, linked list, cursor based implementations of linear structures.
ADT Tree - tree representation, properties traversal of trees; ADT- Binary Trees – properties and algorithms .

Unit-III

ADT Priority Queue - Heaps; heap-based implementations; applications of heaps - sorting;
Search Tree - Binary search tree; balanced binary search trees - AVL tree; Applications of
Search Trees - TRIE; 2-3-4 tree; concept of B-Tree. ADT Dictionary - array based and tree based
implementations; hashing - definition and application.

Unit-IV

Introduction to time complexity. Big-O, worst case complexity, polynomial classifications.
Satisfiability, NP Complete and NP Hard (Definitions only).

Unit-V

Graphs algorithms: ADT- Data structure for graphs - Graph traversal- Transitive Closure-
Directed Acyclic graphs - Weighted graphs – Shortest Paths - Minimum spanning tree – Greedy
Methods for MST. Travelling salesman problem.

Text Books:

1. Goodrich M T, Tamassia R and Michael H. Goldwasser, “Data Structures and Algorithms in Python++”, Wiley publication, 2013.

Reference Books:

1. Goodrich M T and Tamassia R, “Data Structures and Algorithms in Java”, Fifth edition, Wiley publication, 2010.
2. Tremblay J P and Sorenson P G, “An Introduction to Data Structures with Applications”, Second Edition, Tata McGraw-Hill, 2002.

21MAT506 Optimization Techniques 3 0 0 3

Unit I

Single Variable Non-Linear Unconstrained Optimization

One dimensional Optimization methods, Uni-modal function, Region elimination methods - interval halving, Golden section search, Point estimation method - successive quadratic search, Gradient based methods – Newton’s method, secant method.

Unit II

Multi Variable Non-Linear Unconstrained Optimization

Direct search method – Univariant Method, Pattern search methods – Powell’s, search method, Gradient methods – Steepest decent method.

Unit III

Constrained optimization

Lagrangian method, Kuhn-Tucker conditions - Transformation methods.

Unit IV

Integer Programming:

Introduction – LPP, formulation, simplex method. ILP – Gomory cutting plane algorithm – Zero or one algorithm, branch and bound method.

Unit V

Specific Search Algorithms

Hill Climbing, Simulated Annealing, Genetic Algorithms, Ant Colony Optimization.

Text Books:

1. Hamdy A. Taha (1987): Operations Research – An Introduction, 4/e, Prentice Hall of India, Private Ltd, New Delhi.
2. Kanti Swarup, P. K. Gupta and Man Mohan (2004): Operations Research, Sultan Chand and Sons, New Delhi.
3. S.S. Rao, “Optimization Theory and Applications”, Second Edition, New Age International (P) Limited Publishers, 1995.

Reference Books:

1. Kapoor V.K. (2008): Operations Research, 8/e, Sultan Chand & Sons
2. Kalyanmoy Deb, “Optimization for Engineering Design Algorithms and Examples”, Prentice Hall of India, New Delhi, 2004.

21CSA507A Introduction to Data Science 2023

Course Outcomes

- CO1: Exploring and implementing exploratory data analytics
- CO2: Understanding correlation and regression and visualising them using R
- CO3: Understanding supervised learning through linear and logistic regressions
- CO4: Understanding and implementing classifiers for unsupervised data
- CO5: Exploring Massive data sets and implementing classification algorithms

Unit I

Data Collection, classification and analysis - Sampling methods, classification of data and representation of data- bar and pie charts – histogram frequency polygon – Box plot. Data Analysis Measures of Central tendency and dispersion - Mean, median, mode, absolute, quartile and standard deviations, skewness and kurtosis for both grouped and ungrouped data. Association of attributes.

Unit II

Curve fitting and interpolation - Fitting of straight lines and curves - fitting of simple linear lines, polynomials and logarithmic functions.

Unit III

Supervised Learning (Regression/Classification): Basic methods: Distance-based methods, Nearest-Neighbors, Decision Trees, Naïve Bayes. Linear models: Linear Regression, Logistic Regression.

Unit IV

Unsupervised Learning: Clustering: K-means/Kernel K-means. Dimensionality Reduction: PCA and kernel PCA. Matrix Factorization and Matrix Completion.

Unit V

Algorithms for Massive data problems, Clustering, CURE algorithm –ROCK algorithm -The Chameleon Algorithm –DBSCAN Algorithm --DENCLUE Algorithm –Clustering algorithms for high dimensional data.

Text Books /Reference books:

1. Douglas C. Montgomery and George C. Runger, Applied Statistics and Probability for Engineers, John Wiley and Sons Inc., 2005
2. Kevin Murphy, Machine Learning: A Probabilistic Perspective, MIT Press, 2012
3. Data Science and big data analytics: Discovering, analyzing, visualizing and presentating data ,EMC Education Services,John Wiley 2015
4. John Hopcroft and Ravi Kannan, “Foundations of Data Science”, ebook, Publisher, 2013.
5. Amir D Aczel, JayavelSoundarapandian, Complete Business Statistics, Seventh edition, McGraw Hill,NewDelhi

21CSA583APython Programming Lab0 0 21

Unit I

Introduction:History of Python, Need of Python Programming, Applications Basics of Python Programming, Running Python Scripts, Installing Python on Your Computer, Using the Terminal Command Prompt, IDLE, and Other IDEs, Variables, Assignment, Keywords, Input-Output, Indentation.Types, Operators and Expressions: Types - Integers, Strings, Booleans; Operators- Arithmetic Operators, Comparison (Relational) Operators, Assignment

Operators, Logical Operators, Bitwise Operators, Membership Operators, Identity Operators, Expressions and order of evaluations Control Flow- if, if-elif-else, for, while, break, continue, pass .Case Study: An Investment Report and Approximating Square Roots.

Unit II

Data Structures: Lists - Operations, Slicing, Methods; Tuples, Sets, Dictionaries, Sequences Comprehensions.Case Study: Nondirective Psychotherapy

Unit III

Functions: Defining Functions, Calling Functions, Passing Arguments, Keyword Arguments, Default Arguments, Variable-length arguments, Anonymous Functions, Fruitful Functions (Function Returning Values), Scope of the Variables in a Function - Global and Local Variables.

Modules: Creating modules, import statement, from. Import statement, name spacing.Python packages: Introduction to PIP, Installing Packages via PIP, Using Python Packages.Text Files: Text Files and Their Format, Writing Text to a File , Writing Numbers to a File , Reading Text from a File , Reading Numbers from a File, Accessing and Manipulating Files and Directories on Disk.Case Study: Gathering Information from a File System

Unit IV

Data Gathering and Cleaning:Cleaning Data, Checking for Missing Values, Handling the Missing Values, Reading and Cleaning CSV Data, Merging and Integrating Data, Reading Data from the JSON Format, Reading Data from the HTML Format, and Reading Data from the XML Format.

Regular expressions: Character matching in regular expressions, Extracting data using regular expressions, Combining searching and extracting and Escape character.Case Study:Detecting the e-mail addresses in a text file.

Unit V

Popular Libraries for Data Visualization in Python: Matplotlib, Seaborn, Plotly, Geoplotlib, and Pandas.Data Visualization: Direct Plotting, Line Plot, Bar Plot, Pie Chart, Box Plot, Histogram Plot, Scatter Plot, Seaborn Plotting System , Strip Plot , Box Plot, Swarm Plot, Joint Plot , Matplotlib Plot , Line Plot Bar Chart ,Histogram Plot ,Scatter Plot , Stack Plot and Pie Chart.

Coding Simple GUI-Based Programs: Windows and Labels, Displaying Images, Command Buttons and Responding to Events, Viewing the Images of Playing Cards, Entry Fields for the Input and Output of Text, and Using Pop-up Dialog Boxes. Case Study: A GUI-Based ATM

Text Books:

1. Chun, W. (2006) Core python programming. Prentice Hall Professional.
2. Embarak, O. (2018). Data Analysis and Visualization Using Python: Analyze Data to Create Visualizations for BI Systems. Apress.
3. Lambert, K. A. (2011). Fundamentals of Python: First Programs. Cengage Learning.

4. Severance, C. (2013). Python for informatics: Exploring information. CreateSpace.

Reference Books

1. <https://www.w3schools.com/python>
2. Learning Python, Mark Lutz, Orielly
3. Python Programming: A Modern Approach, VamsiKurama, Pearson
4. VanderPlas, J. (2016). Python data science handbook: Essential tools for working with data. " O'Reilly Media, Inc."

SEMESTER-II

21MAT517A Statistical Inference and Design of Experiments 30 0 3

CO1: To understand the concept of testing of hypothesis of various parameters using single sample and apply to engineering, science and business problems.

CO2: To know to apply goodness of fit tests and nonparametric tests

CO3: To understand statistical inference for two samples and apply to engineering, science and business problems.

CO4: To develop experiments and analyse the variance to conclude on the parameters of the population involved

CO5: To construct factorial experiments and to use for various real time problems

Unit I

Tests of Hypothesis based on Single Sample-Tests of Statistical Hypotheses, One-Sided and Two-Sided Hypotheses, *P*-Values in Hypothesis Tests, General Procedure for Hypothesis Tests, Tests on the Mean of a Normal Distribution, Variance Known- Tests on the Mean of a Normal Distribution, Variance Unknown- Tests on the Variance and Standard Deviation of a Normal Distribution.

Unit II

Tests of Hypothesis based on Two Samples -Inference on the Difference in Means of Two Normal Distributions, Variances Known- Inference on the Difference in Means of two Normal Distributions, Variances Unknown- Paired *t*-Test.

Unit III

Goodness of Fit Tests. Tests of Independence of attributes - Nonparametric Procedures -The Sign Test, The Wilcoxon Signed-Rank Test.

Unit IV

Design of Experiments: One and Two way Analysis of Variance-Designing Engineering Experiments.

Unit V

Design of Experiments: Latin Square Design-Factorial Experiments with two Factors.

Text Books /Reference Books

1. Douglas C. Montgomery and George C. Runger, Applied Statistics and Probability for Engineers, (2005) John Wiley and Sons Inc.
2. Sheldon. M. Ross : Probability and Statistics for Engineers and Scientists , MaGraw-Hill , 2004.
3. Amir D Aczel, Soundarapandian Jayavel, Complete Business statistics, Boston : McGraw-Hill/Irwin, 2009
4. Ravichandran, J. Probability and Statistics for engineers, First Reprint Edition, Wiley India, 2012.
5. Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers and Keying Ye, Probability and Statistics for Engineers and Scientists, 8th Edition, Pearson Education Asia, 2007.

21MAT516AMultivariate Statistics and Regression Analysis 30 0 3

Course Outcomes

CO1: To understand the basics of multivariate random variables and sampling distributions.

CO2: To apply multivariate techniques for classification of distributions

CO3: To understand the concept of PCA and its application in clustering analysis

CO4: To gain knowledge on simple linear regression, estimation and testing of model parameters

CO5: To gain knowledge on multiple linear and nonlinear regression and estimation of model parameters

Unit-I:

Multivariate Random variables and Distribution functions – Variance - covariance matrix – correlation - Bivariate normal distribution, Multivariate normal density and its properties - Definition of Wishart matrix and its properties, Mahalanobis Distance. Sampling distributions of \bar{X} and S , Large sample behaviour of \bar{X} and S .

Unit-II:

Classification for two populations, classification with two multivariate normal populations

Unit-III:

Principal components analysis, Dimensionality reduction, Factor Analysis- factor loadings using principal component analysis.

Unit-IV:

Simple Linear Regression- Properties, Least Squares Estimation of parameters, Hypothesis Tests in Simple Linear Regression.

Unit-V:

Multiple Linear Regression: Estimation of model parameters.

Text books/ Reference books:

1. Anderson, T. W. (1983): An Introduction to Multivariate Statistical Analysis. 3rdEd. Wiley.
2. Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers and Keying Ye. Probability and Statistics for Engineers and Scientists, Eighth Edition, Pearson Education Asia, 2007.
3. Douglas C. Montgomery and Elizabeth A. Peck and G. Geoffrey Vining. Introduction to Linear Regression Analysis”, Third Edition, John Wiley & Sons, Inc
4. Amir D. Aczel and Jayavel Sounderpandian. Complete Business Statistics, Sixth Edition, Tata McGraw-Hill Publishing Company, New Delhi. 2006.

21CSA517A

Machine Learning 3 0 2 4

Course Objectives:

CO1:To be able to formulate machine learning problems corresponding to different applications.

CO2:To understand a range of machine learning algorithms along with their strengths and weaknesses.

CO3 :To understand the basic theory underlying machine learning.

CO4:To be able to apply machine learning algorithms to solve problems of moderate complexity.

Unit-I

Introduction: Well-Posed Learning Problems, Designing a Learning System. The Inductive Learning Hypothesis and Concept Learning as Search (definitions). Machine Learning Basics: Learning Algorithms, Capacity, Over fitting and Under fitting, Hyper parameters and Validation Sets, No Free Lunch theorem, Estimators, Bias and Variance, Bayesian statistics, Supervised Learning Algorithms, Unsupervised Learning Algorithms, Stochastic Gradient Descent, Building a Machine Learning Algorithm and Issues in Machine Learning.

Unit-II

Decision Tree learning: Introduction, Decision tree representation, Appropriate problems for decision tree learning, The basic decision tree learning algorithm, Hypothesis space search in decision tree learning, Inductive bias in decision tree learning, Issues in decision tree learning. Implementation aspects of the Decision Tree and Classification Example.

Unit-III

Instance-Based Learning: Introduction, k-Nearest Neighbour Learning, Locally Weighted Regression, Radial Basis Functions, Case-Based Reasoning, Remarks on Lazy and Eager Learning.

Support Vector Machines: Optimal Separation: The Margin and Support Vectors, a Constrained Optimization Problem, Slack Variables for Non-Linearly Separable Problems. KERNELS: Choosing Kernels. The Support Vector Machine Algorithm and Multi-Class Classification. Case Study.

Unit-IV

Genetic Algorithms: Motivation, Genetic Algorithms, Elitism, Tournaments, and Niching, Using Genetic Algorithms and An illustrative Example. Hypothesis Space Search, Genetic Programming, Models of Evolution and Learning, Parallelizing Genetic Algorithms.

Unit-V

Reinforcement Learning: Introduction, the Learning Task, Q Learning, Non-Deterministic, Rewards and Actions, Temporal Difference Learning, Generalizing from Examples, Relationship to Dynamic Programming. Case Study.

Text Books:

1. Machine Learning – Tom M. Mitchell, - MGH
2. Machine Learning: An Algorithmic Perspective, Stephen Marsland, Taylor & Francis (CRC)
3. Haroon, D. (2017). Python Machine Learning Case Studies: Five Case Studies for the Data Scientist Apress.

Reference Books:

1. Harrington, P. (2012). Machine learning in action. Manning Publications Co..
2. Richard o. Duda, Peter E. Hart and David G. Stork, pattern classification, John Wiley & Sons
3. Machine Learning by Peter Flach , Cambridge.

21CSA515A Big Data Analytics and Hadoop 3 0 2 4

Course Outcomes

CO1: Understanding the concepts of Big Data

CO2: Understanding the aspects of managing, cleaning and sampling of Data

CO3: Understanding Hadoop architecture and implement Map Reduce concept

CO4: To understand the aspects of Data base management and Querying system

CO5: Understanding and executing HDFS using PIG and HIVE

Unit I

Introduction to Big Data: Types of Digital Data-Characteristics of Data – Evolution of Big Data - Definition of Big Data - Challenges with Big Data - 3Vs of Big Data - Non Definitional traits of Big Data - Business Intelligence vs. Big Data - Data warehouse and Hadoop environment.

Unit II

Big Data Analytics: Classification of analytics - Data Science - Terminologies in Big Data Data science process – roles, stages in data science project – working with data from files — exploring data – managing data – cleaning and sampling for modeling and validation. working with relational databases - NoSQL: Types of Databases – Advantages – NewSQL - SQL vs. NOSQL vsNewSQL.

Unit III

Introduction – distributed file system – Hadoop Components – Architecture – HDFS - algorithms using map reduce, Matrix-Vector Multiplication by Map Reduce – Hadoop - Understanding the Map Reduce architecture - Writing Hadoop Map Reduce Programs - Loading data into HDFS - Executing the Map phase - Shuffling and sorting - Reducing phase execution. Hadoop 2 (YARN): Architecture - Interacting with Hadoop Eco systems.

Unit IV

No SQL databases: Mongo DB: Introduction – Features - Data types - Mongo DB Query language - CRUD operations – Arrays - Functions: Count – Sort – Limit – Skip – Aggregate - Map Reduce. Cursors – Indexes - Mongo Import – Mongo Export. Cassandra: Introduction – Features - Data types – CQLSH - Key spaces - CRUD operations – Collections – Counter – TTL - Alter commands - Import and Export - Querying System tables.

Unit V

Hadoop Eco systems: Hive – Architecture - data type - File format – HQL – SerDe - User defined functions - Pig: Features – Anatomy - Pig on Hadoop - Pig Philosophy - Pig Latin overview - Data types - Running pig - Execution modes of Pig - HDFS commands - Relational operators - Eval Functions - Complex data type - Piggy Bank - User defined Functions - Parameter substitution - Diagnostic operator.

Text Books / Reference Books:

1. Seema Acharya, SubhashiniChellappan, “Big Data and Analytics”, Wiley Publication, 2015.
2. Judith Hurwitz, Alan Nugent, Dr. Fern Halper, Marcia Kaufman, “Big Data for Dummies”, John Wiley & Sons, Inc., 2013.
3. Data Science and big data analytics : Discovering, analyzing , visualizing and presentating data ,EMC Education Services,John Wiley 2015
4. Tom White, “Hadoop: The Definitive Guide”, O’Reilly Publications, 2011.
5. Kyle Banker, “Mongo DB in Action”, Manning Publications Company, 2012.
6. Russell Bradberry, Eric Blow, “Practical Cassandra A developers Approach“, Pearson Education, 2014.

21CSA516A Data Mining30 0 3

Course Objectives:

CO1:Learn data mining basic concepts and understand association rules mining.

CO2:Capable of grouping data using clustering techniques.

CO3:Able to identify the outliers of the given dataset.

CO4:Capable of minimizing dimensionality of the data with minimum loss of information.

CO5:Able to prioritize the web links and advertisements

Unit - I

Introduction to Data Mining: Introduction, What is Data Mining, Definition, KDD, Challenges, Data Mining Tasks, Data Preprocessing, Data Cleaning, Missing data, Dimensionality Reduction, Feature Subset Selection, Discretization and Binaryzation, Data Transformation; Measures of Similarity and Dissimilarity- Basics.

Unit - II

Association Rules: Problem Definition, Frequent Item Set Generation, The APRIORI Principle, Support and Confidence Measures, Association Rule Generation; APRIORI Algorithm.

Unit-III

Clustering: Problem Definition, Clustering Overview, Evaluation of Clustering Algorithms, Partitioning Clustering-K-Means Algorithm, K-Means Additional issues, PAM Algorithm; Hierarchical Clustering-Agglomerative Methods and divisive methods, Key Issues in Hierarchical Clustering, Strengths and Weakness;

Unit-IV

Outlier Detection: Outliers and Outlier Analysis -What Are Outliers?, Types of Outliers ,Challenges of Outlier Detection, Outlier Detection Methods, Statistical Approaches, Parametric Methods, Nonparametric Methods, Proximity-Based Approaches, Clustering-Based Approaches,

Unit-V

Dimensionality Reduction: Principal-Component Analysis, Singular-Value Decomposition, and CUR Decomposition. Link Analysis: Page Rank, Efficient Computation of Page Rank, Topic-Sensitive Page Rank, Link Spam, Hubs and Authorities.

Text Books:

1. Han, J., Pei, J., & Kamber, M. (2011). Data mining: concepts and techniques. Elsevier.
2. Rajaraman, A., & Ullman, J. D. (2011). Mining of massive datasets. Cambridge University Press.

Reference Books and websites:

1. <https://nptel.ac.in/courses/106/105/106105174/>
2. https://nptel.ac.in/content/storage2/nptel_data3/html/mhrd/ict/text/110105083/lec52.pdf
3. Ngo, T. (2011). Data mining: practical machine learning tools and technique, by ian h. witten, eibe frank, mark a. hell. ACM SIGSOFT Software Engineering Notes, 36(5), 51-52.
1. Romeo Kienzler, "Mastering Apache Spark 2.x", Packt Publishing Limited; 2nd Revised edition, 2017.

SEMESTER-III

21MAT606 STATISTICAL QUALITY CONTROL AND RELIABILITY 3 0 0 3

- CO1 To understand the basic concepts of quality control and to construct variable and attribute control charts.
- CO2 To understand and construct EWMA and CUSUM charts, analyse the process capability and Six Sigma quality metrics.
- CO3 To gain knowledge about acceptance sampling methods and their properties
- CO4 To gain knowledge about reliability and properties
- CO5 To study reliability distributions and analyse reliability of systems and maintenance

Unit- I

Basic concept of quality control, process control and product control, Statistical process control, theory of control charts, Shewhart control charts for variables- R, s charts, attribute control charts - p, np, c, u charts, modified control charts.

Unit- II

Moving average control charts, EWMA charts, CUSUM charts, –process capability analysis, process capability indices – C_p and C_{pk} .

Unit- III

Acceptance sampling for attributes, single sampling, double sampling, measuring performance of the sampling plans- OC, AQL, LTPD, AOQ, ATI curves.

Unit -IV

Introduction to Reliability and its needs; Different Approaches to Reliability Analysis, Application Areas, State Variable, Time to Failure, Failure Rate Function, Mean Time to Failure, Relationship between the Functions $F(t)$, $f(t)$, $R(t)$, and $z(t)$, Bath tub curve, Mean time to failure, Residual time

Unit - V

Parametric families of life distribution –Exponential.Weibull and Gamma and its characterization-Reliability estimation of parameters in these models. Fault Tree Analysis, Reliability Block Diagrams, Systems of Independent Components -System Reliability.Types of Maintenance, Downtime and Downtime Distributions, System Availability Assessment.

Text Books/References:

1. Montgomery D. C. (2005) Introduction to Statistical Quality control, 5th edition, Wiley.
2. Ravichandran, J. Probability and Statistics for engineers, First Reprint Edition, Wiley India, 2012.
3. Schilling E. G. (1982) Acceptance Sampling in Quality Control, Marcel Decker.
4. Marvin Rausand and ArnljotHoyland ,(2003): System Reliability Theory : Models, Statistical methods and applications, 2nd edition, John Wiley and Sons Inc., publications.

21CSA606A Introduction to Deep Learning 3 0 2 4

Course outcomes

CO1: Understand the basics concepts of artificial neural networks.

CO2: Gain knowledge about activation functions and understand the multi layer neural network

CO3: Know the importance of regularization, bagging and ensemble methods

CO4: To gain knowledge convolution neural network and case studies

CO5: To gain knowledge about recurrent neural networks, adversarial neural networks, Spectral CNN and deep reinforcement learning

Unit-I

Biological neuron, idea of computational units, McCulloch – pitts unit and thresholding logic, linear perceptron, perceptron learning algorithm, convergence theorem for Perceptron learning algorithm, logistic regression, gradient descent.

Unit-II

Feed forward neural network, activation functions, non-linear activation functions. multi-layer neural network.

Unit-III

Practical aspects of deep Learning: training, testing, regularization –dataset augmentation, Noise robustness, multitask learning, bagging and other ensemble methods, dropout- generalization.

Unit-IV

Convolution neural networks, backpropagation convolutions and pooling – optimization algorithms: mini-batch gradient descent, -convolutional nets case studies using Keras/Tensorflow.

Unit-V

Neural network architectures – recurrent neural networks, adversarial neural networks Spectral CNN, self-organizing maps, restricted boltzmann machines, long short-term memory networks, deep meta learning - deep reinforcement learning.

TextBooks/ Reference Books

1. Ian Goodfellow, Yoshua Bengio and Aeron Courville, Deep Learning, MIT Press, First Edition, 2016.
2. Gibson and Josh Patterson, Deep Learning A practitioner's approach, Adam O'Reilly, First Edition, 2017.
3. Francois Chollet, Deep Learning with Python, Manning Publications Co, First Edition, 2018.
4. Bishop C.M. Neural Networks for Pattern Recognition, Oxford University Press, 1995.

Elective Courses

21CSA631A Taguchi Techniques

3 0 0 3

Taguchi loss functions – mean square error loss function, average loss function, higher the better and lower the better loss functions – two-way analysis of variance with interactions – factorial experiments with two and three-level factors – orthogonal array experiments with two and three-level factors – methods of interpretation of experimental results - parameter and tolerance design experiments – signal-to-noise ratios – inner and outer array experiments.

Text/Reference Books

1. Taguchi Techniques for Quality Engineering
2. Taguchi G, (1991). Introduction to Quality Engineering: Designing Quality into Products and Processes. Asian Productivity Organization Second Edition,. Wiley

21CSA632A Special Distribution Functions

3 0 0 3

Inverted Beta Distribution, Noncentral Beta Distribution, Beta Binomial Distribution, Cauchy Distribution, Noncentral Chi-Squared Distribution, Dirichlet Distribution, Empirical Distribution Function, Erlang Distribution, Error Distribution, Generalized Exponential Distributions, Noncentral F-distribution, Inverted Gamma Distribution, Normal Gamma Distribution, Generalized Gamma Distribution, Inverse Gaussian (Wald) Distribution, Lognormal

Distribution, Pareto Distribution, Power Function Distribution, Power Series (Discrete) Distribution, Wishart (Central) Distribution.

Text/Reference books

Catherine Forbes, Merran Evens, Nicholas Hastings and Brian Peacock. (2010). *Statistical Distributions*, Fourth Edition, Wiley & Sons Publication, USA.

Karl Bury (1999) :Statistical distributions in Engineering , Cambridge University Press.

Thomopoulos, Nick T(2017): Statistical Distributions:Applications and Parameter Estimates, Springer.

21CSA633A Pattern Recognition

3 0 0 3

Pattern recognition systems – the design cycle – learning and adaptation – Bayesian decision theory – continuous features – Minimum error rate classification – discriminant functions and decision surfaces – the normal density based discriminant functions. Bayesian parameter estimation – Gaussian case and general theory – problems of dimensionality – components analysis and discriminants- Nonparametric techniques – density estimation – Parzen windows – nearest neighborhood estimation – rules and metrics - decision trees – CART methods – algorithm-independent machine learning – bias and variance for regression and classification – resampling or estimating statistics- Unsupervised learning and clustering – mixture densities and identifiability – maximum likelihood estimates – application to normal mixtures – unsupervised Bayesian learning – data description and clustering – criterion functions for clustering – hierarchical clustering – k-means clustering.

Text Reference Book:

1. Richard O. Duda, Peter E. Hart and David G. Stork, “*Pattern Classification*”, Second Edition, 2003, John wily & sons.
2. Earl Gose, Richard Johnsonbaugh and Steve Jost, “*Pattern Recognition and Image Analysis*”, 2002, Prentice Hall of India.

21CSA634A Stochastic Process

3 0 0 3

Random processes: General concepts and definitions - stationarity in random processes - strict sense and wide sense stationary processes - autocorrelation and properties- special processes – Poisson points, Poisson and Gaussian processes and properties , spectrum estimation , ergodicity, mean ergodicity, correlation ergodicity, Power spectrum density functions –

properties, Markov process and Markov chain, transition probabilities, Chapman Kolmogrov theorem, limiting distributions classification of states.

Text Books:

1. J. Ravichandran, “*Probability and Random Processes for Engineers*”, First Edition, IK International, 2015
2. Douglas C. Montgomery and George C. Runger, *Applied Statistics and Probability for Engineers*, (2005) John Wiley and Sons Inc.

Reference Books:

1. A. Papoulis, and Unnikrishna Pillai, “*Probability, Random Variables and Stochastic Processes*”, Fourth Edition, McGraw Hill, 2002.
2. Scott L. Miller, Donald G. Childers, “*Probability and Random Processes*”, Academic press, 2012.

21CSA635A Queuing Theory

3 0 0 3

Queuing Models: Basic characteristics of a Queueing Model – Role of Poisson and Exponential distributions, Stochastic Processes, Markov chains, Poisson Processes, Poisson Queuing Models with single server: Descriptions of the model, Assumptions, Probability distributions for number of Units (steady state), waiting time distribution, simple numerical problems on (M/M/1): (/FIFO) and (M/M/1): (N/FIFO) Models.

Poisson Queuing Models with multiple server: Descriptions of the model, Assumptions, Probability distributions for number of Units (steady state), waiting time distribution, simple numerical problems on (M/M/C): (/FIFO), (M/M/C): (N/FIFO) and (M/M/C): (C/FIFO) Models, M/M/G Models.

Text Books

1. Donald Gross & Carl M Harris (1998): Fundamentals of Queuing theory, John Wiley & Sons, Inc
2. Hamdy A. Taha (2006): Operations Research – An Introduction, 8/e, Prentice Hall of India Private Ltd., New Delhi

Reference Books

1. S.D. Sharma (2003) Operations Research, Kedar Nath Ram Nath & Co, Meerut, India
2. Kanthi Swarup, P.K. Gupta and Man Mohan (2004), Operations Research, Sultan Chand & Sons, New Delhi

21CSA636A Market Analytics 3 0 0 3

Business Analytics Basics: Definition of analytics, Evolution of analytics, Need of Analytics, Business analytics vs business analysis, Business intelligence vs Data Science, Data Analyst Vs Business Analyst, Business Analytics at the Strategic Level, Functional Level, Analytical Level, Data Warehouse Level. Market Segmentation Variables, Market Segmentation Types, Marketing Data Landscape, Analyzing the trend of data in Marketing – case studies.

Time series as a discrete parameter stochastic process, Auto - covariance, Auto-correlation functions and their properties. Exploratory time series analysis, Test for trend and seasonality, Exponential and moving average smoothing, forecasting based on smoothing.

Linear time series models: Autoregressive, Moving Average, autoregressive Moving Average models, Autoregressive Integrated Moving Average models. Estimation of ARMA models: Yule-Walker estimation for AR Processes, Maximum likelihood and least squares estimation for ARMA Processes.

Text / References Books:

1. Grigsby Gert H.N Laursen and Jesper Thorlund : *Business analytics for managers taking business intelligence beyond reporting*, second edition 2016.
2. Wayne L. Winston: *Marketing Analytics: Data-Driven Techniques with Microsoft Excel*, Wiley, 2014.
3. Mike Grigsby : *Marketing Analytics: A Practical Guide to Improving Consumer Insights Using Data Techniques*, Kogan Page; 2 edition ,2018
4. Mike Anderson, T.W : *The Statistical Analysis of Time Series*, John Wiley, New York, 1971.
5. Kendall, Sir Maurice and Ord, J.K. : *Time Series*, Edward Arnold, London, 1990.

21CSA637A Survival Analysis

3 0 0 3

Survival Analysis: Functions of survival times, survival distributions and their applications Censoring Schemes: Type I, Type II and progressive or random censoring with biological examples. Estimation of mean survival time and variance of the estimator for Type I and Type II censored data with numerical examples.

Non-parametric methods: Actuarial and Kaplan-Meier methods for estimating survival function and variance of the Estimator.

Competing Risk Theory: Indices for measurement of probability of death under competing risks and their inter-relations. Estimation of probabilities of death using maximum likelihood principle and modified minimum Chi-square methods.

References

1. Miller, R.G. *Survival analysis*, John Wiley ,1981
2. Collet, D. *Statistical analysis of life time data*, 1984
3. Cox, D.R. and Oakes, D.: *Analysis of survival data*, Chapman & Hall, New York, 1984
4. Gross, A.J. and Clark, V.A.: *Survival distribution: Reliability applications in the Biomedical sciences*, John Wiley and Sons, 1975
5. Elandt-Johnson, R.E. Johnson, N.L. : *Survival models and data analysis*, John Wiley & sons.

21CSA638A

Sampling Techniques

3 0 0 3

Preliminary concepts – schedules and questionnaires, pilot survey, non-sampling errors, use of random numbers. Simple random sampling with and without replacements, random number generation– estimates of population mean and population proportion and their standard errors, Probability proportional to size sampling, estimates of these standard errors. Stratified random sampling – estimates of sample statistic and estimates of their standard errors. Allocation of sample size in stratified random sampling. Linear and circular systematic sampling. Cluster sampling : Two stage sampling (equal first stage units). Ideas of ratio and regression estimators – only estimates of sample mean..

References

1. Cochran, W.G. : *Sampling Techniques*, 3rd Ed., Wiley Eastern. 1984
2. Murthy, M.N. : *Sampling Theory & Statistical Methods*, Statistical Pub. Society, Calcutta, 1977
3. Des Raj and Chandhok P. : *Sample Survey Theory*, Narosa Publishing House, 1988.

21CSA639A Demography and Actuarial Statistics 3 0 0 3

Demographic data – Sources, Coverage and Content errors in demographic data. Measures of fertility period and cohort measures. Use of birth order Statistics and child - Woman ratio. Brass technique to estimate current-fertility levels Estimation of TFR age pattern of fertility. Measures of mortality - standard death rates, neo-natal, perinatal death rates, maternal and infant mortality rates standardization of mortality rates.

Life table: Basic definitions, probabilities, construction of life tables, life expectancy, Life annuities: calculating annuity premium, interest and survivorship discount unction, guaranteed payments, deferred annuities.

Life insurance: Introduction, calculation of life insurance premiums, types of life insurance, combined benefits, insurances viewed as annuities, Insurance and annuity reserves: General pattern reserves, recursion, detailed analysis of an insurance.

Contingent Functions: Contingent probabilities, assurances. Decrement tables. Pension funds: Capital sums on retirement and death, widow's pensions, benefits dependent on marriage.

Text Books:

1. Ramkumar. R : *Technical Demography*, Wiley eastern Ltd, New Delhi, 1986.
2. Rogers. A : *Introduction to Mathematical Demography*, Johnwiley, New york, 1975
3. Biswas. S. : *Stochastic processes in Demography and applications*, Wiley eastern limited, 1988
4. Atkinson, M.E. and Dickson, D.C.M.: *An Introduction to Actuarial Studies*, Elgar Publishing, 2000
5. Philip, M. et. al : *Modern Actuarial Theory and Practice*, Chapman and Hall, 1999.

21CSA640A Official Statistics 3 0 0 3

Introduction to Indian Statistical systems- Role, function and activities of Central Statistical organization and State Agencies. Role of National Sample Survey Organization. General and special data dissemination systems. Scope and Contents of population census of India. statistics, their reliability and limitations. Role of Ministry of Statistics & Program Implementation

(MoSPI), Central Statistical Office (CSO), National Sample Survey Office (NSSO), Registered General Office and National Statistical Commission.

Population growth in developed and developing countries, Evaluation of performance of family welfare programmes. Statistics related to Industries, foreign trade, balance of payment, cost of living, inflation, educational and other social statistics.

Economic development: Growth in per capita income and distributive justice indices of development, human development index. National income estimation- Product approach, income approach and expenditure approach. Measuring inequality in income: Gini Coefficient, Theil's measure; Poverty measurements: Different issues, measures of incidence and intensity; Combined Measures: Indices due to Kakwani, Sen etc.

Text Books:

1. Guide to Official Statistics (CSO) 1999.
2. Principles and Accommodation of National Population Census, UNEDCO
3. CSO (1989)a: National Accounts Statistics- Sources and Methods.
4. Guide to current Indian Official Statistics, Central Statistical Office, GOI, and New Delhi. <http://mospi.nic.in/>

21CSA641A Healthcare Analytics 3 0 0 3

Introduction to Healthcare Data Analytics- Electronic Health Records–Components of EHR- Coding Systems- Benefits of EHR- Barrier to Adopting HER Challenges- Phenotyping Algorithms. Challenges in Healthcare Data Analysis, Acquisition Challenges, Pre-processing, Transformation, Social Media Analytics for Healthcare.

Advanced Data Analytics for Healthcare : Review of clinical trials, Prediction Models. Statistical Prediction Models, Alternative Clinical Prediction Models, Survival Models, Predictive Models for Integrating Clinical and Genomic Data, Data Analytics for Pervasive Health, Fraud Detection in Healthcare, Pharmaceutical Discoveries and Clinical Decision Support Systems.

Text / References books :

1. Chandan K. Reddy and Charu C Aggarwal, “*Healthcare data analytics*”, Taylor & Francis, 2015
2. Hui Yang and Eva K. Lee, “*Healthcare Analytics: From Data to Knowledge to Healthcare Improvement*”, Wiley, 2016.

21CSA642A Computational Biology 3 0 0 3

Introduction to Bioinformatics - applications of Bioinformatics - challenges and opportunities - introduction to NCBI data model- Various file formats for biological sequences.

Bioinformatics resources – Importance of databases - Biological databases- Primary & Secondary databases.

Sequence alignment methods: Sequence analysis of biological data-Significance of sequence alignment- pairwise sequence alignment methods- Use of scoring matrices and gap penalties in sequence alignments- PAM and BLOSUM Scoring Matrices. Introduction to Dynamic Programming, Global alignments: Needleman Wunsch Algorithm, Local Alignments: Smith Waterman Algorithm, Gap Penalties.

Multiple sequence alignment methods – Tools and application of multiple sequence alignment. Sequence alignment tools Phylogenetic analysis algorithms: Maximum Parsimony, UPGMA, Transformed Distance, Neighbors-Relation, Neighbor-Joining, jackknife.

References/ Textbooks

1. Higgins, Des and Willie Taylor: *Bioinformatics: Sequence , Structure and databanks*, Oxford , University Press, 2000.
2. Baxenvants, AD., *Bioinformatics: A practical guide to the analysis of genes and proteins*, Third edition, John wiley&
3. Sons , 2005
4. Teresa Attwood, *Introduction To Bioinformatics* , Pearson Education Singapore Pte Ltd, 2007
5. S.C. Rastogi et al, *Bioinformatics: Methods and Applications: (Genomics, Proteomics and Drug Discovery)* Kindle Edi

21CSA643A Computer aided drug designing 3 0 0 3

Introduction to Molecular Modeling: Molecular Modeling and Pharmacoinformatics in Drug Design, Phases of Drug Discovery, Target identification and validation. Protein Structure Prediction and Analysis: Protein Structure prediction methods: Secondary Structure Prediction, Tools for Structure prediction; Protein structural visualization; Structure validation tools; Ramachandran Plot. QSAR : Quantitative Structure and Activity Relationship - Historical Development of QSAR, Tools and Techniques of QSAR, Molecular Structure Descriptors. Multivariate Statistical methods in QSAR -Principal Component Analysis (PCA) and Hierarchical Cluster Analysis(HCR). Regression analysis tools - Pincipal Component Regression (PCR), Partial Least Squares (PLS) - Case studies. High Throughput / Virtual screening- Introduction, Basic Steps, Important Drug Databases, Designing Lipinski's Rule of Five, ADMET screening. Docking Studies-Molecular visualization tools: RasMol and Swiss-PdbViewer Molecular docking tools: AutoDock and ArgusLab.

References/ Textbooks

1. Leach Andrew R., Valerie J. Gillet, *An introduction to Chemoinformatics*. Publisher: Kluwer academic , 2003. ISBN: 1402013477.
2. Opera Tudor I, Ed. , *Chemoinformatics in drug discovery*, Wiley-VCH Verlag, 2005.
3. Gasteiger Johann, Engel Thomas. *Chemoinformatics: A Textbook*. Publisher: Wiley, First edition. 2003.
4. Kenneth M Merz, Jr, Dagmar Ringe, Charles H. Reynolds , *Drug design: Structure and ligand based approaches* , Cambridge University press , 2010.

21CSA644A Reinforcement Learning 3 0 0 3

Introduction: Reinforcement Learning, Elements of Reinforcement Learning, Limitations and Scope, An Extended Example- Tic-Tac-Toe. Multi-armed Bandits: A k-armed Bandit Problem, Action-value Methods, The 10-armed Testbed, Incremental Implementation, Tracking a Nonstationary Problem, Optimistic Initial Values, Upper-Confidence-Bound Action Selection, Gradient Bandit Algorithms.

Finite Markov Decision Processes: The Agent–Environment Interface, Goals and Rewards, Returns and Episodes, Unified Notation for Episodic and Continuing Tasks, Policies and Value Functions, Optimal Policies and Optimal Value Functions, Optimality and Approximation.

Review of Markov process and Dynamic Programming.

Temporal-Difference Learning: TD Prediction, Advantages of TD Prediction Methods, Optimality of TD, Sarsa: On-policy TD Control, Q-learning: Policy TD Control. Expected Sarsa. Maximization Bias and Double Learning.

Text/ References Book:

1. Richard S. Sutton and Andrew G. Barto, *Reinforcement Learning: An Introduction*, MIT Press, 2018.
2. Sudharsan Ravichandiran, *Hand-on Reinforcement Learning with Python*, Packt Publications, 2018.
3. Sayon Dutta, *Reinforcement Learning with Tensor Flow: A beginner's guide*, Packt Publications, 2018.

21CSA645A Social Network Analytics

3 0 0 3

An Introduction to social network data analytics: research Issues, statistical properties of social networks, random walks in social networks and their applications: survey, applications, community discovery in social networks, node classification in social networks, evolution in social networks - survey, survey of models and algorithms for social influence analysis, survey of algorithms and systems for expert location in social networks, survey of link prediction in social networks, data mining in social media, text mining in social networks

Text and Reference

1. Charu C. Aggrawal : *Social Network Data Analytics*, Springer, 2011.
2. Cioffi-Revilla, Claudio. *Introduction to Computational Social Science*, Springer, 2014.
2. Matthew A. Russell. *Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, Github, and More*, 2nd Edition, O'Reilly Media, 2013.
3. Robert Hanneman and Mark Riddle. *Introduction to social network methods*. Online Text Book, 2005.

21CSA646A Mining of Massive Datasets

3 0 0 3

Basics of Data Mining - computational approaches - statistical limits on data mining - MapReduce - Distributed File Systems . MapReduce . Algorithms using MapReduce . Extensions to MapReduce. Mining Data Streams: The Stream Data Model - Sampling Data in a Stream - Filtering Streams. Link analysis, Frequent itemsets, Clustering, Advertising on web,

Recommendation system, Mining Social-Network Graphs, Dimensionality Reduction, Large-Scale Machine Learning.

Text / References Book

1. Jure Leskovec ,AnandRajaraman, Jeffrey David Ullman, Mining of Massive Datasets, Cambridge University Press, 2014.
2. Tom White, Hadoop: The Definitive Guide: Storage and Analysis at Internet Scale O'Reilly Media; 4 edition , 2015.

21CSA647A Parallel and Distributed Systems 3 0 0 3

Introduction – parallelism and goals, parallel computing models – RAM, PRAM , CTA. Reasoning about Performance – Introduction -Basic Concepts - Performance Loss - Parallel Structure - Measuring Performance. Shared memory architecture.

Parallel Programming: Task and Data Parallelism with examples –Comparison Programming with Threads - POSIX Threads- Thread Creation and Destruction. Mutual Exclusion-Synchronization - Safety and Performance Issues – Reduction – threads Inter process communication – internet protocols – multicast communication – MPI. Remote invocation: Remote procedure call – remote method invocation -

System models : physical models, architecture models, operating system support. Distributed file systems – introduction- time and global states – synchronization of physical clocks – coordination and agreements: Mutual exclusion, election, consensus.

Text Books

1. George Coulouris , Jean Dollimore , Tim Kindberg , Gordon Blair: *Distributed Aystems : Concepts and Design* , Fifth Edition , Addison Wiley, 2012.
2. Calvin Lin ,Larry Snyder : *Principles of Parallel Programming*, Pearson, 2009

References

1. Bertil Schmidt,Jorge Gonzalez-Dominguez, Christian Hundt , Moritz Schlarb, *Parallel Programming: Concepts and Practice* First Edition, Morgan Kaufmann, 2017.
2. Ajay D. Kshemkalyani,MukeshSinghal , *Distributed Computing: Principles, Algorithms, and Systems*, Cambridge University Press, First edition, 2008.