PROPOSED

COURSE STRUCTURE

OF

M.Sc. Tech.

STATISTICS AND DATA SCIENCE

DEPARTMENT OF MATHEMATICS

SCHOOL OF SCIENCE

SESSION: 2018-20

Course Distribution:

ADAMAS UNIVERSITY								
SCHOOL OF SCIENCE DEPARTMENT OF MATEMATICS – M.Sc. Tech. PROGRAM SEMESTER - I								
Type of the Paper	**Paper Code**	**Theory / Practical**	**Brief Contents**	**Contact Hour Per Week**	**L**	**T**	**P**	**Credit**
CORE		Theory of Probability		4	3	1	0	4
CORE		Discrete Mathematics and Algorithms		6	5	1	0	6
CORE		Introduction to Applied Statistics		4	3	1	0	4
FOUNDATION		Computer Programming		4	3	1	0	4
FOUNDATION		Computer Programming Lab		4	0	0	4	2
FOUNDATION		English Language		3	2	1	0	2
Total				25	14	5	4	22

ADAMAS UNIVERSITY								
SCHOOL OF SCIENCE DEPARTMENT OF MATHEMATICS – M.Sc. Tech. PROGRAM SEMESTER - II								
Type of the Paper	**Paper Code**	**Theory / Practical**	**Brief Contents**	**Contact Hour Per Week**	**L**	**T**	**P**	**Credit**
CORE		Introduction to Statistical Inference		4	3	1	0	4
CORE		Introduction to Statistical Modeling		6	5	1	0	6
CORE		Optimization		6	5	1	0	6
CORE		Numerical Linear Algebra		4	3	1	0	4
CORE		Modern Applied Statistics: Learning		4	0	0	4	4
FOUNDATION		Environmental Sc./ Behavioral Sc./ Law and Social Sc./ Accounting & Finance/Journalism		3	2	1	0	2
Total				27	14	5	4	26

ADAMAS UNIVERSITY SCHOOL OF SCIENCE DEPARTMENT OF MATHEMATICS – M.Sc. Tech. PROGRAM SEMESTER - III								
Type of the Paper	**Paper Code**	**Theory / Practical**	**Brief Contents**	**Contact Hour Per Week**	**L**	**T**	**P**	**Credit**
CORE		Modern Applied Statistics: Data Mining		4	3	1	0	4
CORE		Introduction to parallel computing using MPI, openMP, and CUDA		6	5	1	0	6
CORE		Stochastic Methods in Engineering		4	3	1	0	4
CORE		Advanced Software Development for Scientists and Engineers		6	5	1	0	6
CORE		Parallel Computer Architecture and Programming		4	3	1	0	4
CORE		Advanced Multi-Core Systems		4	3	1	0	4
Total				28	22	6		28

ADAMAS UNIVERSITY SCHOOL OF SCIENCE DEPARTMENT OF MATHEMATICS– M.Sc. Tech. PROGRAM SEMESTER - IV								
Type of the Paper	**Paper Code**	**Theory / Practical**	**Brief Contents**	**Contact Hour Per week**	**L**	**T**	**P**	**Credit**
CORE		Optional I		4	3	1	0	4
CORE		Optional II		4	3	1	0	4
CORE		Optional III		4	3	1	0	4
CORE		Interaction with Industry/Institute			0	0	0	2
CORE		Project/Dissertation		10	0	0	10	8
CORE		Seminar on Project/Dissertation and Viva			0	0	0	2
Total				22	9	3	10	24

List of Optional papers		
Optional I	**Optional II**	**Optional III**
Representations and Algorithms for Computational Molecular Biology	Machine Learning	Geostatistics
Data Driven Medicine	Convolutional Neural Networks for Visual Recognition	Business Intelligence from Big Data
Modern Statistics for Modern Biology	Mining Massive Data Sets	Human Neuroimaging Methods
Social and Information Network Analysis	Topics in Computer Graphics	Paradigms for Computing with Data

***Offering of subjects will vary from year to year subject to availability of faculty**

DETAILED SYLLABI FOR SEMESTER 1

Theory of Probability

Credits- 4 **L: T: P=3-1-0**

Unit-I

Review of Probability concepts, condition probability, Bayes' theorem and its applications, random variables and its types, discrete distribution function, probability density function, various measures of central tendency, Moment generating function and its limitations, Discrete probability distributions: Binomial, Poisson, negative binomial, geometric and hyper-geometric distributions and their moment generating functions and properties, Continuous probability distributions: Triangular, gamma, exponential, Weibul, normal distribution, their moment generating functions and applications.

[22L]

Unit-II

Sampling theory: Introduction, types of sampling, parameter and statistics, test of significance, procedure for testing of hypothesis, tests of significance for large samples, sampling of attributes, sampling of variables, χ^2 – distribution, some theorem on Chi-Square distribution, linear transformation, applications of Chi-Square distribution, Student's 't' distribution and F-distribution and its applications, relations between t & F distribution and F & χ^2 – distributions.

[18L]

Unit-III

Time series analysis and forecasting methods: Introduction and analysis of a time series, trend method, time series method, correlation regression method, End–Use method, exponential smoothing method, Delphi method, demand forecasting for industrial products. [20L]

Text Books:

1. S C Gupta and V K Kapoor, Fundamentals of Mathematical Statistics, S Chand & Sons.
2. Vijay K. Rohatgi, A.K. Md. EhsanesSaleh, An Introduction to Probability and Statistics, Second edition, Wiley.
3. T N Srivastava and ShailagaRego, Statistics for Management, McGraw Hill Education.

Reference Books:

1. Dipak Chatterjee, Elements of Statistics, Scitech publications.

2. R V Hogg, J Mckean and A T Craig, Introduction to Mathematical Statistics, 7e, Pearson Education India.

Discrete Mathematics and Algorithms

Credit: 6 **L: T: P=5-1-0**

Unit 1. Classical Propositional Logic

Propositional formulae.Classical (Boolean) interpretation of propositional formulae. Tautologies and satisfiable formulae. Classical propositional calculus . Resolution method.

Unit 2. Classical Predicate Logic

Predicate formulae;quantifiers. Interpretations (models) of predicate logic.Predicate calculus. Completeness and undecidability of classical predicate logic (without proof). Proof search and proof assistants (overview).

Unit 3. Basic Notions of Graph Theory

Definition of graph. Special types of graphs. Trees. Using graphs for modelling networks etc. Isomorphism of graphs. Euler graphs.

Unit 4. Algorithms on Graphs

BFS. DFS. Dijkstra's algorithm. Checking whether a graph is a Euler graph.

Unit 5. Regular Expressions

The notion of regular expression. Regular expressions and finite automata (Kleene's theorem): sketch of proof, examples. Algorithms for matching with regular expressions. Implementation of regular expressions.

Unit 6. Context-Free Grammars and Parsing Algorithms

Context-free grammars and context-free languages. Parsing with CFG: CYK, LR (overview). The limits of CFG (pumping lemma). Practical parsing using CFG: the Bison/YACC parser generator.

Text/References

• C. C. Leary, L. Kristiansen. A friendly introduction to mathematical logic (2nd edition). Milne Library, SUNY Geneseo, NY, 2015.

• J. E. Hopcroft, R. Motwani, J. D. Ullman. Introduction to automata theory, languages, and computation (3rd edition). Pearson, 2007

• I. Chiswell, W. Hodges. Mathematical logic. Oxford University Press, 2007.

• T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein. Introduction to Algorithms (3rd edition), MIT Press, 2009.

• A. V. Aho, M. S. Lam, R. Sethi. J. D. Ullman. Compilers: principles, techniques, and tools (2nd edition). Addison-Wesley, 2006. Course materials will be made available through

Introduction to Applied Statistics

Credits: 4 **L-T-P: 3-1-0**

Unit 1.

INTRODUCTION TO STATISTICS

Basic Concepts in Statistics. Organizing and presenting data: Tables and frequency distribution and graphic representations.

Unit 2.

MEASURES OF CENTRAL TENDENCY

Concepts of mean, median, mode, quintiles. Relationship between measures. Procedures and applications.

Unit 3.

MEASURES OF DISPERSION

Range, interquartile range, deviation, variance, standard deviation and coefficient of variation.

Unit 4.

FREQUENCY DISTRIBUTION

Theoretical frequency distribution of a variable. Measures of concentration and inequalities: asymmetry and kurtosis.Normal distributions.

Unit 5.

SERIES OF TWO VARIABLES

Definition and graphical representation.Central tendency. Statistical dispersion. Covariance.

Unit 6.

STATISTICAL DEPENDENCE, CORRELATION

Correlation: concept, procedure and application. Pearson's correlation coefficient. Fitting linear regressions amongst two variables. Least squares approach.

Unit 7.

TIME SERIES

Definition and graphical representation. Components of time series. Seasonal variation. Seasonal Indices. Seasonal adjustment.

Unit 8.

A DESCRIPTIVE ANALYSIS OF TIME SERIES

Index numbers. Complex index numbers with and without weights.Paasche, Laspeyres and Fisher indices. Link and changes of base periods. Deflation of economic series.

Text/References:

1. Buglear, J. (2010). Stats means business- Statistics with Excel for business, hospitality & tourism (2nd ed.). New York: Elsevier.

2. Davis, G., & Pecar, B. (2009). Business Statistics using Excel (2nd ed.). Oxford University Press.
   Good, P. I., & Hardin, J. W. (2012). Common errors in statistics (and how to avoid them). [Hoboken, N.J.] : John Wiley.
3. Newbold, P., Carlson, W. L., & Thorne, B. (2013). Statistics for business and economics. Harlow, Essex : Pearson Education.
4. Ross, S. M. (2010). Introductory statistics. Amsterdam [etc.]: Elsevier : Academic Press.
5. Rugg, G. (2007). Using statistics: a gentle introduction. Maidenhead : McGraw-Hill.

Computer Programming

Credits- 3 **L: T: P=3-0-0**

Unit-I

Fundamentals of Computer: History of Computer, Generation of Computer, Classification of Computer, Basic Architecture of Computer, Memory, Input and Output Devices, Binary and other Number System, BCD, ASCII, Binary Arithmetic and Logic Gates, Algorithm and Flowchart.

Unit-II

Basics of C Programming: Characters, Identifiers, Keywords, Data type and sizes, Constant and variables, Various operators and expressions, Standard input and output, Formatted input and output Printf(), Flow of control and Exit() Function.

Unit-III

Functions and Pointers: Definition, Declaration of Function, Various Types of Functions, Call-By-Value, Call-By-Reference, Recursion, Tail Recursion, Pointer, Functions with Pointer.

Unit-IV

Arrays: Definition, Declaration, Initialization of an Array, Dimensionality, Sorting and Searching Algorithms

String: Definition, Declaration, Initialization, String Functions.

Unit-V

Structures: Definition, Declaration, Initialization, Operators used in Structure, Structure within Structures,

Unions: Union, Difference between a Structure and an Union.

Unit-VI

Files: Types of File, File Processing, Handling Characters, Handling Integers, Random File Accessing, Errors during File processing.

Text/ References:

1. Herbert Schildt, The Complete Reference, 4th Edition, Tata Mcgraw Hill Education.
2. Brian W. Kernighan, Dennis M. Ritchie, The C Programming Language, 2nd Edition, PHI.
3. Byron S. Gottfried, Schaum's Outline of Programming with C", 2nd Edition, Mcgraw Hill Education.

DETAILED SYLLABUS OF SEMESTER 2

Introduction to Statistical Inference

Credits: 4 **L-T-P: 3-1-0**

Unit I:

Estimation Theory: Sufficiency, Completeness and exponential family Estimation, Rao-Blackwell theorem, Lehman Schaffer theorem, Cramer-Rao inequality minimal sufficiency, Completeness, Bounded completeness, Ancillary statistics, Basu's theorem on independence of Statistics, Exponential family.

Unit II:

Methods of estimation and criterion for good estimators Bhattacharya bound, Chapman Robbins and Kiefer (CRK) bound, maximum likelihood estimation, Zehna theorem for invariance, Cramer theorem for weak consistence, asymptotic normality, BAN and CAN estimators, asymptotic efficiency.

Unit III:

Confidence Estimation and Hypothesis Testing: Simple and composite hypotheses. Randomized and nonrandomized tests, Most powerful test, Neyman-Pearson Lemma and its applications, Determination of minimum sample size to achieve the desired strengths. Composite hypotheses: Monotone likelihood ratio property, power function of a test, existence of UMP, Tests for one-sided alternatives. UMP tests for two sided alternatives Examples, Their existence and non-existence.

Unit IV:

Generalized Neyman Pearson lemma: Unbiased test, UMPU test and their existence in the case of exponential families, similar tests, test with Neyman structure, problem of confidence intervals, Relation with testing of hypotheses problem, Uniformly Most Accurate (UMA) and UMAU confidence intervals, shortest length confidence intervals, Non-parametric tests, One and two sample problem; tests on U-Statistics for various hypotheses; the asymptotic distributions of the statistics involved under the null hypotheses (Statements of the theorems only), Sign test, Run test, Signed-Rank test, Wilcoxon Signed Rank test, Man-Whiteny test.

Texts/ References:

1. B.K. Kale., A First Course on parametric Inference. Narosa, 2^{nd} Edition.
2. V.K. Rohatgi, Introduction to Probability and Mathematical Statistics. Wiley Eastem Ltd. New Delhi.
3. G Casella and R L Berger, Statistical inference, 2^{nd} edition, Duxbury Thomson Learning.
4. S. C. Gupta and V.K. Kapoor-Fundamentals of Mathematical Statistics, S. Chand and Company Pvt. Ltd., New Delhi.

Introduction to Statistical Modeling

Credits: 6 **L-T-P: 5-1-0**

Unit I

Empirical and mechanistic models. Nonlinear growth models like monomolecular, logistic, Gompertz, Richards. Applications in agriculture and fisheries.

Unit II

Nonlinear estimation: Least squares for nonlinear models, Methods for estimation of parameters like Linearization, Steepest, and LevenbergMarquardt's Reparameterization.

Unit III

Two-species systems. Lotka-Volterra, Leslie-Gower and Holling-Tanner non-linear prey-predator models. Volterra's principle and its applications. Gause competition model.

Unit IV

Compartmental modelling - First and second order input-output systems, Dynamics of a multivariable system. Practical Fitting of mechanistic non-linear models; Application of Schaefer and Fox non-linear models; Fitting of compartmental models.

Text/References:

1. Draper NR & Smith H. 1998. Applied Regression Analysis. 3rd Ed. John Wiley.
2. Efromovich S. 1999. Nonparametric Curve Estimation. Springer.
3. Fan J & Yao Q. 2003. Nonlinear Time Series-Nonparametric and Parametric Methods. Springer.
4. France J & Thornley JHM. 1984. Mathematical Models in Agriculture. Butterworths.
5. Harvey AC. 1996. Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge Univ. Press.
6. Ratkowsky DA. 1983. Nonlinear Regression Modelling: A Unified Practical Approach. Marcel Dekker.
7. Ratkowsky DA. 1990. Handbook of Nonlinear Regression Models. Marcel Dekker.
8. Seber GAF & Wild CJ. 1989. Non-linear Regression. John Wiley.
9. Silverman BW. 1986. Density Estimation for Statistics and Data Analysis. Chapman & Hall.

Optimization

Credits- 6 **L: T: P=5-1-0**

Unit I

Optimization – Introduction, Formulation of LPP, Geometry of LPP and Graphical Solution of LPP, Solution of LPP :Simplex Method, Big - M Method, Two - Phase Method, Special Cases in Simple Applications, Introduction to Duality Theory, Dual Simplex Method, Post Optimaility Analysis, Integer Programming – I, Integer Programming – II.

Unit II

Introduction to Transportation Problems, Solving Various types of Transportation Problems, Assignment Problems, Project Management, Critical Path Analysis, PERT, Shortest Path Algorithm, Travelling Salesman Problem, Classical optimization techniques :Single variable optimization, Unconstarined multivariable optimization, Nonlinear programming with equality constraint, Nonlinear programming KKT conditions, Numerical optimization : Region elimination techniques, Numerical optimization : Region elimination techniques (Contd.), Fibonacci Method, Golden Section Methods, Interpolation Methods, Unconstarined optimization techniques : Direct search method.

Unit III

Unconstarined optimization techniques : Indirect search method, Nonlinear programming : constrained optimization techniques, Interior and Exterior penulty Function Method, Separable Programming Problem, Introduction to Geometric Programming, Constrained Geometric Programming Problem, Dynamic Programming Problem, Dynamic Programming Problem (Contd.), Multi Objective Decision Making, Multi attribute decision making.

Texts/References:

1. Operations Research: An Introduction, Hamdy A. Taha, Pearson publication.

Numerical Linear Algebra

Credits- 4 **L: T: P=3-1-0**

Unit-I

Basics: Summary/recap of basic concepts from linear algebra and numerical analysis: matrices, operation counts, Matrix multiplication, block matrices, Matrix norms, Linear system sensitivity.

Unit-II

Matrix factorizations: Matrix factorizations, Cholesky factorization, QR factorization by Householder matrices and by Givens rotations.

LU factorization and Gaussian elimination partial pivoting: Solving triangular systems by substitution, Solving full systems by factorization, Error analysis, Complete pivoting, rook pivoting, Numerical examples.

Unit-III

Sparse and banded linear systems and iterative methods: Storage schemes for banded and sparse matrices, LU factorization, Markowitz pivoting, 2 Iterative methods: Jacobi, Gauss-Seidel and SOR iterations, Kronecker product, Krylov subspace methods, conjugate gradient method, Preconditioning.

Unit-IV

Linear least squares problem: Basic theory using singular value decomposition (SVD) and pseudoinverse, Perturbation theory, Numerical solution: normal equations, SVD and rank deficiency.

Eigenvalue problem: Basic theory, including perturbation results. Power method, inverse iteration, Similarity reduction, QR algorithm.

Texts/References:

1. Timothy A. Davis. Direct Methods for Sparse Linear Systems. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2006.
2. James W. Demmel. Applied Numerical Linear Algebra. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1997.
3. Gene H. Golub and Charles F. Van Loan. Matrix Computations. Johns Hopkins University Press, Baltimore, MD, USA, fourth edition, 2013.
4. Desmond J. Higham and Nicholas J. Higham. MATLAB Guide. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, third edition, 2017.
5. Nicholas J. Higham. Accuracy and Stability of Numerical Algorithms. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, second edition, 2002.
6. Yousef Saad. Iterative Methods for Sparse Linear Systems. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, second edition, 2003.
7. G. W. Stewart. Introduction to Matrix Computations. Academic Press, New York, 1973. ISBN 0-12- 670350-7. xiii+441 pp.
8. G. W. Stewart. Matrix Algorithms. Volume I: Basic Decompositions. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1998.
9. G. W. Stewart. Matrix Algorithms. Volume II: Eigensystems. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2001.

Modern Applied Statistics: Learning

Credits- 4 **L: T: P=3-1-0**

Unit –I

Basic Commands, Graphics , Indexing Data, Loading Data, Less Flexible vs. More Flexible Methods, Training vs. Test Error Rates , Nearest Neighbors Methods , Bayes Classifier , Bias/Variance ideas, Linear Regression Model, Using Least Squares to Fit the Model.

Unit –II

Testing Statistical Significance, Dealing with Categorical Variables, Using the lmage Function to Fit Linear Regression Models in R, Using the Logistic Function for Classification, Estimating Regression Coefficients, Estimating Probabilities, Bayes Theorem for Classification, Estimating the Bayes Classifier, Confusion Matrices , Quadratic Discriminant Analysis, Best Subset Regression.

Unit –III

Leave Out Samples, BIC and AIC, Cross Validation, Illustrations on Real Estate Data, Introduction to Non-Linear Regression, Polynomial Regression , Splines, Illustrations on S&P and Simulated Data Sets, Using the polynomial Function to Implement Polynomial Regression, Fitting Splines Using the smooth,spline Function, Producing a Generalized Additive Model Using the gamma Function. Illustrations on the S&P Data.

Text Books:

1. An Introduction to Statistical Learning with Applications in R, Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani, Springer

DETAILED SYLLABUS OF SEMESTER 3

Introduction to parallel computing using MPI, open MP and CUDA

Credits- 6 **L: T: P=5-1-0**

Unit I

Introduction , From serial to parallel thinking: common gotchas, A history of parallel computers and lessons learned from them. Performance metrics - speedup, utilization, efficiency, scalability. Models of Parallel Computation SIMD, MIMD, PRAM (EREQ,CREW,CRCW), NC. How useful are these models for modern machines.
Parallel Computer Organization Pipelining and Throughput, Latency and Latency hiding Memory Organization Inter-process communication Inter-connection network
Message passing Shared/Distributed memory.

Unit II

Basic Parallel Algorithmic Techniques, Pointer Jumping, Divide-and-Conquer, Partitioning, Pipelining, Accelerated Cascading, Symmetry Breaking, Synchronization (Locked, Lock-free) Parallel Algorithms Data organization for shared/distributed memory Min/Max,Sum Searching, Merging, Sorting, Prefix operations N-body problems, Matrix operations.

Unit III

Writing Parallel Programs, GPU-Compute Architecture, CUDA, Memory organization in CUDA, Multi-Core CPU programming, MPI, PVM, Performance evaluation and scalability.

Text/References :

1. An Introduction to Parallel Algorithms by Joseph Jaja (Addison-Wesley Professional)
2. Introduction to Parallel Computing by Ananth Grama, George Karypis, Vipin Kumar and Anshul Gupta (Pearson)
3. Parallel Programaming in C with MPI and openMP by Michael J Quinn (McGraw Hill)

Stochastic Methods in Engineering

Credits- 4 **L: T: P=3:1:0**

Unit-I

Review of course content. Definitions of probability experiments, sample space, event space and prob. measure, conditional probability, Law of total probability, Bayes' Theorem, independence of events;

Events that occur “infinitely often” and “eventually”, Borel-Cantelli Lemmas, Definition of a random variable (discrete and continuous), distribution of a random variable (cdf and pdf), commonly used random variables, Joint density of two or more random variables and their properties, random vectors, Conditional distribution/density, Bayes’ rule for pdfs, chain rule for densities, Conditional Mean. Covariance, correlation coefficient.

Unit-II

Markov inequality, Chebyshev inequality, and Chernoff bound, Joint moments, covariance matrices. Characteristic function, Moment Generating Function, Probability Generating Function. MMSE Estimation: definition and estimation by a constant; linear estimation.

Unit-III

MMSE Estimation: unconstrained; Orthogonality principle. Convergence of sequence of real numbers.

Convergence of random variables (almost surely, r^th mean, in probability, in distribution).

Law of large numbers (Weak and Strong) and Central Limit Theorem, Convergence of Binomial Distribution to Poisson. Bivariate Normal random variables, Multi-variate Normal Random Variables, PDF, Covariance Matrix, Characteristic Function, and properties. Transformation of Correlated Random variables into Uncorrelated ones.

Unit-IV

Discrete-time Markov Chains,:definitions, examples. Time-homegenous Markov Chains, Transition probability matrix. Recurrence time, transient and recurrent states, classification of states (open, closed).

Period of a state, stationary distributions, irreducible and reducible Markov chains, ergodicity

Unit-V

Random processes: definitions, mean, auto-correlation, and auto-covariance function. First and higher order density of random processes. Independent and Stationary Increments Property. Gaussian random process, Brownian motion.

Counting processes and Poisson Process. Strict Sense Stationarity, Wide Sense Stationarity, Power Spectral Density, Properties, Examples.

Wiener Filtering, and its general solution. Statement of the causal linear Wiener Filtering Problem, Wiener – Hopf equations. Causal functions and spectral factorization.

Texts/ References:

1. Introduction to Probability Models, 11th Ed. by Sheldon M. Ross, Academic Press, 2014.
2. “Engineering Application of Correlation and Spectral Analysis”, second edition, J. S. Bendat & A. G. Piersol, Wiley Interscience
3. “Random Data Analysis and Measurement Techniques”, third edition, J. S. Bendat & A. G. Piersol, Wiley Interscience

4. “Probability Concepts in Engineering Planning and Design: Volume I – Basic Principals”, A. Ang and W. Tang, John Wiley & Sons

Advanced Software Development for Scientists and Engineers

Credits- 4 **L: T: P=3-1-0 Unit-I**

Embedded software and systems engineering: overview, examples and industrial realities, Project Management - Project Planning and Scheduling, Case studies.

Unit-II

United software development process, software process improvement, software economics, software quality, software metrices.

Unit-III

Architecture description language, pattern oriented software architecture, component based development, distributed software architecture, enterprise application integration, model driven architecture.

Unit-IV

UML extension mechanisms, object constraint language, model checking.

Texts/References :

5. I. Jacobson, J. Rumbaugh, G. Booch: The Unified Software Development Process. Addison Wesley. 1999
6. F. Buschmann, R. Meunier, H. Rohnert, P. Sommerlad, M. Stal: Pattern-Oriented Software Architecture. John Wiley. 1996
7. G. Booch, I. Jacobson, J. Rumbaugh: The Unified Modeling Language User Guide. Addison Wesley. 1999
8. J. B. Warmer, A. G. Kleppe: The Object Constraint Language: Precise Modeling With UML Addison Wesley. 1997
9. E. M. Clarke, O. Grumberg, D. Peled: Model Checking. MIT Press. 2000
10. Finkelstein (ed): The Future of Software Engineering. ACM Press. 2000

Parallel Computer Architecture and Programming

Credits- 4 **L:T:P=3-1-0**

UNIT I

INTRODUCTION TO PARALLEL PROCESSING : Basic concepts – types and level of parallelism - classification of parallel architecture – basic parallel techniques - shared memory multiprocessors – distributed memory multicomputer – parallel Random access machine – VLSI complexity model .

PROCESSORS AND MEMORY HIERARCHY : Advanced processor technology – Super scalar and vector processors – Memory hierarchy technology, virtual memory technology – cache memory organization – shared – memory organization.

UNIT II

PIPELINING AND SUPERSCALAR TECHNIQUES : Linear pipeline processors – Nonlinear pipeline processors – Instruction pipeline design –Arithmetic pipeline design – Superscalar pipeline design

PARALLEL AND SCALABLE ARCHITECTURE : Cache coherence and synchronization mechanisms – coherence problem – snoopy bus and directory based protocol - Vector processing principle Vector instruction types – vector access memory schemes - SIMD computer organization - Implementation models - CM2 – architecture latency hiding techniques

UNIT III

MULTITHREADED & DATA FLOW ARCHITECTURE: Principles of Multithreading – issues and solutions – multiple context processors - Scalable and Multithreaded architectures- Stanford Dash multiprocessor - KSR1 - Dataflow computer-static data flow computer -Dynamic data flow computer

UNIT IV

INTRODUCTION TO PARALLEL ALGORITHMS: Parallel programming paradigms; issues in implementing algorithms on parallel computers; Parallel programming with message passing interface; Performance analysis; Scalability analysis; Theoretical models of parallel computation: variants of the PRAM model, interconnection networks, synchronous and asynchronous models. Basic design techniques for parallel algorithms: balanced trees, recursive doubling, divide and conquer, partitioning, pipelining, accelerated cascading, symmetry breaking.

UNIT V

List ranking, the Euler tour technique, tree contraction. Algorithms for searching, merging and sorting. Graph algorithms: Connected Components, Colouring. Parallel algorithms on interconnection networks and other architectures. Algorithms on asynchronous models. Limits to parallelizability. NC-reductions, P-completeness. Parallel algorithms for selected topics like sorting, searching and merging, matrix algebra, graphs, discrete optimization problems and computational geometry.

Texts/References

1. Kai Hwang, “Advanced Computer Architecture”, Parallelism, Scalability, Programmability”, McGraw Hill, 1993.

2. Hwang Briggs, “Computer Architecture and parallel processing”, McGraw Hill, 1984.

3. Dezso sima, Terence Fountain ,Peter Karsuk , “ Advanced Computer Architectures : A design space approach” , Addison Wesley, 1997.

4. D. E. Culler and J. P. Singh with A. Gupta. Parallel Computer Architecture. Morgan- Kaufmann publishers.

5. J. L. Hennessy and D. A. Patterson. Computer Architecture: A Quantitative

Approach. Morgan-Kaufmann publishers.

6.Parallel Programming in C with MPI and OpenMP by M J Quinn

7. Introduction to Parallel Computing by Ananth Grama, George Karypis, Vipin Kumar, and Anshul Gupta.

8. Programming Massively Parallel Processors by D. Kirk and W. Hwu

Advanced Multi-Core Systems

Credits- 4 **L: T: P=3-1-0**

UNIT-I

Introduction to Multi-core Architecture: Motivation for Concurrency in software, Parallel Computing Platforms, Parallel Computing in Microprocessors, Differentiating Multi-core Architectures from Hyper-Threading Technology, Multi-threading on Single-Core versus MultiCore Platforms Understanding Performance, Amdahl's Law, Growing Returns: Gustafson's Law. System Overview of Threading: Defining Threads, System View of Threads, Threading above the Operating System, Threads inside the OS, Threads inside the Hardware, What Happens When a Thread Is Created, Application Programming Models and Threading, Virtual Environment: VMs and Platforms, Runtime Virtualization, System Virtualization.

UNIT-II

Fundamental Concepts of Parallel Programming: Designing for Threads, Task Decomposition, Data Decomposition, Data Flow Decomposition, Implications of Different Decompositions, Challenges You'll Face, Parallel Programming Patterns, A Motivating Problem: Error Diffusion, Analysis of the Error Diffusion Algorithm, An Alternate Approach: Parallel Error Diffusion, Other Alternatives.

UNIT- III

Threading and Parallel Programming Constructs: Synchronization, Critical Sections, Deadlock, Synchronization Primitives, Semaphores, Locks, Condition Variables, Messages, Flow Control- based Concepts, Fence, Barrier, Implementation-dependent Threading Features. Threading APIs : Threading APls for Microsoft Windows, Win32/MFC Thread APls, Threading APls for Microsoft. NET Framework, Creating Threads, Managing Threads, Thread Pools, Thread Synchronization, POSIX Threads, Creating Threads, Managing Threads, Thread Synchronization, Signaling, Compilation and Linking.

UNIT-IV

OpenMP: A Portable Solution for Threading: Challenges in Threading a Loop, Loop-carried Dependence, Data-race Conditions, Managing Shared and Private Data, Loop Scheduling and Portioning, Effective Use of Reductions, Minimizing Threading Overhead, Work-sharing Sections, Performance-oriented Programming, Using Barrier and No wait, Interleaving Singlethread and Multi-thread Execution, Data Copy-in and Copy-out, Protecting Updates of Shared Variables, Intel Task queuing Extension to

OpenMP, OpenMP Library Functions, OpenMP Environment Variables, Compilation, Debugging, performance.

UNIT-V

Solutions to Common Parallel Programming Problems: Too Many Threads, Data Races, Deadlocks, and Live Locks, Deadlock, Heavily Contended Locks, Priority Inversion, Solutions for Heavily Contended Locks, Non-blocking Algorithms, ABA Problem, Cache Line Pingponging, Memory Reclamation Problem, Recommendations, Thread-safe Functions and Libraries, Memory Issues, Bandwidth, Working in the Cache, Memory Contention, Cacherelated Issues, False Sharing, Memory Consistency, Current IA-32 Architecture, Itanium Architecture, High-level Languages, Avoiding Pipeline Stalls on IA-32,Data Organization for High Performance.

Texts/References

1. Multicore Programming , Increased Performance through Software Multi-threading by Shameem Akhter and Jason Roberts , Intel Press , 2006
2. Introduction to Parallel Processing ,Sashi Kumar,PHI
3. Parallel Programming, Wilkinson, Pearson
4. Elements of Parallel Computing, Rajaraman,PHI

DETAILED SYLLABUS OF OPTIONAL 1

Representations and Algorithms for Computational Molecular Biology

Credits- 4 **L: T:P=3-1-0**

Unit-I

Sequence similarity, homology, and alignment, Probabilities and probabilistic models, Pairwise alignment, The scoring model, Alignment algorithms, Dynamic programming with more complex models, Deriving score parameters from alignment data, Markov chains and hidden Markov models (HMMs), Parameter estimation for HMMs, Numerical stability of HMM algorithms, Pairwise alignment using HMMs, Suboptimal alignment.

[14L]

Unit-II

Biological Data in Digital Symbol Sequences, Genomes—Diversity, Size, and Structure, Proteins and Proteomes, Prediction of Molecular Function and Structure, Machine-Learning Foundations: The Probabilistic Framework, Bayesian Modeling, The Cox Jaynes Axioms, Bayesian Inference and Induction, Model Structures: Graphical Models and Other Tricks. **[14L]**

Unit-III

Probabilistic Modeling and Inference, The Simplest Sequence Models, Statistical Mechanics, Machine Learning Algorithms, Dynamic Programming, Markov-Chain Monte-Carlo Methods, Evolutionary and Genetic Algorithms, Neural Networks, Sequence Encoding and Output Interpretation, Sequence Correlations and Neural Networks, Prediction of Protein Secondary Structure, Applications for DNA and RNA Nucleotide Sequences. **[16L]**

Unit-IV

Hidden Markov Models, Prior Information and Initialization, Likelihood and Basic Algorithms, Protein Applications, DNA and RNA Applications, Advantages and Limitations of HMMs, Markov Models and DNA Symmetries, Markov Models and Gene Finders, Hybrid Models and Neural Network Parameterization ofGraphical Models, Bidirectional Recurrent Neural Networks for Protein SecondaryStructure Prediction. **[16L]**

Texts/References:

1. R. Durbin, S. R. Eddy, A. Krogh and G.Mitchison, Biological sequence analysis: Probabilistic models of proteins and nucleic acids, Cambridge University press.
2. P.Baldi and S.Brunak, Bioinformatics:The Machine Learning Approach, The MIT Press Cambridge

Data Driven Medicine

Credits- 4 **L: T: P=3:1:0**

Unit-I

Statistical Relational Learning: Representing Structure and Uncertainty, Representation: First-Order Logic, Uncertainty: Graphical Models, Relational Dependency Networks, Relational Dependency Networks, Learning in SRL Models,

Boosting (Bi-)Directed Relational Models: Boosting RDNs, UW Data Set, IMDB Data Set, Cora Data Set. **[16L]**

Unit-II

Boosting Undirected Relational Models: Functional Gradients for MLNs, Learning Joint Models, UW Data Set, Cora Data Set, IMDB Data Set.

Boosting in the Presence of Missing Data: Structural EM for Relational Functional Gradients, Gradients for Hidden Groundings, Gradients for Observed Groundings, Algorithm for Boosting in Presence of Hidden Data, Empirical Evaluation. **[18L]**

Unit-III

Boosting Statistical Relational Learning in Action: Adaptation to Sequential Decision Making Problems, Relational Imitation Learning, Predicting Cardiovascular Events, Application to a Real HER, Predicting Mild Cognitive Impairment, NLP Applications, NFL Relation Extraction, Temporal Relation Extraction, Weak Supervision. **[22L]**

Texts/References:

1. S. Natarajan, K. Kersting, T. Khot and J. Shavlik, Boosted Statistical Relational Learners: From Benchmarks to Data-Driven Medicine, Springer

Modern Statistics for Modern Biology

Credits- 4 **L: T: P=3:1:0**

Unit-I

Multiple-Hypothesis Testing Strategy, Statistical Hypothesis Testing, Sources of Multiplicity, Multiple-Testing Taxonomy, Single-Step Procedures, Common Gatekeeper Procedure, Tree Gatekeeping Procedure, Procedures Controlling False Discovery Rate, Interpretation of Multiple-Testing Results, The SpringWater Paradox, A Patient's Dilemma with Multiple-Testing. **[12L]**

Unit-II

Pharmaceutical Decision and Game Theory: Pharmaceutical Decisions, Markov Decision Process, Dynamic Programming, Clinical Development Program, R & D Portfolio Optimization, Prescription Drug Marketing Pharmaceutical Games, The Game Concept, Multiple-Player Game, Queuing Game, Cooperative Games, Sequential Game.

Concept of Adaptive Trial Design, Reasons for Adaptive Design, Hypothesis-Based Adaptive Design, Adaptive Design Methods, Evaluation and Analysis of Adaptive Design, Comparison of Adaptive Design Methods. **[18L]**

Unit-III

Introduction to Survival Data Modeling, Maximum Likelihood Method, Frailty Model, First-Hitting-Time Model, Multistage Model, General Framework of Multistage Model, Covariates and Treatment Switching, Competing Risks in Progressive Disease, Longitudinal Multivariate Model. **[12L]**

Unit-IV

Monte Carlo Simulation: Random Number Generation, Inverse c.d.f Method, Acceptance-Rejection Methods, Markov Chain Monte Carlo, Clinical Trial Simulation, Adaptive Trial Simulation, Dynamic Drug Supply, Molecular Design and Simulation, Biological Pathway Simulation.

Bayesian Methods and Applications: Bayesian Paradigm, Bayesian Inference, Hierarchical Model, Bayesian Decision-Making, Bayesian Approach to Multiplicity, Bayesian Computation, Applications of Bayesian Methods, Clinical Trial Design, Bayesian Adaptive Trial, Safety Signal Detection, Meta-Analysis. **[18L]**

Texts/Refernces:

1. Mark Chang,Modern Issues and Methodsin Biostatistics, Springer

Social and Information Network Analysis

Credits- 4 **L: T: P=3:1:0**

Unit-I

Networks, Relations, and Structure: Social Network Analysis in the Social and Behavioral Sciences, the Social Networks Perspective, fundamental Concepts in Network Analysis, Complexity, Descriptive and Statistical Methods, Theory Driven Methods, Levels of Analysis. Social Network Data: Introduction, Structural and Composition Variables, Modes, Affiliation Variables, Boundary Specification and Sampling, Types of Networks, Network Data, Measurement and Collection. **[14L]**

Unit-II

Mathematical Representations of Social Networks: A Single Relation, OMuItiple Relations, Sociometric Notation, Single Relation, Multiple Relations, Graphs, Sub graphs, Dyads, and Triads, Nodal Degree, Density of Graphs and Subgraphs, Walks, Trails, and Paths, Directed Graphs, Signed Graphs and Signed Directed Graphs, Valued Graphs and Valued Directed Graphs, Multigraphs, Relations, Matrices for Graphs, Matrices for Digraphs, Matrices for Valued Graphs, Matrices for Two-Mode Networks, Basic Matrix Operations, Computing Simple Network Properties. **[14L]**

Unit-III

Dyadic and Triadic Methods: Dyads, The Dyad Census, An Index for Mutuality, ⊗A Second Index for Mutuality, Simple Distributions, Simple Distributions on Digraphs, Statistical Analysis of the Number of Arcs, Testing, Estimation, Statistical Analysis of the Number of Mutuals, Random Models and Substantive Hypotheses, The Triad Census, Distribution of a Triad Census, Mean and Variance of a Triad Census, Mean and Variance of Linear Combinations of a Triad Census, Testing Structural Hypotheses, Testing Structural Hypotheses. **[16L]**

Unit-IV

Statistical Analysis of Single Relational Networks: Single Directional Relations, The V-array, Modeling the V-array, Parameters, Attribute Variables, The W-array, The Basic Model with Attribute Variables, Related Models for Further Aggregated Data, Strict Relational Analysis - The V-array, Ordinal Relational Data, ®Single Relations and Two Sets of Actors, Aggregating Dyads for Two-mode Networks, Goodness-of-Fit Statistics for Blockmodels, Stochastic Blockmodels, Application to Special Probability

Functions, Goodness-of-Fit Indices for Stochastic Blockmodels, Stochastic Blockmodel Representations.

[16L]

Texts/References:

2. S. Wasserman and K. Faust, Social network analysis: methods and applications, Cambridge University press.
3. R. Missaoui, T. Abdessalem and M. Latapy (Eds),Trends in Social Network Analysis: Information Propagation, User BehaviorModeling, Forecasting, and Vulnerability Assessment, Springer.
4. Charu C. Aggarwal, Social Network Data Analytics, Springer.

DETAILED SYLLABUS OF OPTIONAL 2

Machine Learning

Credits- 4 **L: T: P=3:1:0**

Unit 1:

Introduction: Overview of machine learning, related areas, applications, software tools, course objectives. Parametric regression: linear regression, polynomial regression, locally weighted regression, numerical optimization, gradient descent, kernel methods.

Unit 2:

Generative learning: Gaussian parameter estimation, maximum likelihood estimation, MAP estimation, Bayesian estimation, bias and variance of estimators, missing and noisy features, nonparametric density estimation, Gaussian discriminant analysis, naive Bayes.

Unit 3:

Discriminative learning: linear discrimination, logistic regression, logit and logistic functions, generalized linear models, softmax regression.
Neural networks: the perceptron algorithm, multilayer perceptrons, back-propagation, nonlinear regression, multiclass discrimination, training procedures, localized network structure, dimensionality reduction interpretation.

Unit 4:

Support vector machines: functional and geometric margins, optimum margin classifier, constrained optimization, Lagrange multipliers, primal/dual problems, KKT conditions, dual of the optimum margin classifier, soft margins, kernels, quadratic programming, SMO algorithm.

Unit 5:

Graphical and sequential models: Bayesian networks, conditional independence, Markov random fields, inference in graphical models, belief propagation, Markov models, hidden Markov models, decoding states from observations, learning HMM parameters.

Unit 6:

Unsupervised learning: K-means clustering, expectation maximization, Gaussian mixture density estimation, mixture of naive Bayes, model selection.

Unit 7:

Dimensionality reduction: feature selection, principal component analysis, linear discriminant analysis, factor analysis, independent component analysis, multidimensional scaling, manifold learning

Texts/References:

1. Elements of Statistical Learning", T. Hastie, R. Tibshirani and J. Friedman, Springer, 2001.
2. Machine Learning", E. Alpaydin, MIT Press, 2010.
3. Pattern Recognition and Machine Learning", C. Bishop, Springer, 2006.
4. Pattern Classification", R. Duda, E. Hart, and D. Stork, Willey-Interscience, 2000.

Computer Graphics and Visualization

Credits- 4 **L: T: P=3:1:0**

Unit 1:

Computer graphics & graphics systems: Overview of computer graphics, representing pictures, preparing, presenting & interacting with pictures for presentations. Visualization & image processing.RGB color model, direct coding, and lookup table. Storage tube graphics display, Raster scan display, 3D viewing devices, Plotters, printers, digitizers, Light pens etc. Active & Passive graphics devices.

Scan conversion: Points & lines, Line drawing algorithms; DDA algorithm, Bresenham's line algorithm, Circle generation algorithm, Ellipse generating algorithm, scan line polygon, fill algorithm, boundary fill algorithm, flood fill algorithm.

Unit 2:

2D transformation & viewing: Basic transformations: translation, rotation, scaling; Matrix representations & homogeneous coordinates, transformations between coordinate systems; reflection shear; Transformation of points, lines, parallel lines, intersecting lines. Viewing pipeline, Window to view port co-ordinate transformation, clipping operations, point clipping, line clipping, clipping circles, polygons & ellipse. Cohen and Sutherland line clipping, Sutherland-Hodgeman Polygon clipping, Cyrus-beck clipping m

Three Dimension Geometry: Introduction, 3D Geometry, Primitives and Transformation, Rotation about an Arbitrary Axis, Parallel Projection, Perspective Projection, Viewing Parameters, and Conversation to View Plan Coordinate, 3D

Unit 3:

Curves: Curve representation, surfaces, designs, Bezier curves, B-spline curves, end conditions for periodic B-spline curves, rational B-spline curves.

Hidden surfaces : Depth comparison, Z-buffer algorithm, Back face detection, BSP tree method, the Painter's algorithm, scan-line algorithm; Hidden line elimination, wire frame methods , fractal - geometry.

Unit 4:

Color & shading models: Light & color model, interpolative shading model, Texture, Introduction to Ray-tracing: Human vision and color, Lighting, Reflection and transmission models

Texts/References:

1. Computer Graphics with Open GL, 4th Edition, Donald D. Hearn, M. Pauline Baker, Warren Carithers, Pearson Education
2. Computer Graphics using OPENGL, Third Edition,F.S. Hill, Pearson Education.
3. Computer Graphics- Principles and Practice, Third Edition,John F. Hughes, Andries Van Dam, James D. Foley, Steven K. Feiner, Addison-Wesley

Mining Massive Data Sets

Credits- 4 **L: T: P=3-1-0**

Unit 1:

Database system architecture: Data Abstraction, Data Independence, Data Definition Language (DDL), Data Manipulation Language (DML).

Data models: Entity-relationship model, network model, relational and object oriented data models, integrity constraints, data manipulation operations.

Unit 2:

Relational query languages: Relational algebra, Tuple and domain relational calculus, SQL3, DDL and DML constructs, Open source and Commercial DBMS - MYSQL, ORACLE, DB2, SQL server.

Relational database design: Domain and data dependency, Armstrong's axioms, Normal forms, Dependency preservation, Lossless design.

Query processing and optimization: Evaluation of relational algebra expressions, Query equivalence, Join strategies, Query optimization algorithms.

Unit 3:

Storage strategies: Indices, B-trees, hashing.

Unit 4:

Transaction processing: Concurrency control, ACID property, Serializability of scheduling, Locking and timestamp based schedulers, Multi-version and optimistic Concurrency Control schemes, Database recovery.

Unit 5:

Database Security: Authentication, Authorization and access control, DAC, MAC and RBAC models, Intrusion detection, SQL injection.

Unit 6:

Advanced topics: Object oriented and object relational databases, Logical databases, Web databases, Distributed databases, Data warehousing and data mining.

Texts/References:

1. Database System Concepts, 6th Edition by Abraham Silberschatz, Henry F. Korth, S. Sudarshan, McGraw-Hill
2. Principles of Database and Knowledge – Base Systems, Vol 1 by J. D. Ullman, Computer Science Press.
3. Fundamentals of Database Systems, 5th Edition by R. Elmasri and S. Navathe, Pearson Education
4. Foundations of Databases, Reprint by Serge Abiteboul, Richard Hull, Victor Vianu, Addison-Wesley

Convolutional Neural Networks for Visual Recognition

Credits- 4 **L: T: P=3-1-0**

Unit 1:

Course Introduction: Computer vision overview, Historical context, Course logistics**,** Image Classification: The data-driven approach, K-nearest neighbor, Linear classification I, Linear classification II, Higher-level representations, image features ,Optimization, stochastic gradient descent

Unit 2:

Introduction to Neural Networks: Backpropagation, Multi-layer Perceptrons, The neural viewpoint Convolutional Neural Networks: History, Convolution and pooling, ConvNets outside vision**,** Training Neural Networks:Activation functions, initialization, dropout, batch normalization, Update rules, ensembles, data augmentation, transfer learning

Unit 3:

Deep Learning Software: Caffe, Torch, Theano, TensorFlow, Keras, PyTorch CNN
Architectures:AlexNet, VGG, GoogLeNet, ResNet etc

Unit 4:

Recurrent Neural Networks: RNN, LSTM, GRU , Language modeling , Image captioning, visual question answering , Soft attention, Detection and Segmentation: Semantic segmentation, Object detection, Instance segmentation

Unit 5:

Visualizing and Understanding:Feature visualization and inversion, Adversarial examples, DeepDream and style transfer

Unit 6:

Generative Models: PixelRNN/CNN, VariationalAutoencoders, Generative Adversarial Networks
Deep Reinforcement Learning: Policy gradients, hard attention, Q-Learning, Actor-Critic

Texts/References:

1. Hertz, John, Anders Krogh, and Richard G. Palmer. *Introduction to the Theory of Neural Computation*. Redwood City, CA: Addison-Wesley Pub. Co., 1991.
2. Bishop, C. M., Neural Networks for Pattern Recognition, Oxford University Press. 1995
3. I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016.

DETAILED SYLLABUS OF OPTIONAL 3

Geostatistics

Credits- 4 **L: T: P=3:1:0**

Unit 1:

Geo-statistical computing: What is geostatistics? The added value of geostatistics, Feature and geographic spaces, Geostatistical computing: inventory of packages, The R Project for Statistical Computing: what and why? Exercise: Introduction to the R environment and S language.

Unit 2:

Exploring and visualizing spatial data, Visualizing spatial structure: point distribution, postplots, quintile plots, Visualizing regional trends, Visualizing local spatial dependence: h-scatterplots, variogram cloud, experimental variogram, Visualizing anisotropy: variogram surfaces, directional variograms.

Unit 3:

Modelling spatial structure from point samples, Trend surfaces, Theory of random elds, Models of spatial covariance, Variogram analysis; variogram modeling.

Unit 4:

Spatial prediction from point samples (Part 1): A taxonomy of spatial prediction methods, Non-geostatistical prediction, Introduction to Ordinary Kriging, Note: the derivation of the kriging equations is deferred to the next lecture.

Unit 5:

Spatial prediction from point samples (Part 2): Derivation of the Ordinary Kriging (OK) system: (1) regression (2) minimization, Simple Kriging (SK), Block Kriging, Universal Kriging (UK), Derivation of the Universal Kriging system: (1) regression (2) minimization, Kriging transformed variables, Kriging with External Drift (KED) and Regression Kriging (RK), Strati ed Kriging (StK).

Unit 6:

Assessing the quality of spatial predictions, Assessment of model quality: overview, Model evaluation with an independent data set, Cross-validation, Kriging prediction variance, spatial simulation.

Unit 7:

Geostatistical risk mapping, Uncertainty, hazard and risk, Indicator variables, Indicator variograms, Probability kriging with indicator variables.

Texts/References:

1. Geostatistics for Natural Resources Evaluation Goovaerts, Pierre. 1997. Oxford University Press.

2. An Introduction to Applied Geostatistics Isaaks, Edward H., and R. Mohan Srivastava.1989. Oxford University Press.
3. Geostatistics: Modeling Spatial Uncertainty Chiles, J.-P. And Delfiner, P. 1999. Wiley & Sons, Inc.

4. Geostatistics for Environmental Scientists" Webster, R, and M.A. Oliver. 2004. Wiley & Sons, Inc.
5. Statistics for Spatial Data. Cressie, Noel A.C. 1993.Wiley & Sons, Inc.
    11. GSLIB Geostatistical Software Library and User's Guide Deutsch, Clayton V. and Andre G. Journel. 1998. Oxford University Press.
6. Contemporary Statistical Models for the Plant and Soil Sciences. Schabenberger, Oliver, and Francis J. Pierce. 2002. Boca Raton : CRC Press

Business Intelligence from Big data

Credits- 4 **L: T: P=3:1:0**

Unit 1:

BI Basics: BI concepts, components & architecture, previewing the future of BI, End User Assumptions, Setting Up Data for BI, The Functional Area of BI Tools, Query Tools and Reporting, OLAP and Advanced Analytics, Supporting the requirements of senior executives, including performance management.

Unit 2:

Elements of Business Intelligence Solutions: Reports & ad hoc queries; Analyze OLAP data; Dashboards & Scorecards development, Metadata Models; Automated tasks & events; Mobile & disconnected BI; Collaboration capabilities; Real time monitoring capabilities; Software development kit;Consume BI through portals, web applications, Desktop applications.

Unit 3:

Building the BI Project: Planning the BI project, Project Resources; Project Tasks, Risk Management and Mitigation, Cost-justifying BI solutions and measuring success, Collecting User Requirements, Requirements-Gathering Techniques; Prioritizing & Validating BI Requirements, Changing Requirements; BI Design and Development, Best Practices for BI Design; Post-Implementation Evaluations, Maintaining Your BI Environment.

Unit 4:

Report authoring: Building reports with relational vs. Multidimensional data models; Types of Reports - List, crosstabs, Statistics, Chart, map, financial etc.; Data Grouping & Sorting, Filtering Reports, Adding Calculations to Reports, Conditional formatting, Adding Summary Lines to Reports. Drill-up, drill-down, drill-through capabilities. Run or schedule report, different output forms – PDF, excel, csv, xml etc.

Unit 5:

BI Deployment, Administration & Security: Centralized Versus Decentralized Architecture, BI Architecture Alternatives, phased & incremental BI roadmap, System Sizing, Measurements, and Dependencies, System Sizing, Measurements, and Dependencies. Setting Early Expectations and Measuring the Results. End-User Provisos. OLAP Implementations. Expanding BI, Authentication, Authorization, Access Permissions, Groups and Roles, Single-sign on, Server Administration, Manage Status & Monitoring, Audit, Mail server & Portal integration, Back Up and Restore.

Texts/References:

1. Business Intelligence by IBM ICE Publications (Will be provided by IBM Team).
2. Rajiv Sabherwal and Irma Becerra-Fernandez, Business Intelligence, Wiley Publications (2010).
3. Swain Scheps, Business Intelligence for Dummies, Wiley Publications (2011).
4. Arshad Khan, Business Intelligence & Data Warehousing Simplified, Mercury learning & information LLC (2012).

Paradigms for Computing with Data

Credits- 4 **L: T: P=3:1:0**

Unit 1:

Introduction:Data Science history, evolution and current state-of-the-art techniques. What is Big Data, and why Big Data processing is different. Evolution of processing systems. Difference between Data Science/ Data Analysis, Data engineering and Big Data.

Unit 2:

Introduction to Data Analytics: Data Analysis process. Investigate a Data Set. Apply NumPy, Pandas to an example dataset using Python, Anaconda, conda and Jupyter notebook will be installed for analysis.

Unit 3:

Data Wrangling: Data Extraction Fundamentals. Data in complex formats: CSV, JSON, XML, SQL Data. Data Acquisition techniques, Data Quality, Identify dirty data. Data curation and storage techniques. NoSQL Databases. Conclude with an example data set investigation.

Unit 4:

Document Indexing and Searching: Introduce Lucene, Solr -indexing/searching techniques used by document search companies. Install Solr and run a few examples on document indexing.

Unit 5:

Distributed / Parallel Computing and methodologies: MapReduce paradigm HDFS - Hadoop Distributed File System Hadoop YARN

Unit 6:

Data Visualization and Notebooks: Zeppelin, Jupyter etc. Study Datasets. Investigate datasets given by the instructor and share your findings with the class. Discuss final project and review assignment solutions. Scattered across semester.

Texts/ References:

1. Rajkumar Buyya, James Broberg, Andrzej M. Goscinski, Cloud Computing Principles and Paradigms, Wiley, 2010

2. Michael Miller, Cloud Computing: Web-Based Applications That Change the Way You Work and Collaborate Online, Que Publishing, 2009.