B.Sc.(Hons)Mathematics,Semester-V,DSE-Courses

DISCIPLINESPECIFICELECTIVECOURSE–3(i):MATHEMATICALDATASCIENCE

CREDITDISTRIBUTION,ELIGIBILITYANDPRE-REQUISITESOFTHECOURSE

Coursetitle&	Credits	Creditdistributionofthecourse			Eligibility	Pre-requisiteof
Code		Lecture	Tutorial	Practical/ Practice	criteria	thecourse (if any)
Mathematical Data Science	4	3	0	1	ClassXIIpass with Mathematics	Basicknowledgeof R/Python DSC-3:Probability & Statistics

LearningObjectives: Themainobjectiveofthiscourseis to:

- Introducevarioustypesofdataandtheirsources,alongwithstepsinvolvedindatascience case-study, including problems with data and their rectification and creation methods.
- Coverdimensionalityreductiontechniques,clusteringalgorithmsandclassificationmethods.

LearningOutcomes: Thecoursewillenablethestudentsto:

- Gain a comprehensive understanding of data science, its mathematical foundations including practical applications of regression, principal component analysis, singular value decomposition, clustering, support vector machines, and *k*-NN classifiers.
- Demonstrate data analysis and exploration, linear regression techniques such as simple, multiple explanatory variables, cross-validation and regularization using R/Python.
- Use real-world datasets to practice dimensionality reduction techniques such as PCA, SVD, and multidimensional scaling using R/Python.

SYLLABUSOFDSE-3(i)

UNIT-I: Principles of Data Science (12 hours)

Types of Data: nominal, ordinal, interval, and ratio; Steps involved in data science case-study: question, procurement, exploration, modeling, and presentation; Structured and unstructureddata:streams,frames,series,surveyresults,scaleandsourceofdata–fixed, variable,highvelocity,exactandimplied/inferred;Overviewofproblemswithdata–dirty and missing data in tabular formats – CSV, data frames in R/Pandas, anomaly detection, assessing data quality, rectification and creation methods, data hygiene, meta-data for inline data-description-markups such as XML and JSON; Overview of other data-source formats – SQL, pdf, Yaml, HDF5, and Vaex.

Unit-II: Mathematical Foundations (15hours)

Model driven data in R^n, Log-likelihoods and MLE, Chebyshev, and Chernoff-Hoeffding inequalitieswithexamples,Importancesampling;NormsinVectorSpaces–Euclidean,and metric choices; Types of distances: Manhattan, Hamming, Mahalanobis, Cosine and angular distances, KL divergence; Distances applied to sets– Jaccard, and edit distances; Modeling text with distances; Linear Regression: Simple, multiple explanatory variables, polynomial, cross-validation, regularized, Lasso, and matching pursuit; Gradient descent.

Unit-III: Dimensionality Reduction, Clustering and Classification (18hours)

Problem of dimensionality, Principal component analysis, Singular value decomposition (SVD),Best*k*-rankapproximationofamatrix,EigenvectorandeigenvaluesrelationtoSVD, Multidimensional scaling, Linear discriminant analysis; Clustering: Voronoi diagrams, Delaunay triangulation, Gonzalez's algorithm for *k*-center clustering, Lloyd's algorithm for *k*-meansclustering,MixtureofGaussians,Hierarchicalclustering,Density-basedclustering and outliers, Mean shift clustering; Classification: Linear classifiers, Perceptron algorithm, Kernels, Support vector machines, and *k*-nearest neighbors (*k*-NN) classifiers.

EssentialReadings

1. Mertz,David.(2021).CleaningData forEffectiveDataScience,PacktPublishing.
2. Ozdemir,Sinan.(2016).Principlesof DataScience,Packt Publishing.
3. Phillips,JeffM.(2021).MathematicalFoundationsforDataAnalysis,Springer. (https://mathfordata.github.io/).

SuggestiveReadings

- FrankEmmert-Streib,etal.(2022).MathematicalFoundationsofDataScienceUsingR. (2nd ed.). De Gruyter Oldenbourg.
- WesMcKinney.(2022).PythonforDataAnalysis(3rded.).O'Reilly.
- Wickham,Hadley, etal.(2023).RforData Science(2nded.). O'Reilly.

Practical(30hours)-Practicalworktobeperformed in ComputerLabusing R/Python:

1. Toexploredifferenttypesdata(nominal,ordinal,interval,ratio)andidentifytheirproperties.
2. Todealwithdirtyandmissingdata,suchasimputation,deletion,anddatanormalization.
3. Usethereal-worlddatasets(https://data.gov.in/)todemonstratethefollowing:
   a) Data analysis and exploration, linear regression techniques such as simple, multiple explanatory variables, cross-validation, and regularization.
   b) Dimensionality reduction techniques such as principal component analysis, singular value decomposition (SVD), and multidimensional scaling.
   c) Clusteringalgorithmssuchas*k*-means,hierarchical,anddensity-basedclusteringand evaluate the quality of the clustering results.
   d) Classificationmethodssuchas linearclassifiers,supportvector machines (SVM),and *k*-nearestneighbors(*k*-NN).

DISCIPLINESPECIFICELECTIVECOURSE–3(ii):LINEARPROGRAMMINGANDAPPLICATIONS

CREDITDISTRIBUTION,ELIGIBILITYANDPRE-REQUISITESOFTHECOURSE

Coursetitle&Code	Credits	Creditdistributionofthecourse			Eligibility criteria	Pre-requisite ofthecourse (if any)
		Lecture	Tutorial	Practical/ Practice		
Linear Programming and Applications	4	3	1	0	ClassXIIpass with Mathematics	DSC-4:Linear Algebra

LearningObjectives: Primaryobjectiveofthiscourseistointroduce:

- SimplexMethodforlinearprogrammingproblems.
- Duallinearprogrammingproblems.
- TheapplicationsoflinearProgrammingtotransportation,assignment,andgametheory.

LearningOutcomes: Thecoursewillenablethestudentsto:

- Learnaboutthebasicfeasiblesolutionsoflinearprogrammingproblems.
- Understandthetheoryofthesimplexmethodtosolvelinearprogrammingproblems.
- Learnabouttherelationshipsbetweentheprimalanddualproblems.
- Solvetransportationandassignmentproblems.
- Understandtwo-personzerosumgame,gameswithmixedstrategiesandformulation of game to primal and dual linear programing problems to solve using duality.

SYLLABUSOFDSE-3(ii)

UNIT– I: Introduction to Linear Programming (12hours)

Linearprogrammingproblem: Standard, Canonical and matrixforms, Geometric solution; Convex and polyhedral sets, Hyperplanes, Extreme points; Basic solutions, Basic feasible solutions, Correspondence between basic feasible solutions and extreme points.

UNIT–II:Optimalityand DualityTheoryof Linear ProgrammingProblem(18 hours) Simplex method: Optimal solution, Termination criteria for optimal solution of the linear programming problem, Unique and alternate optimal solutions, Unboundedness; Simplex algorithm and its tableau format; Artificial variables, Two-phasemethod, Big-Mmethod. Duality Theory: Motivation and formulation of dual problem, Primal-Dual relationships, Fundamentaltheoremofduality;Complementaryslackness.

UNIT–III: Applications (15hours)

Transportation Problem: Definition and formulation, Northwest-corner,Least-cost, and Vogel's approximation methods of finding initial basic feasible solutions;Algorithm for solvingtransportationproblem.

AssignmentProblem:Mathematicalformulation and Hungarian method of solving. Game Theory: Two-person zero sum game, Games with mixed strategies, Formulation of game to primal and dual linear programming problems, Solution of games using duality.

EssentialReadings

1. Bazaraa,MokhtarS.,Jarvis,JohnJ.,&Sherali,HanifD.(2010).LinearProgrammingand Network Flows (4th ed.). John Wiley and Sons. Indian Reprint.
2. Hillier,FrederickS.&Lieberman,GeraldJ.(2021).IntroductiontoOperationsResearch (11th ed.). McGraw-Hill Education (India) Pvt. Ltd.
3. Taha,HamdyA.(2017).OperationsResearch:AnIntroduction(10thed.).Pearson.

SuggestiveReadings

- Hadley,G.(1997).LinearProgramming.NarosaPublishingHouse.NewDelhi.
- Thie,PaulR.,&Keough,G.E.(2008).AnIntroductiontoLinearProgrammingandGame Theory. (3rd ed.). Wiley India Pvt. Ltd. Indian Reprint 2014.

DISCIPLINE SPECIFICELECTIVECOURSE–3(iii):MATHEMATICALSTATISTICS

CREDITDISTRIBUTION,ELIGIBILITYANDPRE-REQUISITESOFTHECOURSE

Coursetitle&Code	Credits	Creditdistributionofthecourse			Eligibility criteria	Pre-requisiteof the course (ifany)
		Lecture	Tutorial	Practical/		

				Practice		
Mathematical Statistics	**4**	**3**	**1**	**0**	**ClassXIIpass with Mathematics**	**DSC-3:Probability & Statistics DSC-11: Multivariate Calculus**

LearningObjectives: Themain objectiveofthiscourseistointroduce:

- The joint behavior of several random variables theoretically and through illustrative practical examples.
- Thetheoryunderlyingmodernstatisticstogivethestudentasolidgroundingin (mathematical) statistics and the principles of statistical inference.
- Theapplicationofthetheorytothestatisticalmodelingofdatafromrealapplications, including model identification, estimation, and interpretation.
- Theideaof Fisherinformationtofindtheminimum possiblevariancefor an unbiased estimator, and to show that the MLE is asymptotically unbiased and normal.

LearningOutcomes: Thecoursewillenablethestudentsto:

- Understandjointdistributionsofrandomvariablesincludingthebivariatenormal distribution.
- Estimate model parameters from the statistical inference based on point estimation and hypothesis testing.
- Apply Rao-Blackwell theorem for improving an estimator, and Cramér-Rao inequality to find lower bound on the variance of unbiased estimators of a parameter.
- Understandthetheoryof linearregressionmodelsandcontingency tables.

SYLLABUSOFDSE-3(iii)

UNIT–I: Joint Probability Distributions (15hours)

Joint probability mass function for two discrete random variables, Marginal probability mass function, Joint probability density function for two continuous random variables, Marginal probability density function, Independent random variables; Expected values, covariance, and correlation; Linear combination of random variables and their moment generating functions; Conditional distributions and conditional expectation, Laws of total expectation and variance; Bivariate normal distribution.

UNIT-II: Sampling Distributions and Point Estimation (15hours)

Distribution of important statistics such as the sample totals, sample means, and sample proportions,Centrallimittheorem,Lawoflargenumbers;Chi-squared,*t*,and*F*distributions; Distributions based on normal random samples; Concepts and criteria for point estimation, The methods of moments and maximum likelihood estimation (MLE); Assessing estimators: Accuracy and precision, Unbiased estimation, Consistency and sufficiency, The Neyman factorization theorem, Rao-Blackwell theorem, Fisher Information, The Cramér-Rao inequality, Efficiency,

UNIT-III:Confidence Intervals,Tests ofHypotheses andLinear RegressionAnalysis(15hours)

Interval estimation and basic properties of confidence intervals, One-sample *t* confidence

interval, Confidence intervals for a population proportion and population variance. Statisticalhypothesesandtestprocedures,One-sampletestsaboutapopulationmeanand a population proportion, *P*-values for tests; The simple linear regression model and its estimating parameters; Chi-squared goodness-of-fit tests, Two-way contingency tables.

EssentialReading

1. Devore, JayL., Berk, Kenneth N. &Carlton Matthew A. (2021).ModernMathematical Statistics with Applications. (3rd ed.). Springer.

SuggestiveReadings

- Devore,JayL.(2016).ProbabilityandStatisticsforEngineeringandtheSciences.Ninth edition, Cengage Learning India Private Limited, Delhi. Fourth impression 2022.
- Hogg,RobertV.,McKean,JosephW.,&Craig,AllenT.(2019).Introductionto Mathematical Statistics. Eighth edition, Pearson. Indian Reprint 2020.
- Mood, A.M., Graybill, F.A., & Boes, D.C. (1974). Introduction the Theory of Statistics (3rd ed.). Tata McGraw Hill Pub. Co. Ltd. Reprinted 2017.
- Wackerly,DennisD.,MendenhallIII,William&Scheaffer,RichardL.(2008). Mathematical Statistics with Applications. 7th edition, Cengage Learning.