MSc in Computer Science

Department of Computer Science

RKMVERI, Belur Campus

Program Outcomes

Program Specific Outcomes

Course Outcomes

DA210
Advanced Statistics

Time: TBA
Place: IH402 & Bhaskara Lab

Instructor: TBA

Course Description:
DA*** introduce the conceptual foundations of statistical methods and how to apply them to address more advanced statistical question. The goal of the course is to teach students how one can effectively use data and statistical methods to make evidence based business decisions. Statistical analyses will be performed using R and Excel.

Prerequisite(s): NA
Note(s): Syllabus changes yearly and may be modified during the term itself, depending on the circumstances. However, students will be evaluated only on the basis of topics covered in the course.
Course url:
Credit Hours: 4

Text(s):

Statistical Inference;
P. J. Bickel and K. A. Docksum

Introduction to Linear Regression Analysis;
Douglas C. Montgomery

Course Objectives:

Knowledge acquired: Students will get to know
(1) advance statistical concepts and some of their basic applications in real world,
(2) the appropriate statistical analysis technique for a business problem,
(3) the appropriateness of statistical analyses, results, and inferences , and,
(4) advance data analysis in R.

Skills gained: The students will be able to
(1) use data to make evidence based decisions that are technically perfect,
(2) communicate the purposes of the data analyses,
(3) interpret the findings from the data analysis, and the implications of those findings,
(4) implement the statistical method using R and Excel.

Course Outline (tentative) and Syllabus:
The weekly coverage might change as it depends on the progress of the class. However, you must keep up with the reading assignments. Each week assumes 4 hour lectures. Quizzes will be unannounced.

Week	**Content**
Week 1	Point Estimation, Method of moments, Likelihood function, Maximum likelihood equations, Unbiased estimator
Week 2	Mean square error, Minimum variance unbiased estimator, Consistent estimator, Efficiency
Week 3	Uniformly minimum variance unbiased estimator, Efficient estimator, Sufficient estimator, Jointly sufficient Minimal sufficient statistic
Week 4	Interval Estimation, Large Sample Confidence Intervals: One Sample Case
Week 5	Small Sample Confidence Intervals for μ, Confidence Interval for the Population Variance, Confidence Interval Concerning Two Population Parameters
Week 6	Type of Hypotheses, Two types of errors, The level of significance, The p-value or attained significance level,
Week 7	The NeymanPearson Lemma, Likelihood Ratio Tests, Parametric tests for equality of means and variances.
Week 8	Problem Session, Review for Midterm exam
Week 9	Linear Model, Gauss Markov Model
Week 10	Inferences on the Least-Squares Estimators
Week 11	Analysis of variance.
Week 12	Multiple linear regressionn Matrix Notation for Linear Regression
Week 13	Regression Diagnostics, Forward, backward and stepwise regression,
Week 14	Logistic Regression.
Week 15	Problem Session, Review for Final Exam

DA330
Advanced Machine Learning

Tanmay Basu

Email: welcometanmay@gmail.com
URL: https://www.researchgate.net/profile/Tanmay_Basu
Office: IH 405, Prajna Bhavan, RKMVERI, Belur, West Bengal, 711 202
Office Hours: 11 pm-5 pm
Phone: (+91)33 2654 9999

Course Description: DA330 deals with topics in supervised and unsupervised learning methodologies. In particular, the course will cover different advanced models of data classification and clustering techniques, their merits and limitations, different use cases and applications of these methods. Moreover, different advanced unsupervised and supervised feature engineering schemes to improve the performance of the learning techniques will be discussed.

Prerequisite(s): (1) Machine Learning, (2) Linear Algebra and (3) Basic Statistics.
Note(s): Syllabus changes yearly and may be modified during the term itself, depending on the circumstances. However, students will be evaluated only on the basis of topics covered in the course.
Course URL:
Credit Hours: 4

Text(s):

Introduction to Machine Learning
E. Alpaydin
ISBN: 978-0262-32573-8

The Elements of Statistical Learning
J. H. Friedman, R. Tibshirani, and T. Hastie
ISBN: 978-0387-84884-6

Neural Networks and Learning Machines
S. Haykin
ISBN: 978-0-13-14713-99

Deep Learning
I. Goodfellow, Y. Bengio and A. Courville
ISBN: 978-0262-03561-3

Pattern Recognition and Machine Learning
C. M. Bishop
ISBN: 978-0387-31073-2

Probabilistic Graphical Models: Principles and Techniques
D. Koller and N. Friedman
ISBN: 978-0262-01319-2

Introduction to Information Retrieval
C. D. Manning, P. Raghavan and H. Schutze
ISBN: 978-0-521-86571-5

Course Objectives:

Knowledge acquired: (1) Different advanced models of learning techniques,
(2) their merits and limitations, and,
(3) applications.

Skills gained: The students will be able to
(1) analyze complex characteristics of different types of data,
(2) knowledge discovery from high dimensional and large volume of data efficiently, and,
(3) creating advanced machine learning tools for data analysis.

Grade Distribution:

Assignments 50%, Midterm Exam 20%, Endterm Exam 30%

Course Outline (tentative) and Syllabus:

The weekly coverage might change as it depends on the progress of the class. However, you must keep up with the reading assignments. Each week assumes 4 hour lectures.

Week	Contents
Week 1	• Overview of machine learning: concept of supervised and unsupervised learning • Decision tree classification: C4.5 algorithm
Week 2	• Random forest classifier • Discussion on overfitting of data. Boosting and bagging techniques
Week 3	• Non linear support vector machine (SVM): Method and Applications • Detailed discussion on SVM using kernels
Week 4	• Neural network: overview, XOR problem, two layer perceptrons • Architecture of multilayer feedforward network
Week 5	• Backpropagation algorithm for multilayer neural networks • Neural network using radial basis function: method and applications
Week 6	• Design and analysis of recurrent neural networks • Deep learning: a case study
Week 7	• Assignment 1: design of efficient neural networks for large and complex data of interest • Overview of data clustering and expectation maximization method
Week 8	• Spectral clustering method • Non negative matrix factorization for data clustering • Review for midterm exam
Week 9	• Fuzzy c-means clustering technique • Overview of recommender systems
Week 10	• Different types of recommender systems and their applications • Probabilistic graphical model: an overview
Week 11	• Learning in Bayesian networks • Markov random fields
Week 12	• Hidden markov model: methods and applications • Temporal data mining
Week 13	• Conditional random fields (CRF) • Overview of named entity recognition (NER) in text: A case study
Week 14	• Named entity recognition: Inherent vs contextual features, rule based method • Rule based text mining using regular expressions
Week 15	• Gazetteer based and CRF based method for NER • Assignment 2: Automatic de-identification of protected information from clinical notes • Review for endterm exam

CS312
Approximation and Online Algorithms

Instructor
Prof. Subir Kumar Ghosh

Prerequisite(s): CS241: DAA

Credit Hours: 4

Text(s):

1. M. R. Garey and D. S. Johnson, Computers and Intractibility: A guide to the theory of NP-completeness, W. H. Freeman, 1979.
2. R. Motwani, Lecture Notes on Approximation Algorithms, Volume 1, No. STAN-CS-92-1435, Stanford University, 1992.
3. D. P. Williamson and D. B. Shmoys, The Design of Approximation Algorithms, Cambridge University Press, 2011.
4. Vijay Vazirani, Approximation algorithms, Springer-Verlag, 2001.
5. S. Albers, Competitive Online Algorithms, Lecture notes, Max Plank Institute, Saarbrucken, 1996.
6. S. K. Ghosh and R. Klein, Online algorithms for searching and exploration in the plane, Computer Science Review, vol. 4, pp. 189-201, 2010.

Course Outline (tentative) and Syllabus:

The weekly coverage might change as it depends on the progress of the class. However, you must keep up with the reading assignments. Each week assumes 4 hour lectures.

Approximation Algorithm:
Performance Measure, Greedy Algorithm, Unweighted Vertex Cover Problem
Minumum-Degree Spanning Tree, Minimum Weight Spanning Tree, The Traveling-Salesman Problem,The k-Center Problem, Multiway Cut and K-Cut Problems, Scheduling Jobs with Deadlines on a Single Machine, Scheduling Jobs on Identical Parallel Machines, The Set Cover Problem, An Application of Set Cover to Art Gallery problems, Shortest Superstring Problem
Rounding Data and Dynamic Programming, The Knapsack Problem, The Bin-Packing Problem,
The Primal-Dual Method, Weighted Vertex Cover Problem

Online Algorithms:
Competitive Analysis, The Paging Problem, Amortized Analysis, List Update Problem, Scheduling Jobs on Identical Parallel Machines, Graph Colouring, Machine Learning, K-Server Problem, Target Searching in an Unbounded Region and Target Searching in Streets

Syllabus for the Computer Architecture Course

Class.no Course Materials to be taught

1	Fundamental Concepts and ISA The von Neumann Model Von Neumann vs Dataflow ISA vs. Microarchitecture
2	ISA Tradeoffs -I
3	ISA Tradeoffs -II
4	Intro to Microarchitecture: Single-Cycle
5	Multi-Cycle and Microprogrammed Microarchitectures
6	Pipelining
7	Introduction to Verilog
8	LAB
9	Branch Prediction I
10	Introduction to TEJAS simulatorr
11	LAB
12	Branch Prediction II
13	Out-of-Order Execution
14	Memory Hierarchy and Caches
15	High Performance Caches
16	Virtual Memory

- Few Homework and Lab assignment are also included.
- Few topics after *pipelining* are very intense , so it may be that if students are not very comfortable ,due to limited class and time , we may trim down the syllabus.

CS301

Theory of Computation

Time: Wed & Fri (12 noon—2 pm)
Place: MB215

Sarvottamananda

`sarvottamananda@rkmvu.ac.in, sarvottamananda@gmail.com`
url: `http://cs.rkmvu.ac.in/s̃arvottamananda/`
Office: MB115, Medhabhavan, RKMVERI, Belur
Office Hours: 11 pm—12 noon, 4 pm—5 pm
`(+91) 98740 94516`

Course Description: CS301 deals with topics in computability theory and computational complexity. In particular, the course will cover different models of computation, their associated complexity classes, undecidability, intractability, space and time complexity classes, oracle turing machines, circuit complexity, and related topics.

Prerequisite(s): (1) Discrete Mathematics and (2) Automata Theory.
Note(s): Syllabus changes yearly and may be modified during the term itself, depending on the circumstances. However, students will be evaluated only on the basis of topics covered in the course.
Course url: `http://cs.rkmvu.ac.in/courses/cs301/`
Credit Hours: 4

Text(s):

Introduction to Automata Theory, Languages, and Computation, third edition
John E. Hopcroft, Rajeev Motwani & Jeffery D. Ullman
ISBN-13: 978-8131720479

Introduction to the Theory of Computation, second edition
Michael Sipser
ISBN-10: 81-315-0162-0

Computational Complexity: A Modern Approach, first edition
Sanjeev Arora & Boaz Barak
ISBN-13: 978-0-521-42426-4

Course Objectives:

Knowledge acquired: (1) Different models of computation,
(2) their associated complexity classes, and,
(3) reducibility.
Skills gained: The students will be able to
(1) analyze the complexity classes of problems closely related to those discussed in the class,
(2) analyze intractability and undecidability of some practical problems, and,
(3) do reductions based on knowledge gained in the class.

Grade Distribution:

Assignments	20%
Quizzes	20%
Midterm Exam	20%
Final Exam	40%

Grading Policy: There will be relative grading such that the cutoff for A grade will not be less than 75% and cutoff for F grade will not be more than 34.9%. Grade distribution will follow normal bell curve (usually, A: $\geq \mu + 3\sigma/2$, B: $\mu + \sigma/2 \ldots \mu + 3\sigma/2$ C: $\mu - \sigma/2 \ldots \mu + \sigma/2$, D: $\mu - 3\sigma/2 \ldots \mu - \sigma/2$, and F: $< \mu - 3\sigma/2$)

Approximate grade assignments:

>= 90.0	A+
75.0 – 89.9	A
60.0 – 74.9	B
50.0 – 59.9	C
about 35.0 – 49.9	D
<= 34.9	F

Course Policies:

- **General**

  1. Computing devices are not to be used during any exams unless instructed to do so.
  2. Quizzes and exams are closed books and closed notes.
  3. Quizzes are unannounced but they are frequently held after a topic has been covered.
  4. **No makeup quizzes or exams will be given.**

- **Grades**

  Grades in the **C** range represent performance that **meets expectations**; Grades in the **B** range represent performance that is **substantially better** than the expectations; Grades in the **A** range represent work that is **excellent**.

- **Labs and Assignments**

  1. Students are expected to work independently. **Offering** and **accepting** solutions from others is an act of dishonesty and students can be penalized according to the *Academic Honesty Policy*. Discussion amongst students is encouraged, but when in doubt, direct your questions to the professor, tutor, or lab assistant. Many students find it helpful to consult their peers while doing assignments. This practice is legitimate and to be expected. However, it is not acceptable practice to pool thoughts and produce common answers. To avoid this situation, it is suggested that students not write anything down during such talks, but keep mental notes for later development of their own.
  2. **No late assignments will be accepted under any circumstances**.

- **Attendance and Absences**

  1. Attendance is expected and will be taken each class. Students are not supposed to miss class without prior notice/permission. Any absences may result in point and/or grade deductions.
  2. Students are responsible for all missed work, regardless of the reason for absence. It is also the absentee's responsibility to get all missing notes or materials.

Course Outline (tentative) and Syllabus:

The weekly coverage might change as it depends on the progress of the class. However, you must keep up with the reading assignments. Each week assumes 4 hour lectures. Quizzes will be unannounced.

Week	**Content**
Week 1	• Finite Automata: Basic definitions, equivalence of finite automata, mille and muller automata, definition and acceptance criteria of timed and hybrid automata • Reading assignment: Chapter 2, HMU
Week 2	• Regular Expressions and Languages: definition of regular expressions and regular languages, relationship with finite automata, regulation expression algebra • Reading assignment: Chapter 3, HMU • Home assignment 1 • Quiz 1
Week 3	• Properties of Regular Languages: Pumping lemma for regular languages, Myhill-Nerode theorem and minimization of finite automata, closure properties of regular languages, decision problems and algorithms for regular languages • Reading assignment: Chapter 4: HMU
Week 4	• Context Free Grammar and Languages: Definition of context free grammars and context free languages, parse trees, ambiguity in grammars and inherent ambiguity in languages, context sensitive languages • Reading assignment: Chapter 5, HMU
Week 5	• Pushdown Automata: Definition of pushdown automata, languages of pushdown automata, equivalence of pushdown automata and context free grammars, deterministic pushdown automata and its language class • Reading assignment: Chapter 6, HMU • Home assignment 2 • Quiz 2
Week 6	• Properties of Context Free Languages: Normal forms, pumping lemma for context free languages, closure properties of context free languages, decision properties of context free languages • Reading assignment: Chapter 7, HMU
Week 7	• Turing Machines: Halting problem, definition of Turing machines, its extensions, restrictions and their equivalences, linear bounded automata and its relationship with context sensitive languages • Reading assignment: Chapter 8, HMU • Home assignment 3 • Quiz 3
Week 8	• Undecidability and Intractability: class of recursive languages and recursively enumerable languages, non recursively enumerable diagonalization language L_d, undecidable recursively enumerable language L_u, Rice's theorem, Post's correspondence problem, reductions, classes P and NP, NP-complete problem 3SAT • Reading Assignment: Chapter 9 & 10, HMU • Review for Midterm Exam

Week	Content
Week 9	• Computational Model for space and time complexity classes: Defining Turing Machine model, efficiency and running time, machine representation, universal turing machine, efficient simulation of universal turing machine, class P • Reading assignment: Chapter 1, AB • Home assignment 4
Week 10	• NP and NP-completeness: Definition of class NP, reducibility, NP-completeness, Cook-Levin theorem, web of reductions, definitions of coNP, EXP, and NEXP • Reading assignment: Chapter 2, AB • Home assignment 5 • Quiz 4
Week 11	• Diagonalization: Time hierarchy theorem, nondeterministic time hierarchy theorem, Ladner's theorem, oracle machines and limits of diagonalization, Baker-Gill-Solovay theorem • Reading assignment: Chapter 3, AB • Home assignment 6
Week 12	• Space complexity: Space hierarchy thorems, class PSPACE, PSPACE-completeness, class NL and coNL, NL-completeness, NL=coNL, Savitch'e theorem • Reading assignment: Chapter 4, AB • Home assignment 7 • Quiz 5
Week 13	• Polynomial Hierarchy and Alternations: Class Σ_2^p, polynomial hierarchy, class PH, alternating Turing machines • Reading assignment: Chapter 5, AB • Home assignment 8
Week 14	• Boolean Circuits: Boolean circuits, class $P_{/poly}$, uniformly generated circuits, Turing machines with advice, circuit lower bounds, nonuniform hierarchy theorem, circuits of exponential size • Reading assignment: Chapter 6, AB
Week 15	• Randomized Computation: Probabilistic Turing machines, classes RP, coRP, ZPP, BPP, relationships between BPP and other classes, randomized reductions, randomized space bound computations • Reading assignment: Chapter 7, AB • Review for Final Exam

DA102
Basic Statistics

Time: TBA
Place: IH402 & Bhaskara Lab

Dr. Sudipta Das

jusudipta@gmail.com
Office: IH404, Prajnabhavan, RKMVERI, Belur
Office Hours: 11 pm—12 noon, 3 pm—4 pm
(+91) 99039 73750

Course Description:
DA102 is going to provide an introduction to some basic statistical methods for analysis of categorical and continuous data. Students will also learn to make practical use of the statistical computer package R.

Prerequisite(s): NA
Note(s): Syllabus changes yearly and may be modified during the term itself, depending on the circumstances. However, students will be evaluated only on the basis of topics covered in the course.
Course url:
Credit Hours: 4

Text(s):

Statistics;
David Freedman, Pobert Pisani and Roger Purves

The visual display of Quantitative Information;
Edward Tufte

Mathematical Statistics with Applications;
Kandethody M. Ramachandran and ChrisP.Tsokos

Course Objectives:

Knowledge acquired: Students will get to know
(1) fundamental statistical concepts and some of their basic applications in real world.
(2) organizing, managing, and presenting data,
(3) how to use a wide variety of specific statistical methods, and,
(4) computer programming in R.

Skills gained: The students will be able to
(1) apply technologies in organizing different types of data,
(2) present results effectively by making appropriate displays, summaries, and tables of data,
(3) perform simple statistical analyses using R
(4) analyze the data and come up with correct interpretations and relevant conclusions.

Course Outline (tentative) and Syllabus:

The weekly coverage might change as it depends on the progress of the class. However, you must keep up with the reading assignments. Each week assumes 4 hour lectures. Quizzes will be unannounced.

Week	**Content**
Week 1	Introduction, Types of Data, Data Collection, Introduction to R, R fundamentals, Arithmetic with R
Week 2	Tabular Representation: Frequency Tables, Numerical Data Handling, Vectors, Matrices, Categorical Data Handling
Week 3	Data frames, Lists, R programming, Conditionals and Control Flow, Loops, Functions
Week 4	Graphical Representation: Bar diagram, Pie-chart, Histogram, Data Visualization in R, Basis R graphics, Different plot types, Plot customizations
Week 5	Descriptive Numerical Measures:- Measures of Central Tendency, Measures of Variability, Measure of Skewness, Kurtosis Quiz 1
Week 6	Descriptive Statistics using R:- Exploring Categorical Data, Exploring Numerical Data
Week 7	Numerical Summaries, Box and Whiskers Plot
Week 8	Problem Session, Review for Midterm exam
Week 9	Concept of sample and population, Empirical distribution, Fitting probability distribution
Week 10	Goodness of fit, Distribution fitting in R
Week 11	Analysis of bivariate data:- Correlation, Scatter plot Representing bivariate data in R
Week 12	Simple linear regression
Week 13	Linear Regression in R Quiz 2
Week 14	Two-way contingency tables, Measures of association, Testing for dependence
Week 15	Problem Session, Review for Final Exam

DA101

Computing for Data Science

Time: TBA

Place: MB212 / Vijnana Computing Lab

Instructor: Dhyanagamyananda

`dhyangamyananda@gmail.ac.in, swathyprabhu@gmail.com`
url: `http://cs.rkmvu.ac.in/s̃wat/`
Office: MB205, Medhabhavan, RKMVERI, Belur
Office Hours: 10 pm—12 noon, 3 pm—5 pm
`(+91) 033-2654 9999`

Course Description: DA101 is an introductory course in Data Science giving an overview of programming, and computing techniques. This course is specially designed for students of Mathematics, Physics, and Statistics.

Prerequisite(s): (1) Basic logic and mathematics.
Note(s): Syllabus changes yearly and may be modified during the term itself, depending on the circumstances. However, students will be evaluated only on the basis of topics covered in the course.
Moodle url: `http://moodle.rkmvu.ac.in/course/view.php?id=58`
Credit Hours: 4

Text(s):

Algorithms in Data Science, First edition
Brian Steele, John Chandler, & Swarna Reddy

How to proram in Python
Louden & Louden

How to proram in Java
Louden & Louden

Relevant Internet resources

Course Objectives:

Knowledge acquired: .

(1) Turing machine model of computing.
(2) Computer programming in python and java.
(3) Algorithm design and analysis
(4) Simulation.

Skills gained: The students will be able to

1. distinguish between computing and non-computing tasks.
2. read and understand a program written in Python, and Java.
3. represent basic data as data structures suited to computing.
4. break down a computing problem into individual steps and code them in python or java.
5. measure the performance and efficiency of an algorithm in terms of time and space complexity.
6. understand graph theoritical concepts applied to algorithm.
7. interact with relational database using sql.
8. use simulation techniques in solving computational problems.

Grade Distribution:

Assignments	20%
Quizzes	10%
Midterm Exam	20%
Final Exam	40%

Grading Policy: There will be relative grading such that the cutoff for A grade will not be less than 75% and cutoff for F grade will not be more than 34.9%. Grade distribution will follow normal bell curve (usually, A: $\geq \mu + 3\sigma/2$, B: $\mu + \sigma/2 \ldots \mu + 3\sigma/2$ C: $\mu - \sigma/2 \ldots \mu + \sigma/2$, D: $\mu - 3\sigma/2 \ldots \mu - \sigma/2$, and F: $< \mu - 3\sigma/2$)

Approximate grade assignments:

>= 90.0	A+
75.0 – 89.9	A
60.0 – 74.9	B
50.0 – 59.9	C
about 35.0 – 49.9	D
<= 34.9	F

Course Policies:

- **General course policies, Grades, Labs and assignments, Attendance and Absences** These clauses are common to all courses. And it can be found in the program schedule.

Course Outline (tentative) and Syllabus:

The weekly coverage might change as it depends on the progress of the class. However, you must keep up with the reading assignments. Each week assumes 4 hour lectures. Quizzes will be unannounced.

Week	Content
Week 1	• Definition of computing, Binary representation of numbers intergers, floating point, text. • Reading assignment:
Week 2	• Unconventional / application specific file formats, like media. Bitmap representation for monochromatic image and generalizing the representation for RGB. File metadata, Speed of CPU, Memory, Secondary storage, DMA. Hardisk organization into Cylinder, Track, and Sectors for storing data. • Reading assignment: XBitmap from Wiki. • Programming assignment 1: • Quiz 1
Week 3	• Using and understanding the basics of Linux. • Lab activity.
Week 4	• Learning programming using Python. arrays([], [][]), conditional structures (if), and iterative structures (while, for), defining functions, using library functions. • Programming assignment:
Week 5	• Dictionary data structure in python, File access in python, Sorting and Searching algorithms, appreciating complexity of algorithms. Programming using numerical methods. • Programming assignment: • Quiz 2
Week 6	• Basics of Turing machine as a model of computing, analysing the performance of a program, time complexity, space complexity, difference between efficiency and performance, Analyse the first sorting algorithm. • Home assignment:
Week 7	• Basic notations of complexity like Big Oh, omega etc, and their mathematical definitions, given a programme to compute the complexity measures. • Reading assignment: Chapter 2.4, BJS • Home assignment: • Quiz 3
Week 8	• Discussion on the reading assignment, and implementing in the lab. • Review for Midterm Exam

Week	Content
Week 9,10,11	• Programming in SQL (Structured query language) to query relational databases. • Home assignment 4 • Quiz at the end of three weeks.
Week 12	• Representation of graphs, basic algorithms like minimum spanning tree, matching etc. • Home assignment 7 • Quiz 5
Week 13	• Monte-Carlo simulation • Reading assignment: • Home assignment 8
Week 14,15,16	• Object oriented programming using Java

CS123

Concepts of Programming Languages

Time: TBA

Place: MB212

Instructor: Dhyanagamyananda

`dhyangamyananda@gmail.ac.in, swathyprabhu@gmail.com`
url: `http://cs.rkmvu.ac.in/s̃wat/`
Office: MB205, Medhabhavan, RKMVERI, Belur
Office Hours: 10 pm—12 noon, 3 pm—5 pm
`(+91) 033-2654 9999`

Course Description: CS123 deals with analysing the relevance, benefit, and limitations of various features that have been implemented in important and widely used programming languages. It introduces the student to various programming paradigms. With C programming language as a case study, the student is introduced to the different stages in compilation, namely Lexical analysis, Semantic Analysis, and intermediate code generation.

Prerequisite(s): (1) Good working knowledge of C, and C++/Java
Note(s): Syllabus changes yearly and may be modified during the term itself, depending on the circumstances. However, students will be evaluated only on the basis of topics covered in the course.
Moodle url: `http://moodle.rkmvu.ac.in/course/view.php?id=58`
Credit Hours: 4

Text(s):

Principles of programming languages, third edition
Kenneth Louden

Understanding Programming Languages -ebook
M. Ben-Ari

The anatomy of programming languages
Alice. E. Fisher, & Frances. S. Grodzinsky

Compilers: Principles, Techniques, & Tools
Aho, Lam, Seith, & Ullman

Course Objectives:

Knowledge acquired: (1) Different models of computation,
(2) their associated complexity classes, and,
(3) reducibility.

Skills gained: The students will be able to

1. classify different langauges based on the programming paradigms, like imperative, functional, logic, procedural, object oriented, declarative.
2. critically analyse the programming language design criterion like readability, writeability, orthogonality, generality etc.
3. differentiate between the syntactic and semantic notions of programming languages.
4. discern the relative merit and demerit in the choice of programming language to solve a given computing problem.
5. explain equivalence checking, conversion, polymorphism for PL Data types.
6. conceptualize the PL Procedure environments, activations and allocations.
7. understand how memory is dynamically managed, and exception handling is implemented.
8. understand the differences among operational semantics, denotational semantics, and axiomatic semantics.

Grade Distribution:

Assignments	20%
Quizzes	10%
Midterm Exam	20%
Final Exam	40%

Grading Policy: There will be relative grading such that the cutoff for A grade will not be less than 75% and cutoff for F grade will not be more than 34.9%. Grade distribution will follow normal bell curve (usually, A: $\geq \mu+3\sigma/2$, B: $\mu+\sigma/2 \ldots \mu+3\sigma/2$ C: $\mu-\sigma/2 \ldots \mu+\sigma/2$, D: $\mu - 3\sigma/2 \ldots \mu - \sigma/2$, and F: $< \mu - 3\sigma/2$)

Approximate grade assignments:

>= 90.0	A+
75.0 – 89.9	A
60.0 – 74.9	B
50.0 – 59.9	C
about 35.0 – 49.9	D
<= 34.9	F

Course Policies:

- **General course policies, Grades, Labs and assignments, Attendance and Absences** These clauses are common to all courses. And it can be found in the program schedule.

Course Outline (tentative) and Syllabus:

The weekly coverage might change as it depends on the progress of the class. However, you must keep up with the reading assignments. Each week assumes 4 hour lectures. Quizzes will be unannounced.

Week	Content
Week 1	• Definition of programming languages, their elements, environments, and design criteria, • Reading assignment: Chapter 1,2, KL
Week 2	• Lexical structure of PL, scope of lexical analysis, tools for implementing lexical analysis. • Reading assignment: Chapter 6, KL, Ref: Ch 5 ASUL • Programming assignment 1: Building a lexical analyser for C–. • Quiz 1
Week 3	• Context free grammars, Parse trees, Abstract Syntax trees, Ambiguity, Associativity and precedence of operators. Understanding the C grammar. • Reading assignment: Chapter 6: KL, Chapter 4.2,3 ASUL, C-Grammar from KR
Week 4	• Overview of various Parsing Techinques, Top-Down parsing • Reading assignment: Chapter 2.4, 4.4 ASUL, • Programming assignment: Building a top down parser for expression grammar.
Week 5	• Bottom-up parsing: Reductions, Handle pruning, Shift-reduce parsing, handling conflicts. • Reading assignment: Chapter 4.5 ASUL • Quiz 2
Week 6	• LR Parsing: Items, LR(0) Automaton, SLR parsing tables, Viable prefixes • Reading assignment: Chapter 4.6 ASUL
Week 7	• LR(1) items, construction of LR(1) automaton, LR(1) parsing tables, LALR parsing tables. • Reading assignment: Chapter 4.7, ASUL • Quiz 3
Week 8	• Parser generator tool: Yacc/Bison • Reading Assignment: Internet resources, Ref: Bison, Shroff Publishers. • Proramming assignment: Building AST for C– using yacc/bison • Review for Midterm Exam

Week	Content
Week 9	• Syntax directed translation: Inherited and Synthesized attributes, S-attributed and L-attributed definitions • Reading assignment: Chapter 5.1,2 ASUL • Home assignment 4
Week 10	• SDT-contd: structure of a Type, postfix translation schemes, Parser-stack implementation of postfix SDT's. • Reading assignment: Chapter 5.3,4 ASUL • Home assignment 5 • Quiz 4
Week 11	• Intermediate code generation: Tranlation of expressions, Type checking • Reading assignment: Chapter 6.4,5, ASUL • Home assignment 6
Week 12	• ICG-contd: Boolean expressions, short-circuit code, flow-of-control statements, avoiding redudant gotos, boolean values, and jumping code. • Reading assignment: Chapter 6.6, ASUL • Home assignment 7 • Quiz 5
Week 13	• ICG-contd: Backpatching, switch statements, procedures. • Reading assignment: Chapter 6.7,8, ASUL • Home assignment 8
Week 14	• Types revisited: Type Constructors, Type Equivalence, Type Checking, Type Conversion, Hindley-milner Polymorphic Type Checking. • Reading assignment: Chapter 8, KL
Week 15	• Dynamic memory management • Reading assignment: Chapter 7, AB • Review for Final Exam
Week 16,17	• Logic programming: Horns clauses, resolution and unification, Prolog: a case study. • Reading assignment: Chapter 7, AB • Programming assignment:
Week 18,19	• Functional programming: A study of Haskell • Reading assignment: Programming in Haskell. • Programming assignment:

CS241
Design and Analysis of Algorithms

Joydeep Mukherjee joydeep.m1981@gmail.com

Course Description: This course deals with topics in design and analysis of algorithms. In particular, the course will cover different techniques of algorithm design illustrating them with several examples and also highlight some of the lower bounding techniques in algorithm design such as NP-Completeness.

Prerequisite(s): (1) High School Mathematics.
Note(s): Syllabus changes yearly and may be modified during the term itself, depending on the circumstances. However, students will be evaluated only on the basis of topics covered in the course.
Course url:
Credit Hours: 4

Text(s):

Introduction to Automata Theory, Languages, and Computation, third edition
John E. Hopcroft, Rajeev Motwani & Jeffery D. Ullman
ISBN-13: 978-8131720479

Introduction to Algorithms, third edition
Thomas H. Cormen, Charles E. Leiserson, Ronald Rivest, Clifford Stein
ISBN: 9788120340077

Algorithm Design, first edition
Eva Tardos, Jon Kleinberg
ISBN: 9789332518643

Course Objectives:

Knowledge acquired: (1) Asymptotic analysis of running time of algorithms,
(2) different techniques of algorithm design, and,
(3) polynomial time reducibility.
Skills gained: The students will be able to
(1) compare different algorithms in terms of their running time ,
(2) design algorithms for some practical problems, and,
(3) do polynomial time reductions based on knowledge gained in the class.

Course Outline (tentative) and Syllabus:

The weekly coverage might change as it depends on the progress of the class. However, you must keep up with the reading assignments. Each week assumes 4 hour lectures. Quizzes will be unannounced.

Week	**Content**
Week 1	• Different order notations like $O, \Theta, \Omega, o, \theta, \omega$ and compare two different functions using order notation. • Methods to calculate and state running time of algorithms using order notations.
Week 2	• Introduction of the Divide and Conquer paradigm of algorithm design. • Devising algorithms using divide and conquer for merge sort, counting inversions, finding closest pair of points in a plane, fast integer multiplication etc. • Home assignment 1
Week 3	• Fast Fourier Transform and its application. • Quiz 1
Week 4	• Introducing the concept of Dynamic Programming and use of memoization. • Devising algorithms using dynamic programming for the problems like longest increasing subsequence, edit distance, knapsack, matrix chain multiplication, independent sets in trees etc.
Week 5	• Greedy methods of algorithm design. • Studying few techniques for proving the correctness of greedy algorithm.
Week 6	• Devising greedy algorithm for various problems like minimum spanning tress, Huffman codes, Horn clauses etc. • Home assignment 2 • Quiz 2
Week 7	• Breadth First Search (BFS)in graphs. • Depth First Search (DFS) in graphs.
Week 8	• Topological sorting of a directed acyclic graph. • Finding all strongly connected components of a directed graph. • Finding articulation points,bridges and biconnected component of a graph. • Finding Eulerian tour in a Eulerian graph. • Home assignment 3
Week 9	• Union Find data structure. • Kruskal and Prim's algorithm for minimum spanning trees. • Home assignment 4
Week 10	• Algorithms for single source shortest paths in a directed graph like Bellman-Ford algorithm, Dijkstra's algorithm. • Home assignment 5 • Quiz 4
Week 11	• Few applications of Single Source Shortest Paths algorithms • Home assignment 6
Week 12	• Algorithms for all pair shortest paths. • Matrix multiplication based procedure. • Floyd-Warshall algorithm. • Johnson's algorithm for sparse graphs. • Home assignment 7 • Quiz 5
Week 13	• String Matching algorithms • Home assignment 8
Week 14	• Introduction to the concept of P, NP, NP-Completeness, • Circuit satisfiability, Boolean satisfiability
Week 15	• NP-Completeness reduction for few problems. • Review for Final Exam

Database management System:

Course content:

1. a) Introduction to Database Systems – Advantages over file management system, Functions of Database Administrator and Database Manager, Basic structure of a data model and different types of data models. Object based and Record based models. Hierarchical and Network models.
   b) Introduction to relational model - Definition of relation as a mathematical entity, Database schema and database instance. Physical, Conceptual and View levels.
   (3 Lecture)
2. Entity relationship model – Database design - the software engineering approach, requirement analysis, discovery of entities and their relationships, identification of keys - candidate keys, super-keys and primary key, representation of entities, relationships and attributes, mapping of ER-diagram to relations, concept of specialization and generalization, extended ER model and EER to relational mapping. Exercises and assignments on creating ER/EER diagrams from problem description and mapping to relations. (6 Lectures)
3. Relational schema design and query – Data definition and data manipulation languages, Relational Algebra – Unary operators - selection, projection, Binary operators – Cartesian product, Joins – Theta join, Equi join, Natural join, Outer joins, Semi join, Anti join, Division. Tuple and Domain Relational Calculus. Exercises and assignments.
   (6 Lectures)
4. Normalization – Concept of Universal relation, dependency and inter-attribute relationships, Difference between database design from ER-diagram and normalization approach, Attribute atomicity and First normal form (1NF), Functional dependency, Partial dependency and Second Normal Form (2NF), Transitive dependencies, Dependency derivation and Armstrong's axioms, Closure of functional dependencies, Transitive closure of attributes, Discovery of Candidate keys from functional dependencies, Finding the Canonical cover of functional dependencies, Decomposition of Universal relation based on Partial and Transitive dependencies, Boyce-Codd normal form (BCNF), Concept of Lossless decomposition and dependency preservation, Third normal form (3NF), Concept of Multivalued dependencies and Fourth normal form (4NF), Join dependencies and Fifth normal form (5NF). Exercises and assignments.
   (10 Lectures)
5. Transaction management – Concept of a transaction, ACID properties (Atomicity, Consistency, Isolation and Durability), Different states of a transaction – active state to committed, rolledback or aborted state, Role and Look-ahead feature of transaction log, Requirements of Crash recovery facility for Atomicity & Consistency, Log analysis, Redo and Undo processes, Immediate update and Deferred update principles, Role of check points, Recovery for nested transactions, Exercises and assignments.
   (3 Lectures)
6. Concurrency control – Transactions and schedules, Concurrent execution of transactions, Concept of serializability, Detection of conflicts among concurrent transactions using precedence graph. Prevention of conflicts – Locking protocols, two phase locking protocol and its variations, Optimistic concurrency control and timestamp based concurrency control. (5 Lectures)

7. Distributed database- Need and advantage of distribution. Fragmentation, Replication and Allocation principles. Horizontal and Vertical Fragmentation methods. Need for creation of multi-database and federated database. (5 Lectures)

Recommended Textbooks:
Database Management Systems – R. Ramakrishnan & J. Gehrke – McGraw Hill.
Principles of Distributed database Systems – M. T. Ozsu & P. Valduriez

Additional Reference:
Database Systems, The complete book - H G Molina, J D Ullman & J Widom, Pearson Education Inc.

Evaluation: Assignment and Class Test: 40%, Semester Exam: 60%.

Expected Outcome:
1. Students should be able to design simple database schemas from User description / SRS document.
2. Student should be able to find design defects and suggest modifications for simple database schemas.

Suggested extension / modification:
Each student will be encouraged to deliver a seminar lecture each on related topics like: Temporal data model, Spatial data model, Different aspects of Query Processing & Optimization, Database Tuning etc.
Modified evaluation: Assignment and Class Test: 30%, Seminar: 20%, Semester Exam: 50%

CS250 Database Management Systems

Instructor:

Course Description: CS250 deals with a detailed study of principles of RDBMS.

Prerequisite(s): The student must know about a typical file system, data types like integer, float, and string, basic computer arithmetic, venn diagram regpresentation of union, intersection, and complement of sets.
Note(s): Syllabus changes yearly and may be modified during the term itself, depending on the circumstances. However, students will be evaluated only on the basis of topics covered in the course.
Course url:
Credit Hours: 4

Text(s):

- H. F. Korth and A. Silberschatz: Database System Concepts, McGraw Hill, New Delhi, 1997.
- R. A. Elmasri and S. B. Navathe: Fundamentals of Database Systems, 3rd ed., Addison-Wesley, 1998.
- R. Ramakrishnan: Database Management Systems, 2nd ed., McGraw Hill, New York, 1999.
- C. J. Date, A. Kannan and S. Swamynathan, An Introduction to Database Systems, Pearson Education, Eighth Edition, 2009
- J D Ullman : Principles of Database Systems, Computer Science Press; 2nd edition (December 1982)

Course Objectives:

Knowledge acquired: At the finish of this course, students will be quite empowered and will know
(1) basic concepts of the database approach, the underlying models and organizational issues
(2) the relational database model takes a logical view of data
(3) data modelling
(4) the theoretical underpinnings of the relational database, including concepts like functional dependence, entity integrity, and relational integrity.
(5) how a flawed data model can impact relational database implementation and manipulation
(6) relational database operators, the data dictionary, and the system catalog
(7) the various relational algebra operations that provide the basis for relational database manipulation
(8) concurrency control and locking protocols.

Skills gained: The students will be able to
(1) interpret the modeling symbols for the most popular ER modeling tools.
(2) model the RDBMS schema with the help of ER models given a problem statement in English.
(3) construct queries in SQL to manipulate live RDBMS
(4) analyze database requirements and determine the entities involved in the system and their relationship to one another sophisticated database applications

Competence Developed: The student will be able to
(1) tackle the design, development, and implementation of databases in an organization.
(2) assume any role in the database design and implementation process
(3) identify computational bottlenecks in the performance of an algorithm

Course Outline (tentative) and Syllabus: The weekly coverage might change as it depends on the progress of the class. However, you must keep up with the reading assignments. Each week assumes 4 hour lectures.

1. Introduction and Conceptual Modeling Database system concepts, three-schema Architecture, data independence, database administrator, database user, Client/Server Architecture, E-R diagram, mapping constraints, Keys, Generalization, Aggregation, Reducing E-R diagram to tables.
2. Relational Model: Concepts, constraints and Languages . Structure of Relational database, Entity Integrity, Referential Integrity, Foreign Keys, Query languages, Relational algebra and relational calculus, SQL, views.
3. Database Design Theory and Methodology Functional dependencies, Closure of a set of functional dependencies, Canonical cover, closure of attribute sets, Lossless decomposition, Dependency preservation, 1 NF, 2 NF, 3 NF, BCNF, Multivalued dependencies and 4 NF, Join dependencies and 5 NF.
4. Data Storage, Indexing and Query Processing File organization, Sequential file, B+ tree index files, B-tree index file, Static hash Functions, Dynamic hash functions, Query processing and Query optimization.
5. Transaction Processing Concepts Transaction, Properties of transaction, database recovery, shadow paging, recoverable schedule, serializable schedule; Concurrency control: Lock-Based protocol, Timestamp-Based protocol, Multiple granularity, Multiversion schemes; Deadlock Handling.
6. Database Security Discretionary access control, Mandatory access control and multi-level security, statistical database security, Introduction to flow control, Encription and public key infrastructures, privacy issues and preservation.

CS220
Data and File Structures

Course Description: This course introduces the study of internal and external data structures and algorithms with an on-going emphasis on the application of software engineering principles. Trees, graphs and the basic algorithms for creating, manipulating and using them will be covered. Various types of hash and indexed random access file structures will be discussed and implemented. B-trees and external file sorting will be introduced. Internal and external data/file organizations and algorithms will be compared and analyzed.

Prerequisite(s): (1) Programming in C/C++/JAVA/Python.
Note(s): Syllabus changes yearly and may be modified during the term itself, depending on the circumstances. However, students will be evaluated only on the basis of topics covered in the course.
Credit Hours: 4

Text(s): *Data Structures and Algorithms in JAVA*
Michael T. Goodrich, Roberto Tamassia, and Michael H. Goldwasser

Fundamentals of Data structures
Horowitz, E., and Sahni.S:

File Structures in C++
Folk & Zoellick & Riccardi

Data structures and algorithm analysis in C
Mark Allen Weiss

Course Objectives: Having completed this course successfully, the student should:

1. Be familiar with the use of data structures as the foundational base for computer solutions to problems.
2. Become introduced to and investigate the differing logical relationships among various data items.
3. Understand the generic principles of computer programming as applied to sophisticated data structures.
4. Comprehend alternative implementations using the differing logical relationships and appreciate the significance of choosing a particular logical relationship for implementation within real-world setting.
5. Demonstrate the ability to plan, design, execute and document sophisticated technical programs to handle various sorts of data structures.
6. become introduced the most important high-level file structures tools which include indexing, co-sequential processing, B trees, Hashing.

7. know the techniques for organization and manipulation of data in secondary storage including the low level aspects of file manipulation which include basic file operations, secondary storage devices and system software.

Course Outline (tentative) and Syllabus:
The weekly coverage might change as it depends on the progress of the class. Each week assumes 4 hour lectures.

Week	**Content**
Week 1	• Introduction to algorithm analysis: pseudo code, algorithm efficiency, asymptotic and empirical analysis of algorithms.
Week 2	• Introduction to data structures. Linear data structures: arrays, stacks, queues, linked lists (operations, implementations, applications.)
Week 3	• Non-linear data structures: binary trees and general trees (operations, implementations and applications). Binary search trees.
Week 4	• Priority queues and heaps: using a heap to implement a priority queue. Heap sort.
Week 5	• Balanced search trees: AVL trees
Week 6	• (2,4) and red-black trees.
Week 7	• B-trees and B^+ trees
Week 8	• Amortized Analysis, Splay tree
Week 9	• Hashing, Hash fucntions and collision resolution techniques -linear probing
Week 10	• Hashing and collision resolution techniques - quadratic probing, Double hashing
Week 11	• Graphs and elementary Graph operations - Breadth First Search, Depth First Search
Week 12	• Spanning Trees, Shortest paths
Week 13	• File Structure: Concepts of fields, records and files, Sequential, Indexed and Relative/Random File Organization
Week 14	• Indexing structure for index files, hashing for direct files
Week 15	• Multi-Key file organization and access methods.

DA205 Data Mining

Instructor: **Prof. Aditya Bagchi**

Course Description: The quantity and variety of online data is increasing very rapidly. The data mining process includes data selection and cleaning, machine learning techniques to "learn" knowledge that is "hidden" in data, and the reporting and visualization of the resulting knowledge. This course will cover these issues.

Prerequisite(s): First course in DBMS,
Credit Hours: 2

Text(s):

- Data Mining Concepts and techniques, J. Han and M. Kamber, Morgan Kaufmann.
- Mining of Massive datasets, A. Rajaraman, J. Leskovec, J.D. Ullman
- Mining the WEB, S. Chakrabarti, Morgan Kaufmann.

Course Objectives:

Knowledge acquired: At the finish of this course, students will be quite empowered and will know
(1) standard data mining problems and associated algorithms.
(2) how to apply and implement standard algorithms in similar problem.
Competence Developed: The student will be able to
(1) Understand a data environment, extract relevant features and identify necessary algorithms for required analysis.
(2) Accumulation, extraction and analysis of Social network data.

Course Outline (tentative) and Syllabus: The weekly coverage might change as it depends on the progress of the class. However, you must keep up with the reading assignments. Each week assumes 4 hour lectures.

1. Introduction to Data Mining concept, Data Cleaning, transformation, reduction and summarization. (1 lecture = 2 hours)
2. Data Integration - Multi and federated database design, Data Warehouse concept and architecture. (2 lectures = 4 hours)
3. Online Analytical Processing and Data Cube. (2 lectures =4 Hours)
4. Mining frequent patterns and association of items, Apriori algorithm with fixed and variable support, improvements over Apriori method - Hash-based method, Transaction reduction method, Partitioning technique, Dynamic itemset counting method. (2 Lectures = 4 Hours)
5. Frequent Pattern growth and generation of FP-tree, Mining closed itemsets. (1 Lecture = 2 Hours)
6. Multilevel Association rule, Association rules with constraints, discretization of data and association rule clustering system. (1 Lecture = 2 Hours)
7. Association mining to Correlation analysis. (1 Lecture = 2 Hours)
8. Mining time-series and sequence data. (2 Lectures = 4 Hours)
9. Finding similar items and functions for distance measures. (4 Lectures = 8 Hours)
10. Recommendation system, content based and collaborative filtering methods. (5 Lectures = 10 Hours)
11. Graph mining and social network analysis. (5 Lectures = 10 Hours)

DA230
Enabling Technologies for Big Data Computing

Instructor

Sudeep Mallick, Ph.D.
Sudeep.mallick@gmail.com

Course Description:

DA230 deals with technologies and engineering solutions for enabling big data processing and analytics . More specifically, it deals with the tools for data processing, data management and programming in the distributed programming paradigm using techniques of MapReduce programming, NoSQL distributed databases, streaming data processing, data injestion, graph processing and distributed machine learning for big data use cases.

Prerequisite(s): (1) Basic knowledge of python and Java programming languages (2) Tabular data processing / SQL queries. (3) Basic knowledge of common machine learning algorithms.
Credit Hours: 4

Text(s):

Hadoop: The Definitive Guide, fourth edition
Tom White
ISBN: 978-1-491-90163-2

Hadoop in Action, edition: 2011
Chuck Lam
ISBN: 978-1-935-18219-1

Spark in Action, edition: 2017
Petar Zecevic & Marko Bonaci
ISBN: 978-93-5119-948-9

Data-Intensive Text Processing with MapReduce, edition: 2010
Jimmy Lin & Chris Dyer
ISBN: 978-1-608-45342-9

Course Outline (tentative) and Syllabus:

The weekly coverage might change as it depends on the progress of the class. Each week assumes 4 hour lectures.

Week	Content
Week 1	• Big data computing paradigm and Hadoop: big data, hadoop architecture • Reading assignment: Chapter 1, LD & Chapter 1, TW • Lab: setting up Hadoop platform in standalone mode
Week 2	• Hadoop MapReduce (MR): Lab session with simple MR algorithms in Hadoop standalone mode • Reading assignment: Chapter 2, LD & Chapter 2, TW
Week 3	• Hadoop Distributed File System (HDFS), YARN and MR architecture, daemons, serialization concept, command line parameters: Lab session • Reading assignment: Chapter 3-5 & 7, TW
Week 4	• Implementing algorithms in MR - joins, sort, text processing, etc.: Lab session • Reading assignment: Chapter 3, LD & Chapter 7, TW • Lab assignment 1
Week 5	• Hadoop operations in Cluster Mode, Hadoop on AWS Cloud: Lab session • Reading assignment: Instructor notes
Week 6	• Understanding NoSQL using Pig: Lab Session • Reading assignment: Chapter 16, TW • Lab assignment 2
Week 7	• Introduction to Apache Spark platform and architecture, RDD, • Reading assignment: Chapters 1-3, ZB
Week 8	• Mapping, joining, sorting, grouping data with Spark RDD: Lab session • Reading assignment: Chapter 4, ZB • Review for Mid term exam
Week 9	• Advanced usage of Spark API: Lab session • Reading assignment: Chapter 4, ZB • Lab assignment 3
Week 10	• NoSQL queries using Spark DataFrame and Spark SQL: Lab session • Reading assignment: Chapter 5, ZB
Week 11	• Using SQL Commands with Spark: Lab session • Reading assignment: Chapter 5, ZB
Week 12	• Machine Learning using Spark MLib: Lab session • Reading assignment: Chapter 7, ZB
Week 13	• Machine Learning using Spark ML: Lab session • Reading assignment: Chapter 8, ZB • Lab assignment 4
Week 14	• Spark operations in Cluster Mode, Spark on AWS Cloud: Lab session • Reading assignment: Chapter 11, ZB
Week 15	• Graph processing with Spark GraphX: Lab session • Reading assignment: Chapter 9, ZB

CS211
Graph algorithms and Combinatorial optimization

Instructor: Dhyanagamyananda swathyprabhu@gmail.com

Course Description: CS211 is the first course to deal with the topic of this course. This course is a mixed bag of graph algorithms. Some of these algorithms are relevant in the context of optimization. The field of graph algorithms is vast and the kind of problem studied in CS211 are those that are in general difficult to solve but has easy solutions for a sub-class of them.

Prerequisite(s): Design and Analysis of Algorithms, Data and File Structures.
Credit Hours: 4

Text(s):

Algorithm Design, PHI
Kleinberg & Targos

Introduction to Graph Theory
Douglas West

Lecture Notes from University of Waterloo

Draft on Discharging techinque by Douglas West

Course Objectives:

Knowledge acquired: .
(1) Flow networks.
(2) Planar graph theory
(3) Algorithm design and analysis

Grade Distribution:

Assignments 20%, Quizzes 10%, Midterm Exam 20%, Final Exam 40%

Course Outline (tentative) and Syllabus:

The weekly coverage might change as it depends on the progress of the class. However, you must keep up with the reading assignments. Each week assumes 4 hour lectures. Quizzes will be unannounced.

Week	Content
Week 1	• Network Flow: Definition, Basic Idea, Algorithm, Maxflow mincut theorem, Ford Fulkerson Algorithm Analysis, LP formulation of maxflow and proof. • Reading assignment: Chapter 3, KT
Week 2	• Layered Network: Definition, Theorem, Computation of blocking flow (Edmonds, Dinics, MPM) • Reading assignment: XBitmap from Wiki. • Programming assignment 1: • Quiz 1
Week 3	• Student presentation of Tarzan's algorithm
Week 4	• Bipartite matching: Definition, Application, Using Ford Fulkerson Algorithm bipartite matching is obtained in O(—V——E—) time Edge connecting problem. The augmenting path algorithm for bipartite matching. • Reading Assignment:
Week 5	• Matching for Non-Bipartite Graph: Theorem and proof (Edmonds blossom shrinking) • Reading Assignment: • Quiz 2
Week 6, 7	• Max-Cut: NP-Hard problem and its proof, 2-Approximation algorithm, Randomized algorithm for max-cut, De-randomization LP based approximation algo for maxcut • Reading assignment:
Week 8,9	• Interval Graph: Intersection graph, Perfect elimination order (PEO), Chordal graph (Triangulated Graph), Simplicial vertex, Algorithm MIS, vertex cover, coloring, clique cover for interval graph, Finding a PEO Comparability graph • Reading assignment: Waterloo Lecture Notes • Home assignment: • Quiz 3

Week	Content
Week 10,11,12	• Trees and Friends, Trees, Treewidth, Tree decomposition, Closure properties, Partial k-trees, Partial k-trees to tree decomposition, Tree decomposition to partial k-trees, Dynamic programming MIS algo for partial k-tree • Home assignment 4 • Quiz at the end of three weeks.
Week 13,14	• Perfect Graph, Definition and properties, Perfect graph theorem, Triangulated graph is a perfect graph • Home assignment 7 • Quiz 5
Week 15	• Discharging method • Reading assignment: DW on discharging • Home assignment 8

DA103
Linear Algebra

Course Description: CS301 deals with topics in linear algebra. In particular, the course will cover linear equations, vector spaces, linear transformations, eigenvalues and eigenvectors, bilinear forms, introduction to linear programming and related topics.

Prerequisite(s): (1) Highschool mathematics.
Note(s): Syllabus changes yearly and may be modified during the term itself, depending on the circumstances. However, students will be evaluated only on the basis of topics covered in the course.

Credit Hours: 4

Text(s):

Linear Algebra,
K. M. Hoffmann, R. Kunze
Prentice Hall.

Algebra,
M. Artin
Prentice Hall.

Introduction to Linear Algebra,
G. Strang
Wellesley-Cambridge Press.

Linear Programming,
L. I. Gass
Tata McGraw Hills.

Linear Programming,
G. Hadley
Narosa Publishing House.

Course Objectives:

Knowledge acquired: (1) systems of linear equations, their associated matrices and their properties,
(2) characteristic polynomial, eigenvalues and eigenvectors,
(3) bilinear forms, and,
(4) linear programming.
Skills gained: The students will be able to
(1) analyze system of linear equations,
(2) solving linear recurrences, and,
(3) formulating linear programming problems and finding their feasible and optimal solutions.

Grade Distribution:

Assignments	10%
Quizzes	10%
Midterm Exam	30%
Final Exam	50%

Course Outline (tentative) and Syllabus:
The weekly coverage might change as it depends on the progress of the class. However, you must keep up with the reading assignments. Each week assumes 4 hour lectures. Quizzes will be unannounced.

Week	**Content**
Week 1	• Systems of linear equations, Matrices and elementary row operations, Row reduced Echelon matrices,
Week 2	• Matrix multiplication, Invertible matrices, Transpose of a matrix,
Week 3	• Systems of homogeneous equations, Equivalence of row rank and column rank of a matrix, Determinant and volume of the fundamental parallelepiped,
Week 4	• Permutation matrices, Cramers rule, • Home assignment 1
Week 5	• Vector spaces and subspaces, Bases and dimensions, Coordinates and change of bases, Direct sums, • Home assignment 2
Week 6	• The Rank-Nullity theorem, Matrix of a linear transformation, Linear operators and isomorphism of vector spaces, Determinant of a linear operator,
Week 7	• Linear functionals, Annihilators, The double dual, • Home assignment 3
Week 8	• Eigenvalues and eigenvectors of matrices, The characteristic polynomial, Algebraic and geometric multiplicities of eigenvalues,
Week 9	• Diagonalizability, Cayley-Hamilton theorem, Solving linear recurrences, • Home assignment 4
Week 10	• Matrix of a bilinear form, Symmetric and positive definite bilinear forms, Normed spaces,
Week 11	• Cauchy-Schwarz inequality and triangle inequality, Angle between two vectors, Orthogonal complement, Projection,
Week 12	• Gram-Schmidt orthogonalization, Hermitian operators, The Spectral theorem, • Home assignment 5
Week 13	• Bounded and unbounded sets, Convex functions, Convex cone, Interior points and boundary points, Extreme points or vertices,
Week 14	• Convex hulls and convex polyhedra, Supporting and separating hyperplanes, Formulating linear programming problems,
Week 15	• Feasible solutions and optimal solutions, Graphical method, The basic principle of Simplex method, Big-M method, • Home assignment 6

Ramakrishna Mission Vivekananda Educational and Research Institute
Syllabus for Linear Algebra I
Prepared by: Dr. Soumya Bhattacharya

---

1 Linear equations

- Systems of linear equations
- Matrices and elementary row operations
- Row reduced Echelon matrices
- Matrix multiplication
- Invertible matrices
- Transpose of a matrix
- Systems of homogeneous equations
- Equivalence of row rank and column rank of a matrix
- Determinant and volume of the fundamental parallelepiped
- Permutation matrices
- Cramer's rule

2 Vector spaces

- Vector spaces and subspaces
- Bases and dimensions
- Coordinates and change of bases
- Direct sums

3 Linear transformations

- The Rank-Nullity theorem
- Matrix of a linear transformation
- Linear operators and isomorphism of vector spaces
- Determinant of a linear operator
- Linear functionals
- Annihilators
- The double dual

4 Eigenvalues and eigenvectors

- Eigenvalues and eigenvectors of matrices
- The characteristic polynomial
- Algebraic and geometric multiplicities of eigenvalues
- Diagonalizability
- Cayley-Hamilton theorem
- Solving linear recurrences

5 Bilinear forms

- Matrix of a bilinear form
- Symmetric and positive definite bilinear forms
- Normed spaces
- Cauchy-Schwarz inequality and triangle inequality
- Angle between two vectors
- Orthogonal complement
- Projection
- Gram-Schmidt orthogonalization
- Hermitian operators
- The Spectral theorem

6 Introduction to linear programming

- Bounded and unbounded sets
- Convex functions
- Convex cone
- Interior points and boundary points
- Extreme points or vertices
- Convex hulls and convex polyhedra
- Supporting and separating hyperplanes
- Formulating linear programming problems
- Feasible solutions and optimal solutions
- Graphical method
- The basic principle of Simplex method
- Big-M method

Reference books

1. M. Artin, *Algebra*, Prentice Hall.
2. K. M. Hoffmann, R. Kunze, *Linear Algebra*, Prentice Hall.
3. G. Strang, *Introduction to Linear Algebra*, Wellesley-Cambridge Press.
4. L. I. Gass, *Linear Programming*, Tata McGraw Hills.
5. G. Hadley, *Linear Programming*, Narosa Publishing House.

The students by the end of the course will be able to explain:

- How to check whether a given system of linear equations has any solution or not.
- How to find the solutions (if any) of a system of linear equations.
- Why a system of linear equations with more variables than equations always has a solution, whereas a system of such equations with more equations than variables may not have any solution at all.
- How to find the rank and nullity of a matrix.
- Why each permutation matrix is of full rank.

- Why a matrix is invertible if and only if it has nonzero determinant and how to find the inverse of such a matrix.
- Why a matrix with more columns than rows (resp. more rows than columns) does not have a left (resp. right) inverse.
- How to extend a basis of a subspace of a vector space V to a basis of V.
- How a change of basis affects the coordinates of a given vector.
- Why both the ranks of a matrix A and its transpose A^{T} are the same as that of $A^{\mathrm{T}}A$.
- Why the determinant of the matrix of a linear operator does not depend on the choice of the basis of the ambient space.
- Why the sum of the dimension of a subspace W of a vector space V and the dimension of the annihilator of W is the dimension of V.
- Why the double dual of a vector space V is canonically isomorphic to V itself.
- Why the fact that a certain conjugate of a given matrix A is diagonal is equivalent to the fact that the space on which A acts by left multiplication is a direct sum of the eigenspaces of A.
- Why every idempotent matrix is diagonalizable.
- Why conjugate matrices have the same eigenvalues with the same algebraic and geometric multiplicities.
- What Cayley-Hamilton theorem states and why replacing the variable t by the square matrix A in $\det(tI - A)$ does not lead to a proof of this theorem.
- How to solve a linear recurrence whose associated matrix is diagonalizable.
- Why the determinant of an upper or lower triangular matrix is the product of its diagonal entries.
- Why two diagonalizable matrices commute if and only if they are simultaneously diagonalizable.
- Why for a matrix which represent the dot product with respect to some basis, it is necessary and sufficient to be symmetric and positive definite.
- Why for a symmetric matrix to be positive definite, it is necessary and sufficient for it to have strictly positive eigenvalues.
- What is the role of the Cauchy-Schwarz inequality in defining the angle between two vectors.
- Why the elements in a basis a subspace W of V and the elements in a basis of the orthogonal complement of W are linearly independent.
- How to orthogonalize a given basis of an inner product space.

- Why each inner product on a real vector space V induces an isomorphism between V and its dual.
- Why any symmetric matrix is diagonalizable and why all its eigenvalues are real.
- Why in a closed and bounded convex region, a convex function attains its maximum at the boundary.
- Why it suffices to check only the corner points to find a solution to a given linear programming problem, whose feasible region is a convex polyhedron.

Sample questions

Linear equations

1. Let A be a square matrix. Show that the following conditions are equivalent:

(i) The system of equations $AX = 0$ has only the trivial solution $X = 0$.

(ii) A is invertible.

2. Show that a matrix with more columns than rows (resp. more rows than columns) does not have a left (resp. right) inverse.

3. Explain why a system of linear equations with more variables than equations always has a solution, whereas a system of such equations with more equations than variables may not have any solution at all.

4. Let $A^n = 0$. Let I denote the identity matrix of the same size as that of A. Compute the inverse of $A - I$.

5. Prove that if A is invertible, then $(A^t)^{-1} = (A^{-1})^t$.

6. Compute the determinant of the following matrix:

$$\begin{pmatrix} 2 & 1 & & & & \\ 1 & 2 & 1 & & \text{\Large 0} & \\ & 1 & 2 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ & \text{\Large 0} & & 1 & 2 & 1 \\ & & & & 1 & 2 \end{pmatrix}_{n\times n}.$$

7. Let n be a positive integer and let

$$A = \begin{pmatrix} 2 & -1 & & & & \\ -1 & 2 & -1 & & \text{\huge 0} & \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & \text{\huge 0} & & -1 & 2 & -1 \\ & & & & -1 & 2 \end{pmatrix}_{n\times n}.$$

Find the value of the determinant of the matrix A.

8. Show that every permutation matrix is of full rank.

9. Compute the determinant of the following matrix:

$$\begin{pmatrix} 2 & -2 & & & & & \\ -1 & 5 & -2 & & & \text{\huge 0} & \\ & -2 & 5 & -2 & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & -2 & 5 & -2 & \\ & \text{\huge 0} & & & -2 & 5 & -1 \\ & & & & & -2 & 2 \end{pmatrix}_{n\times n}.$$

10. Compute the determinant of the following matrix:

$$\begin{pmatrix} 3 & 2 & & & & \\ 1 & 3 & 2 & & \text{\huge 0} & \\ & 1 & 3 & 2 & & \\ & & \ddots & \ddots & \ddots & \\ & \text{\huge 0} & & 1 & 3 & 2 \\ & & & & 1 & 3 \end{pmatrix}_{n\times n}.$$

11. If possible, find all the solutions of the equation $XY - YX = I$ in 3×3 real matrices X, Y.

12. Let $A \in M_{n,n}(\mathbb{R})$. Show that

$$(\det A)^2 \leq \prod_{i=1}^{n} \left(\sum_{k=1}^{n} A_{k,i}^2 \right),$$

where $A_{k,i}$ denotes the k, i-th entry of A.

13. Let

$$A = \begin{pmatrix} 2 & -2 & -4 \\ -1 & 3 & 4 \\ 1 & -2 & -3 \end{pmatrix} \in M_{3,3}(\mathbb{R}).$$

Find the inverse of the matrix $\left(37 \cdot A^{372} + 2 \cdot I\right)$.

Vector spaces and linear transformations

14. Let f and g be two nonzero linear functionals on a finite dimensional real vector space V such that their nullspaces (i.e. kernels) coincide. Show that there exists a $c \in \mathbb{R}$ such that $f = cg$.

15. Show that if the product of two $n \times n$ matrices is 0, then sum of their ranks is less than or equal to n.

16. The cross product of two vectors in $\mathbb{R}^3$ can be generalized for $n \geq 3$ to a product of $n-1$ vectors in $\mathbb{R}^n$ as follows: For $x^{(1)}, \dots, x^{(n-1)} \in \mathbb{R}^n$, define

$$x^{(1)} \times \ldots \times x^{(n-1)} := \sum_{i=1}^{n} (-1)^{i+1} (\det A_i) \cdot e_i,$$

where $A \in M_{n-1,n}(\mathbb{R})$ is the matrix, whose rows are $x^{(1)}, \dots, x^{(n-1)}$ and A_i is the submatrix of A obtained by deleting the i-th column of A. Similarly as in the case $n = 3$, the cross product $x^{(1)} \times \cdots \times x^{(n-1)}$ is given by the formal expansion of

$$\det \begin{pmatrix} e_1 & e_2 & \cdots & e_n \\ x_1^{(1)} & x_2^{(1)} & \cdots & x_n^{(1)} \\ \vdots & \vdots & & \vdots \\ x_1^{(n-1)} & x_2^{(n-1)} & \cdots & x_n^{(n-1)} \end{pmatrix}$$

w.r.t. the first row. Show that the following assertions hold for the generalized cross product:

a) $x^{(1)} \times \ldots \times x^{(i-1)} \times (x+y) \times x^{(i+1)} \times \ldots \times x^{(n-1)} = x^{(1)} \times \ldots \times x^{(i-1)} \times x \times x^{(i+1)} \times \ldots \times x^{(n-1)} + x^{(1)} \times \ldots \times x^{(i-1)} \times y \times x^{(i+1)} \times \ldots \times x^{(n-1)}$.

b) $x^{(1)} \times \ldots \times x^{(i-1)} \times \lambda x \times x^{(i+1)} \times \ldots \times x^{(n-1)} = \lambda \left(x^{(1)} \times \ldots \times x^{(i-1)} \times x \times x^{(i+1)} \times \ldots \times x^{(n-1)}\right)$.

c) $x^{(1)} \times \ldots \times x^{(n-1)} = 0 \;\Leftrightarrow\; x^{(1)}, \dots, x^{(n-1)}$ are linearly dependent.

d) $\langle x^{(1)} \times \ldots \times x^{(n-1)}, y \rangle = \det \begin{pmatrix} y_1 & y_2 & \cdots & y_n \\ x_1^{(1)} & x_2^{(1)} & \cdots & x_n^{(1)} \\ \vdots & \vdots & & \vdots \\ x_1^{(n-1)} & x_2^{(n-1)} & \cdots & x_n^{(n-1)} \end{pmatrix}$.

e) $\langle x^{(1)} \times \ldots \times x^{(n-1)}, x^{(i)} \rangle = 0$ for $i \in \{1, \dots, n-1\}$.

17. For any matrix A, show that the ranks of A and $A^{\mathrm{T}}A$ are the same.

18. Let $n \geq 3$, $A \in \mathrm{O}_n$ and $x^{(1)}, \ldots, x^{(n-1)} \in \mathbb{R}^n$. Define the linear map $\varphi_A : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ by $\varphi(v) = Av$ and let the generalized cross product of $n-1$ vectors in $\mathbb{R}^n$ be defined as in the last exercise. Show that:

$$\varphi_A\big(x^{(1)}\big) \times \cdots \times \varphi_A\big(x^{(n-1)}\big) = \det A \cdot \varphi_A\big(x^{(1)} \times \cdots \times x^{(n-1)}\big).$$

19. Let V and W be finite dimensional vector spaces and let $i_V : V \to V$ and $i_W : W \to W$ be identity maps. Let $\phi : V \to W$ and $\psi : W \to V$ be two linear maps. Show that $i_V - \psi \circ \phi$ is invertible if and only if $i_W - \phi \circ \psi$ is invertible.

20. If W_1 and W_2 are two subspaces of a vector space V, then show that

$$(W_1 + W_2)^0 = W_1^0 \cap W_2^0.$$

21. If W_1 and W_2 are two subspaces of a vector space V, then show that

$$(W_1 \cap W_2)^0 = W_1^0 + W_2^0.$$

22. Let $V = \mathbb{R}^3$ and let $\mathbb{B} = \left\{ \begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right\}$ be a basis of V. Compute the dual basis $\mathbb{B}^*$ of V^*.

23. Let V, W finite dimensional vector spaces over a field $\mathbb{K}$ and let $\varphi : V \to W$ be a linear map.
(1) Show that $\varphi^* : W^* \to V^*$ is a linear map.
(2) Show that $\psi : \mathrm{Hom}_{\mathbb{K}}(V, W) \longrightarrow \mathrm{Hom}_{\mathbb{K}}(W^*, V^*)$, $\varphi \mapsto \varphi^*$ is an isomorphism.

24. Let V, W be finiet dimensional vector spaces over a field $\mathbb{K}$ and let $\varphi : V \to W$ be a linear map.
(1) Show that if φ is surjective, then φ^* injective.
(2) Show that if φ is injective, then φ^* is surjective.

Eigenvalues and eigenvectors

25. Let A be a diagonalizable matrix. Show that A and A^{T} are conjugate.

26. Let $v, w \in \mathbb{R}^n$ are eigenvectors of a matrix $A \in M_{n,n}(\mathbb{R})$ with corresponding eigenvalues λ and μ respectively. Show that if $v + w$ is also an eigenvector of A, then $\lambda = \mu$.

27. Let $V = \mathbb{R}^n$ and $A \in M_{n,n}(\mathbb{R})$ be a diagonalizable matrix. Show that:

$$V = (\ker \varphi_A) \oplus (\mathrm{Im}\ \varphi_A),$$

where the map $\varphi_A : V \longrightarrow V$ is defined by $\varphi_A(v) := Av$ for all $v \in V$.

28. Find a closed formula for the n-th term of the linear recurrence defined as follows: $F_0 = 0, F_1 = 1$ and

$$F_{n+1} = 3F_n - 2F_{n-1}.$$

29. Let $A \in \mathrm{O}_n$ with $\det A = -1$. Show that -1 is an eigenvalue of A with an odd algebraic multiplicity.

30. Let n be a positive odd integer and let $A \in \mathrm{SO}_n$. Show that 1 is an eigenvalue of A.

31. If each row sum of a real square matrix A is 1, show that 1 is an eigenvalue of A.

32. Let A be a 2017×2017 matrix with all its diagonal entries equal to 2017. If all the rest of the entries of A are 1, find the distinct eigenvalues of A.

33. Let λ be an eigenvalue of the $n \times n$ matrix $A = (a_{ij})$. Show that there exists a positive integer $k \leq n$ such that

$$|\lambda - a_{kk}| \leq \sum_{j=1, j \neq k}^{n} |a_{jk}|.$$

34. Let A be a diagonalizable matrix. Show that A and A^{T} have the same eigenvalues with the same algebraic and geometric multiplicities.

35. (*a*) Let A be a 3×3 matrix with real entries such that $A^3 = A$. Show that A is diagonalizable.
(*b*) Let n be a positive integer. Let A be a $n \times n$ matrix with real entries such that $A^2 = A$. Show that A is diagonalizable.

36. Let A be a diagonalizable matrix. Show that A and A^{T} have the same eigenvalues with the same algebraic and geometric multiplicities.

37. Let A be a 3×3 matrix with positive determinant. Let $\mathcal{P}_A(t)$ denote the characteristic polynomial of A. If $\mathcal{P}_A(-1) > 1$, show that A is diagonalizable.

38. Let A be a 3×3 matrix with real entries. If $\mathcal{P}_A(-1) > 0 > \mathcal{P}_A(1)$, where $\mathcal{P}_A(t)$ denotes the characteristic polynomial of A, show that A is diagonalizable.

39. (*a*) Show that similar matrices (i.e. conjugate matrices) have the same eigenvalues with the same algebraic and geometric multiplicities.
(*b*) Give examples of two matrices with the same characteristic polynomial but with an eigenvalue which does not have the same geometric multiplicity.

40. Let A be a 3×3 matrix with real entries such that $A^3 = A$. Show that A is diagonalizable.

41. Let n be a positive integer and let A be a $n \times n$ matrix with real entries such that $A^3 = A$. Show that A is diagonalizable.

42. For an $n \times n$ matrix A and be the characteristic polynomial $\mathcal{P}_A(t)$ of A, is the following a correct proof of Cayley-Hamilton theorem?

$$\mathcal{P}_A(A) = \det(A \cdot I_n - A) = \det(A - A) = 0.$$

Justify your answer.

43. Determine the eigenvalues of the orthogonal matrix

$$A = \frac{1}{2} \cdot \begin{pmatrix} 1+\frac{1}{\sqrt{2}} & -1 & \frac{1}{\sqrt{2}}-1 \\ 1-\frac{1}{\sqrt{2}} & 1 & -\frac{1}{\sqrt{2}}-1 \\ 1 & \sqrt{2} & 1 \end{pmatrix}.$$

44. (*a*) Find a closed formula for the n-th term of the linear recurrence defined as follows: $F_0 = 0, F_1 = 1$ and

$$F_{n+1} = 2F_n + F_{n-1}$$

by diagonalizing the matrix $\begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix}$.

(*b*) Explain why the above method fails to help us in finding a closed formula for the n-th term of the linear recurrence defined as follows: $F_0 = 0, F_1 = 1$ and

$$F_{n+1} = 2F_n - F_{n-1}.$$

45. Let A be a 5×5 real matrix with negative determinant. If $\mathcal{P}_A(\pm 2) > 0 > \mathcal{P}_A(\pm 1)$, where $\mathcal{P}_A(t)$ denotes the characteristic polynomial of A, show that A is diagonalizable.

46. We say that two matrices A and B are *simultaneously diagonalizable* if there exists an invertible matrix P such that both PAP^{-1} and PBP^{-1} are diagonal. Show that two diagonalizable matrices A and B commute with each other if and only if they are simultaneously diagonalizable.

47. Find a closed formula for the n-th term of the linear recurrence defined as follows: $F_0 = 0$, $F_1 = 1$ and

$$F_{n+1} = 3F_n - 2F_{n-1}.$$

48. Solve the following equation for a 2×2 matrix X:

$$X^2 = \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}.$$

49. Let

$$A = \begin{pmatrix} 3 & -1 & 1 \\ 1 & -1 & 1 \\ 1 & -1 & 3 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 3 & 1 & 1 \\ -1 & -1 & -1 \\ 1 & 1 & 3 \end{pmatrix}.$$

Without doing any calculations, explain for which one of the matrices $A + B$ and AB, the eigenvectors form a basis of $\mathbb{R}^3$.

(*b*) (3 points) Determine that basis of eigenvectors of $\mathbb{R}^3$ for one of the matrices $A + B$ or AB.

50. Construct and example of the scenario where $\alpha, \beta, \gamma \in \mathbb{R}^n$ such that $\alpha \perp \beta$, $\gamma \neq 0$ and A, B are $n \times n$ matrices such that $A \cdot \alpha = a\gamma$ and $B \cdot \beta = b\gamma$, where a is a nonzero eigenvalue of A and b is a nonzero eigenvalue of B.

Bilinear forms

51. How many $n \times n$ real matrices are both symmetric and orthogonal? Justify your answer.

52. We call a linear map $\mathbb{R}^n$ an *isometry* if it preserves the dot product on $\mathbb{R}^n$. Show that left multiplication by a real square matrix A defines an isometry on $\mathbb{R}^n$ if and only if A is orthogonal.

53. How many $n \times n$ complex matrices are there which are positive definite, self-adjoint as well as unitary?

54. For any complex square matrix A, show that the ranks of A and A^* are equal.

55. Show that if the columns of a square matrix form an orthonormal basis of $\mathbb{C}^n$, then its rows do too.

56. Let $B \in M_{n,n}(\mathbb{R})$. Show that

$$\ker \varphi_B := (\mathrm{Im}\ \varphi_{B^{\mathrm{T}}})^{\perp},$$

where the map $\varphi_B : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is defined by $\varphi_B(v) = Bv$.

57. Let $V = \mathbb{R}^4$ and let $f : V \longrightarrow V$ such that $f^2 = 0$. Show that for each triplet $v_1, v_2, v_3 \in \mathrm{Im}\, f$, we have

$$\mathrm{Vol}(v_1, v_2, v_3) = 0.$$

58. Let $V = \mathbb{C}^2$ and let s be a symmetric bilinear form on V. Let $q : V \longrightarrow \mathbb{R}$ be the quadratic form corresponding to s. Suppose, for all $z_1, z_2 \in \mathbb{C}$, we have

$$q\left(\begin{pmatrix} z_1 \\ z_2 \end{pmatrix}\right) = |z_1|^2 + |z_2|^2 + i(\overline{z_1} z_2 - z_1 \overline{z_2}).$$

Compute the determinant of the matrix representing s with respect to the basis $\mathbb{B} = \left\{ \begin{pmatrix} 1 \\ i \end{pmatrix}, \begin{pmatrix} 1+i \\ 1 \end{pmatrix} \right\}$.

59. Let V be a real vector space with inner product s and let $v_1, \ldots, v_n \in V \setminus \{0\}$ such that $s(v_i, v_j) = 0$ for all $i, j \in \{1, \ldots, n\}$. For $v \in V$, we define $\|v\| = \sqrt{s(v, v)}$.
(1) Show that for all $v \in V$, we have

$$\sum_{i=1}^{n} \frac{s(v, v_i)^2}{\|v_i\|^2} \leq \|v\|^2 . \tag{1}$$

(2) Determine all the cases when the equality holds in (1).

60. Let V be a finite dimensional vector space and let P and Q be projection maps from V to V. Show that the following are equivalent:

(*a*) $P \circ Q = Q \circ P = 0$.

(*b*) $P + Q$ is a projection.

(c) $P \circ Q + Q \circ P = 0$.

61. Let $V = \mathbb{R}^3$ be the three dimensional euclidean space with the usual dot product and let U be the subspace of V which is spanned by $\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$. Determine the matrix of the orthogonal projection P_U with respect to the standard basis of V.

62. Do the following exercise without using the Spectral Theorem:
(1) Let $A = \begin{pmatrix} a & b \\ b & d \end{pmatrix} \in M_{2,2}(\mathbb{R})$. Show that A is diagonalizable.
(2) Let $B \in M_{3,3}(\mathbb{R})$ be a symmetric matrix. Show that B is diagonalizable.

63. Let V be a finite dimensional real vector space. For $v, w \in V \setminus \{0\}$, we define the *angle* $\measuredangle(v, w)$ between the vectors v und w as the uniquely determined number $\vartheta \in [0, \pi]$, for which

$$s(v, w) = \cos\vartheta \, \|v\| \|w\|.$$

We call $\varphi \in \mathrm{End}(V)$ *conformal* if φ is injective and if

$$\measuredangle(v, w) = \measuredangle(\varphi(v), \varphi(w)) \text{ for all } v, w \in V \setminus \{0\}.$$

Show that a linear map φ is conformal if and only if there exists an isometry $\psi \in \mathrm{End}(V)$ and a $\lambda \in \mathbb{R} \setminus \{0\}$ such that $\varphi = \lambda \cdot \psi$.

64. Find all the unitary matrices A such that $s(v, w) := \langle v, Aw \rangle$ defines an inner product on $\mathbb{C}^n$, where $\langle \ , \ \rangle$ denotes the canonical inner product on $\mathbb{C}^n$.

65. Let V be a finite dimensional vector space over $\mathbb{R}$. Show that each bilinear form on V can be uniquely written as the sum of a symmetric and a skew-symmetric bilinear form.

66. Let s be a symmetric bilinear form on a vector space V. If there are vectors $v, w \in V$, such that $s(v, w) \neq 0$, show that there is a vector $v \in V$, such that $s(v, v) \neq 0$.

67. Let V be the vector space of the complex-valued continuous functions on the unit circle in $\mathbb{C}$. a) Show that

$$\langle f, g \rangle := \int_0^{2\pi} f(e^{i\theta})\overline{g(e^{i\theta})} d\theta$$

defines an inner product on V.
b) Define the subspace $W \subseteq V$ by $W := \{f(e^{i\theta}) : f(x) \in \mathbb{C}[x] \text{ and } \deg(f) \leq n\}$. Find an orthonormal basis of W w.r.t. the above inner product.

68. Let A be the following 3×3 matrix:

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & 1 \end{pmatrix}.$$

(a) Without any computation, explain why there must exist a basis of $\mathbb{R}^3$ consisting only of the eigenvectors of A.
(b) Find such a basis of $\mathbb{R}^3$.
(c) Determine whether or not the bilinear form $s : \mathbb{R}^3 \to \mathbb{R}$ given by $s(u, v) := u^{\mathrm{T}} A v$ defines an inner product on $\mathbb{R}^3$.

69. (a) Let V be a finite dimensional vector space over $\mathbb{R}$ and let f and g be two linear functionals on V such that $\ker f = \ker g$. Show that there exists an $r \in \mathbb{R}$ such that $g = rf$.
(b) Let $\varphi_1, \varphi_2, \ldots, \varphi_5$ be linear functionals on a vector space V such that there does not exist any vector $v \in V$ for which $\varphi_1(v) = \varphi_2(v) = \cdots = \varphi_5(v)$. Show that $\dim V \leq 5$.

70. Let $w = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ and let the linear map $f : \mathbb{R}^3 \to \mathbb{R}$ be defined by

$$f(v) = v^{\mathrm{T}} w$$

for all $v \in \mathbb{R}^3$.
a) Find an orthonormal basis of $\operatorname{Ker} f$ w.r.t. dot product.
b) Extend this orthonormal basis of $\operatorname{Ker} f$ to an orthonormal basis of $\mathbb{R}^3$.

71. Let $P_2(\mathbb{R})$ denote the set of polynomials of degree ≤ 2 with real coefficients. Define the linear map $\phi : P_2(\mathbb{R}) \to \mathbb{R}$ by $\phi(f) = f(1)$. Determine $(\operatorname{Ker} \phi)^{\perp}$ with respect to the following inner product:

$$s(f, g) = \int_{-1}^{1} f(t)g(t)dt.$$

72. Let $P_3(\mathbb{R})$ denote the set of polynomials of degree ≤ 3 with real coefficients. On $P_3(\mathbb{R})$, we define the symmetric bilinear form s by

$$s(f, g) = \int_{-1}^{1} f(t)g(t)dt.$$

a) Determine the matrix representtation of s w.r.t. the basis $\{1, t, t^2, t^3\}$.
b) Show that s is positive definit.
c) Determine an orthonormal basis of $P_3(\mathbb{R})$.

73. Show that the eigenvectors associated with distinct eigenvalues of a self-adjoint matrix are orthogonal.

74. Let $A \in M_{n,n}(\mathbb{R})$ have eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{R}$ which are not necessarily distinct. Suppose $v_1, v_2, \ldots, v_n \in \mathbb{R}^n$ are eigenvectors of A associated with the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ respectively, such that $v_i \perp v_j$ if $i \neq j$. Show that A is symmetric.

75. Let $A \in M_{n,n}(\mathbb{R})$ a skew symmetric matrix. Let v und w be two eigenvectors of A corresponding respectively to the distinct eigenvalues λ_1 and λ_2. Show that v and w are orthogonal to each other (w.r.t. the dot product).

76. Let $A \in M_{n,n}(\mathbb{C})$ be a self-adjoint matrix. Show that the eigenvalues of A are real.

77. How many orthonormal bases (w.r.t. the dot product) are there in $\mathbb{R}^n$, so that all the entries of the basis vectors are integers?

78. Let $V = \mathbb{C}^n$, let $A \in M_{n,n}(\mathbb{C})$ a self-adjoint Matrix and let the linear operator $\phi_A : V \longrightarrow V$ be defined by $\phi_A(v) = Av$. Let W be a subspace of V, so that $\phi_A(W) \subseteq W$ (i.e. $\phi_A(w) \in W$ for all $w \in W$). Show that

$$\phi_A(W^{\perp}) \cap W = \{0\}.$$

79. Let $V = \mathbb{R}^2$ and let s a symmetric bilinear form on V. let $q : V \longrightarrow \mathbb{R}$ be the quadratic form corresponding to s given by

$$q\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = x^2 + 5xy + y^2.$$

Determine the matrix of s w.r.t. the basis $\mathbb{B} = \left\{\begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 2 \end{pmatrix}\right\}$ of $\mathbb{R}^2$.

80. Let V be a finite dimensional vector space over $\mathbb{R}$ with an inner product $\langle\ ,\ \rangle$ and let $f : V \to \mathbb{R}$ be a linear map. Show that there is an uniquely determined vector v_f such that for all $v \in V$, we have

$$f(v) = \langle v, v_f \rangle.$$

81. Given

$$A = \begin{pmatrix} 3 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & -1 \end{pmatrix} \in M_{3,3}(\mathbb{R}),$$

find a matrix $g \in \mathrm{GL}_3(\mathbb{R})$, such that $g^{\mathrm{T}} A g$ is of the form

$$\begin{pmatrix} I_k & & \\ & -I_l & \\ & & \mathrm{O} \end{pmatrix}.$$

82. Draw the curve $C := \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 \,\middle|\, 3x^2 + 4xy + 3y^2 = 5 \right\}$.

83. Let $X \in M_{n,n}(\mathbb{C})$ be a self-adjoint matrix and suppose m be a positive integer such that $X^m = I$. Show that $X^3 - 2X^2 - X + 2I = 0$.

84. Let $n \in \mathbb{Z}_{\geq 2}$. Show that $s(A, B) := \mathrm{tr}(A \cdot B^{\mathrm{T}})$ defines an inner product on $V = M_{n,n}(\mathbb{R})$. Let $\varphi \in \mathrm{End}(V)$ be defined by

$$\varphi(A) = A^{\mathrm{T}}.$$

(1) Show that φ is hermitian.
(2) Show that φ is an isometry.
(3) Find the eigenvalues of φ.
(4) Find an orthonormal basis $\mathbb{B}$ of V, made up of the eigenvectors ef φ.
(5) Find the algebraic multiplicities ef the eigenvalues of φ.

85. Let for $x \in \mathbb{R}$, the matrix A_x defined by

$$A_x := \frac{1}{1 + x + x^2} \begin{pmatrix} -x & x + x^2 & 1 + x \\ 1 + x & -x & x + x^2 \\ x + x^2 & 1 + x & -x \end{pmatrix}.$$

(1) Show that for all $x \in \mathbb{R}$, we have $A_x \in SO_3$.
(2) Conclude from (1) that for all real $x \neq \pm 1$, there exists a $g_x \in O_3$ and an $\alpha_x \in (0, \pi) \cup (\pi, 2\pi)$ such that

$$g_x A_x g_x^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha_x & -\sin \alpha_x \\ 0 & \sin \alpha_x & \cos \alpha_x \end{pmatrix}.$$

(3) Determine the complex eigenvalues of A_x for $x = 1 + \sqrt{2} + \sqrt{3} + \frac{1+\sqrt{3}}{\sqrt{2}}$.

86. (1) Find a matrix $g \in O_2$ which diagonalizes the matrix $A = \begin{pmatrix} 13 & 12 \\ 12 & 13 \end{pmatrix}$.
(2) Find a matrix $X \in M_{2,2}(\mathbb{R})$, which defines a scalar product through $s(v, w) = \langle v, Xw \rangle$ on $\mathbb{R}^2$ and which satisfies the following equation:

$$X^2 - A = 0.$$

87. Let $A \in M_{n,n}(\mathbb{R})$ be a symmetric matrix and let $B \in M_{n,n}(\mathbb{R})$ be a skew-symmetric matrix. Let $M = A + iB$ and let $v := \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$, where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of M. Show that

$$\|v\| = \sqrt{\sum_{j,k=1}^{n} |M_{jk}|^2},$$

w.r.t. the canonical norm on $\mathbb{C}^n$.

88. Let $\phi : \mathbb{C}^n \to \mathbb{C}^n$ be a nilpotent, hermitian endomorphism. Show that: $\phi = 0$.

89. Let $A, B \in M_{n,n}(\mathbb{C})$ be two self-adjoint matrices. Show that the following are equivalent:
(1) There is an unitary matrix g such that both gAg^{-1} and gBg^{-1} are diagonal matrices.
(2) The matrix AB is self-adjoint.
(3) $AB = BA$.

90. (1) Let $A, B \in M_{n,n}(\mathbb{C})$ be nilpotent matrices such that $AB = BA$ holds. Show that $A+B$ is nilpotent.
(2) Let $A, B \in M_{n,n}(\mathbb{C})$ and $r, s \in \mathbb{Z}_{>0}$ such that $A^r = I$, $B^s = 0$ and $AB = BA$. Show that $A - B$ is invertible.

91. Let

$$A = \begin{pmatrix} 1 & -2 & 2 \\ 0 & -2 & 1 \\ -2 & 1 & -2 \end{pmatrix} \in M_{3,3}(\mathbb{R}).$$

(1) Find a decomposition $A = D+N$, where D is a diagonal matrix an N is a nilpotente Matrix.
(2) Berechnen Sie A^{2012}.

92. Let $A \in M_{n,n}(\mathbb{R})$ be a nilpotent matrix and let $V = M_{n,n}(\mathbb{R})$. Let $\varphi \in \mathrm{End}(V)$ defined by

$$\varphi(B) = AB - BA \quad \text{for } B \in V.$$

Show that φ is nilpotent on V.

93. Let $V = \mathbb{R}^n$ with $s = \langle \cdot, \cdot \rangle$ and let $\mathbb{B} = \{v_1, \dots, v_n\}$ an orthonormal basis of V. Let $U_i = (\mathrm{span}\{v_i\})^\perp$ for $i \in \{1, \dots, n\}$. Show that

$$S_{U_i} \circ S_{U_j} = S_{U_j} \circ S_{U_i}$$

for $i, j \in \{1, \dots, n\}$, where S_{U_i} and S_{U_j} are the reflections in U_i and U_j.

94. Let V be a finite dimensional vectore space and let $P \in \mathrm{End}(V)$ be a projektion. Let $\mathrm{Id} \in \mathrm{End}(V)$ the identity map of V (i.e. $\mathrm{Id}(v) = v$ for all $v \in V$). Show that
(1) $\mathrm{Id} - P$ is a projektion.
(2) $\mathrm{Id} - 2P$ is bijective.
(3) $E_0 \oplus E_1 = V$, where E_0 and E_1 are respectively the eigenspaces of P corresponding to the eigenvalues 0 and 1.

95. Let $A \in M_{n,n}(\mathbb{C})$ and let $B = A - A^*$. Show that B is diagonalizable and the real parts of all the eigenvalues of B are zero.

96. Let $A \in \mathrm{SO}_2$. Show that there is a skew symmetric matrix $X \in M_{2,2}(\mathbb{R})$, such that

$$\exp(X) = A.$$

97. Let $V = \mathbb{R}^5$ and let $\ell \in V^*$ be given by $\ell(v) = v_1 + 2v_2 + 3v_3 + 4v_4 + 5v_5$ für $v = \begin{pmatrix} v_1 \\ \vdots \\ v_5 \end{pmatrix} \in V$.

(1) Find an orthonormal basis of $\ker \ell$ w.r.t. the dot product.
(2) Extend this basis of $\ker \ell$ to an orthonormal basis of V.

98. Let $V = \mathbb{R}^4$, let

$$A = \frac{1}{2} \begin{pmatrix} 2 & 1 & 2 & -3 \\ 1 & 2 & -3 & 2 \\ 2 & -3 & 2 & 1 \\ -3 & 2 & 1 & 2 \end{pmatrix} \in M_{4,4}(\mathbb{R})$$

and let s be the symmetric bilinear form whose associated matrix is A.
(1) Determine a basis $\mathbb{A}$ of V, such that $M_{\mathbb{A}}(s)$ is a diagonal matrix.
(2) Determine a basis $\mathbb{B}$ of V, such that

$$M_{\mathbb{B}}(s) = \begin{pmatrix} I_k & & \\ & -I_l & \\ & & \mathrm{O} \end{pmatrix}.$$

99. Let $V = \mathbb{R}^3$ with $s = \langle \cdot, \cdot \rangle$ (the dot product), let $U = \operatorname{span}\left\{ \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \right\}$ be a subspace of V and let S_U be the reflection in in U.
(1) Determine a matrix representation of S_U, w.r.t. the canonical basis $\mathbb{A}$ of V.
(2) Show that $M(S_U)_{\mathbb{A}} \in \mathrm{O}_3$ and decide whether $M(S_U)_{\mathbb{A}} \in \mathrm{SO}_3$ oder $M(S_U)_{\mathbb{A}} \notin \mathrm{SO}_3$ or not.

Introduction to linear programming

100. Maximize $f(x, y, z) := 6x + 3y + 10z$ using Simplex method under the following constraints:

$$4x + y + z \leq 5,$$

$$2x + y + 4z \leq 5,$$

$$x + 5y + z \leq 6,$$

where x, y and z are non-negative rational numbers.

101. Minimize $f(x, y, z) := x + 2y + 9z$ using big-M method under the following constraints:

$$2x + y + 4z \geq 5,$$

$$2x + 3y + z \geq 4,$$

where x, y and z are non-negative rational numbers.

102. (a) A convex linear combination of $v_1, v_2, \ldots, v_n \in \mathbb{R}^m$ is a a linear combination of the form $t_1v_1 + \cdots + t_nv_n$, where $t_1 + \cdots + t_n = 1$.For example, the points on the straight line connecting v_1 and v_2 is given by $tv_1 + (1-t)v_2$, where t lies in the interval $[0,1] \subset \mathbb{R}$. Show that any arbitrary point in a triangle in $\mathbb{R}^m$ with vertices v_1, v_2 and v_3 is given by a convex linear combination of its vertices.
(b) Show that any arbitrary point in a tetrahedron in $\mathbb{R}^m$ with vertices v_1, v_2, v_3 and v_4 is given by a convex linear combination of its vertices.

103. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by $f(x,y) := 2x + 3y$. Find the maximum value attained by f in the region where $2y - x \leq 10$, $3x + 2y \leq 9$ and $2x + 5y \geq 8$.

104. Maximize $f(x,y,z) := 2x + 5y + 3z$ using Simplex method under the following constraints:

$$14x + 8y + 5z \leq 15,$$

$$12x + 7y + 8z \leq 14,$$

$$3x + 17y + 9z \leq 16,$$

where x, y and z are non-negative rational numbers.

105. Minimize $f(x,y,z) := x + 9y + 9z$ using big-M method under the following constraints:

$$6x + y + 5z \geq 11,$$

$$4x + 7y + 2z \geq 9,$$

where x, y and z are non-negative rational numbers.

106. (a) Recall that any arbitrary point in a convex polyhedron is given by a convex linear combination of its vertices. Using this, show that the minimum and the maximum values attained by a linear functional $f : \mathbb{R}^n \to \mathbb{R}$ in a convex polyhedron $\mathcal{P} \subset \mathbb{R}^n$ is the same as the minimum and the maximum values attained by f at the set of the vertices of $\mathcal{P}$.
(b) Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by $f(x,y) := 5x - 3y$. Find the maximum value attained by f in the region where $4y - 3x \leq 10$, $7x + 2y \leq 9$ and $2x + 5y \geq 8$.

107. Maximize $f(x,y,z) := 3x + y + 3z$ using Simplex method under the following constraints:

$$2x + y + z \leq 2,$$

$$x + 2y + 3z \leq 5,$$

$$2x + 2y + z \leq 6,$$

where x, y and z are non-negative rational numbers.

108. Maximize $f(x,y,z) := 3x + y + 4z$ using big-M method under the following constraints:

$$x + 3y + 4z \leq 20,$$

$$2x + y + z \geq 8,$$

$$3x + 2y + 3z = 18,$$

where x, y and z are non-negative rational numbers.

DA220 Machine Learning

Instructor: **Tanmay Basu**

Course Description:
DA220 deals with topics in supervised and unsupervised learning methodologies. In particular, the course will cover different advanced models of data classification and clustering techniques, their merits and limitations, different use cases and applications of these methods. Moreover, different advanced unsupervised and supervised feature engineering schemes to improve the performance of the learning techniques will be discussed.

Prerequisite(s): (1) Linear Algebra and (2) Probability and Stochastic processes
Credit Hours: 4

Text(s):

Introduction to Machine Learning E. Alpaydin ISBN: 978-0262-32573-8
The Elements of Statistical Learning J. H. Friedman, R. Tibshirani, and T. Hastie ISBN: 978-0387-84884-6
Pattern Recognition S. Theodoridis and K. Koutroumbas ISBN: 0-12-685875-6
Pattern Classification R. O. Duda, P. E. Hart and D. G. Stork ISBN: 978-0-471-05669-0
Introduction to Information Retrieval C. D. Manning, P. Raghavan and H. Schutze ISBN: 978-0-521-86571-5

Course Objectives:

Knowledge Acquired:

1) The background and working principles of various supervised learning techniques viz., linear regression, logistic regression, bayes and naive bayes classifiers, support vector machine etc. and their applications.
2) The importance of cross validation to optimize the parameters of a classifier.
3) The idea of different kinds of clustering techniques e.g., k-means, k-medoid, single-linkage, DBSCAN algorithms and their merits and demerits.
4) The significance of feature engineering to improve the performance of the learning techniques and overview of various supervised and unsupervised feature engineering techniques.
5) The essence of different methods e.g., precision, recall etc. to evaluate the performance of the machine learning techniques.

Skills Gained: The students will be able to

1) pre-process and analyze the characteristics of different types of standard data,
2) work on scikit-learn, a standard machine learning library,
3) evaluate the performance of different machine learning techniques for a particular application and validate the significance of the results obtained.

Competence Developed:

1) Build skills to implement different classification and clustering techniques as per requirement to extract valuable information from any type of data set.
2) Can train a classifier on an unknown data set to optimize its performance
3) Develop novel solutions to identify significant features in data e.g., identify the feedback of potential buyers over online markets to increase the popularity of different products.

Evaluation:
Assignments 50% Midterm Exam 25% Endterm Exam 25%

Course Outline (tentative) and Syllabus:
The weekly coverage might change as it depends on the progress of the class. However, you must keep up with the reading assignments. Each week assumes 4 hour lectures.

Week	Contents
Week 1	• Overview of machine learning: idea of supervised and unsupervised learning, regression vs classification, concept of training and test set, classification vs clustering and significance of feature engineering • Linear regression: least square and least mean square methods
Week 2	• Bayes decision rule: bayes theorem, bayes classifier and error rate of bayes classifier • Minimum distance classifier and linear discriminant function as derived from Bayes decision rule
Week 3	• Naive bayes classifier: gaussian model, multinomial model, bernoulli model • k-Nearest Neighbor (kNN) decision rule: idea of kNN classifier, distance weighted kNN decision rule and other variations of kNN decision rule
Week 4	• Perceptron learning algorithm: incremental and batch version, proof of convergence • XOR problem, two layer perceptrons to resolve XOR problem, introduction to multilayer perceptrons
Week 5	• Discussion on different aspects of linear discriminant functions for data classification • Logistic regression and maximum margin classifier
Week 6	• Support vector machine (SVM): hard margin • Soft margin SVM classifier
Week 7	• Cross validation and parameter tuning • Different techniques to evaluate the classifiers e.g., precision, recall and f-measure
Week 8	• The basics to work with Scikit-learn: a machine learning repository in python • How to implement different classifiers in scikit-learn, tune the parameters and evaluate the performance
Week 9	• Text classification(case study for data classification): overview of text data, stemming and stopword removal, tf-idf weighting scheme and n-gram approach. • How to work with text data in scikit-learn
Week 10	• Assignment 2: Evaluate the performance of different classifiers to classify a newswire e.g., Reuters-21578. • Review for midterm exam • Data clustering: overview, cluster validity index
Week 11	• Partitional clustering methods: k-means, bisecting k-means • k-medoid, buckshot clustering techniques
Week 12	• Hierarchical clustering techniques: single linkage, average linkage and group average hierarchical clustering algorithms • Density based clustering technique e.g., DBSCAN
Week 13	• Feature engineering: overview of feature selection, supervised and unsupervised feature selection techniques • Overview of principal component analysis for feature extraction
Week 14	• How to work with Wordnet, an English lexical database • Sentiment analysis (case study for data clustering): overview, description of a data set of interest for sentiment identification, sentiment analysis using Wordnet
Week 15	• Assignment 2: Sentiment analysis from short message texts • Practice class for the second assignment • Review for endterm exam

DA321 Modeling for Operations Management

Instructor

Sudeep Mallick, Ph.D.
Sudeep.mallick@gmail.com

Course Description:

DA321 deals with the topics in modelling techniques for accomplishing operations management tasks for business. In particular, the course will cover advanced techniques of operations research and modelling along with their applications in various business domains with a special focus on supply chain management and supply chain analytics.

Prerequisite(s): Basic course in Operations Research covering Linear Programming fundamentals.
Credit Hours: 4

Text(s):

Operations Research, seventh revised edition (2014)
P K Gupta and D S Hira
ISBN: 81-219-0218-9

Introduction to Operations Research, eighth edition
Frederick S. Hillier & Gerald J. Lieberman
ISBN: 0-07-252744-7

Operations Research: An Introduction, ninth edition
Hamdy A. Taha
ISBN: 978-93-325-1822-3

AMPL: A Modeling Language for Mathematical Programming, second Edition
www.ampl.com

Course Objectives:

Knowledge acquired:

1. Different operations research modelling techniques.
2. Application of the modelling techniques in business domains.
3. Hands-on implementation of the models using computer software such as MS-EXCEL, CPLEX solvers.

Skills acquired: Students will be able to

1. apply the appropriate operations research technique to formulate mathematical models of the business problem
2. implement and evaluate alternative models of the problem in computer software

Grade Distribution:

Assignments 20%, Internal Test 20%, Mid-term exam 30%, Final exam 30%

Course Outline (tentative) and Syllabus:

Week	Content
Week 1	• Advanced Linear Programming: Duality theory, Dual Simplex method • Reading assignment: Chapter 6, GH / Chapter 4, HT
Week 2	• Lab session on Linear Programming and Sensitivity Analysis with AMPL (CPLEX solver) • Lab assignment 1, Reading assignment: AMPL manual
Week 3	• Supply chain management modelling: supply chain management definition, modelling, production planning decisions • Reading assignment: Instructor notes
Week 4	• Lab session on modelling aggregate planning problems
Week 5	• Transportation problem: transportation model, solution techniques, variations. • Reading assignment: Chapter 3, GH / Chapter 5, HT • Transportation problem Lab sessions • Lab instructions: Instructor notes
Week 6	• Multi-stage transportation problem: formulation, solution techniques, truck allocation problem, Traveling Salesman Problem, vehicle routing problem • Reading assignment: Instructor notes • Internal test 1
Week 7	• Assignment problem: assignment, solution techniques • Reading assignment: Chapter 4, GH / Chapter 5. HT • Lab assignment 2
Week 8	• Integer programming: problem formulation and solution techniques • Reading assignment: Chapter 6, GH / Chapter 9, HT • Review for Midterm Exam
Week 9	• Non-linear Programming: problem formulation and solution techniques • Reading assignment: Chapter 16, GH / Chapter 21, HT • Lab assignment 3
Week 10	• Inventory management: deterministic inventory models, cycle inventory models • Reading assignment: Chapter 12, GH / Chapter 13, HT • Internal test 2
Week 11	• Inventory management: stochastic inventory models, safety stock models • Reading assignment: Chapter 12, GH / Chapter 13, HT • Lab session: Inventory management modeling • Reading assignment: Instructor notes
Week 12	• Lab Session: Supply chain management beer game
Week 13	• Queueing theory: pure birth and death models • Reading assignment: Chapter 10, GH / Chapter 18, HT • Reading assignment: Chapter 10, GH / Chapter 18, HT
Week 14	• Queueing theory: general poisson model, specialised poisson queues • Lab session: queueing theory • Reading assignment: Chapter 10, GH / Chapter 18, HT • Lab assignment 4
Week 15	• Queueing theory: queueing decision models • Reading assignment: Chapter 10, GH / Chapter 18, HT

CS 244 : Introduction to Optimization Techniques

Course Overview: The process of making optimal judgement according to various criteria is known as the science of decision making. A mathematical programming problem, also known as an optimization problem, is a special class of problem where we are concerned with the optimal use of limited resources to meet some desired objective(s). Mathematical models (simulation based and/or analytical based) are used in providing guidelines for making effective decisions under constraints. This course covers three major analytical topics in mathematical programming [linear, nonlinear and integer programming]. On each topic, the theory and modeling aspects are discussed first, and subsequently solution techniques or algorithms are covered.

Prerequisite(s): Linear Algebra
Credit Hours: 4
Course Objectives: Optimization techniques are used in various fields like machine learning, graph theory, VLSI design and complex networks. In all these applications/fields, mathematical programming theory supplies the notion of optimal solution via the optimality conditions, and mathematical programming algorithms provide tools for training and/or solving large scale models. Students will have knowledge of theory and applications of several classes of math programs.

Text(s): The course material will be drawn from multiple book chapters, journal articles, reviewed tutorials etc. However, the following two books are recommended texts for this course.

- *Linear programming and Network Flows*, Wiley-Blackwell; 4th Edition, 2010
  M. S. Bazaraa, John J. Jarvis and Hanif D. Sheral, **ISBN-13:** 978-0470462720
- *Nonlinear Programming: Theory and Algorithms*, Wiley-Blackwell; 3rd Edition (2006)
  M. S. Bazaraa, Hanif D. Sherali, C. M. Shetty, **ISBN-13:** 978-0471486008

Course Policies:

- **Grades**

  Grades in the **C** range represent performance that **meets expectations**; Grades in the **B** range represent performance that is **substantially better** than the expectations; Grades in the **A** range represent work that is **excellent**.

- **Assignments**

  1. Students are expected to work independently. Discussion amongst students is encouraged but offering and accepting solutions from others is an act of dishonesty and students can be penalized according to the *Academic Honesty Policy*.
  2. **No late assignments will be accepted under any circumstances**.

- **Attendance and Absence**

  Students are not supposed to miss class without prior notice/permission. Students are responsible for all missed work, regardless of the reason for absence. It is also the absentee's responsibility to get all missing notes or materials.

Grade Distribution:

Assignments	40%
Midterm Exam	20%
Final Exam	40%

Grading Policy: *Approximate grade assignments:*

>= 90.0 %	A+
75.0 – 89.9 %	A
60.0 – 74.9 %	B
50.0 – 59.9 %	C
about 35.0 – 49.9 %	D
<= 34.9%	F

Table 1: **Topics Covered**

Mathematical Preliminaries

- Theory of Sets and Functions,
- Vctor spaces,
- Matrices and Determinants,
- Convex sets and convex cones,
- Convex and concave functions,
- Generalized concavity

Linear Programming

- The (Conventional) Linear Programming Model
- The Simplex Method: Tableau And Computation
- Special Simplex Method And Implementations
- Duality And Sensitivity Analysis

Integer Programming

- Formulating Integer Programing Problems
- Solving Integer Programs (Branch-and-Bound Enumeration, Implicit Enumeration, Cutting Plane Methods)

Nonlinear Programming: Theory

- Constrained Optimization Problem (equality and inequality constraints)
- Necessary and Suffiecent conditions
- Constraint Qualification
- Lanrangian Duality and Saddle Point Optimality Criteria

Nonlinear Programming: Algorithms

- The concept of Algorithm
- Algorithms for Uconstrained Optimization
- Constraint Qualification
- Algorithms for Constrained Optimization (Penalty Function, Barrier Function, Feasible Direction)

Special Topics (if time permits)

- Semi-definite and Semi-infinte Programs
- Quadratic Programming
- Linear Fractional programming
- Separable Programming

CS111: Foundations of Statistical Learning: Probability

Instructor
Brahmachari Aditya

Course Description:
The course covers basic principles of probability, counting methods, discrete and continuous random variables, joint distributions, expectation values and other features of statistical distributions and limit theorems. Based on time availability and interest, the course may also include a brief discussion on stochastic processes.

Prerequisite(s): (1) Basic knowledge of algebra, calculus, set theory.

Credit Hours: 2

Text(s):
1. A first course in Probability, Sheldon M. Ross, Pearson Education, 2010
2. P. G. Hoel, S. C. Port and C. J. Stone: Introduction to Probability Theory, University Book Stall/ Houghton Mifflin, New Delhi/New York, 1998/1971.
3. Introduction to Probability Models (11^{th} Edition), Sheldon M. Ross, Academic Press, 2014

Course Objectives:

Knowledge acquired: The students are expected to gain a basic understanding of the mathematical approach to dealing with uncertainty. By the end of the course, the students may be conversant in the basic principles of probability, random variables and statistical approach to dealing with them. They may also learn some of the basic features of stochastic processes.

Skills acquired: Students will be conversant in the mathematical framework for understanding probability, including counting methods involving permutations and combinations. They would also gain familiarity in handling basic statistical principles and tools that are employed in the analysis of random variables and probability distributions.

Grade Distribution:

Assignments	30%
Mid-term exam	30%
Final exam	40%

Grading Policy:
There will be absolute grading given by the following:

>=90% marks	A+
>=75%	A
>=60%	B
>=50%	C
>=35%	D
< 35%	F

Course Policies:

Course Policies:

- General course policies, Grades, Labs and assignments, Attendance and Absences These clauses are common to all courses. And it can be found in the program schedule

Probability and Stochastic Process

1. Basic Probability

a. Introduction
b. Sample Spaces
c. Probability Measures
d. Computing Probabilities: Counting Methods
  i. The Multiplication Principle
  ii. Permutations and Combinations
e. Conditional Probability
f. Independence

2. Random Variables

a. Discrete Random Variables
  i. Bernoulli Random Variables
  ii. The Binomial Distribution
  iii. Geometric and Negative Binomial Distributions
  iv. The Hypergeometric Distribution
  v. The Poisson Distribution
b. Continuous Random Variables
  i. The Exponential Density
  ii. The Gamma Density
  iii. The Normal Distribution
  iv. The Beta Density
c. Functions of a Random Variable

3. Joint Distributions

a. Introduction
b. Discrete Random Variables
c. Continuous Random Variables
d. Independent Random Variables
e. Conditional Distributions
  i. The Discrete Case
  ii. The Continuous Case
f. Functions of Jointly Distributed Random Variables
  i. Sums and Quotients
  ii. The General Case

4. Expectation Values

a. The Expectation Value of a Random Variable

i. Expectations of Functions of Random Variables
ii. Expectation of Linear Combinations of Random Variables

b. Variance and Standard Deviation
c. Covariance and Correlation
d. Conditional Expectation
e. Definitions and Examples
f. The Moment-Generating Function

5. Limit Theorems

a. Introduction
b. The Law of Large Numbers
c. Convergence in Distribution and the Central Limit Theorem

6. Stochastic Process (depending on time availability)

a. Markov chain
   i. State transition matrix
   ii. Hitting time
   iii. Different States
b. Poisson process

DA104 Probability and Stochastic Processes

Instructor

Dr. Arijit Chakraborty (ISI Kolkata)

Course Description:

DA104 deals with technologies and engineering solutions for enabling big data processing and analytics . More specifically, it deals with the tools for data processing, data management and programming in the distributed programming paradigm using techniques of MapReduce programming, NoSQL distributed databases, streaming data processing, data injestion, graph processing and distributed machine learning for big data use cases.

Prerequisite(s): (1) Basic knowledge of python and Java programming languages (2) Tabular data processing / SQL queries. (3) Basic knowledge of common machine learning algorithms.
Credit Hours: 4

Text(s):

1. Introduction to time series analysis; PJ Brockwell and RA Davis
2. Time Series Analysis and Its Applications; Robert H. Shumway and David S. Stoffer
3. Introduction to Statistical time series; WA Fuller
4. A first course in Probability, Sheldon Ross, Pearson Education, 2010
5. Time Series Analysis; Wilfredo Palma
6. P. G. Hoel, S. C. Port and C. J. Stone: Introduction to Probability Theory, University Book Stall/Houghton Mifflin, New Delhi/New York, 1998/1971.

Syllabus

1. Basic Probability
a. Introduction
b. Sample Spaces
c. Probability Measures
d. Computing Probabilities: Counting Methods
  i. The Multiplication Principle
  ii. Permutations and Combinations
e. Conditional Probability
f. Independence

2. Random Variables
a. Discrete Random Variables
  i. Bernoulli Random Variables
  ii. The Binomial Distribution
  iii. Geometric and Negative Binomial Distributions
  iv. The Hypergeometric Distribution
  v. The Poisson Distribution
b. Continuous Random Variables

i. The Exponential Density
ii. The Gamma Density
iii. The Normal Distribution
iv. The Beta Density
c. Functions of a Random Variable

3. Joint Distributions
a. Introduction
b. Discrete Random Variables
c. Continuous Random Variables
d. Independent Random Variables
e. Conditional Distributions
i. The Discrete Case
ii. The Continuous Case
f. Functions of Jointly Distributed Random Variables
i. Sums and Quotients
ii. The General Case

4. Expected Values

a. The Expected Value of a Random Variable
i. Expectations of Functions of Random Variables
ii. Expectation of Linear Combinations of Random Variables
b. Variance and Standard Deviation
c. Covariance and Correlation
d. Conditional Expectation
e. Definitions and Examples
f. The Moment-Generating Function

5. Limit Theorems
a. Introduction
b. The Law of Large Numbers
c. Convergence in Distribution and the Central Limit Theorem

6. Stochastic Process
a. Markov chain
i. State transition matrix
ii. Hitting time
iii. Different States
b. Poisson process

DA311

Time Series

Time: TBA
Place: IH402 & Bhaskara Lab

Dr. Sudipta Das

jusudipta@gmail.com
Office: IH404, Prajnabhavan, RKMVERI, Belur
Office Hours: 11 pm—12 noon, 3 pm—4 pm
(+91) 99039 73750

Course Description: DA311 is going to provide a broad introduction to the most fundamental methodologies and techniques used in time series analysis.

Prerequisite(s): (1) Probability & Stochastic Process and (2) Linear Algebra.
Note(s): Syllabus changes yearly and may be modified during the term itself, depending on the circumstances. However, students will be evaluated only on the basis of topics covered in the course.
Course url:
Credit Hours: 4

Text(s):

Introduction to time series analysis;
PJ Brockwell and RA Davis

Time Series Analysis and Its Applications;
Robert H. Shumway and David S. Stoffer

Introduction to Statistical time series;
WA Fuller

Time Series Analysis;
Wilfredo Palma

Course Objectives:

Knowledge acquired: Students will get to know
(1) Different time series models MA, AR, ARMA, ARIMA
(2) Autocorrelation and Partial Autocorrelation functions,
(3) Method of time series modelling, in presence of seasonality, and,
(4) Different non-linear time series models such as ARCH and GARCH.
Skills gained: The students will be able to
(1) explore trend and seasonality in time series data by exploratory data analysis,
(2) implement stationary as well as non-stationary models through parameter estimation,
(3) compute forecast for time series data.

Grade Distribution:

Assignments	20%
Quizzes	10%
Midterm Exam	20%
Final Exam	50%

Grading Policy: There will be relative grading such that the cutoff for A grade will not be less than 75% and cutoff for F grade will not be more than 34.9%. Grade distribution will follow normal bell curve (usually, A: $\geq \mu + 3\sigma/2$, B: $\mu + \sigma/2 \ldots \mu + 3\sigma/2$ C: $\mu - \sigma/2 \ldots \mu + \sigma/2$, D: $\mu - 3\sigma/2 \ldots \mu - \sigma/2$, and F: $< \mu - 3\sigma/2$)

Approximate grade assignments:

>= 90.0	A+
75.0 – 89.9	A
60.0 – 74.9	B
50.0 – 59.9	C
about 35.0 – 49.9	D
<= 34.9	F

Course Policies:

- **General**
  1. Computing devices are not to be used during any exams unless instructed to do so.
  2. Quizzes and exams are closed books and closed notes.
  3. Quizzes are unannounced but they are frequently held after a topic has been covered.
  4. **No makeup quizzes or exams will be given.**
- **Grades**

  Grades in the **C** range represent performance that **meets expectations**; Grades in the **B** range represent performance that is **substantially better** than the expectations; Grades in the **A** range represent work that is **excellent**.
- **Labs and Assignments**
  1. Students are expected to work independently. **Offering** and **accepting** solutions from others is an act of dishonesty and students can be penalized according to the *Academic Honesty Policy*. Discussion amongst students is encouraged, but when in doubt, direct your questions to the professor, tutor, or lab assistant. Many students find it helpful to consult their peers while doing assignments. This practice is legitimate and to be expected. However, it is not acceptable practice to pool thoughts and produce common answers. To avoid this situation, it is suggested that students not write anything down during such talks, but keep mental notes for later development of their own.
  2. **No late assignments will be accepted under any circumstances**.
- **Attendance and Absences**
  1. Attendance is expected and will be taken each class. Students are not supposed to miss class without prior notice/permission. Any absences may result in point and/or grade deductions.
  2. Students are responsible for all missed work, regardless of the reason for absence. It is also the absentee's responsibility to get all missing notes or materials.

Course Outline (tentative) and Syllabus:
The weekly coverage might change as it depends on the progress of the class. However, you must keep up with the reading assignments. Each week assumes 4 hour lectures. Quizzes will be unannounced.

Week	Content
Week 1	• The Nature of Time Series Data • Financial, Economic, Climatic, Biomedical, Sociological Data. • Reading assignment: Chapter 1, BD
Week 2	• Time Series Statistical Models • Components of time series: Trend, Seasonality and randomness • Whiteness Testing • Quiz 1
Week 3	• Stationary time series • Linear process • Strong and weak stationarity • Causality, invertibility and minimality • Reading assignment: Chapter 2, BD
Week 4	• Auto Regressive model • Moving Average model • Auto Regressive model • Moving Average models
Week 5	• Auto-covariance Function • Auto-correlation Function • Partial Auto-correlation Function • Reading assignment: Chapter 3, BD
Week 6	• Estimating Sample mean, • Estimating Auto-correlation function • Estimating Partial autocorrelation functions • Quiz 2
Week 7	• YuleWalker estimation • Burgs algorithm • Maximum Likelihood Estimation • Reading assignment: Chapter 5, BD
Week 8	• Order Selection • The AIC, BIC and AICC criterion • Review for Midterm Exam

Week	Content
Week 9	• Forecasting • Minimum MSE Forecast • Forecast Error
Week 10	• Forecasting Stationary Time Series • The DurbinLevinson Algorithm • The Innovations Algorithm
Week 11	• Non-stationarity time series • Unit root tests • Reading assignment: Chapter 6, BD
Week 12	• ARIMA Processes • Forecasting ARIMA Models • Quiz 3
Week 13	• Modelling seasonal time series • Seasonal ARIMA Models • Forecasting SARIMA Processes
Week 14	• Nonlinear Time Series • Testing for Linearity • Heteroskedastic Data
Week 15	• Auto-regressive conditional heteroskedastic model • Generalized auto-regressive conditional heteroskedastic model • Reading assignment: Chapter 5, SS • Review for Final Exam