University of the Witwatersrand, Johannesburg
School of Computer Science and Applied Mathematics

Computer Science Honours/Big Data Analytics Honours Courses and Syllabuses

January 2023

The following list attempts to offer a general information only. For exact details, one must adhere to the respective course outlines provided by the lecturers.

Please take note that some courses are **compulsory** for your academic plan. The information can be found in the **Rules and Syllabuses 2023** of Science faculty (`https://www.wits.ac.za/students/academic-matters/rules-and-syllabuses/`).

1 Adaptive Computation and Machine Learning

Course code: COMS4030A

Offered: Semester 1

Course description: Building systems that can adapt and learn from their environments has been a core area of research in the broader field of Artificial Intelligence and has thus attracted researchers from fields such as computer science, engineering, neuroscience, cognitive science and many other diverse fields. Machine learning has been a key enabler in building adaptive AI systems. This course aims at providing a broad overview of machine learning principles and practises.

Course prerequisite/prior knowledge: i) Familiarity with Python programming; ii) Familiarity with basic linear algebra, calculus, probability and statistics.

Course contents: See the list below.

- Introduction to Machine Learning
- Supervised Learning (esp. Neural Networks)
- Unsupervised Learning (e.g., Clustering)
- Semi-supervised Learning
- Topics in Deep Learning (e.g., CNNs, RNNs, GANs, Autoencoders)
- Reinforcement Learning

2 High Performance Computing and Scientific Data Management

Course code: COMS4040A

Offered: Semester 1

Course description: High Performance Computing & Scientific Data Management. High performance computing (HPC) deals with implementations of given algorithms, and hardware they run on. Similar concepts include parallel computing and super computing. The problems that are solved using HPC require huge amounts of computing and memory resources that a single computer such as a PC cannot afford. HPC enables us solving problems, doing research, or engineering design using computer modelling, simulation, and analysis. In this course, we will study basic architectural features of modern processors, basics of parallel computing, parallel programming models for shared- and distributed-memory systems, graphics processing unit (GPU)-based computing, analytical modelling of parallel programs, along with management of large datasets in a high performance computing environment.

Course prerequisite/prior knowledge i) Familiarity with programming in C or C++. ii) Familiarity with the basics of data structures, algorithms, analysis of algorithms, and machine architectures. iii) Parallel Computing (COMS3008A) or a similar course.

Course contents: The course covers the following topics (subject to some necessary changes in 2023):

- Modern processors
  - General-purpose cache-based microprocessor architecture
  - Memory hierarchies
  - Multicore processors
- Basic optimization techniques for serial code
- Data access optimizations
- Parallel computers
- Basics of parallelization
- Shared memory parallel programming with OpenMP
- Programming GPUs as co-processors to CPU using CUDA programming model
- Distributed memory parallel programming with MPI
- Hybrid computing
  - Combining OpenMP and CUDA
  - Combining MPI and OpenMP

3 Robotics

Course code: COMS4045A

Offered: Semester 1

Course description: Robotics is one of the world's fastest growing fields, with the potential to revolutionise society. Robots come in a wide range of physical forms, and are designed to operate in many different settings. This course looks at how robots perceive the world, build models and plans, and take actions to manipulate their environment. This is done by focusing on algorithms related to sensing, learning, planning, and control.

Course prerequisite/prior knowledge: Machine Learning.

Course contents: See the list below.

- Characteristics of a robot
- Path planning

- Kinematics and dynamics
- Motion planning
- PID and Optimal Control
- State estimation
- Localization and mapping
- Implementation on ROS (Robot Operating System)

4 Mathematical Foundations of Data Science

Course code: COMS4055A

Offered: Semester 1

Course description: Advanced areas of data science require a deeper understanding of the fundamental mathematics pertaining to the field. This course aims to bridge the mathematical gap that many data scientists face. The broad areas of mathematics under consideration are linear algebra, analytical geometry, matrix decompositions, vector calculus, fundamental probability theory, and continuous optimization. Additionally, selected classic machine learning problems are considered as an application of the presented mathematics.

It should be clear that all of the aforementioned areas of mathematics cannot be condensed into one course, however the parts directly related to data science most certainly can. While this module will be relatively dense in content it is imperative for a data scientist to have sufficient mastery of the mathematics presented to be capable of developing new approaches, accurately implementing, and using state of the art techniques.

Course prerequisite/prior knowledge:

Course contents: See the list below.

- Linear Algebra
- Analytic Geometry
- Matrix Decompositions
- Vector Calculus
- Probability and Distributions
- Continuous Optimization
- Central Machine Learning Problems and Concepts

5 Introduction to Research Methods

Course code: COMS4057A

Offered: Semester 1

Course description: The Introduction to Research Methods (IRM) course introduces the core components of an academic research project and aims to equip students with the subject-neutral skills required to plan and execute such a project within the field of computer science. Students go on to apply the skills and knowledge acquired in the course in the Honours Research Project course in the second half of the year. Topics covered include the formulation of a research question or problem, identification and analysis of literature and the methods relevant to a research problem, and the preparation of a research proposal.

Course prerequisite/prior knowledge: None.

6 Digital Image Processing

Course code: APPM4058A

Offered: Semester 1

Course description: Digital Image Processing (DIP) involves processes such as image acquisition, noise reduction, contrast enhancement, image sharpening, segmentation, image coding, image description, and object recognition (or classification). In this course we study a broad range of techniques for various DIP processes. These DIP processes often form the building blocks for higher level of image analysis, or image understanding tasks.

Course prerequisite/prior knowledge: Equivalent of a second year course in mathematics, and familiarity with programming in Matlab or Python.

Course contents: The course covers the following topics:

- Introduction to digital image processing, its scope and applications
- Intensity transforms and spatial filtering
- Filtering in the frequency domain
- Image restoration
- Mathematical morphology, morphological image processing, and their applications
- Image segmentation
- Colour image processing
- Image coding and compression
- Image representation of data

7 Research Project for Computer Science Honours

Course code: COMS4059A

Offered: Full year.

Course description: This course introduces candidates to research activities. A student is expected to work with a specific problem in computer science under the guidance of a supervisor. The work includes essential components of doing research in a thorough manner. It includes conducting literature survey, formulating the research problem and hypothesis, designing and implementing a feasible and effective research methodology to solve the proposed problem, and present the research work in a report, as well as to an audience in a format such as a poster.

8 Research Project for Big Data Analytics Honours

Course code: COMS4058A

Offered: Full year.

Course description: This course introduces candidates to research activities. A student is expected to work with a specific problem in data science or computer science under the guidance of a supervisor. The work includes essential components of doing research in a thorough manner. It includes conducting literature survey, formulating the research problem and hypothesis, designing and implementing a feasible and effective research methodology to solve the proposed problem, and present the research work in a report, as well as to an audience in a format such as a poster.

9 Data Analysis and Exploration

Course code: COMS4048A

Offered: Semester 1

Course description: Applied statistical modelling is part art and part science. Briefly speaking, statistics is the set of principles and procedures for how we generalise results from a finite set of observations into "facts" about a subject. In this course, we develop statistical principles and apply necessary tools to successfully navigate common data science problem: developing an interpretable linear statistical model. The course has two main purposes: firstly, to provide foundational skills for careers in data analytics, and secondly, to introduce students to the field of statistics.

Course prerequisite/prior knowledge: Mathematics II.

---

10 Applications of Algorithms

Course code: COMS4032A

Offered: Semester 2

Course description: This course is about designing and analysing algorithms. Computational problems for various types of applications are considered and algorithmic approaches to solving them are investigated. The computational complexity of algorithms is investigated using some mathematical techniques. The course closely follows the book Introduction to Algorithms, by Cormen, Leiserson, Rivest and Stein.

Course prerequisite/prior knowledge: A degree in computer science that includes some introduction to analysis of algorithms is required. Familiarity with programming in either Java or C++ is required.

Course contents: See the list below.

- Introduction to algorithms
- Growth of Functions
- Divide and Conquer
- Probabilistic Analysis and Randomized Algorithms
- Sorting in Linear Time
- Medians and Order Statistics
- Binary Search Trees
- Red-Black Trees
- Augmenting Data Structures
- Amortized Analysis
- Computational Geometry
- NP-completeness

11 Artificial Intelligence

Course code: COMS4033A

Offered: Semester 2

Course description For as long as people have made machines, they have wondered whether machines could be made intelligent. This course will introduce you to the various approaches that have been developed with the aim of creating generally intelligent machines. We will cover everything from the invention of the field to the present, ranging from logical inference to modern machine learning. We will also touch on the question of what intelligence actually means, how we might recognise whether a machine is truly intelligent and whether AI is, in fact, a good idea.

Course prerequisite/prior knowledge: i) Familiarity with Python/C++ at a level sufficient to create a non-trivial program; ii) A background in graph theory and algorithms; iii) Some knowledge of statistics and optimisation.

Course contents: See the list below.

- Search
- Adversarial search
- Game theory
- Knowledge representation and reasoning
- Uncertain Knowledge and Bayesian Networks
- Classical planning
- Reinforcement Learning
- Unsupervised Learning
- Supervised Learning
- Computer Vision
- Natural Language Processing
- Philosophy and Ethics of AI

12 Computer Vision

Course code: COMS4036A

Offered: Semester 2

Course description: This course will provide an introduction to computer vision, with topics including image formation, feature detection, motion estimation, image mosaics, 3D shape reconstruction, and object and face detection and recognition. Applications of these techniques include building 3D maps, creating virtual characters, organizing photo and video databases, human-computer interaction, video surveillance, automatic vehicle navigation, and mobile computer vision.

We consider techniques in Bayesian Inference and Deep Learning, as well as more traditional methods involving handcrafted features, transformations and shape models.

Course prerequisite/prior knowledge: i) Familiarity with Python or C++ at a level sufficient to write a reasonably non-trivial computer program. ii) Familiarity with calculus.

Course contents: See the list below.

- Image Processing and Feature Extraction

- Probability and Distributions
- Maximum Likelihood, Maximum a Posteriori, Baysian Inference
- Modelling Visual Data
- The Pinhole Camera
- Geometric Transformations
- Shape Models
- Deep Learning & Convolutional Neural Networks (Common Networks, Classification, Object Detection, Semantic Segmentation, Embeddings, Style Transfer, Generative Models)

13 Multi-agent systems

Course code: COMS4043A

Offered: Semester 2

Course description The course is devoted to the game-theoretic and logical foundations of multi-agent systems.

Course prerequisite/prior knowledge: A degree in computer science that includes some introduction to discrete mathematics and analysis of algorithms

Course contents: See the list below.

- Introduction to multi-agent systems.
- Game theory in multi-agent systems.
- Games in normal form.
- Optimality and equilibria in games.
- Solution concepts for normal-form games.
- Perfect-information extensive form games.
- Imperfect-information extensive form games.
- Repeated and stochastic games.
- Bayesian games.
- Coalitional game theory.11. Logic in multi-agent systems: formal verification.
- Propositional logics of knowledge.
- Propositional logics of strategic ability.

14 Introduction to Data Visualisation & Exploration

Course code: COMS4060A

Offered: Semester 2

Course description: Understanding, cleaning and preparing data is fundamental to any data analysis project: data scientists spend 80%+ of their time cleaning and manipulating data, and only 20% of their time actually analysing it. This guides which data mining and modelling techniques are appropriate for the situation, and should be undertaken before these steps are applied.

Exploratory data analysis is used as a tool to understand the underlying structure of data whilst uncovering features and relationships, typically without an a priori understanding of what the data might contain. This aids in identifying anomalies, missing values, and evaluating data integrity. In increasingly large datasets, the relationships and patterns can become more complex and less intuitive, requiring specialised techniques to uncover the structure. A key aspect of data exploration is reducing the complexity of the data by finding suitable ways of summarising the data, or reducing the dimensionality, whilst minimising information loss. Visualising the data is a powerful way of doing this, providing effective communication of information using an intuitive representation of key insights into complex datasets, which is particularly useful for humans to understand — visualisation allows for analysis, exploration and presentation of the data and insights.

Course prerequisite/prior knowledge: i) Familiarity with basic statistics: descriptive statistics, probability, sampling, distributions, correlation, regression. This will not be covered in detail, but some understanding of the concepts is required. ii) Familiarity with Python will be useful but is not required. Any language may be used, but all examples and resources are provided in Python and Jupyter Notebooks.

Course contents: See the list below.

- Data cleaning
- Using statistics and plots to describe data
- Plotting in Python
- Feature engineering
- Feature selection
- Data augmentation
- Linear dimensionality reduction
- Time series data
- Clustering techniques
- Non-linear dimensionality reduction
- Geospatial data
- Presenting data

15 Discrete Optimization

Course code: COMS4050A

Offered: Semester 2

Course description:

Course prerequisite/prior knowledge:

Course contents: See the list below.

Problems The basic concepts and definitions from graph theory, various types of graphs, Eulerian and Hamiltonian Circuit;

Graph partitioning such as max-k-cut, max/min cut problem, graph bisection, and graph coloring problem; shortest path and transportation problems; Max flow Problem; minimum spanning tree; travelling salesman problem; the knapsack problem (linear and quadratic); the satisfiability problem (SAT); the independent set problem; facility location problem and p-median problem; linear assignment and quadratic assignment problem; Vertex covering, Set covering, and Winner Determination Problems; integer linear and integer quadratic programming problems, mix-integer programming problems.

Methods: Methods1: Max-flow Min Cut Algorithm (Ford-Fulkerson Algorithm), 2-Opt Heuristics, Approximate Algorithms (Christofides Approximation Algorithm); Iterative Improvement Local Search for Discrete Optimization; Exact Algorithms: branch and bound, and cutting plane.

Methods2: Metaheuristics: Simulated Annealing, Tabu Search, Genetic Algorithm

Game theory: Zero-sum, non-zero sum games, dominance, Nash equilibrium, prisoner's dilemma and battle of sexes.

16 Natural Language Processing

Course code: COMS4054A

Offered: Semester 2

Course description: "Language is not just words. It's a constructed self, an orientation towards the world, an attempt to make sense of external phenomena. Language is the hands with which our minds reach into the world and try to make sense of it." — Nanjala Nyabola, Digital Democracy and Analogue Politics.

In this course, students will gain a thorough introduction to the most important applications and techniques in Natural Language Processing (NLP). Through lectures, assignments and a final project, students will learn the necessary skills to design, implement, and understand their own natural language processing models, as well as curate and collect data for the NLP pipeline.

Course prerequisite/prior knowledge: i) Familiarity with Python at a level sufficient to write reasonably a non-trivial computer program. ii) Familiarity with calculus. iii) Familiarity with probability theory. iv) Familiarity with machine learning fundamentals.

Course contents: See the list below.

- Word Embeddings
- Text Classification
- Language Modelling
- Seq2Seq and Attention
- Transfer Learning
- Data Collection, Curation and Statements
- African Language NLP
- NLP Ethics
- NLP in Production

17 Reinforcement Learning

Course code: COMS4061A

Offered: Semester 2

Course description: Reinforcement learning (RL) is the area of machine learning concerned with how intelligent agents can learn to behave optimally in their environments. This means learning about how the environment works, and using this knowledge to choose actions at each point in time, considering the future implications of these decisions. RL has received considerable attention in the last five years in the research community, particularly with notable successes in domains such as the boardgame Go and in learning to play video games ranging from Pac-man to StarCraft. The applications of this field are widespread, including healthcare, online advertising, and robotics.

In this course we will discuss the background of the field as well as the mathematical formulation of the RL problem. We will consider different classes of algorithms, suitable to various problems, culminating in discussions around some recent papers and breakthroughs. Practical components of the course will involve implementing several of the methods and using them to learn behaviours in several benchmarks and games.

Course prerequisite/prior knowledge: Machine learning

Course contents: See the list below.

- Problem setup
- Multi-armed bandits
- Markov Decision Processes
- Value iteration and policy iteration
- TD methods, including SARSA and Q-learning
- Function approximation
- Policy search methods
- Model-based RL
- Hierarchical RL

18 Special Topics in Computer Science

Course code: COMS4047A

Offered: Semester 2

Course description: Probabilistic Graphical Models (PGMs) are perhaps one of two of the most powerful modelling tools available today for reasoning and decision-making under uncertainty, the other being Deep Learning. While Deep Learning has enjoyed a large amount of growth in recent years, establish limitations of deep learning include being data hungry (requiring millions or even billions of training samples); compute-intensive to train and deploy; poor at representing uncertainty; non-trivial to incorporate prior knowledge and symbolic representations; easily fooled by adversarial examples; finicky to optimise; and are seen largely as uninterpretable black boxes. The language of probabilistic graphical models is well established and has been extensively used in machine learning to express complex joint multivariate probabilistic distributions. Even the most basic probabilistic graphical model representations, such as Gaussian mixture models, factor analysis, hidden Markov models, naïve Bayes, Factorial Hidden Markov Models, latent Dirichlet allocation etc, are the cutting-edge model-of-choice to solve many domain-specific problems. Research in probabilistic graphical models has been enjoying an enormous amount of attention because of several unique properties. These properties include: (a) a calibrated model and prediction under uncertainty (i.e. a system that "knows when it doesn't know"; (b) automatic model complexity control and structure learning; (c) building systems that make rational decisions; (d) a way of building prior knowledge into learning systems, and making sure that knowledge is updated coherently and robustly as more data is provided; and (e), provides learning opportunities in the presence of both small and large datasets. These factors make PGMS a sustainable and cutting-edge mathematical framework for data science and machine learning. PGMs have many application domains, such as perception (e.g. computer vision), automated reasoning, natural language processing, knowledge representation, and robotics.