**STRATHMORE UNIVERSITY**

**COURSE OUTLINE AND DELIVERY PLAN**

### DSA 8302 Computational Techniques in Data Science

**Lecturer: Prof. Thomas N. O. Achia**

**Office: 0707410979/0702117190**

**Office Extension No:**

**Email:** [**achiathomas@gmail.com**](mailto:achiathomas@gmail.com)**;** [**yrb2@cdc.gov**](mailto:yrb2@cdc.gov)**;** [**tachia@strathmore.edu**](mailto:tachia@strathmore.edu)

**Zoom:** <https://strathmore.zoom.us/j/91382378774?pwd=YXJtVkQ3dUsycHdjZkxLdEszVk9vQT09>

**Module Leader: Prof. Thomas Achia**

**Office Hours: CDC-KEMRI MBAGATHI**

*Meeting times by appointment may be arranged as well*

**Purpose of the course**

This course introduces numerical methods for data scientists and delves into the mathematical and practical formulations of such methods. The goal of this course is to expose students to the role of computation in solving data related problems and develop algorithms that can solve these problems using numerical techniques

**Expected Learning Outcomes**

At the end of this course, the student should be able to:

1. Provide an understanding of the role computation can play in solving problems
2. Develop insight into the process of moving from an ambiguous problem statement to a computational formulation of a method for solving a problem
3. Apply algorithmic and problem reduction techniques,
4. Using simulations to solve complex equations that do not have closed form solutions.

**Content**

|  |  |  |  |
| --- | --- | --- | --- |
| **Credit Hours**: 40 Lecture hours, 60 Lab hours (Credits: 4) | | | |
| **Prerequisite:** DSA 8202- Time Series Analysis and Forecasting, DSA 8203- Principles of Data Science | | | |
| **Content: From the course structure/description**  **Root Finding**: Bisection Method; Newton’s Method; Secant Method; Order of convergence  **Multivariate root finding methods:** Newton’s Method and Fisher Scoring; Iteratively Reweighted Least Squares; Newton-Like Methods; Quasi-Newton Methods; Gauss–Newton Method  **Least-squares approximation**: Fitting a straight line; Fitting to a linear combination of functions. Sampling distributions; Maximum likelihood estimation; The likelihood function; the score function; the Hessian.  **Interpolation**: Polynomial interpolation; Vandermonde polynomial; Lagrange polynomial; Newton polynomial; Piecewise linear interpolation; Cubic spline interpolation; Multidimensional interpolation  **Integration**: Midpoint rule; Trapezoidal rule; Simpson’s rule; Composite rules; Local versus global error; Adaptive integration  **Monte-carlo simulations**: Random Variable Generation: Uniform simulation; The inverse transform; General transformation methods: A normal generator; Discrete distributions; Mixture representations; Accept-reject methods  Classical Monte Carlo integration: Importance sampling  Monte Carlo Optimization: Numerical optimization methods; Stochastic search; Stochastic gradient methods; Stochastic approximation  **Bootstrapping:** The Bootstrap Principle: Nonparametric Bootstrap; Parametric Bootstrap: Bootstrapping Regression  Bootstrap Bias Correction; Bootstrap Inference; Percentile Method.  **Expectation-Maximization approaches**  Practical implementation of methods using commercial data mining software. Case studies address applications of current significance. | | | |
| **Lecture/Dates** | **Topic** | **Intended Learning Outcomes** | **Activities** |
| **Lecture 1-2** | **Univariate Root finding methods** | At the end of these sessions the student will be able to :   * Describe and apply the Bisection Method; Newton’s Method; and Secant Method; * Determine the Order of convergence of the methods * Use software to implement these techniques | * Practical exercises in R |
| **Lecture 3-6** | **Multivariate Root finding methods and Least squares methods** | At the end of these sessions the student will be able to :   * Describe and implement the Newton’s Method and Fisher Scoring * Describe and implement the Iteratively Reweighted Least Squares; Newton-Like Methods; * Describe and implement the Quasi-Newton Methods; Gauss–Newton Method * Fitting a straight line; Fitting to a linear combination of functions. Sampling distributions; Maximum likelihood estimation; The likelihood function; the score function; the Hessian. * Implement the approaches using software | * Practical exercises in R |
| **Lecture 7-9** | **Interpolation techniques** | At the end of these sessions the student will be able to :   * Describe and implement Polynomial interpolation; Vandermonde polynomial; Lagrange polynomial; Newton polynomial; Piecewise linear interpolation; Cubic spline interpolation; Multidimensional interpolation * Implement these methods in software | * Practical exercises in R |
| **Lecture 10-12** | **Integration techniques** | At the end of these sessions the student will be able to:   * Describe and implement the Midpoint rule; Trapezoidal rule; Simpson’s rule; Composite rules; Local versus global error; Adaptive integration   Implement these methods in software |  |
| **Lecture 12-16** | **Monte Carlo techniques** | At the end of these sessions the student will be able to:   * Describe and implement the monte carlo simulations for different types of statistical distributions * Implement these methods in software | * Practical exercises in R |
| **Lecture 17-19** | **Bootstrapping and EM algorithms** | At the end of these sessions the student will be able to:   * Describe and implement the Bootstrapping and EM algorithms * Implement these methods in software | * Practical exercises in R |

**Course Delivery Methodology**

**Academic Assessment**

|  |  |
| --- | --- |
| **Type** | **Weighting (%)** |
| **End of semester Examination (Final Exam)** | **60** |
| **Coursework**   * CAT 1 (Supervised CAT) * CAT 2 (Group/Individual Research Assignment, Quizzes) | **40**  20  20 |
|  |  |
| **Total** | **100 %** |

**Course Materials**

Reference Text

1. Robert, C. P., Casella, G., & Casella, G. (2010**). *Introducing monte carlo methods with r*** (Vol. 18). New York: Springer.
2. Givens, G. H., & Hoeting, J. A. (2012). ***Computational statistics*** (Vol. 703). John Wiley & Sons.
3. **Numerical Methods** Jeffrey R. Chasnov. <https://www.math.hkust.edu.hk/~machas/numerical-methods.pdf>

**Recommended Reading Materials:**

1. James P Howard II, (2017). *Computational Methods for Numerical Analysis with R (Chapman & Hall/CRC Numerical Analysis and Scientific Computing Series)* 1st Edition, CRC Press.
2. Venkateshan, S. P., & Swaminathan, P. (2013). *Computational methods in engineering*. Elsevier.

**Journals**

1. Frontiers in Applied Mathematics and Statistics
2. Mathematics of Computation and Data Science

**Policies**

Classes

1. **Punctuality** is fundamental.
2. Active participation in class discussions is encouraged

Assignments and/or Course Work

1. **Plagiarism** is a serious offence. If detected in any form in course work and assignments, the following will apply:
   1. In partial or non-serious cases (such as not citing whole word-for-word quotes), half the total possible marks of the assignment are duly struck off.
   2. In serious cases (such as whole duplication of a paper), a zero policy will apply i.e., all offending assignments will be awarded a mark of zero.

Note: The level of seriousness referred to above is at the discretion of the lecturer. Appeals are certainly possible through the relevant channels

1. **Referencing:** APA System
2. Notwithstanding the above, **collaboration** in course work is certainly encouraged as this promotes team spirit and group synergy as long provided originality is preserved.

**Communication Channels**

* *E-mail*
* *Module Leader*