Assumption University
Vincent Mary School of Science and Technology
Department of Computer Science

CSX 2006 – Mathematics and Statistics for Data Science
Semester 1/2023

Course Status :	Specialized Course
Number of Credits :	3 Credits
Pre-requisite(s) :	N/A
Instructor :	Tapanan Yeophantong
Office Location :	VMS 0604
E-mail Address :	tapanan@scitech.au.edu
Class Schedule :	Wed 09:00-12:00
Class Venue :	VMS0404

Course Description

Linear algebra such as vector spaces in n-space, inner product, norm and distance, orthogonal vector, vector product, Orthogonal functions and Fourier series, including various techniques in multivariate data analysis such as multiple regression analysis, discriminant analysis, logistic regression analysis, principal component analysis, factor analysis, and cluster analysis.

Text & References

- Peter Bruce & Andrew Bruce. Practical Statistics for Data Scientists: 50+ Essential Concepts Using R and Python. O'Reilly Media, Inc., 2017. (ISBN: 9781491952962)
- Thomas Nield. Essential Math for Data Science: Take Control of Your Data with Fundamental Linear Algebra, Probability, and Statistics. O'Reilly Media, Inc., 2022. (ISBN: 9781098102937)
- Additional reading materials to be announced in class.

Mark Allocation

+ Workshops & Exercises:	60%
+ Final Examination:	40%
Total :	**100%**

Course Content & Schedule

Week	Date	Topics
Module I – Basic Probabilities & Statistical Methods		
1	07 Jun 2023	**Data Science** + Overview of data science & its applications + Setting up a Python workspace
2	14 Jun 2023	**Basic Probability Theory** + Probabilities & probability spaces + Conditional probabilities & Bayes theorem
3	21 Jun 2023	**Descriptive Statistics** + Point & variability estimations + Working with percentiles, boxplots & histograms
4	28 Jun 2023	**Probability Distributions** + Overview of probability distributions & their applications + Understanding joint probability distributions
5	05 Jul 2023	**Inferential Statistics** + Confidence intervals + Hypothesis testing & statistical significance
Module II – Mathematics for Data Analysis & Visualisation		
6	12 Jul 2023	**Exploratory Data Analysis** + Components of structured data + Working with numpy & matplotlib libraries
7	19 Jul 2023	**Analysis of Correlations** + Understanding correlations + Working with scatterplots & correlation matrices
8	26 Jul 2023	**Regression Models** + Regression techniques & their applications + Metrics for evaluating regression models
9	16 Aug 2023	**Cluster Analysis** + Clustering techniques & their applications + Metrics for evaluating clusters
10	23 Aug 2023	**Principal Component Analysis** + Overview of principal component analysis + Computing & interpreting principal components
Module III – Mathematics for Machine & Deep Learning		
11	30 Aug 2023	**Fundamentals of Machine Learning** + Overview of machine learning & its applications + Working with scikit-learn libraries
12	06 Sep 2023	**Decision Tree Learning** + Decision tree learning & its applications + Computing entropy, information gain & gini ratio
13	13 Sep 2023	**Bayesian Learning** + Bayesian learning & its applications + Constructing conditional probability tables
14	20 Sep 2023	**Nearest Neighbours Learning** + Nearest neighbour learning & its applications + Computing similarities & distances
15	27 Sep 2023	**Artificial Neural Networks** + Artificial neural networks & its applications + Perceptron & gradient descent algorithms
Final Examination		