

Subject Description Form

Subject Code	DSAI5103
Subject Title	Advanced High Dimensional Data Analysis
Credit Value	3
Level	5
Pre-requisite/ Co-requisite/ Exclusion	Pre-requisite: AMA563/DSAI5102
Objectives	To provide a comprehensive introduction to high dimensional data analysis. To present fundamental concepts and algorithms for selected topics in data mining, to provide the students with the necessary background for the applications to real problems, and to provide a starting point for students who are interested in pursuing research in data science or related fields.
Intended Learning Outcomes	Upon completion of the subject, students will be able to: (a) Demonstrate mastery of the principles of high dimensional data analysis. (b) Develop quantitative skills of data analysis and interpret the outcomes of algorithms. (c) Identify, define, and formulate problems of high dimensional data analysis in real applications and generate workable solutions to problems.
Subject Synopsis/ Indicative Syllabus	Multivariate normal distribution, Estimation of the mean vector and covariance matrix, Multiple and partial correlation coefficients; Estimation of high dimensional sparse parameters (Regularization, threshold estimators); Principal component analysis; Classification and Clustering; Regularized linear regression; Least absolute shrinkage and selection operator (LASSO); Choice of tuning parameters; Selected topics (e.g. FDR, Screening, Knockoffs).
Teaching/Learning Methodology	The subject will mainly be delivered through lectures and tutorials in computer lab. The theoretical background and the real applications of learning algorithms are both emphasized.

Assessment Methods in Alignment with Intended Learning Outcomes	Specific assessment methods/tasks	% weighting	Intended subject learning outcomes to be assessed (Please tick as appropriate)		
			a	b	c
	1. Assignments	24%	✓	✓	✓
	2. Mid-term test/Project	16%	✓	✓	✓
	3. Examination	60%	✓	✓	
	Total	100%			
	Explanation of the appropriateness of the assessment methods in assessing the intended learning outcomes:				
	Assignments: used to help students review basic mathematics and the details of the algorithms.				
	Mid-term test/Project: a part of continuous assessment for theory and training students’ ability to organize data mining algorithms for real problems				
	Examination: an overall examination of the mathematics and algorithms studied in the whole semester.				
Student Study Effort Required	Class contact:				
	▪ Lectures/Tutorials		39 Hrs.		
	Other student study effort:				
	▪ Assignments/Projects		58 Hrs.		
	▪ Self-study		30 Hrs.		
	Total student study effort		127 Hrs.		
Reading List and References	<u>Textbooks:</u>				
	James G., Witten D., Hastie T., Tibshirani R.	An Introduction to Statistical Learning	Springer 2013		
	<u>References:</u>				
	Hastie, T., Tibshirani, R., and Friedman, J.	The Elements of Statistical Learning	Springer 2009		

Reading List and References (Cont'd)	Kelleher, J.D., Namee M.B., D'Arcy, A.	Fundamentals of Machine Learning for Predictive Data Analysis	The MIT Press 2015
	Steinwart, I., Christmann A.	Support Vector Machines	Springer 2008
	Buhlmann, P., & De Geer, S.	Statistics for high- dimensional data: methods, theory, and applications	Springer Sciences & Van Business Media 2011