

Subject Description Form

Subject Code	AMA4002																
Subject Title	High Dimensional Data Analysis																
Credit Value	3																
Level	4																
Pre-requisite	Applied Statistical Methods (AMA2631/AMA2631A) or Applied Linear Models for Finance Analytics (AMA2602) or Probability & Distributions (AMA2691) or equivalent																
Objectives	This subject is to enable students to understand the theory of multivariate and high dimensional data analysis and apply it to real data analysis. The use of computer software such as R and MATLAB will be required in completing the assignments and mini-projects.																
Intended Learning Outcomes	<p>Upon completion of the subject, students will be able to:</p> <ol style="list-style-type: none"> master the basic techniques for high dimensional data analysis; produce presentable statistical analysis for high dimensional data; interpret analysis results and make recommendations for actions based on analysis results; 																
Subject Synopsis/ Indicative Syllabus	<p>Multivariate normal distribution, Estimation of the mean vector and covariance matrix, Multiple and partial correlation coefficients</p> <p>Discrimination and classification, Clustering, Principal component analysis</p> <p>Estimation of high dimensional sparse parameters (mean and covariance matrix): Regularized/threshold estimators.</p> <p>High dimensional linear regression: least absolute shrinkage and selection operator (LASSO); least angle regression (LARS); Choice of tuning parameters.</p> <p>High dimensional discriminant analysis, classification and clustering.</p>																
Teaching/Learning Methodology	The subject will be delivered mainly through lectures and tutorials. The lectures will be conducted to introduce the concepts of high dimensional data analysis methods in the syllabus, which are then reinforced by learning activities involving self-reading, demonstration, tutorial exercise, assignments and mini-project.																
Assessment Methods in Alignment with Intended Learning Outcomes	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th rowspan="2">Specific assessment methods/tasks</th> <th rowspan="2">% weighting</th> <th colspan="3">Intended subject learning outcomes to be assessed (Please tick as appropriate)</th> </tr> <tr> <th>a</th> <th>b</th> <th>c</th> </tr> </thead> <tbody> <tr> <td>1. Assignments/Projects</td> <td>20%</td> <td style="text-align: center;">✓</td> <td style="text-align: center;">✓</td> <td style="text-align: center;">✓</td> </tr> </tbody> </table>				Specific assessment methods/tasks	% weighting	Intended subject learning outcomes to be assessed (Please tick as appropriate)			a	b	c	1. Assignments/Projects	20%	✓	✓	✓
Specific assessment methods/tasks	% weighting	Intended subject learning outcomes to be assessed (Please tick as appropriate)															
		a	b	c													
1. Assignments/Projects	20%	✓	✓	✓													

	2. Quizzes/Mid-Term	20%		✓	✓
	3. Examination	60%	✓		✓
	Total	100 %			
	<p>Explanation of the appropriateness of the assessment methods in assessing the intended learning outcomes: The subject focuses on knowledge, skills and understanding of High Dimensional Data Analysis, thus, Exam-based assessment is the most appropriate assessment method, including 60% examination. Continuous Assessment comprises of individual assignments/project (20%) and quizzes/mid-term (20%) are included so as to keep the students in progress. A written examination is held at the end of the semester.</p>				
Student Study Effort Expected	Class contact:				
	• Lecture		26 Hrs.		
	• Tutorial		13 Hrs.		
	Other student study effort:				
	• Assignment		20 Hrs.		
	• Project		20 Hrs.		
	• Self-study		30 Hrs.		
	Total student study effort:		109 Hrs.		
Reading List and References	References: Bühlmann, P., & Van De Geer, S. Statistics for high-dimensional data: methods, theory and applications Springer Sciences & Business Media 2011				