Subject Description Form

Subject Code	AMA4602
Subject Title	High Dimensional Data Analysis
Credit Value	3
Level	4
Pre-requisite	Applied Linear Models for Finance Analytics (AMA2602) or Applied Linear Models (AMA3602) or Statistics for Data Science (AMA3631) **and** Linear Algebra (AMA1751) or Mathematical Methods for Data Science (AMA3001/AMA3701) or Further Mathematical Methods for Finance (AMA3723) or Further Mathematical Methods (AMA3724) or equivalent
Exclusion	High Dimensional Data Analysis (AMA4002)
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	Upon completion of the subject, students will be able to: a. master the basic techniques for high dimensional data analysis; b. produce presentable statistical analysis for high dimensional data; c. interpret analysis results and make recommendations for actions based on analysis results; d. be aware of technological social responsibilities and academic integrity.
Subject Synopsis/ Indicative Syllabus	Multivariate normal distribution; Estimation of the mean vector and covariance matrix; Multiple and partial correlation coefficients Discrimination and classification; Principal component analysis; Estimation of high dimensional sparse parameters (mean and covariance matrix): Regularized/ threshold estimators. High dimensional linear regression: Ridge regression; Least absolute shrinkage and selection operator (LASSO); Coordinate descent algorithm; Choice of tuning parameters. Feature screening; Multiple testing methods Ethical handling of data; reproducibility in research
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

<table>
<tr><td></td><td>methods in the syllabus, which are then reinforced by learning activities involving self-reading, demonstration, tutorial exercise, assignments and mini-project.</td></tr>
<tr><td>Assessment Methods in Alignment with Intended Learning Outcomes</td><td>
<table>
<tr><td>Specific assessment methods/tasks</td><td>% weighting</td><td colspan="4">Intended subject learning outcomes to be assessed</td></tr>
<tr><td></td><td></td><td>a</td><td>b</td><td>c</td><td>d</td></tr>
<tr><td>1. Assignments/Projects</td><td>20%</td><td>✓</td><td>✓</td><td>✓</td><td>✓</td></tr>
<tr><td>2. Quizzes/Mid-term</td><td>20%</td><td>✓</td><td>✓</td><td>✓</td><td></td></tr>
<tr><td>3. Examination</td><td>60%</td><td>✓</td><td></td><td>✓</td><td></td></tr>
<tr><td>Total</td><td>100%</td><td colspan="4"></td></tr>
</table>
Explanation of the appropriateness of the assessment methods in assessing the intended learning outcomes:
The subject focuses on knowledge, skills and understanding of <u>High Dimensional Data Analysis,</u> thus, <u>Exam-based assessment</u> 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.
</td></tr>
<tr><td>Student Study Effort Expected</td><td>
<table>
<tr><td>Class contact:</td><td></td></tr>
<tr><td>• Lecture</td><td>26 Hrs.</td></tr>
<tr><td>• Tutorial</td><td>13 Hrs.</td></tr>
<tr><td>Other student study effort:</td><td></td></tr>
<tr><td>• Assignment</td><td>20 Hrs.</td></tr>
<tr><td>• Project</td><td>20 Hrs.</td></tr>
<tr><td>• Self-study</td><td>30 Hrs.</td></tr>
<tr><td>Total student study effort:</td><td>109 Hrs.</td></tr>
</table>
</td></tr>
<tr><td>Reading List and References</td><td>References:
Bühlmann, P., & Van De Geer, S. Statistics for high-dimensional data: methods, theory and applications Springer Sciences & Business Media 2011</td></tr>
</table>