Annexe A: New/Revised Course Content in OBTL+ Format

Course Overview

The sections shown on this interface are based on the templates UG OBTL+ or PG OBTL+
If you are revising/duplicating an existing course and do not see the pre-filled contents you expect in the subsequent sections e.g. Course Aims, Intended Learning Outcomes etc. please refer to Data Transformation Status for more information.

Expected Implementation in Academic Year	AY2022-2023
Semester/Trimester/Others (specify approx. Start/End date)	Semester 2
Course Author * Faculty proposing/revising the course	Lai Ru Ying
Course Author Email	rylai@ntu.edu.sg
Course Title	Statistics & Computational inference to Big Data
Course Code	CB4247
Academic Units	3
Contact Hours	38
Research Experience Components	Not Applicable

Course Requisites (if applicable)

Pre-requisites	CH2010 Engineering Statistics
Co-requisites	
Pre-requisite to	
Mutually exclusive to	
Replacement course to	
Remarks (if any)	

Course Aims

The advent of the big data era has highlighted great new opportunities and challenges for statistical inference in manufacturing and daily life. To embrace big data (from an industrial manufacturing perspective), there is an urgent need to truly understand the core concepts and become capable of leveraging key algorithms/techniques/methodologies pertaining to data (big-data) statistics and computational inference, which is essential for extracting useful and valuable information for informed decision-making. This course will start with the core principles of data analytics and will equip you with the statistics and computational inference (including regression, dimensionality reduction, modeling) suitable for coping with big data case scenarios. This course is expected to help students develop interpretation of easy-to-use techniques/algorithms/methods and equip the students with essential skills in addressing big data inference problems in the chemical and biomedical industries.

Course's Intended Learning Outcomes (ILOs)

Upon the successful completion of this course, you (student) would be able to:

ILO 1	Understand the concept of big data, and apply concepts of probability and probability distributions. Identify the different type of statistical distribution (including normal distribution, Chi-square and F distribution) and describe the key characteristics of these distributions.
ILO 2	Master big data pre-processing techniques, including how to deal with the missing data, detection and processing methods of outliers, and resampling methods.
ILO 3	Revisit and apply ordinary least square (OLS) and nonlinear least-squares.
ILO 4	Learn and apply weighted least-square (WLS) methods to estimate the parameters in a regression model.
ILO 5	Learn and apply machine learning methods (e.g., principal components analysis (PCA), PCA based least squares methods, partial least squares methods
ILO 6	Learn and apply Lasso regression methods and Ridge regression methods for big data regression.
ILO 7	Use commonly used programming-based computing platforms (e.g., MATLAB or Python) to process data, conduct regression, build regression model, and visualize and analyze the obtained results.

Course Content

Key topics taught: 1. Review of probability and probability distributions 2. Data Pre-processing for big data analytics, regression/data-driven predictive modeling 3. Fundamentals of regression (ordinary least-squares, weighted least-squares) 4. Nonlinear regression 5. Principal component analysis (PCA) for dimensionality reduction 6. PCA-based regression and Partial least-squares 7. Other variants of least-squares in the context of big data (e.g., Lasso regression and Ridge regression) 8. Applications of the methods/techniques to real-world problems for analysis and regression/modeling

Reading and References (if applicable)

The following textbooks may be used as references. 1. D. C. Montgomery, G. C. Runger, Applied Statistics and Probability for Engineers. 2. G. James, D. Witten, T. Hastie, R. Tibshirani. An Introduction to Statistical Learning. 3. J. Warren, N. Marz. Big Data: Principles and Best Practices of Scalable Realtime Data Systems.

Planned Schedule

Week or Session	Topics or Themes	ILO	Readings	Delivery Mode	Activities
1	Introduction to big data analytics and probability distributions	1			
2	Data pre-processing for big datasets	2			
3	Revisiting ordinary least-squares	3			
4	Revisiting ordinary least-squares, Revisiting nonlinear least-squares regression	3			
5	Revisiting nonlinear least-squares regression, Weighted least-squares regression	3, 4			
6	Weighted least-squares regression	4			
7	CA1	1-4			
8	Introduction to principal component analysis (PCA) and PCA-based regression	5			

Week or Session	Topics or Themes	ILO	Readings	Delivery Mode	Activities
9	Partial least-squares regression	5			
10	Lasso regression	6			
11	Ridge regression	6			
12	CA2	1-7			
13	Applications in Engineering Problems	1-7			

Learning and Teaching Approach

Approach	How does this approach support you in achieving the learning outcomes?
Lecture	Course materials covering all the topics
Tutorial	12 classroom discussion sessions on tutorial questions and related topics

Assessment Structure

Assessment Components (includes both continuous and summative assessment)

No.	Component	ILO	Related PLO or Accreditation	Weightage	Team/Individual	Rubrics	Level of Understanding
1	Summative Assessment (EXAM): Final exam(Final Examination(2hrs, Closed-book exam))	1-6	EAB SLO* a, b, c, d	50	Individual	Analytic	Multistructural
2	Continuous Assessment (CA): Test/Quiz(CA1: Quiz)	1-4	EAB SLO* a, b, c, d	20	Individual	Analytic	Multistructural
3	Continuous Assessment (CA): Project(CA2: Project)	1-7	EAB SLO* a,b,c,d,e,h	30	Individual	Analytic	Multistructural

Description of Assessment Components (if applicable)

Formative Feedback

During tutorials, the instructor will articulate expected learning outcomes in detail and use examples/case studies to better illustrate the methods introduced in the previous lectures.

After each CA, the instructor will go through the problems during tutorials. Common mistakes and misunderstanding in concepts will also be addressed.

Specific feedback to the progress of project work may be returned to students via email.

General feedback to project work will be published online.

NTU Graduate Attributes/Competency Mapping

This course intends to develop the following graduate attributes and competencies (maximum 5 most relevant)

Attributes/Competency	Level
Critical Thinking	Basic
Systems Thinking	Intermediate

Course Policy

Policy (Academic Integrity)

Good academic work depends on honesty and ethical behaviour. The quality of your work as a student relies on adhering to the principles of academic integrity and to the NTU Honour Code, a set of values shared by the whole university community. Truth, Trust and Justice are at the core of NTU's shared values. As a student, it is important that you recognize your responsibilities in understanding and applying the principles of academic integrity in all the work you do at NTU. Not knowing what is involved in maintaining academic integrity does not excuse academic dishonesty. You need to actively equip yourself with strategies to avoid all forms of academic dishonesty, including plagiarism, academic fraud, collusion and cheating. If you are uncertain of the definitions of any of these terms, you should go to the academic integrity website for more information. On the use of technological tools (such as Generative AI tools), different courses / assignments have different intended learning outcomes. Students should refer to the specific assignment instructions on their use and requirements and/or consult your instructors on how you can use these tools to help your learning. Consult your instructor(s) if you need any clarification about the requirements of academic integrity in the course.

Policy (General)

Students are responsible for meeting all course requirements, observing all deadlines,

examination times, and other course procedures.

You will be awarded ZERO mark for being absence from quizzes unless it is due to the

following reasons:

· Illness (valid medical certificate is required, not from Chinese doctor)

· Passing away of immediate family member (parents, siblings or

grandparents)

· Participate in an activity representing NTU (support letter from

participating organization)

·

There will be no makeup given for missed quizzes. Final grade will be determined based

on the participated quiz and final examination.

You are responsible for following the university regulations for final examination.

You are responsible for being on time for all lectures and tutorials. Sufficient efforts

should be put into solving or attempting the tutorial problems prior to attending the

respective tutorial classes.

You might be awarded an “F” for a component or expelled from the university if you are

caught cheating.

You are responsible for seeking academic help in a timely fashion.

Policy (Absenteeism)

In-class activities make up a significant portion of your course grade. Absence from class without a valid reason will affect your participation grade. Valid reasons include falling sick supported by a medical certificate and participation in NTU’s approved activities supported by an excuse letter from the relevant bodies. There will be no make-up opportunities for in-class activities.

Policy (Others, if applicable)

Last Updated Date: 21-11-2024 03:55:41

Last Updated By: Yin Xunyuan (Asst Prof)

Appendix 1: Assessment Criteria for the Project

Criteria	Exceed Expectations 71%-100%	Meet Expectations 41% - 70%	Meet Baseline Expectations 26% – 40%	Below Expectations 0 – 25%
Identify and formulate a good regression/predi ctive modeling problem, pre- process the data; explore and choose relevant and appropriate algorithm(s) to conduct regression or build predictive models (LO 1-6)	Systematically and most appropriately pre-process raw data to retain usable variables and extract the most valuable information from the raw data to facilitate further analysis and results. Truly understand the nature of the considered problem. Through exploration, choose the most appropriate regression/mod eling methodology (may also explore some algorithms/conc epts that are relevant to this course but not discussed in detail in lectures) for the specific problems and well justify the selection and adoption of the method using clear and concrete language. Showcase the obtained results in effective ways (visualization, tables, etc.),	Appropriately pre-process raw data to mitigate the negative effect of missing data/poor-quality samples in the raw dataset on further analysis and results. Select appropriate regression/mod eling methodology for the specific regression/mod eling problems and present explanations about the selection and adoption of the method. Visualize the obtained results, discuss and interpret the obtained results, discuss the advantages and limitations of the proposed solution, and justify the conclusions based on the results. In the report, present clear figures with necessary legends and labels, and prepare well-written project	Apply necessary algorithms/techni ques to pre- process raw data so that the processed dataset may be acceptable for regression and modeling. Apply regression/modeli ng methodology covered in this course to the processed data to generate some regression/modeli ng results Present reasonable and acceptable interpretation of the results but without drawing solid contributions and presenting convincing justifications of the conclusions. Submit a complete project report, but without appropriately presenting visualization results. Not well written with ambiguous statements and typos and grammatical errors seen throughout the report.	Unable to formulate a well-defined regression/mod eling problem. Unclear ultimate goals. Unable to apply appropriate data pre-processing and regression/mod eling algorithms to pursue the objectives. No result and/or interpretation to showcase.

Criteria	Exceed Expectations 71%-100%	Meet Expectations 41% - 70%	Meet Baseline Expectations 26% – 40%	Below Expectations 0 – 25%
	discuss in detail and interpret the obtained results, discuss the advantages and limitations of the adopted algorithms and the developed solution, discuss current perspectives and future directions, and justify the conclusions based on the results. In the report, present very nicely-plotted, easy to interpret figures with labels and legends, draw a schematic of the proposed solution, conduct extensive literature review, and present very well- structured project report with (almost) no grammatical errors/typos.	report with minimal grammatical errors/typos.		

Appendix 2

Mapping of Course ILOs to EAB Graduate Attributes

Course Code & Title	CB4247
Course Type	MPE/BDE

Overview											
(a)	●	(b)	●	(c)	◐	(d)	●	(e)	◐	(f)	
(g)		(h)	O	(i)		(j)		(k)			

Legend:

● Fully consistent (contributes to more than 75% of Student Learning Outcome)
◐ Partially consistent (contributes to about 50% of Student Learning Outcome)
O Weakly consistent (contributes to about 25% of Student Learning Outcome)
Blank Not related to Student Learning Outcome

Course ILOs		**EAB Graduate Attributes**
1)	Understand the concept of big data, and apply concepts of probability and probability distributions. Identify the different type of statistical distribution (including normal distribution, Chi-square and *F* distribution) and describe the key characteristics of these distributions.	a
2)	Master big data pre-processing techniques, including how to deal with the missing data, detection and processing methods of outliers, and resampling methods.	a,b,d
3)	Revisit and apply ordinary least square (OLS) and nonlinear least-squares.	a,b,c,d
4)	Learn and apply weighted least-square (WLS) methods to estimate the parameters in a regression model.	a,b,c,d,e
5)	Learn and apply machine learning methods (e.g., principal components analysis (PCA), PCA based least squares methods, partial least squares methods	a,b,c,d,e
6)	Learn and apply Lasso regression methods and Ridge regression methods) for big data regression.	a,b,c,d,e
7)	Use commonly used programming-based computing platforms (e.g., MATLAB or Python) to process data, conduct regression, build regression model, and visualize and analyze the obtained results.	a,b,c,d,e,h

EAB GRADUATE ATTRIBUTES	KNOWLEDGE PROFILE
a) **Engineering Knowledge**: Apply the knowledge of mathematics, natural science, computing and engineering fundamentals, and an engineering specialisation as specified in WK1 to WK4 respectively to the solution of complex engineering problems.	**WK1** A systematic, theory-based understanding of the natural sciences applicable to the discipline and awareness of relevant social sciences.
b) **Problem Analysis**: Identify, formulate, research literature, and analyse complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences with holistic considerations for sustainable development. (WK1 to WK4)	**WK2** Conceptually-based mathematics, numerical analysis, data analysis, statistics and formal aspects of computer and information science to support detailed analysis and modelling applicable to the discipline.
c) **Design / Development of Solutions**: Design creative solutions for complex engineering problems and design systems, components or processes that meet identified needs with appropriate consideration for public health and safety, whole-life cost, net zero carbon as well as resource, cultural, societal, and environmental considerations as required. (WK5)	**WK3** A systematic, theory-based formulation of engineering fundamentals required in the engineering discipline.
d) **Investigation**: Conduct investigations of complex problems using research-based knowledge (WK8) and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.	**WK4** Engineering specialist knowledge that provides theoretical frameworks and bodies of knowledge for the accepted practice areas in the engineering discipline; much is at the forefront of the discipline.
e) **Modern Tool Usage**: Create, select and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modelling to complex engineering problems, with an understanding of the limitations. (WK2 and WK6)	**WK5** Knowledge including efficient resource use, environmental impacts, whole-life cost, re-use of resources, net zero carbon, and similar concepts that supports engineering design and operations in a practice area.
f) **The Engineer and the World**: When solving complex engineering problems, analyse and evaluate sustainable development impacts to: society, the economy, sustainability, health and safety, legal frameworks and the environment (WK1, WK5, and WK7).	**WK6** Knowledge of engineering practice (technology) in the practice areas in the engineering discipline.
g) **Ethics**: Apply ethical principles and commit to professional ethics and responsibilities and norms of engineering practice and adhere to relevant national and international laws. Demonstrate an understanding of the need for diversity and inclusion (WK9).	**WK7** Knowledge of the role of engineering in society and identified issues in engineering practice in the discipline such as the professional responsibility of an engineer to public safety and sustainable development.
h) **Individual and Collaborative Team Work**: Function effectively as an individual, and as a member or leader in diverse and inclusive teams and in multidisciplinary, face-to-face, remote and distributed settings (WK9).	**WK8** Engagement with selected knowledge in the current research literature of the discipline, awareness of the power of critical thinking and creative approaches to evaluate emerging issues.
i) **Communication**: Communicate effectively and inclusively on complex engineering activities with the engineering community and with society at large, such as being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions, taking into account cultural, language, and learning differences.	**WK9** Knowledge of professional ethics, responsibilities, and norms of engineering practice. Awareness of the need for diversity by reason of ethnicity, gender, age, physical ability etc with mutual understanding and respect, and of inclusive attitudes.
j) **Project Management and Finance**: Demonstrate knowledge and understanding of engineering management principles and economic decision-	

making, and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments. k) **Life-long Learning**: Recognise the need for, and have the preparation and ability to (i) engage in independent and life-long learning, and (ii) adapt to new and emerging technologies, and (iii) think critically, in the broadest context of technological change (WK8).	

Reference: EAB Accreditation Manual