City University of Hong Kong
Course Syllabus

offered by School of Data Science
with effect from Semester A 2022/23

Part I Course Overview

Course Title:	Statistical Machine Learning I
Course Code:	SDSC5001
Course Duration:	One Semester
Credit Units:	3
Level:	P5
Medium of Instruction:	English
Medium of Assessment:	English
Prerequisites: *(Course Code and Title)*	Nil
Precursors: *(Course Code and Title)*	Nil
Equivalent Courses: *(Course Code and Title)*	Nil
Exclusive Courses: *(Course Code and Title)*	Nil

Part II Course Details

1. Abstract

This course focuses on the theoretical foundation and fundamental methods in statistical machine learning, covering the key concepts and results of the probability theory and statistical inference for machine learning, classical and cutting-edge methods and theories for regression and classification, and theoretical foundation of kernel methods.

2. Course Intended Learning Outcomes (CILOs)

(CILOs state what the student is expected to be able to do at the end of the course according to a given standard of performance.)

No.	CILOs	Weighting (if applicable)	Discovery-enriched curriculum related learning outcomes (please tick where appropriate)		
			A1	*A2*	*A3*
1.	Familiarize the key concepts of statistical inference theory for machine learning	20%	✓		
2.	Familiarize fundamental statistical regression and classification models	20%	✓		
3.	Able to formulate kernel machines and regularization forms	20%	✓	✓	
4.	Comprehend key concepts and results in learning theory	20%	✓	✓	
5.	Able to apply taught machine learning methods to conduct real life data analytics problem via software	20%		✓	✓
		100%			

A1: Attitude
Develop an attitude of discovery/innovation/creativity, as demonstrated by students possessing a strong sense of curiosity, asking questions actively, challenging assumptions or engaging in inquiry together with teachers.

A2: Ability
Develop the ability/skill needed to discover/innovate/create, as demonstrated by students possessing critical thinking skills to assess ideas, acquiring research skills, synthesizing knowledge across disciplines or applying academic knowledge to self-life problems.

A3: *Accomplishments*
Demonstrate accomplishment of discovery/innovation/creativity through producing /constructing creative works/new artefacts, effective solutions to real-life problems or new processes.

3. Teaching and Learning Activities (TLAs)

(TLAs designed to facilitate students' achievement of the CILOs.)

TLA	Brief Description	CILO No.					Hours/week (if applicable)
		1	2	3	4	5	
Lecture	Introduce fundamental theories and principles in statistical learning to students	✓	✓	✓	✓		26 hours/sem
Laboratory work	Assist students to develop the data analytics skill base and use learned knowledge to address real problems through lab activities		✓	✓	✓	✓	13 hours/sem

Lectures cover not only the narrowly focused techniques in engineering economy but also the wider issues of the environment that affect engineering economic decision making. Students are expected to participate in class discussion when needed.

4. Assessment Tasks/Activities (ATs)

(ATs are designed to assess how well the students achieve the CILOs.)

Assessment Tasks/Activities	CILO No.					Weighting	Remarks
	1	2	3	4	5		
Continuous Assessment: 50 %							
Group Project		✓	✓		✓	0-30%	
Individual Coursework	✓	✓	✓	✓		20-30%	
Examination: 50 % (duration: 2 hours , if applicable)							
Examination	✓	✓	✓	✓	✓	50%	
						100%	

5. Assessment Rubrics

(Grading of student achievements is based on student performance in assessment tasks/activities with the following rubrics.)

Applicable to students admitted in Semester A 2022/23 and thereafter

Assessment Task	Criterion	Excellent (A+, A, A-)	Good (B+, B)	Marginal (B-, C+, C)	Failure (F)
1. Group Project	0-30%	High	Significant	Basic	Not even reaching marginal levels
2. Individual Coursework	20-30%	High	Significant	Basic	Not even reaching marginal levels
3. Examination	50%	High	Significant	Basic	Not even reaching marginal levels

Applicable to students admitted before Semester A 2022/23

Assessment Task	Criterion	Excellent (A+, A, A-)	Good (B+, B, B-)	Fair (C+, C, C-)	Marginal (D)	Failure (F)
1. Group Project	0-30%	High	Significant	Moderate	Basic	Not even reaching marginal levels
2. Individual Coursework	20-30%	High	Significant	Moderate	Basic	Not even reaching marginal levels
3. Examination	50%	High	Significant	Moderate	Basic	Not even reaching marginal levels

Part III Other Information (more details can be provided separately in the teaching plan)

1. Keyword Syllabus

(An indication of the key topics of the course.)

- Review of Probability Theory
- Statistical Inference: Method of Moments; Maximum Likelihood Estimators; Bayes Estimator; EM Algorithm; Bootstrap; Asymptotic Theory
- Regression: Linear and Nonlinear Regression; Kernel Smoothing; Local Polynomial; Cubic Splines; Regression Splines; Asymptotic Inference
- Classification: Misclassification Error; Discriminant Analysis; Logistic Regression; CART; Bagging; Random Forest; Estimation Consistency; Asymptotic Normality
- Kernel Machines: Bayes Rule; Fisher Consistency; Loss Functions; Empirical Risk Minimization; Regularization Form; SVM; Reproducing Kernel Hilbert Space; Function Approximation

2. Reading List

2.1 Compulsory Readings

(Compulsory readings can include books, book chapters, or journal/magazine articles. There are also collections of e-books, e-journals available from the CityU Library.)

1.	Statistical Inference by George Casella and Roger L. Burger
2.	The Elements of Statistical Learning by Hastie, Tibshirani, and Friedman, Springer
3.	Lecture Notes

2.2 Additional Readings

(Additional references for students to learn to expand their knowledge about the subject.)

NIL