**City University of Hong Kong**
**Course Syllabus**

**offered by School of Data Science**
**with effect from Semester A 2024/25**

---

**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 of the probability theory and statistical inference for machine learning, classical and cutting-edge methods and theories for regression and classification, and popular methods for unsupervised learning.

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. | Summarize the key concepts of statistical inference theory for machine learning | 20% | ✓ | | |
| 2. | Build fundamental statistical regression and classification models | 20% | ✓ | | |
| 3. | Formulate kernel machines and regularization forms | 20% | ✓ | ✓ | |
| 4. | Describe the underlying bias-variance tradeoff in each learning method | 20% | ✓ | ✓ | |
| 5. | Apply taught machine learning methods to solve 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 real-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. Learning and Teaching Activities (LTAs)

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

| LTA | Brief Description | CILO No. | | | | | Hours/week (if applicable) |
|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | |
| Lecture | In lectures, students will learn fundamental theories and principles in statistical machine learning. | ✓ | ✓ | ✓ | ✓ | | 26 hours/semester |
| Laboratory | In labs, students will learn software of statistical machine learning and apply taught methods to address real-world problems. | | ✓ | ✓ | ✓ | ✓ | 13 hours/semester |

## 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 | | ✓ | ✓ | | ✓ | 20% | |
| Homework assignments | ✓ | ✓ | ✓ | ✓ | | 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 before Semester A 2022/23 and in Semester A 2024/25 & thereafter**

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

**Applicable to students admitted from Semester A 2022/23 to Summer Term 2024**

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

- Regression: Linear and Nonlinear Regression, Kernel Smoothing, Local Polynomial, Cubic Splines, Regression Splines, Gaussian Process Regression

- Classification: Misclassification Error, Discriminant Analysis, Logistic Regression, CART, Bagging, Random Forest, Boosting

- Kernel Machines: Loss Functions, Regularization Form, SVM

- Unsupervised Learning: Principle Component Analysis, Clustering

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