**The Chinese University of Hong Kong (Shenzhen)**

**Course Outline**

Spring Term, 2023/2024

1. **Course Identity**
2. **Course as listed in CUHK (SZ)**

|  |  |
| --- | --- |
| Course code | DDA2001 |
| Course title (English) | Introduction to Data Science |
| Course title (Chinese) | 数据科学导论 |
| Units | 3 |
| Description (English) | This course is to provide a comprehensive introduction to data science. Topics will include probability, statistics, optimization, and machine learning. These four topics will be taught in the above-mentioned order. For each topic, we selectively choose the content to fit the first-year undergraduate students' background. |
| Description (Chinese) | 本课程是数据科学的入门级课程。分为四个部分，将依次讲解概率，统计，优化，和机器学习的基础知识。这门课将辅以大量实例，从具体例子出发讲述核心内容和思想，从而激发学生的兴趣和学习热情。 |

1. **Corresponding course in CUHK**
2. **Prerequisites / Co-requisites**

**A. Prerequisites**

CSC1001: Introduction to Computer Science: Programming Methodology; AND MAT1001 or MAT1011 or MAT1010: Calculus I.

**B. Co-requisites**

None

1. **Learning Outcomes**

After taking this course, the students are expected

* Knowledge Outcomes:
* Basic knowledge of discrete and continuous random variables and their properties.
* Basic knowledge of simulations, such as Monte Carlo.
* Basic knowledge of point estimator, such as MLE.
* Basic knowledge of regression analysis.
* Basic knowledge of convex optimization.
* Basic knowledge of machine learning, including supervised learning, unsupervised learning, etc.
* Value Outcomes:
* Be prepared to acquire necessary and adequate knowledge for further studies of more advanced theory and methods of data science.

1. **Course syllabus**
2. **Assessment Scheme**

|  |  |
| --- | --- |
| **Component/ method** | **% weight** |
| Assignments | 15% |
| Quizzes | 15% |
| Mid-Term Exam | 30% |
| Final Exam | 40% |

**Course Policies:**

1. Missing the midterm or final exam without prior notification to and approval of the instructors will automatically result in the **"0"** grade for the exam.
2. No make-up exam for Mid-Term Exam. If the Mid-Term Exam is missed under the approval of the instructor, the Final Exam will take 70% of the assessment.
3. Maximum one make-up exam for the Final Exam.
4. There are about 5 assignments. No late homework will be accepted.
5. Random quizzes. No make-up quizzes.

Remark: the number of assignments/quizzes is subject to change as the instructor sees fit.

1. **Descriptor**

|  |  |
| --- | --- |
| **Grades** | Overall course |
| **A** | * Hand in 100% assignments, AND * Demonstrate fully understanding of the course materials in all assignments, quizzes, midterm and final, AND * Outstanding ideas are shown in the assignments, midterm or final |
| **A-** | * Hand in 100% assignments AND, * Demonstrate fully understanding of the course materials in all assignments, quizzes, midterm and final |
| **B** | * Hand in 100% assignments AND, * Demonstrate fully understanding of the course materials in MOST of the assignments, quizzes, midterm and final |
| **C** | * Hand in 100% assignments AND, * Demonstrate fully understanding of the course materials in SOME parts of the assignments, quizzes, midterm and final |
| **D** | * Hand in 100% assignments AND, * Demonstrate fully understanding of the course materials in LIMITED parts of the assignments, quizzes, midterm and final |
| **F** | * Overall score from assignments is 0 OR, * Demonstrate fully understanding of the course materials in VERY LIMITED parts of the assignments, quizzes, midterm and final |

1. **Feedback for evaluation**

- Formal Course and Teaching Evaluation

- Feedback from office hour discussions

- Feedback after class

- Feedback from tutorial sessions

1. **Reading**

None.

1. **Course components**

|  |  |
| --- | --- |
| **Activity** | **Hours/week** |
| Lecture | 3 hours per week |
| Tutorial | 1 hour per week |

1. **Indicative teaching plan**

|  |  |
| --- | --- |
| **Weeks** | **Content/ topic/ activity** |
|  | **Course Introduction** |
| 1 | Course plan, grading, overview, examples, and applications, etc. |
|  | **Probability** |
| 3 | 1 Terminologies: sample spaces, probability function, event, etc..  2. Main Properties: mean, variance, CDF.  3. Common Discrete RV: Bernoulli, Binomial, Poisson, Geometric.  4. Common Continuous RV: uniform, normal, exponential.  5\*. Advanced Concepts: correlation and conditional probability (Bayes rule).  6#. Additional examples and exercises. |
|  | **Statistics** |
| 3 | 1. Sampling: Monte Carlo (integration, MDP, etc.).  2. Sampling: generating uncommon RV.  3. Point estimator: MLE.  4. MLE Application: Regression Analysis.  5\*. Confidence interval  6#. Additional examples and exercises. |
| 0.5 | **Midterm Review** |
|  | **Optimization** |
| 3.5 | 1. Problem formulation (objective function, constraints, decision variables, etc.) Application examples in revenue management, online advertising, etc.  2. Optimization: convex set.  3. Convex Optimization: convex function.  4. Convex Optimization: convex function Property.  5. Convex Optimization: quasi-convex function.  6\*. Gradient Descent and Stochastic Gradient Descent  7#. Additional examples and exercises. |
|  | **Machine Learning** |
| 2.5 | 1\*. Supervised learning: K-Nearest Neighbors (K-NN).  2\*. Supervised learning: Logistic regression, Neural Nets.  3\*. Unsupervised learning: Clustering (K means).  4\*\*. Cross Validation.  5\*\*: Deep Learning and Reinforcement Learning. |
| 0.5 | **Final Review** |

# Problem exercise.

\* Only concepts will be examined.

\*\* Contents that will be covered if time permits. If covered, only concepts will be examined.

1. **Implementation plan (2023-24)**
2. **Approval**
3. **Any other information**
4. **Version date**

|  |  |
| --- | --- |
| Version number | V001 |
| As of (date) | 2023/11/22 |

Please save file as XXXXxxxx v-nnn yymmdd

XXXXxxxx = course code, e.g., MATH1212

nn = version number, e.g., 001 for version 1

yymmdd = date of this version, e.g., 131210