

Academic Org: Div of Computer Science & Engg – Subject: Computer Science

<b>Course:</b> CSCI5620	<b>Course ID:</b> 013174	<b>Eff Date:</b> 2024-07-01	<b>Crse Status:</b> Active	<b>Apprv. Status:</b> Approved	<b>[New Course]</b>
Algorithms for Data Science 數據科學算法					

This is a graduate level course which provides the mathematical and algorithmic foundations for data science. Target audience are students interested in doing research in algorithms, statistics, machine learning or data mining. Topics to be covered will be probability and concentration bounds, high-dimension space and its properties, nearest neighbouring search and projection methods, singular value decomposition and its algorithms, random walks and Markov chains, constrained convex optimization, online stochastic gradient descent, boosting, streaming and sampling algorithms, sketches (e.g., count-min), graph sketching, random graphs.

Advisory: Students are expected to have taken CSCI3160 and CSCI3320

這是一門研究生課程，為數據科學提供數學和算法基礎。目標是對算法，統計，機器學習或數據挖掘研究感興趣的學生。涉及的主題將是概率和集中度邊界，高維空間及其屬性，最近鄰搜索和投影方法，奇異值分解及其算法，隨機遊走和馬爾可夫鏈，約束凸優化，在線隨機梯度下降，增強，流和採樣算法，草圖（例如，最小計數），圖形草圖，隨機圖形。

建議:學生應曾學習CSCI3160 及CSCI3320

**Grade Descriptor:** A

EXCELLENT – exceptionally good performance and far exceeding expectation in all or most of the course learning outcomes; demonstration of superior understanding of the subject matter, the ability to analyze problems and apply extensive knowledge, and skillful use of concepts and materials to derive proper solutions.

有關等級說明的資料，請參閱英文版本。

B

GOOD – good performance in all course learning outcomes and exceeding expectation in some of them; demonstration of good understanding of the subject matter and the ability to use proper concepts and materials to solve most of the problems encountered.

有關等級說明的資料，請參閱英文版本。

C

FAIR – adequate performance and meeting expectation in all course learning outcomes; demonstration of adequate understanding of the subject matter and the ability to solve simple problems.

有關等級說明的資料，請參閱英文版本。

D

MARGINAL – performance barely meets the expectation in the essential course learning outcomes; demonstration of partial understanding of the subject matter and the ability to solve simple problems.

有關等級說明的資料，請參閱英文版本。

F

FAILURE – performance does not meet the expectation in the essential course learning outcomes; demonstration of serious deficiencies and the need to retake the course.

有關等級說明的資料，請參閱英文版本。

**Equivalent Offering:**

**Units:** 3 (Min) / 3 (Max) / 3 (Acad Progress)

**Grading Basis:** Graded

**Repeat for Credit:** N

**Multiple Enroll:** N

**Course Attributes:** MPhil-PhD Computer Sci & Erg

**Topics:**

**COURSE OUTCOMES**

**Learning Outcomes:**

- At the end of the course of studies, students will have acquired the ability to
1. Be familiar with the concept of concentration and various tail bounds
  2. Be familiar with data in high dimension space and its properties
  3. Be familiar with singular value decomposition (SVD) and its algorithms
  4. Be familiar with random walks, Markov chains and their applications
  5. Be familiar with online learning and VC-dimension

6. Be familiar with streaming, sketching and sampling techniques
7. Be familiar about graph structures and their applications

**Course Syllabus:**

This is a graduate level course which provides the mathematical and algorithmic foundations for data science. Target audience are students interested in doing research in algorithms, statistics, machine learning or data mining. Topics to be covered will be probability and concentration bounds, high-dimension space and its properties, nearest neighbouring search and projection methods, singular value decomposition and its algorithms, random walks and Markov chains, constrained convex optimization, online stochastic gradient descent, boosting, streaming and sampling algorithms, sketches (e.g., count-min), graph sketching, random graphs.

**Assessment Type:**

Examination	: 50%
Homework or assignment	: 20%
Project	: 30%

**Feedback for Evaluation:**

1. Quiz and examinations
2. Course evaluation and questionnaire
3. Question-and-answer sessions during class
4. Student consultation during office hours or online

**Required Readings:**

To be provided by course instructor.

**Recommended Readings:**

1. Foundations of Data Science, by Avrim Blum, John Hopcroft, and Ravindran Kannan
2. Various research papers

**OFFERINGS**

- |             |  |
|-------------|--|
| 1. CSCI5620 | Acad Organization=CSEGV; Acad Career=RPG |
|-------------|--|

**COMPONENTS**

LEC : Size=30; Final Exam=Y; Contact=3

**ENROLMENT REQUIREMENTS**

1. CSCI5620

**Enrollment Requirement Group:**

For students in MPhil-PhD Computer Science & Engineering

**Additional Information**

VTL-Onsite face-to-face hrs    0  
VTL-Online synch. hrs        0  
VTL-Online asynch. hrs      0

---

---

**< E N D O F R E P O R T >**