

Cluster Analysis Syllabus

Course Description

In this course you will learn concepts and methodologies for cluster analysis, which is also known as clustering, data segmentation, or unsupervised learning. Cluster analysis is an important subfield of data mining. We will introduce the basic concepts of cluster analysis and then study a set of typical clustering methodologies, algorithms, and applications. This includes partitioning methods such as k-means, hierarchical methods such as BIRCH, density-based methods such as DBSCAN, and grid-based methods such as CLIQUE. We will also discuss methods for clustering validation. The learning will be enhanced by some clustering software and some programming assignments.

Course Goals and Objectives

Upon successful completion of this course, you will be able to:

- Recall basic concepts, methods, and applications of cluster analysis, including the concept of clustering, the requirements and challenges of cluster analysis, a multi-dimensional categorization of cluster analysis, and an overview of typical clustering methodologies
- Learn multiple distance or similarity measures for cluster analysis, including Euclidean and Minkowski Distances, proximity measures for symmetric and asymmetric binary variables, distance measures between categorical attributes, ordinal attributes, and mixed types, proximity measures between two vectors, Cosine similarity, and correlation measures between two variables: covariance and correlation coefficient
- Learn popular distance-based partitioning algorithms for cluster analysis, including K-Means, K-Medians, K-Medoids, and the Kernel K-Means algorithms
- Learn hierarchical clustering algorithms, including basic agglomerative and divisive clustering algorithms, BIRCH: a micro-clustering-based approach, CURE which explore well-scattered representative points, CHAMELEON which explores graph partitioning on the KNN Graph of the data, and a probabilistic hierarchical clustering approach
- Learn density-based approach to cluster analysis which can group dense regions of arbitrary shape, such as DBScan and OPTICS
- Learn grid-based approach which organizes individual regions of the data space into a grid-like structure, such as STING and CLIQUE
- Study concepts and methods for clustering evaluation and validation by introducing clustering validation using external measures and internal measures and the measures for evaluating cluster stability and clustering tendency

Textbook and Readings

Although the lectures are designed to be self-contained, we recommend (but do not require) that students refer to the book:

Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques* (3rd ed.). Waltham: Morgan Kaufmann.

You can download a PDF version of [Chapters 2, 10, 11, and 13](#) from *Data Mining: Concepts and Techniques*, 3rd Edition for free.

If you would like to purchase the entire textbook, the publisher has an exclusive offer just for Coursera students. You can save 30% on either the print or eBook version of *Data Mining: Concepts and Techniques*, 3rd Edition and receive free shipping on all orders. Here is how it works:

- [Add the book to your cart.](#)
- Enter code **COMP315** and click **Apply**.
- The discount will be applied to the list price and cannot be combined with other promotions.

Course Outline

This course consists of **4 weekly modules, with 6 lessons**.

Please note: there are no required readings for this course. All readings listed below are optional.

Module	Key Concepts	Suggested Readings from <i>Data Mining: Concepts and Techniques</i>
Week 1 (Lesson 1 - 2)	• Cluster Analysis: An Introduction • Similarity Measures for Cluster Analysis	• Chapter 2: Getting to Know Your Data • Chapter 10: Cluster Analysis: Basic Concepts and Methods
Week 2 (Lesson 3 - 4 part 1)	• Partitioning-Based Clustering Methods • Hierarchical Clustering Methods (Part I)	• Chapter 10: Cluster Analysis: Basic Concepts and Methods
Week 3 (Lesson 4 part 2 - 5)	• Hierarchical Clustering Methods (Part I) • Density-Based and Grid-Based Clustering Methods	• Chapter 10: Cluster Analysis: Basic Concepts and Methods
Week 4 (Lesson 6)	• Cluster Validation	• Chapter 10: Cluster Analysis: Basic Concepts and Methods

Elements of This Course

The course is comprised of the following elements:

- **Lecture videos.** Each week your instructor will teach you the concepts you need to know through a collection of short video lectures. You may either stream these videos for playback within the browser by clicking on their titles, or you can download each video for later offline playback by clicking the download icon.
- **In-video questions.** Some lectures have questions associated with them to help verify your understanding of the topics. These questions will automatically appear while watching the video if you stream the video through your browser. The questions are available as a separate downloadable text file for those who prefer to download the videos. These questions do not contribute toward your final score in the class.
- **Lesson Quizzes.** Each lesson will include one graded quiz. Your cumulative score will be used when calculating your final score in the class. There is no time limit for how long you take to complete each quiz. The

deadline for all quizzes is the last day of the course.

- **Programming Assignments.** There are two required programming assignments for this class. The first programming assignment is about implementing the K-means clustering algorithm, and the second programming assignment is about implementing clustering validation measures. **Both programming assignments are required for the Course Certificate.**

How to Pass This Course

To qualify for a Course Certificate, simply start verifying your coursework at the beginning of the course, **get a 75% or higher on all graded lesson quizzes, an 80% or higher on all programming assignments**, and pay the fee. Coursera [Financial Aid](#) is available to offset the registration cost for learners with demonstrated economic needs. If you have questions about Course Certificates, [please see the help topics here](#).

Also note that this course is the fourth course in the [Data Mining Specialization](#) offered by the University of Illinois at Urbana-Champaign. By earning a Course Certificate in this course, you are on your way toward earning a [Specialization Certificate in Data Mining](#). You may also choose to pre-pay for the entire Specialization at a discount. See more information about [Specialization payments](#) here.

If you choose not to pay the fee, you can still audit the course. You will still be able to view all videos and view required assessments. Auditing does not include the option to submit required assessments. As such, you will not be able to earn a grade or a Course Certificate.

The following table explains the breakdown for what is required in order to pass the class and qualify for a Course Certificate. You must pass each and every required activity in order to pass this course.

Activity	Required?	Number per Lesson	Estimated time per module	% required to pass
Lecture Videos	Yes	Varies	One hour	N/A
Lesson Quizzes	Yes	1	One hour	75% or higher
Programming Assignments	Yes	1 in Lesson 3 and 1 in Lesson 6	3 hours	80% or higher

Getting and Giving Help

You can get/give help via the following means:

- Use the [Learner Help Center](#) to find information regarding specific technical problems. For example, technical problems would include error messages, difficulty submitting assignments, or problems with video playback. If you cannot find an answer in the documentation, you can also report your problem to the Coursera staff by clicking on the **Contact Us!** link available on each topic's page within the Learner Help Center.
- Use the [Content Issues](#) forum to report errors in lecture video content, assignment questions and answers, assignment grading, text and links on course pages, or the content of other course materials. University of Illinois staff and Community Mentors will monitor this forum and respond to issues.

Note: Due to the large number of learners enrolled in this course, I am not able to answer emails sent directly to my account. Rather, all questions should be reported as described above.