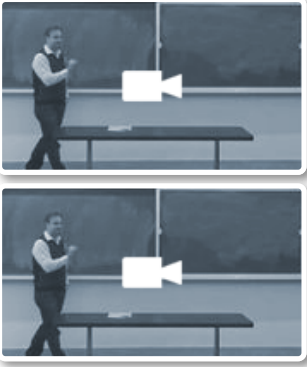

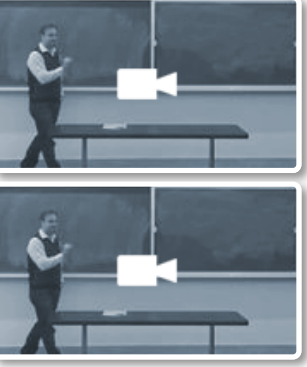


# CS6220/DS5230 Unsupervised Data Mining, SUMMER 2022

[About CS6220](#) [Home](#) [Schedule](#) [Piazza](#) [Gradescope](#) [DM Resources](#)

*\* Schedule and materials subject to change*

-->

Module / Live Stream	Topic / Recorded.Lecture	Other Reading	Assignment
<p>5/9 - 5/16</p> <p>Module 1 : Data Basics, Similarity, KNN Week 1 : Intro, Data Features, Mining Rules</p>	<p>Slides: DM Intro Slides: ML Intro Slides: Frequent Sets Mining</p> 	<p>Background:</p> <ul style="list-style-type: none"> <li>• Probabilities</li> <li>• Linear algebra</li> <li>• Programming: <a href="#">MATLAB</a>, Java, Python, R</li> </ul>	<p><a href="#">HW 1</a> Due: 5/23</p>
<p>5/16 - 5/23</p> <p>Week 2 : Similarity, KNN <a href="#">Lecture 3 Notes</a></p>	<p>Slides: Distance and Similarity Paper: Distance / Similarity Measures Slides: kNearestNeighbors predictions</p> 	<p>[A] ch 3</p>	
<p>5/23 - 5/30</p> <p>Module 2: Clustering Week 3 : KMeans <a href="#">Lecture 4 Notes</a></p>	<p>Slides: Intro to Clustering</p> 		<p><a href="#">HW 2A</a> Due: 6/6</p>
<p>5/30 - 6/6</p> <p>Week 4 : soft KMeans / Gaussian Mixture EM <a href="#">Lecture 5 Notes</a> <a href="#">Lecture 6 Notes</a></p>	<p>Notes: Gaussian Mixtures Mixture Matlab code</p>		<p><a href="#">HW 2B</a> Due: Fri 6/17</p>



6/6 - 6/13

Week 5 :  
Hierarchical,  
DBScan  
Lecture 7 Notes

Notes: PCA

6/13 - 6/20

Module 3: Dim  
Reduction,  
Features,  
Classification  
Week 6 : PCA,  
feature Selection  
Lecture 8 Notes

Class notes  
(handwritten+ DHS  
book): PCA

Paper: Harr Features    HW3A  
Paper: Feature Section    Due: Fri 7/1  
for Gaussian Mixtures

6/20 - 6/27

Week 7 : Feature  
Selection  
Lecture 9 Notes

Notes: ChiSquare\_FeatureSelection  
Wikipedia: Mutual Information

StanfordNLP:  
ChiSquare Feature  
Sealection  
StanfordNLP: Mutual  
Information Feature  
Sealection

6/27 - 7/4

Week 8 :  
Supervised  
Classification  
Lecture 10 Notes  
Linear Regression

Notes: Linear  
Regression  
Notes: Logistic  
Regression  
Notes: Regression  
Regularization

HW4  
HW4 due: Optional, No  
credit

7/4 - 7/11

Notes: Decision Trees

Module 3:  
Classification  
Week 8 :  
Supervised  
Classification

Notes: Perceptrons,  
Neural Networks  
Slides (Mitchell book):  
Neural Networks

Neural Networks

Decision Trees  
Lecture 12 Notes  
Decision Notes  
(Virgil)  
Boosting Notes



7/11 - 7/18

Slides: NMF  
paper: NMF  
Slides: LDA

Module 4: Text  
Modeling  
Week 9 : Topic  
Models, LDA  
Lecture 13 Notes  
Lecture 14 Notes



paper: LDA  
More Slides: LDA  
paper: Bayesian  
Parameter Estimation  
for text

HW5  
Due: 7/25

7/18 - 7/25

Sampling Basics  
(Matlab)  
Sampling MC/ Gibbs  
Demo

Week 10 :  
Sampling  
Lecture 15 NMF  
Lecture 17  
Markov chains  
Stevens Method:  
Sample Non-  
uniform Without  
Repetition

Rejection Sampling  
Inverse Transform  
Sampling

Book: Un-uniform  
Sampling Procedures



7/25 - 8/1

Paper: Text  
Summarization Survey  
Paper: Topic Modeling  
Summarization  
Paper: ROUGE  
Evaluation for  
Summaries  
Slides: ROUGE

Week 11 :  
Summarization  
Lecture 15 NMF  
Lecture 16  
Summarization



IR/Linguistics old  
paper: Automatic  
Abstracts

Summarization basics

8/1 - 8/8

Textbook: Aggarwal, Data Mining, ch 18-19  
Slides: Girvan - Newman Algorithm

Module 5: Graphs/  
Social Mining

Paper1: Girvan -  
Newman Algorithm  
Paper2: Girvan -  
Newman Algorithm

HW6  
Due: 8/16

Week 12 : Social  
Graphs  
Lecture 19 Graph  
Intro/Communities  
Lecture 20 Graph  
Communities

8/8 - 8/15

Textbook: Aggarwal,  
Data Mining, ch 18-19

Notes: collaborative  
filtering basic formula  
Slides: Netflix User  
Profiles

Week 13 : Social  
Mining  
Lecture 18 Collab  
Filtering  
Lecture 21 KB-  
QA

FINAL EXAM 8/19 in class

You will need a computer for the exam problems, and might be  
called to explain/demo your code after.

Paper3: Girvan -  
Newman Algorithm

