

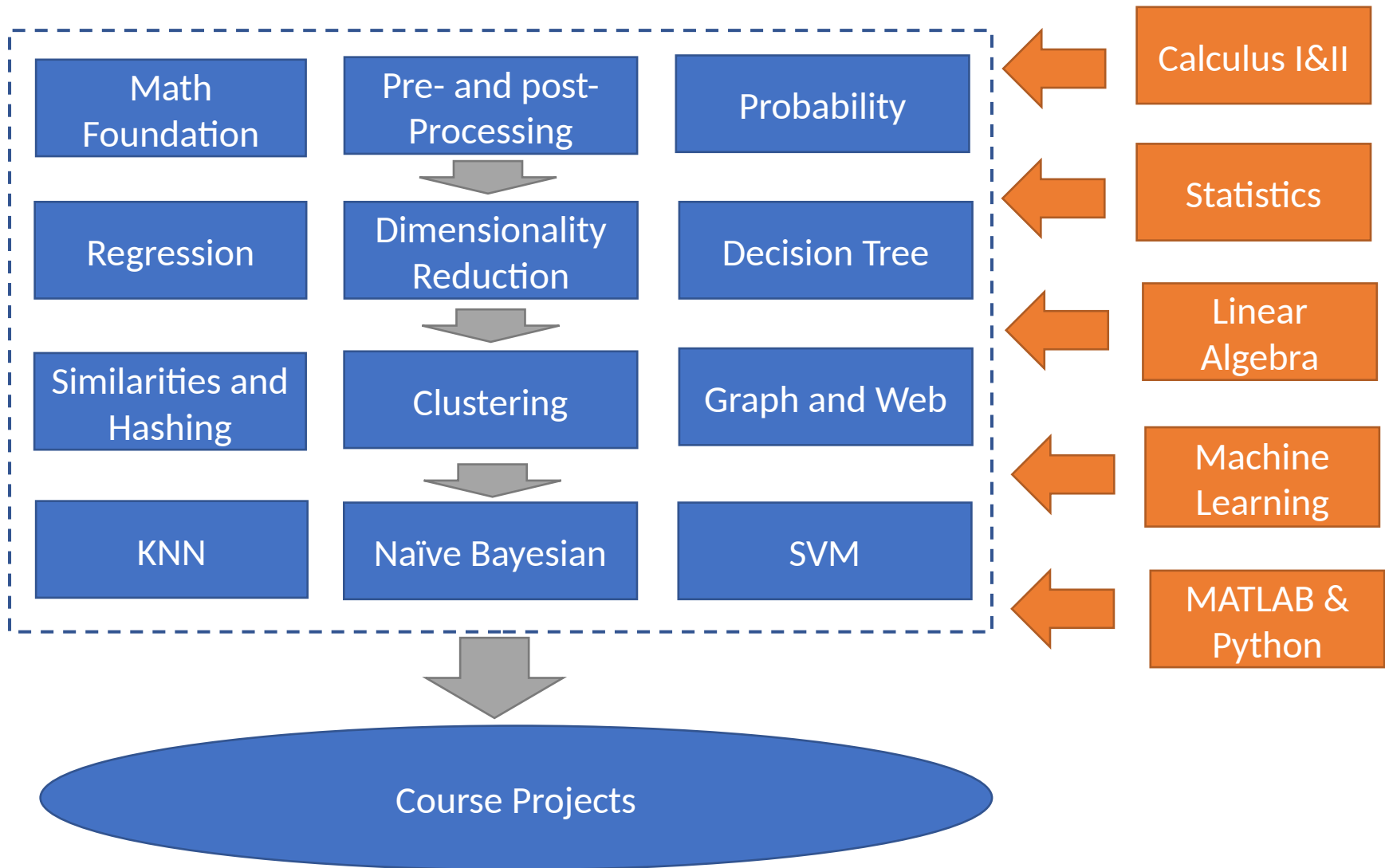
CIS 530—Advanced Data Mining

1- Course Introduction

Thomas W. Gyeera, Assistant Professor
Computer and Information Science
University of Massachusetts Dartmouth

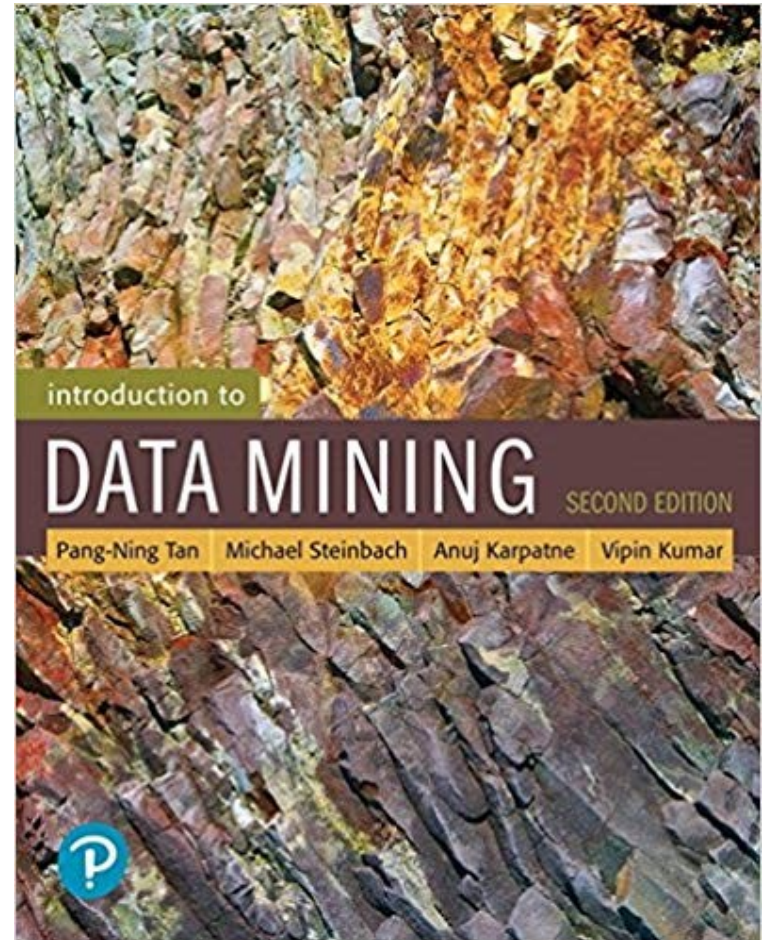


Class Roadmap



Textbook

- Pang-Ning Tan, Michael Steinbach, Anuj Karpatne, Vipin Kumar, *Introduction to Data Mining*, 2nd Edition, published by Pearson, Copyright © 2019, ISBN-13: 978-0-13-312890-1



Course Details

CIS 530—Advanced Data
Mining

Spring 2023

Where: SENG 115

TA: Anudeepsri Bathina
abathina@umassd.edu

Preferred Skills Set in the Class

Primary Programming tool(s)

- MATLAB, Python

Algorithms and data structure

- Divide and conquer, dynamic programming, ...

Linear algebra, statistics

- Vector, matrix, rank, eigen-decomposition, SVD
- Expectation, variance, maximum likelihood

Visualization

And more...

Objectives

Compare the learned data mining algorithms and find the strength and weakness of them

Able to apply the appropriate data mining algorithms for real-world problems of different data types

Able to improve the data mining algorithms for real-world problems of different data types

How to Grade

Weighted scores

- 25% Assignments
- 25% Midterm exam
- 50% Project

Expect decent scores

- Only if you follow the course announcements and instructions

Zero tolerance to cheating!!!!

- One time ☾ **C**
- Two times ☾ **F**

How to Grade

- Attendance
 - Attendance is advised and checked randomly
- Mid-term
 - 3 hours exam
 - Everything in the course before the exam
- Assignment (usually due in one week)
 - Five assignments
 - Will be released and graded in myCourses
- Team up ASAP!!
 - Make your life easier
 - 3 students
 - **Same** for assignment, and project

Project - Proposal + Project presentation + Report



Proposal presentation

Each team 5 mins



Final presentation

Each team 10 mins
Cover everything in the
report



Report

IEEE Standard paper +
code/demo

Doc to enter your Project team details

• <https://forms.office.com/r/mJYBNYEfML>

\$ 530 - Advanced Data Mining



Sample Assignments

- 1.3** (★★) Suppose that we have three coloured boxes r (red), b (blue), and g (green). Box r contains 3 apples, 4 oranges, and 3 limes, box b contains 1 apple, 1 orange, and 0 limes, and box g contains 3 apples, 3 oranges, and 4 limes. If a box is chosen at random with probabilities $p(r) = 0.2$, $p(b) = 0.2$, $p(g) = 0.6$, and a piece of fruit is removed from the box (with equal probability of selecting any of the items in the box), then what is the probability of selecting an apple? If we observe that the selected fruit is in fact an orange, what is the probability that it came from the green box?

Q2. (15 pts) Consider the below joint probability distribution between x and y .

		y	
		0	1
x	0	$1/3$	$1/3$
	1	0	$1/3$

- Compute the marginal probabilities $p(x)$ and $p(y)$ for each value x and y take on.
- Are x and y independent? Prove your result.
- Compute conditional probabilities $p(x|y)$ and $p(y|x)$ for each value x and y take on.

Sample Projects

Hot topics in Data Mining

- Link prediction
 - Friend suggestions
- Recommendation system
 - Product recommendation
- Prediction
 - Disease prediction using EHR
- Community discovery
 - Social network analytics

Where to find data and problems??

- Kaggle

Sample Projects

- A good report may have:
 - Background + Related work
 - Problem definition
 - Better a new problem
 - Methodology
 - Existing work or your own method (preferred!!)
 - Experimental setting
 - Experimental results and analysis
 - Your own implementations!
 - Conclusion