Part 1 Review

Part 1, session 6a of Data Mining Intro

2025-01-19 21:02 GMT

Abstract

Part 1 review: audience, content, pedagogy, and goals.

Table of Contents

Introduction

The aim of this session is to review sessions of the previous week, in order to prepare to present this material going forward. The primary reference is `eda_v3`:

- Notes on Exploratory Data Analysis (EDA)
  - MATH4350 version 3
  - by Karen Trageser

The subsequent material for the course is based on the following `mlci` reference:

- Machine Learning: a Concise Introduction
  - by Steven W. Knox
  - next edition pending publication by Wiley

We first compare the `eda_v3` content to the selection of last week's topics, and then discuss the intended audience, the intended learning outcomes, selection and organization of topics, and the use of time during class sessions.

Overview of Topics

Here are the chapter titles of `eda_v3`.

EDA v3 chapters

chpt	title
1	Introduction
2	Exploratory Data Analysis (EDA)
3	Unsupervised Learning
4	Some Linear Algebra
5	Dimension Reduction
6	Topic Modeling
7	Sampling
8	Time Series
9	App-A: Probability Review
10	App-B: SVD Notes by Carla Martin

This represents a broad range of topics to be covered in just a few days. Each chapter can easily take up a one-semester course at the undergraduate or graduate level. Consequently we must select and organize topics very carefully to equip participants with essential tools and prepare them for the remainder of the course

Here are the topics we discussed last week.

Topics discussed Jan 13-17

date	time	topic
2025-01-13	AM	Exploratory Data Analysis
2025-01-13	PM	Conditional Distributions
2025-01-14	AM	Clustering: EDA in Higher Dimensions
2025-01-14	PM	Text Analysis
2025-01-15	AM	Sampling and Study Design
2025-01-15	PM	Linear Algebra
2025-01-16	AM	Dimension Reduction
2025-01-17	AM	Time Series
2025-01-17	PM	Time Series & Point Processes: Frequency Analysis

Class Exercises

For each of the following class exercises, form a small group with one or two classmates. Use the alloted time to prepare to report out to the class.

Clarification Needed (10 minute prep)

Which topics presented last week need clarification? Which topics from `eda_v3` would you like to know more about? Take 10 minutes to prepare to report out to the class. We'll then discuss the top one or two topics of greatest interest.

Selection of Topics (15 minute prep)

As you prepare to teach MATH4350, which topics and key points from `eda_v3` would you emphasize? Based on your experience with the material discussed last week, how would you organize classroom sessions? What guidance would you offer a friend preparing to teach this material? Take 15 minutes to prepare to report out to the class.

Intended Participants (10 minute prep)

Last week several of us noted that the EDA content was presented largely from a mathematical viewpoint, prompting us to consider whether and how to engage those potential participants whose expertise was more computational than mathematical. Take 10 minutes to report out to the class on the questions below, or on related questions of your own.

- The US version of the course lays out pre-requisites in mathematics, statistics, and computation, citing established textbooks.
- For the UK version of the course, should those pre-requisites be modified? How?
- How should the readiness of a potential participant be assessed?
- When should a potential participant be advised to shore up needed pre-requisites on their own? What guidance should we give on courses or other learning resources? Are high-side refresher courses feasible?

Learning Outcomes (15 minute prep)

From last week's discussions, we seem to be agreed that the goal of the EDA material should be to equip participants with tools they can (responsibly) use in their work, either immediately or eventually. Take 15 minutes to consider the questions below and then report to the class how you would express your goals for participant learning of the EDA content.

- Should this first week of content also be used to level-set, to ensure that participants share some common knowledge and vocabulary needed for the remainder of the course? If so, how should we define that baseline?
- As an instructor, should you attempt to detect gaps in needed participant knowledge? How would you go about that? How would you address such gaps?

- Responsible use of data science methods requires knowledge of the capabilities and limitations of the methods, and of one's current level of proficiency in a wide-ranging set of competencies. How would you help participants get a clear picture of both?

- How would you summarize your goals as an instructor for participants with respect to the EDA content?

Teaching Style (10 minute prep)

Each of us goes about learning a new area in our own distinct manner. As instructors we naturally emphasize what we know or enjoy most, and tend to present content in a way that we would want that content presented to us. How should we take advantage of instructor strengths while maintaining some degree of consistency in content and delivery? Take 10 minutes to consider the following questions and then report out to the class.

- In the US `eda_v3` followed by `mlci` are the reference materials on which MATH4350 is based. Other instructors contribute content on specific topics, thereby enabling each instructor to assemble content as seems best. Is this a model for the UK version of the course? Where and how should the content be maintained?

- In your experience as an instructor, have you ever encountered a course participant who seemed to learn differently from the way you learn?

- As a community of instructors, how should we share our successes and challenges?

- How, if at all, should we vet potential instructors?