2025

CS 4407
Data Mining & Machine Learning

LEARNING JOURNAL UNIT 1

SANA UR REHMAN

INSTRUCTOR: NIRMAL ADHIKARI

INTRODUCTION

Basics of Data Mining and Machine Learning were first taught to me this week. I delved into the definitions, their uses and their differences, primarily in supervised vs. unsupervised learning. Data mining is about extracting patterns from large data sets; machine learning is about building predictive models which are refined and improved over time. I also learned about supervised learning, where algorithms are trained using labeled data, and unsupervised learning, which finds hidden patterns in unlabeled data, from the readings and active parts of the course. I also learnt and became familiarized with R programming, which I will be using as a toolbox when applying algorithms in the future HomeWorks. These ideas really are the groundwork for seeing how data-driven methods can fix real-world issues.

DIFFICULTIES FACED

One problem I encountered was differentiating data mining from machine learning. There does seem to be a fair amount of overlap between the two fields, but I had an initial hunch that data mining tends to be more about knowledge discovery, whereas machine learning is more geared to predictive modelling. The next challenge was to get to know R programming language. While the examples were simple, I found the syntax to be completely new to me, and I quickly realized that I would need a lot of practice to get good. Also, in reading about supervised and unsupervised learning, I discovered that when we tune and apply unsupervised algorithms, it's less intuitive to gauge the performance of the model because you don't have explicit labels that you can check it against.

ACTIVITIES PERFORMED

I was involved in many activities that all became clear at the end of the week. I created a discussion post that compared supervised and unsupervised learning; in this post, I defined supervised learning as learning from labeled data to make predictions about unseen or future data, and unsupervised as discovering

patterns in data: (Han et al., 2022; Aggarwal, 2015). Writing out this post helped me articulate my learning and criticize the material on my own. I took the self-quiz, too, which also challenged me to test myself on important differences and facts. On attempting the quiz questions, it struck me that I could now, for the first time, consciously come up with representative examples for each type of learning, e.g. spam-fighting-for-supervised vs cluster-the-patients-for-unsupervised.

REFLECTION

One of the surprises for me was how well-supervised and unsupervised complemented each other. I thought they were kind of two different entities, but I guess semi-supervised learning can use both forms together (particularly if we don't have enough labeled examples). This made me think about how ubiquitous hybrid approaches are in everyday industry.

The most difficult part was figuring out how to use R. It's more than just getting the syntax right, it's about understanding how R stores and works with data. This seemed like a lot at first, but it also illustrated to me the need to be a bit elastic when you are learning a new programming tool.

I am aware that I am gaining technical and analytical skills. Technically, I am gaining an understanding of how algorithms function and how programming utilities can be used to make them do what needs to be done. Analytically, I am gaining the expertise to evaluate what type of learning is appropriate to a problem at hand. This ideal combination of expertise helps me realize that effective data analysis requires coding expertise as much as it requires critical thinking.

As a learner, I am realizing that I tend to focus more on the "why" rather than the "how." I need to understand the general purpose of an approach before I can apply it in practice. It makes me capable of transforming theory into practice, and I think this will be a strength as the course becomes more practice oriented.

In practice, I already see how clustering techniques can be used to segment customers, or how classification techniques can support medical decision-making. Thinking about these practical applications motivates me since I can relate ideas to actual issues.

CONCLUSION

Learning experience this week reinforced my conceptual knowledge in machine learning and data mining. I not only understand the difference between supervised learning and unsupervised learning, but I also understand how to apply each method to real-life situations. I also understood that it would take getting used to practicing constantly when using R programming, but I am motivated to improve. What was most noticeable to me is how closely related data mining and machine learning are but what different purposes they serve in data analysis. In the future, I wish to continue developing my ability to apply these subjects beyond theory by experimenting with real datasets.

REFERENCES

Aggarwal, C. C. (2015). *Data mining: The textbook*. Springer. https://doi.org/10.1007/978-3-319-14142-8

Han, J., Pei, J., & Tong, H. (2022). *Data mining: Concepts and techniques* (4th ed.). Morgan Kaufmann. https://shop.elsevier.com/books/data-mining/han/978-0-12-811760-6

Word count: 784