Alhassane Samassekou

ITAI 1371

Reflective Journal: From Theory to Practice in the ML Workflow

The biggest concept that was interesting in this lab for me was data splitting. I had heard about training and testing sets, but it always seemed a bit abstract. Seeing the train_test_split function in action made it all make sense. It's like studying for a final exam. If the professor gives you the exact exam questions and the answer key to study, you'll ace the test, but you will not have actually learned the material. You just memorized the answers. Testing the model on data it has never seen before is the only true way to know if it has learned to generalize, or if it just "cheated" by memorizing the training data. This idea of holding out a portion of the data for a final, unbiased evaluation seems so simple now, but it is clearly one of the most important steps to building a trustworthy model.

It was also really interesting to train two different models, Logistic Regression and a Decision Tree, on the exact same data and see them produce slightly different results. It showed me that there is not a single "best" algorithm for every problem. The Decision Tree performed a bit better, and that got me thinking about why. It is not just about plugging data in, it is about understanding how these models "think" and which one might be a better fit for the patterns in your data. Going beyond just the accuracy score to look at the confusion matrix was another major step. Seeing which specific wine classes the model was mixing up is so much more insightful than a single percentage. It gives you clues about where to focus your efforts to improve the model next.

This lab also did a great job of connecting the different types of machine learning to the practical exercise. We were obviously doing supervised learning because we had the chemical features and the correct wine class labels. This hands on experience made the definitions of the other types feel much more real. I can now imagine how you could use unsupervised learning on this same dataset to discover natural clusters of wines without using the labels, maybe finding a new way to categorize them. And I can see how reinforcement learning is a totally different beast, focused on learning through trial and error, which makes perfect sense for games or robotics.