ANNETTE BAZAN

ITAI 2372:

DEEP LEARNING

FEBUARY 25, 2025

LAB 05: AWS ML

UNIVERSITY MODULE 2

LAB EXPLORATION

Module 2 of the AWS Machine Learning University course spanned from lab 1 through lab 4. Introducing foundational and advanced techniques in NLP (natural language processing), these labs covered text vectorization. Text vectorization using the BoW method (Bag-of-Words), GloVe word embeddings, cosine similarity, and recurrent neural networks (RNNs) for text classification. This journal hopes to reflect my learning experience as I challenge my way through these 4 labs, record any growth and a critical evaluation of these experiences.

The AWS ML learning labs provided a structured introduction in blending theoretical concepts with practical applications. Each lab built upon the previous one creating cohesive learning experience to help build up on learning and retaining new information and experience. Lab 1 introduces the groundwork by introducing the distinction between supervised and unsupervised learning. Supervised learning uses labeled data to train models for prediction while unsupervised learning identifies patterns without the predefined labels. This foundational knowledge was being reinforced through the overview from the AWS Sage Maker, this platform streamlines the machine learning process. The Sage Maker is impressive with how it integrates data preparation, model training, and deployment into a single interface, making machine learning feel approachable despite its complexity. Lab 2 tasks had me built a linear regression model to predict house prices, diving into feature engineering and data preprocessing. Feature engineering involved selecting relevant variables like square footage or number of bedrooms. Then transforming them to improve model accuracy, this hands-on experience helped me build more confidence in handling missing values and normalize data, steps that ensure the model performs reliably. Splitting the dataset into training and testing sets was a key takeaway as it taught me to assess a model's ability to generalize to new data. Predicting house prices has made the abstract concept of regression tangible showing me how machine learning applies to everyday problems. Lab 3 introduced more advanced algorithms such as decision trees and random forests and exploring hyperparameter tuning. By working with classification tasks which required understanding how algorithms differ from regression models. Hyperparameter tuning using techniques like grid search and cross-validation revealed the iterative nature of machine learning. Adjusting the parameters like tree depth or the number of estimators in a random forest, was trail and error process that deepened my appreciation for the balance model complexity and performance. Lab 4 focused on evaluating model performance and deploying it into production. It teaches metrics like accuracy, precision, recall, and F1 score, each offering a different lens on model effectiveness. Deploying the model via Sage Maker endpoint was another highlight of this group of labs. It demonstrated how machine learning transitions from experimentation to real-world utility. This last lab tied together the entire pipeline from data preparation to deployment and underscored the practical value of skills I enjoyed developing and enforcing. These labs collectively taught the end-to-end process of

machine learning emphasizing the importance of data quality, algorithm selection, and evaluation. The real-world examples- predicting house prices and classifying the data.

Each lab had its challenges, and each challenge was helping my growth. There were more challenges from lab 2 through lab 4. Lab 2 exposed what I needed to work on when data was preprocessing. I accidentally skipped checking for missing values only to encounter errors during the model training. This mistake caused me to lose time and taught me to explore the dataset thoroughly and carefully before proceeding. It taught me to create a preprocessing checklist. Verify missing data, encoding categorical variables, scale features that I wrote down to help remember it. This structured approach prevented similar issues as I proceed to the next labs. In lab 3, understanding why we use a decision tree versus a random forest. After reading more the theoretical differences made more sense. Random forest generally outperformed decision trees in terms of accuracy. This hands-on help to understand and clarify their practical applications. The hyperparameter tuning proved daunting as well. The grid search process, which tests multiple combinations of parameters, was time intensive and I wasn't sure which parameters mattered the most. I spent some time tweaking minor settings before realizing the number of trees in a random forest had a bigger impact than I'd assumed. By using step-by-step approach by starting with a smaller parameter grid, analyzing the results and then refining my search. Which made tuning more efficient and less overwhelming. In lab 4 interpreting evaluation metrics for classification models was confusing. I did end up mixing precision which the accuracy of positive predictions and recall the ability to find all positive cases because it was unclear which to prioritize. To clarify, I manually calculated these metrics using a small confusion matrix which helped me see how it relates to real world trade-offs. If you think about medical diagnosis scenario, recall might matter more than precision to avoid missing cases.

Before these labs my knowledge was limited to the concepts from lectures and PowerPoint slides. The linear regression equations, for instance, were just formulas on a page. Building the model in lab 2 to predict house prices brought those equations to practical understanding for me. Being able to adjust features and seeing their impact on predictions made the math more understandable. The iterative approach was complex at first but as it was a syntax error more than anything, tuning hyperparameters wasn't just a linear process. The labs equip confidence in my skills to perhaps become a data scientist if these labs keep boosting my confidence and provide great structured hands-on labs.

The labs ignited an interest in advanced topics like deep learning and NLP a lot more. Lab 4 mentioned deploying models made me question if a neural network could enhance predictions while classification tasks that hint at NLP applications like sentiment analysis. Covering supervised learning and deployment but only scratched the tip of the iceberg of unsupervised learning, reinforcement learning and ethical issues like bias in models. Recognizing this, the labs serve as a foundation for more complex projects, so they provide the basics for real-world challenges.

In conclusion, this reflective journal goes over my journey of labs 1 through 4. A transformative experience that spans technical learning, personal challenges, and critical insights. Being able to overcome preprocessing to deploying models, it helps me gain practical understanding of machine learning's core components. The struggles facing between choosing algorithms, tuning parameters, while the hands-on approach to help foster my growth. As we move forward through more labs the skills, we gain from previous labs help to build upon the future labs.