Faiza Abdullah

ITAI 1371 Intro to Machine Learning

Professor: Viswanatha Rao

Reflective Journal: Lab 03 - ML Workflow & Learning Types

GitHub:

https://github.com/AbdullahFaiza/ITAI_ML_FirstRepo_FaizaAbdullah/tree/main/ITAI-1371-ML-Labs/L03%20Lab%20ITAI%201371

Key Insights: The lab illustrates the structured ML workflow's power in transforming raw data into actionable predictions, shifting my view from abstract concepts to a practical, iterative process. Distinguishing between supervised, unsupervised, and reinforcement learning clarified their unique applications, supervised learning's labeled predictions felt most intuitive for classification tasks like the Wine dataset. Building and evaluating models, such as Logistic Regression outperforming Decision Tree with ~88.9% accuracy, underscored evaluation metrics' role in model selection. Experimenting in the hands-on section with features like ['alcohol', 'color_intensity', 'proline'] had lesser accuracy to 0.833, however by trying alternatives like flavanoids and total_phenols while keeping alcohol intact improved accuracy to 0.944. Adding inline comments to sections, such as noting class imbalances in EDA or misclassifications in confusion matrices, enhanced reproducibility and deepened my understanding of data-model interplay.

Challenges Encountered: Grasping feature selection's nuances during the hands-on exercise, where initial combinations yielded varying accuracies, emphasizing trial-and-error's role. This conceptual hurdle, like deciding between correlated features in EDA's heatmap, required patience to avoid overfitting. Documenting observations via inline comments, such as "# Better features improved performance," helped clarify steps and mitigate errors propagating to evaluation. Navigating the full workflow (splitting data, training models, interpreting results), added complexity, especially balancing stratification for class ratios. The assessment on ML types tested application, but mapping scenarios like house prices to supervised reinforced learning through practice.

Connections to Real-World Applications: The focus on supervised classification connected to healthcare scenarios, like predicting patient readmissions using features such as age and lab results, mirroring the Wine dataset's chemical analysis. Evaluating models with accuracy and confusion matrices evoked fraud detection in banking, where precision/recall balance prevents costly errors.

Comparisons Between Approaches: Comparing supervised (labeled, predictive) to unsupervised (pattern-finding) and reinforcement (trial-error) approaches highlighted supervised learning's accessibility for beginners, as seen in the lab's Wine classification versus hypothetical clustering. Logistic Regression's linear efficiency versus Decision Tree's interpretability but lower performance taught context-based choice; trees suit explainability, regression handles probabilities.

Questions That Arose: While experimenting with features and adding inline comments, questions emerged about optimization: how to systematically select the best features beyond trial? I wondered: what trade-offs exist between model complexity and interpretability, like deeper trees versus simple regression? While assessing types, I pondered handling imbalanced datasets effectively in real workflows.

Conclusion: This lab consolidated previous course learnings, elevating them from steps to integrated cycles through model building, evaluation, and type distinctions. Challenges like feature selection honed my analytical skills, while real-world ties affirmed their practicality. Documenting via inline comments clarified synergies, emphasizing iterative experimentation, like hands-on trials, for robust pipelines.