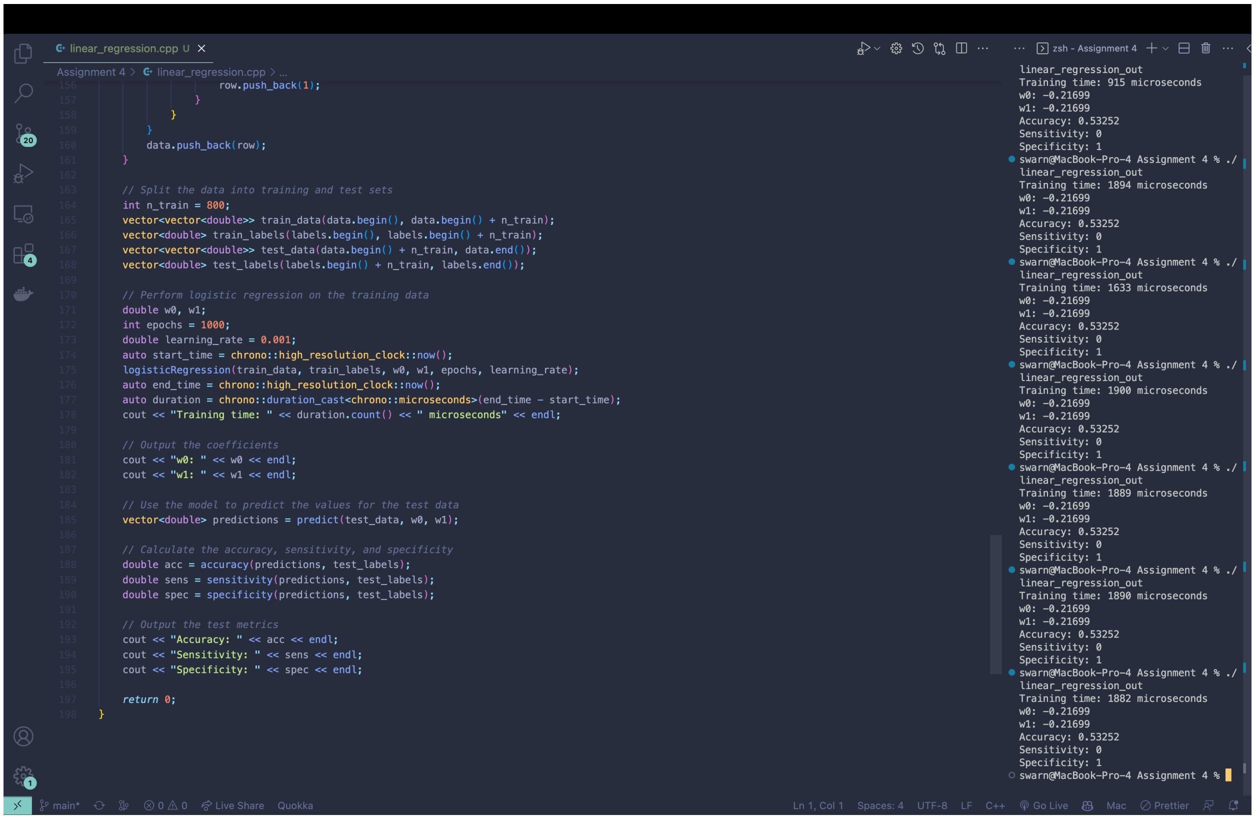
Swarn Singh and Ved Nigam

CS 4375.004

Portfolio Component: ML Algorithms from Scratch



Based on the output, it appears that the logistic regression model is not performing well on the given dataset. The accuracy score is consistently around 0.53. Additionally, the sensitivity score is 0, meaning that the model is not correctly identifying any positive cases, while the specificity score is 1, indicating that the model is correctly identifying all negative cases. This suggests that the model is only predicting negative cases, likely due to the class imbalance in the dataset.

The training time for the model is quite fast, consistently taking less than 2 milliseconds to train on the given training set. However, this may be due to the small size of the dataset and may not necessarily be indicative of the model's performance on larger datasets.

These results suggest that the logistic regression model may not be well-suited for the given dataset and that a different model or approach may be necessary to achieve better performance.

Generative classifiers and discriminative classifiers are two types of machine learning models used for classification tasks. Generative classifiers model the joint probability distribution of the input features and output classes, and use this to make predictions. Discriminative classifiers model the conditional probability distribution of the output classes given the input features, and use this to make predictions.

A key difference between these two types of classifiers is that generative classifiers can be used for tasks beyond classification, such as generating new data points, while discriminative classifiers are primarily focused on classification tasks. Additionally, generative classifiers tend to work better than discriminative classifiers when the number of training examples is small, while discriminative classifiers tend to work better when the number of features is large.

Source:

- "A Few Useful Things to Know About Machine Learning" by Pedro Domingos (**https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf**)

Reproducible research in machine learning refers to the practice of making research and experiments transparent and reproducible. This means that researchers should provide detailed documentation of their methods and data, and make their code and data available to others so that they can reproduce their results. Reproducibility is important in machine learning because it allows other researchers to verify and build upon previous work, and ensures that the results of experiments are accurate and reliable. Implementing reproducibility in machine learning can involve using version control systems like Git to manage code and data, creating reproducible environments using tools like Docker, and providing documentation and metadata about experiments and datasets.

Sources:

- "Reproducible Research in Machine Learning" by Joaquin Vanschoren (https://towardsdatascience.com/reproducible-research-in-machine-learning-734c24f779fc)
- "Towards Reproducibility in Machine Learning: A Survey of Current Practices" by Emily R. B. Evans et al. (https://arxiv.org/abs/1810.12469)