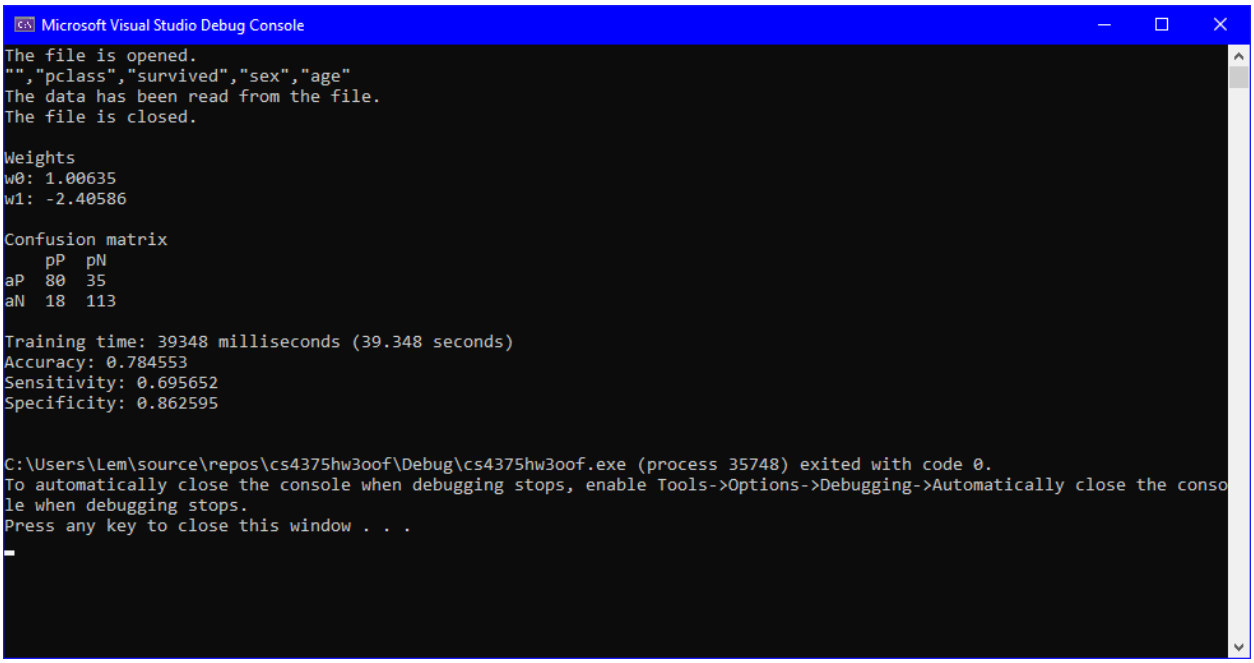
Zuhayr Ali

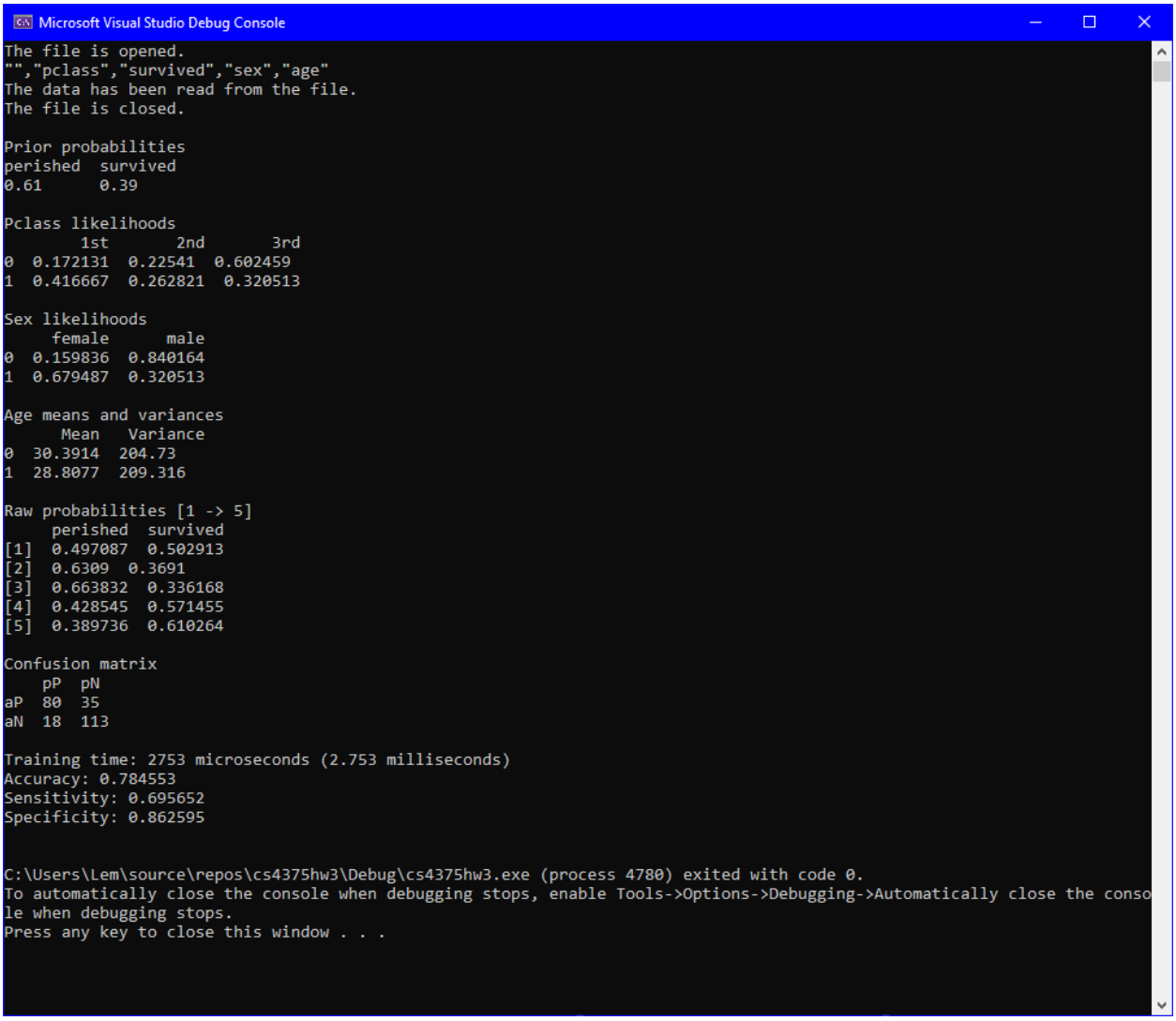
Dr. Karen Mazidi

CS 4375

12 September 2022

ML Algorithms from Scratch





Both models produced identical results and metrics, but the naïve Bayes model was dramatically quicker, computing in a couple milliseconds when the logistic regression model took dozens of seconds to achieve the same accuracy. The naïve Bayes assumption that all factors are independent could be responsible for this, as the training ultimately reduces to finding either the likelihoods or the mean and variance of each factor and so for a model with only three factors iterates through less loops than logistic regression by multiple orders of magnitude.

Naïve Bayes is an example of generative classification. Generative classification is designed to understand all classes known in the domain before assigning an observation to one of

the classes based on closest match. It is especially useful on smaller datasets, and takes less work to implement. However, it can also be inefficient because it will compare an observation against every class to find the best match, which spends longer and longer amounts of time with increasingly multiclassed classification regardless of how strongly an observation aligns with one class.

Logistic regression is an example of discriminative classification. Discriminative classification is designed to directly find the class of an observation based on the differences between classes. Although it is more work to implement, discriminative classification is more accurate on larger datasets. Also, like in the case of logistic regression where an observation is plugged into a standardized model, this style of classification is more efficient as it compares only the differences between classes to where an observation sits relative to a boundary condition.

Reproducibility in machine learning is the capacity of a study to be repeated with the same materials to produce the same results. It's very important as it not only validates the findings of a study but provides transparency into the inner mechanisms of what is being studied. It proves a study to be reliable. Types of reproducibility include methods reproducibility, which emphasizes the replication of the procedures, data, and tools to produce the same results, results reproducibility, which emphasizes the replication of the results for following the same experimental procedures, and inferential reproducibility, which emphasizes an agreement in findings from either an independent study or a peer review.

References

S. N. Srihari, “Machine Learning: Generative and Discriminative Models,” in *Introduction to Machine Learning: Course Materials*.

Z. Ding, A. Reddy, and A. Joshi, “5 – Reproducibility,” *MLCMU*, 31-Aug-2020. .