Alekhya Pinnamaneni and Aloksai Choudari

Dr. Karen Mazidi

CS 4375.003

2 October 2022

ML Algorithms from Scratch

a.

```
Welcome to the Logistic Regression algorithm from scratch!

Reading in Titanic data...

Calculating coefficients for logistic regression model...

Coefficients:
        Intercept = 0.9999
        Sex = -2.4108

Predicting values using test data...

Calculating test metrics...

Test metrics:
        Accuracy = 0.784553
        Sensitivity = 0.695652
        Specificity = 0.862595

Algorithm runtime = 3.0803s
```

```
Welcome to the Naive Bayes algorithm from scratch!

Reading in Titanic data...

Calculating prior probabilities...

Prior probabilities:
        Perished = 0.61
        Survived = 0.39

Calculating likelihoods for pclass, sex, and age...

Likelihood for p(pclass|survived):
                1               2               3
Perished        0.172131        0.22541         0.602459
Survived        0.416667        0.262821        0.320513

Likelihood for p(sex|survived):
                Female          Male
Perished        0.159836        0.840164
Survived        0.679487        0.320513

Age:
                Mean            Standard Deviation
Perished        30.4182         14.3085
Survived        28.8261         14.439

Calculating test metrics...

Test metrics:
        Accuracy = 0.784553
        Sensitivity = 0.695652
        Specificity = 0.862595

Algorithm runtime = 0.0043647s
```

b. Our algorithms on the Titanic data produced the same results as the built-in functions in R produced on the data. In addition, naïve bayes and logistic regression had the same accuracy, sensitivity, and specificity. Although the runtime for logistic regression was much longer than naïve bayes due to the iterative approach, using gradient descent, taken by logistic regression, they have the same results at the end. Nonetheless, in regard to performance, naïve bayes is a better approach.
c. This assignment allowed us to learn more about the advantages, disadvantages, and the differences between generative classifiers and discriminative classifiers. In this assignment, the Naïve Bayes algorithm is an example of generative classifiers, and the logistic regression algorithm is an example of discriminative classifiers. Generative classifiers involve observing the environment to note what conditions contribute to a certain outcome. This often requires calculating conditional probabilities to make a prediction of what the most likely outcome is. Algorithms using generative classifiers try to model the class and the features of the class. They try to figure out what features the input data should have to generate that particular class. Discriminative classifiers on the other hand focuses more on the input data. These algorithms learn which features of the data are the best predictors of the various potential outcomes. These algorithms calculate posterior probability, rather than conditional probability like generative classifiers, to create an accurate model. Rather than focusing on observing the environment, discriminative classifiers algorithms search for direct maps between an input x and an outcome class of y. Discriminative classifiers are more commonly used because they often tend to be more accurate than generative classifiers. This is due to the fact that the discriminative classifier tries to find direct links between x and y, rather than taking in indirect approach by modeling classes like the generative classifier.
d. Reproducibility in machine learning is the ability to run an algorithm on various datasets and get the same, or similar results on a specific project. It is important to machine learning, and projects in general, because it gives value to the cycled use of data management and programming. As data changes are bound to happen, reproducibility is vital to keep the success of a software environment. Another important aspect of reproducibility is the accuracy and validity it provides to some research. It allows an experiment to be more reliable in a sense and decreases the chances for gambling with results and data analysis.

Works Cited

"The Importance of Reproducibility in Machine Learning Applications." *DecisivEdge*, 14 Oct. 2020, https://www.decisivedge.com/blog/the-importance-of-reproducibility-in-machine-learning-applications/.

Improving Reproducibility in Machine Learning Research. May 2021, https://jmlr.org/papers/volume22/20-303/20-303.pdf.

Malhotra, Akanksha. "Generative Classifiers v/s Discriminative Classifiers." *Medium*, Medium, 16 Oct. 2019, https://medium.com/@akankshamalhotra24/generative-classifiers-v-s-discriminative-classifiers-1045f499d8cc.