Valari Graham, Jeffrey Li

Dr. Mazidi

CS 4375

ML from Scratch

a. Copy/paste runs of your code showing the output and run times

Logistic Regression:

```
Intercept coefficient: 0.999869
Sex coefficient:       -2.41085

Accuracy:    0.784553
Sensitivity: 0.695652
Specificity: 0.862595

Algorithm run time: 0.413002s
```

Naive Bayes:

```
Survivability for pclass:
1               2               3
0.172131        0.22541 0.602459
0.416667        0.262821        0.320513
Survivability for sex:
0               1
0.159836        0.840164
0.679487        0.320513

Run Time Results: 0.002904 seconds

        RESULTS
-----------------------
Accuracy: 0.784553

Sensitivity: 0.695652

Specificity: 0.862595
```

b. Analyze the results of your algorithms on the Titanic data

Based on the results from above, we can see that logistic regression had a run time slower than naïve bayes. The training for naïve bayes was faster due to no coefficients needing to be calculated by optimization procedures. The outputs for accuracy, sensitivity, and specificity were congruent in both programs.

c. Write two paragraphs comparing and contrasting generative classifiers versus discriminative classifiers. Cite any sources you use.

According to Analytics Vidhya, generative models are presumed to be a "class of statistical models that can generate new data instances." Models like these are seen primarily in unsupervised machine learning. Popular tasks performed with such models are, probability estimation, modeling data points, and differing between classes. A common example of generative modeling is, Naïve Bayes. On the other hand, discriminative models are used in statistical classification. Contrary to generative modeling, discriminative models are used for supervised learning. Models of these types grow to understand the boundaries between labels in a dataset.

The largest difference between these two models is that discriminative modeling creates boundaries in the data space, and generative modeling portrays how data is distributed throughout the space. The generative models can work well with missing data while discriminative models cannot because all features must be observed. The greatest similarity between both models is that they are both very useful in machine learning and the decision between which to use comes down to the task at hand and the data available.

d. Google this phrase: reproducible research in machine learning. Using 2-3 sources, at least one of which should be academic, write a couple of paragraphs of what this means, why it is important, and how reproducibility can be implemented. Cite your sources using any format.

One key aspect of implementing a machine learning program is making sure the research is reproducible. Reproducible research can be defined as research results that can be replicated by a different researcher from the original author while using the same methods (Shenouda & Bajwa, 2022). Reproducibility in machine learning is also important for a few key reasons. Firstly, it guarantees that the results are correct when multiple different researchers can come to the same conclusions. Additionally, reproducibility ensures credibility of the research and allows other researchers to be confident in the results (Ding et al., 2020).

Reproducibility in machine learning can be implemented in multiple ways. The goal of methods reproducibility is to implement the machine learning procedures with the same data and tools to achieve the exact same results as the original (Ding et al., 2020). However, this can be hard to achieve as it is difficult to reproduce every step of the original experiment in the exact same manner. Another method for reproducible research is with results reproducibility, where a new experiment with the same experimental procedures arrives at the same results (Ding et al., 2020). Both of these approaches implement reproducible research with their own strategies and challenges.

Citations

Ding, Z., Reddy, A., & Joshi, A. (2020, August 24). *5 - Reproducibility*. CMU ML Blog. Retrieved October 3, 2022, from https://blog.ml.cmu.edu/2020/08/31/5-reproducibility/

Goyal, C. (2021, July 19). *Deep understanding of discriminative and generative models*. Analytics Vidhya. Retrieved October 3, 2022, from https://www.analyticsvidhya.com/blog/2021/07/deep-understanding-of-discriminative-and-generative-models-in-machine-learning/#:~:text=Discriminative%20models%20draw%20boundaries%20in,the%20labels%20of%20the%20data

Shenouda, J., Bajwa, W. (2022, February 15). *A Guide to Computational Reproducibility in Signal Processing and Machine Learning*. Cornell University. Retrieved October 3, 2022, from https://arxiv.org/abs/2108.12383