Huy Nguyen
HNN190000
CS 4375.004
Portfolio Component 2: ML Algorithms from Scratch

a. copy/paste runs of your code showing the output (coefficients and metrics) and run times.

Logistic Regression

```
"C:\Users\Huy Nguyen\CLionProjects\LogisticRegression\main.exe"
Opening file titanic_project.csv.
heading: "","pclass","survived","sex","age"
Closing file titanic_project.csv.

duration: 11647 milliseconds

Coefficients and Accuracy
Weight 1: -8.49158
Weight 2: 16.9413

Accuracy: 53

Sensitivity: 0.304348
Specificity: 0.137405

pred   0   1
   0  18  80
   1 113  35


Program terminated.
Process finished with exit code 0
```

Naïve Bayes

```
"C:\Users\Huy Nguyen\CLionProjects\LogisticRegression\cmake-build-debug\ML.exe"
Opening file titanic_project.csv.
heading: "","pclass","survived","sex","age"
Closing file titanic_project.csv.


Duration: 4 milliseconds
Original Probabilities:
0: 0.61 1: 0.39

Pclass Probabilities:
0.172131 0.22541 0.602459
0.416667 0.262821 0.320513

Sex Probabilities:
0.159836        0.840164
0.679487        0.320513

Age:
Mean: 30.4182, Variance: 205.153
Mean: 28.8261, Variance: 209.155

pred   0   1
   0 113  35
   1  18  18

Accuracy: 0.784553
Sensitivity: 0.695652
Specificity: 0.862595


Program terminated.
Process finished with exit code 0
```

Huy Nguyen
HNN190000
CS 4375.004
Portfolio Component 2: ML Algorithms from Scratch

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

Logistic Regression
Our accuracy is only 53 so that means it got about half of our predictions incorrect.
The pred table shows we mostly had 113 false negatives and 80 false positives which shows that this did not do well predicting. The low rates for sensitivity and specificity shows that it is not that accurate when predicting.

Naïve Bayes
The original probabilities shows the chances for 0 and 1 for survival.
Pclass probabilities shows the chances of survival with respect to the 3 classes.
Sex probabilities shows the chances of survival with respect to the sex of the person.
There are statistics for ages with the first row being a 0 for survival and the second being 1 for survival. The pred table here that we have 113 true negatives and 18 true positives. This time the accuracy was much higher being 0.78. The sensitivity and specificity are much larger showing that we had a better rate of getting true negatives and true positives.

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

Generative classifiers models the distribution between classes. Naïve Bayes is an example of generative classifiers. Discriminative classifiers learn the boundaries between the classes. Logistic regression is an example of discriminative classifiers.

When trying to create models they will most of the time provide similar outputs. The main difference is how they get the output. Generative would decide classes based on similarities while discriminative would decide based on how they are different. Generative tends to have higher bias and lower variance while discriminative does the opposite having lower bias and higher variance.

https://towardsdatascience.com/generative-vs-discriminative-classifiers-in-machine-learning-9ee265be859e

Huy Nguyen
HNN190000
CS 4375.004
Portfolio Component 2: ML Algorithms from Scratch

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.

Reproducible means to recreate your algorithm on datasets and getting same or similar results. Being able to reproduce the same results reduces the number of errors and makes it easier to debug. Having consistency with the outputs of your algorithm allows you to show that your algorithm is performing correctly. Reproducibility can be implemented by providing proper documentation on your algorithm. Documentation allows others to be able to understand what your thought process was when creating the algorithm and how to use it.

J. (2018, May 25). *Reproducibility in ML: why it matters and how to achieve it*. Determined AI. https://www.determined.ai/blog/reproducibility-in-ml

Reproducible research: a minority opinion. (n.d.). Taylor & Francis. https://www.tandfonline.com/doi/full/10.1080/0952813X.2017.1413140

The Importance of Reproducibility in Machine Learning applications. (2022, December 7). DecisivEdge. https://www.decisivedge.com/blog/the-importance-of-reproducibility-in-machine-learning-applications/