Assignment 3

Snip of code run:

Program 1:

```
The weights for the trained Logistical Model
[0.999877,-2.41086],

Time taken for training logistic Model: 856
The Logistical Model Stats:
Accuracy: 0.784553
Sensitivity: 0.695652
Specificity: 0.862595
```

Program 2:

```
The Likelihood values for p(pclass | survived)
[[0.172131,0.22541,0.602459,],
[0.416667,0.262821,0.320513,],
]

The Likelihood values for p(sex | survived)
[[0.159836,0.840164,],
[0.679487,0.320513,],
]

The Naive Bayes Model Stats:
Accuracy: 0.784553
Sensitivity: 0.695652
Specificity: 0.862595

Time taken for whole program: 856
```

My Results: In the end both of logistic and naïve bayes regression have the same results. This could be because there are certain data that are outliers from the data. This data could be potentially confusing for both models and causing the issues. This is shown in the low sensitivity of both models. One reason for this could have been that even if a specific class had a much higher chance of surviving but there was a lack of boats to take them we could be getting a lot of false positives. The number of false negatives or our specificity is much lower. Meaning that most people we expected to perish perished.

Generative Versus Discriminative:

The main difference between discriminative and generative models is that discriminative defines a decision boundary upon which if one is above the line it is one result and if it is below then it is the other. Generative on the other hand estimates probability distributions over which the probability is estimated from the training data and then we attempt to guess the result from those probabilities. Logistical regression falls under the discriminative line because it forms a line upon which results are chosen from. Naïve Bayes is Generative because we guess the probability assuming one variable implies another before deducing the final result.

This difference makes these models suited for different tasks. Generative is generally used for models where we want to learn everything possible from the predictors. This requires more data though and can be can be prone to not learning from outliers in the data. Discriminative is good for tasks where we want to learn the differences between any two predictors. For example if we have a model where a line can easily separate two types of data then a discriminative would be good for drawing that line between the points. The downsides though are that we need data to be separable from one another.

Reproducible Research in AI

The term Reproducible Research in AI describes the importance of machine learning engineers to formulate models and save data in a form that makes it so that anyone can follow the same steps and reach the same conclusions. This can mean many things but the simplest are good data structure, sharing the parameters that a model was trained on, and copious amounts of justification wherever data was used or changed. These steps make it so that when the community looks at a model or result we can understand how the result of a persons train of thought affected the final product.

This is important in industry because many companies use data and models that will affect real people, so having models that are based on bad data or poorly written models with biases can affect real people in negative ways. This is furthered by research groups that publish their results to conventions and seminars. If these results are bad, then they will eventually be discovered by other groups who attempt to further their research on other topics. All round it benefits groups to make research that is effective and reproducible to ensure that we each stay accountable for our

Works Cited

The importance of reproducibility in machine learning applications. DecisivEdge. (2022, December 7). Retrieved March 4, 2023, from https://www.decisivedge.com/blog/the-importance-of-reproducibility-in-machine-learning-applications/#:~:text=Reproducibility%20with%20respect%20to%20machine,reporting%2C%20data%20analysis%20and%20interpretation.

LeVeque, R. (2022). *Reproducible research for scientific computing: Tools and strategies ...* Reproducable Research in Computer Science. Retrieved March 4, 2023, from https://staff.washington.edu/rjl/pubs/cise12/CiSE12.pdf