Logistic Regression

```
PS C:\Users\gerun\Desktop\stuff\School\UTD FALL 2022\CS4
Opening file titanic_project.csv
Reaing line 1
heading: "","pclass","survived","sex","age"
Closing file titanic_project.csv.
Beginning training on data with descent of 500000...
Training completed in 18 seconds.
Weight = -1.41099

Accuracy: 0.784553
Sensitivity: 0.325203
Specificity: 0.45935
```

Naïve-Bayes

```
PS C:\Users\gerun\Desktop\stuff\School\UTD FALL 2022\CS4375\pfc3>
Opening file titanic_project.csv
Reaing line 1
heading: "","pclass","survived","sex","age"
Closing file titanic_project.csv.
Completed in 0 seconds.
Accuracy: 0.536585
Sensitivity: 0.00406504
Specificity: 0.53252
PS C:\Users\gerun\Desktop\stuff\School\UTD FALL 2022\CS4375\pfc3>
```

Analysis

In my implementation of the two different algorithms, Logistic Regression was slower but more

accurate while Naïve-Bayes was much faster but much less accurate. Admittedly, the

implementation of Naïve-Bayes is faulty, as the intermediate values didn't quite match the values

that R would produce. Surprisingly, Naïve-Bayes was more accurate in one aspect – identifying

Specificity (that is, how accurately it identifies True Negatives – non-survivors). Additionally,

the Logistic Regression only utilized one predictor while Naïve-Bayes used all the predictors.

Generative vs. Discriminative Classifiers

Generative and discriminative classifiers both attempt to find probability distributions between two observations in hopes of being able to use one or more to predict an observation column. They both are generally supervised learning problems that are used in a multitude of machine learning algorithms.

However, they do differ on multiple accounts; for starters, generative models learn about joint probability distribution whereas discriminative models learn about conditional probability distribution. Moreover, generative models "explicitly model the actual distribution of each class" whereas discriminative models the decision boundary between classes. The logistic regression algorithm implemented in this activity is a discriminative classifier while Naïve-Bayes is a generative classifier.

-https://medium.com/@mlengineer/generative-and-discriminative-models-af5637a66a3

reproducible research in machine learning

Reproducibility is the ability for other researchers to help confirm research findings, essentially being able to recreate the behavior of a certain algorithm or program multiple times. This term is similar to things such as reliability in that it measures the ability to observe the same observations across multiple trials.

Reproducibility is important in that it ensures that results are accurate and correct as well as encourages transparency in the process – there's no reason not to share findings and the intellectual process, as it furthers expands the scope and ceiling that the field can reach. Additionally, it reduces the risk of errors, as being able to reproduce results consistently generally does not happen if the algorithm or model is error-ridden; rather, errors generally contribute to inconsistency. Reproducibility also encourages the use of good practice and treating

results with caution, as faulty conclusions and inconsistencies can lead to incorrect assumptions and ideas that are overall bad for progress in the field of machine learning.

Reproducibility can be implemented in a multitude of ways; one way that has been suggested by ML researchers at CMU is the implementation and normalization of a data analysis flowchart, ensuring that researchers go through a variety of steps to verify and substantiate their findings. This flow of thought helps not only with improving reproducibility but also helps with ensuring that research and reports are well thought out and overall accurate and solid. Additionally, acknowledgement of bias and possibilities for false positives and false negatives helps to ensure that reports and research are not narrow, rather it encourages knowledge and recognition of edge cases and outliers.