CS 4375.003
Matthew Naughton

```
B0=-0.361827 B1=-0.0968872 Error= 0
B0=-0.363281 B1=-0.0954333 Error= 0
B0=-0.362288 B1=-0.0954333 Error= 0
B0=-0.361368 B1=-0.0963531 Error= 0
B0=-0.362795 B1=-0.0963531 Error= 0
B0=-0.364249 B1=-0.094899 Error= 0
B0=-0.365676 B1=-0.094899 Error= 0
B0=-0.367131 B1=-0.0934446 Error= 0
B0=-0.366141 B1=-0.0934446 Error= 0

Ending Values are: B0=-0.366141 B1=-0.0934446 Error=0

Enter 0(female) or 1(male) for the prediction:
```

```
Enter 0(female) or 1(male) for the prediction: 1
The value our model predicts is: -0.459586
Our model predicted: 0
Accuracy is 45.7676%
The sensitivity is: 0.723
The specificity is: 0.638
The Running Time was: 0.682467 seconds.

Program terminated.
Process returned 0 (0x0)   execution time : 10.025 s
Press any key to continue.
```

The results of our data demonstrate that logistic regression was somewhat able to predict the data, and that naïve bayes was mostly able to tell where the data came from based on the titanic data.

A generative model is able to predict the conditional probability using the Bayes Theorem. It does this by comparing data around it and trying to match it as best as it can. This is effective when there is large chunks of descriptive data available to the model.

A discriminative model will model the decision boundaries between classes. While both models are predicting conditional probability, discriminative is doing it based off of the properties it learned, rather than comparing it with other data. A discriminative model is more representative of machine learning because the model actually learns from the data and makes a prediction based on the properties it learned, while a generative model simply compares the data with other data and matches it.

Reproducible research in machine learning is the idea that you can repeatedly run your model and get the same or very similar results each time. This is very important, because research with only one trial can produce different results than 50 other trials with the same variables. Some of the articles I read discussed how it can actually be very difficult to get reproducible results even with code, and that because the source code used in several trials is not open-sourced, it is difficult for people outside the study to verify the data and the claims. One solution to this is to change the techniques we are using, one example is to use a baseline that has the same accuracy as our original research. Another solution is to prove correctness by verifying that other people are also able to produce the same results.

https://blog.ml.cmu.edu/2020/08/31/5-reproducibility/

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.

CS 4375.003
Matthew Naughton

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