Ryan Donaldson
ML Algorithms from Scratch Write-up

Naive Bayes Output

```
Opening file titanic_project.csv...
Populating vectors...
Closing file titanic_project.csv...
Running Naive Bayes and calculating coefficients, please wait...

Training Time: 1874 Microseconds

Class = Negative:
  Coefficient age: Mean = 7415.5, Stdev = 163143
  Coefficient pclass: Mean = 593, Stdev = 13046.1
  Coefficient sex: Mean = 205, Stdev = 4510.05
Class = Positive:
  Coefficient age: Mean = 4494, Stdev = 78871.4
  Coefficient pclass: Mean = 297, Stdev = 5212.46
  Coefficient sex: Mean = 50, Stdev = 877.553

Accuracy: 0.46748
Sensitivity: 1
Specificity: 0
Program ended with exit code: 0
```

Logistic Regression Output

```
Opening file titanic_project.csv...
Populating vectors...
Closing file titanic_project.csv...
Running logistic regression and calculating coefficients, please wait...

Training Time: 18.108 Seconds

Estimate        Std. Error      z value     Pr(>|z|)
0.999877        0.414447        2.41256     1.01584     (Intercept)
-2.41086        0.317002        -7.6052     1       sex

Accuracy: 0.46748
Sensitivity: 1
Specificity: 0
Program ended with exit code: 0
```

The output for our Titanic dataset from the Naive Bayes model shows the mean and standard deviation for age, pclass, and sex (predictor variables) for the two Negative and Positive classes. The model's accuracy shows that it correctly predicted the correct class, Negative or Positive, for 47% of the observations in the test set. Because the specificity is 0, the model incorrectly classified all Negative observations, which indicates a high bias for the Positive class. Comparing the mean and standard deviation for age in the Positive and Negative classes, we can conclude that young people were more likely to survive. Likewise, let us compare the mean and standard deviation for the pclass in the Positive and Negative classes. People in the lower class fares were more likely not to survive. Finally, let us compare the mean and standard deviation of sex in the Positive and Negative classes. Women were more likely to survive than men. Similarly, the logistic regression model shows the same accuracy and specificity, and we can draw the same conclusions about the passengers.

The Naive Bayes model uses the generative classifier to calculate the joint distribution of a feature x on the target variable y. This allows the generative classifier to predict the conditional probability on new data by transforming the joint probability into a conditional probability. In contrast, the discriminative classifier used by the logistic regression directly learns the conditional probability and tries to find boundaries that separate classes. However, generative classifiers are more computationally expensive than discriminative classifiers. They are also sensitive to outliers as they can affect the distribution significantly [1].

Works Cited:

[1] Yıldırım, Soner. “Generative vs Discriminative Classifiers in Machine Learning.” *Medium*, Towards Data Science, 14 Nov. 2020, https://towardsdatascience.com/generative-vs-discriminative-classifiers-in-machine-learning-9ee265be859e.