Kevin Cordano
C++ Algorithms from Scratch
CS 4375.004

Logistic Regression Output

```
Weights: 0.999877, -2.41086
Accuracy: 0.784553
Sensitivity: 0.695652
Specificity: 0.862595

Algorithm Run Time: 49.3448s
```

Naïve Bayes Output

```
Accuracy: 0.784553
Sensitivity: 0.695652
Specificity: 0.862595

Algorithm Run Time: 7.3e-05s
```

How I got the exact same accuracy, sensitivity, and specificity, I have no idea, but the performance of the Logistic Regression would likely begin to have a better performance around this point of dataset size, since LR generally performs better on larger data sets. However, Logistic Regression is significantly slower due to having to find optimal weights, which is a costly operation. In contrast, Naïve Bayes is able to acquire all the information needed with one read through of the data.

Generative classifiers base their estimations on parameters for the likelihood (P(X|Y)) and the prior (P(Y)), while discriminative classifiers base their estimations on parameters for the posterior (P(Y|X)) [1]. Generative classifiers base their learning on the environment, trying to understand what elements are used to generate data within the different classes; they focus on the distribution of a class [2]. Discriminative classifiers base their learning on the distinguishing characteristics between classes, focusing on the defining traits that separate elements of one class from another [2].

Both generative classifiers and discriminative classifiers utilize conditional probability to make decisions, meaning they operate on related sets of parameters [2]. This can be seen in Bayes theorem, where the posterior and prior are two sides of the same equation. As such, Naïve Bayes and Logistic Regression converge on similar classifiers as the training set approaches infinity [1].

Reproducibility is the ability to replicate experiment results across multiple trials and related test sets. In a field where false claims can be supported by biased information and weak test data, it is important to have a form of reproducibility to provide credit and verification to a scientific conclusion.

Naturally, researchers can often be hesitant to be completely open on their research as a means of protecting their intellectual property. As such, parameter values, function sequences, and other computational details that are typically omitted are crucial in the ability to reproduce computational results [3]. With the necessity of reproducibility, standardization of published research has been established to provide pertinent information for third-party researchers in order to replicate the results as accurately as possible.

Although not possible in all cases, presentation of the raw data used in research should be provided to allow for analysis and reproduction. When the data is publicly available, researchers are provided with learning opportunities to explore and extend computational concepts and techniques [4]. Along those lines, virtual review of the programming environment can be used to provide narration on the research process [4]. By doing so, it will allow for a simpler approach to reproduction since the physical environment will be made available; this is a huge assist in any field regarding some form of computer science, including machine learning.

Kevin Cordano
C++ Algorithms from Scratch
CS 4375.004

Sources

[1] K. Mazidi. Machine Learning Handbook. https://learn-us-east-1-prod-fleet02-xythos.content.blackboardcdn.com/6093de9f52e1b/38795318?X-Blackboard-S3-Bucket=learn-us-east-1-prod-fleet01-xythos&X-Blackboard-Expiration=1677974400000&X-Blackboard-Signature=KIpRhHaNf9G7OjyMZLMQIeQ1ZCUfKyniihNgX4neieE%3D&X-Blackboard-Client-Id=100946&X-Blackboard-S3-Region=us-east-1&response-cache-control=private%2C%20max-age%3D21600&response-content-disposition=inline%3B%20filename%2A%3DUTF-8%27%27ML%2520kjg.pdf&response-content-type=application%2Fpdf&X-Amz-Security-Token=IQoJb3JpZ2luX2VjEDQaCXVzLWVhc3QtMSJHMEUCIEtH2uQ3UqhQ3evw71yTyMdGfyz7jcWKCVyPiFXxRJ9FAiEA5PPvcUfp0xnbqc3s7GzltsIXsHU5Z1gE%2BidEgAIS7rAqvgUI3f%2F%2F%2F%2F%2F%2F%2F%2F%2F%2FARACGgw2MzU1Njc5MjQxODMiDOh1X%2Bbh8lVyXJ4%2B1SqSBXJfOxT8cXO7GCNnzXpUrfMlncDVvwelP9%2FF%2B21ZXvVA0L7ZFvy6vwo77A%2B5jvvio%2B8AsKBpTEpWPQCLuNNRwQeJG6mDTlsrouLWS12vDOQjNtj3PftfDaqu3PPqCgYHU%2Bu96KV%2BWQyfSRLtpXp0UQ9ZbS18RFyztpNcEF04OsVQkFJyjhzFtwQaS346BWjXDgF%2BXC%2FwIAVUOeiWVVwpsOC4IphefKw3LC770de1wzk0ecuxq6LkWRohU9dMsyySpdTSdylW41qQyXFQwRoYDYRdiJicCC2llgDwW4NbJ%2FvM6GSgJwod9vTyBo8QVCZ5qWhUv8dDaGUh3Ls3vZV7JhQt4nOlklAzSj%2FTZh48ADmn5dJuW7vZDbe93voy%2By7l%2BF%2BKxT4ZvS0eQmWWrSjfnjzrBrQC4zTZ2B9ijLCVL58RYvjb%2BtpiDfjPn6pEfVrjJQ9bVhwT5%2FjORrUVpA5V%2FvtPwcUYHhfzI5q2rrI%2BFUd6l3GDetdGYjtrKrGDhEgn7D7gYRSKPAFiQ7TuimieKE9aiEfTM4p2%2FoaDr1rg3k6h4bW%2BQn163wEpGcFtmdUfLuo7aToQYWmdAmkSZwhCUAERiZLEH4BKlBhpoK9sMXipVgYegBgz1Ah9sYus%2Br%2FMoFizq7ocTLQG7cRVjOB54YtO9Kay0MdPdJeFJLfSifF%2FVDKoAISQWw9NymHkFt3MSQq7blRp39Tb1CicRbvzRtCUUc36vHjCeD2s1L8XYs5Ab%2FUP6FfouCHl4nPSu2efkMWrMxJW9BVOaF7NLVJ51n6i3boETh%2BFFKJMSYGO4DZ7dO%2BVerp6M3uCPd9pbY%2FpVSM3Sn7lhuxynX9BdtXw2t13LQA2tHQV18ueW%2FJmYen0kUQw07mOoAY6sQE166fvox6gOOl%2BTIXCTeqU55aVLDOZrGlKTsVb1VgqB9Cw%2BKM%2FPplGWb5Hh4k9oZV76%2F2NByk%2BgonYQSn0iHq8JGbj2QLnaKPRPUR6VxsSLOlh7YQv4jw5pF7n3aPmWH%2FZmF8AmGISTDAe3ji8CqmkMETHCx39d9yXzWsoEMZE1Z54j44XR%2F9MeGOYFIQ%2BuLaFMAZgt5GGpAXv9VjE%2BB2QfeVQreNLjAJROW3xvDC7p3o%3D&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Date=20230304T180000Z&X-Amz-SignedHeaders=host&X-Amz-Expires=21600&X-Amz-Credential=ASIAZH6WM4PLSR4WBBDD%2F20230304%2Fus-east-1%2Fs3%2Faws4_request&X-Amz-Signature=51633e39e346c537f3ec04f89eaf6fa9cfe0ad81e5397d00a70ce7e0f9863638

[2] A. Malhotra. Generative Classifiers V/S Discriminative Classifiers. https://medium.com/@akankshamalhotra24/generative-classifiers-v-s-discriminative-classifiers-1045f499d8cc#:~:text=Discriminative%20Classifiers%20learn%20what%20the,the%20image%20is%20of%20dog.

[3] R. LeVeque, I. Mitchell, V. Stodden. Reproducible Research for Scientific Computing: Tools and Strategies for Changing the Culture. https://staff.washington.edu/rjl/pubs/cise12/CiSE12.pdf

[4] R. Carter, Z. Attia, F. Lopez-Jimenez, P. Friedman. Pragmatic Considerations for Fostering Reproducible Research in Artificial Intelligence. https://www.nature.com/articles/s41746-019-0120-2