David Favela Corella

Dxf200002

CS 4375.003

ML Algorithms from Scratch

a.

```
[Running] cd "c:\Users\david\OneDrive\Documents\UT Dallas\UT DALLAS\Fall 2022\CS 4375\CS4375_Portfolio\CS4375_Portfolio\Component-3\" && g++ dataexploration.cpp -o dataexploration && "c:\Users\david\OneDrive\Documents\UT Dallas\UT DALLAS\Fall 2022\CS 4375\CS4375_Portfolio\CS4375_Portfolio\Component-3\"dataexploration
Opening titanic.csv
Predicted From Test Matrix:
0 | -1.41099 | -2.41086 | 0.999877 | 0.999877 | 0.999877 | -2.41086 | -2.41086 | -2.41086 | -2.41086 | -1.41099 | -2.41086 | 0.999877 | 0.999877 | 0.999877 | 0.999877 | 0.999877 | 0.999877 | 0.999877 | -2.41086 | -2.41086 | 0 | 0.999877 | -2.41086 | -2.41086 | 0.999877 | -2.41086 | -2.41086 | -2.41086 | 0 | -2.41086 | -2.41086 | -2.41086 | -2.41086 | -1.41099 | -1.41099 | -2.41086 | 0.999877 | 0.999877 | -2.41086 | -2.41086 | -1.41099 | 0.999877 | 0.999877 | 0.999877 | 0.999877 | -2.41086 | 0.999877 | 0.999877 | -2.41086 | -2.41086 | -2.41086 | 0.999877 | -2.41086 | 0.999877 | -1.41099 | -2.41086 | 0.999877 | -1.41099 | -2.41086 | 0 | 0.999877 | -2.41086 | -2.41086 | -2.41086 | 0 | -2.41086 | 0.999877 | -2.41086 | 0 | -2.41086 | 0.999877 | -2.41086 | 0.999877 | 0.999877 | -1.41099 | -1.41099 | 0.999877 | 0.999877 | 0.999877 | 0.999877 | 0 | 0.999877 | -2.41086 | -2.41086 | -1.41099 | 0 | 0.999877 | -1.41099 | -2.41086 | -2.41086 | 0.999877 | 0.999877 | -2.41086 | 0.999877 | -2.41086 | -2.41086 | -2.41086 | 0.999877 | -2.41086 | 0.999877 | -2.41086 | -2.41086 | -2.41086 | -1.41099 | -2.41086 | 0.999877 | -2.41086 | 0.999877 | 0.999877 | -2.41086 | -2.41086 | 0.999877 | 0 | -2.41086 | -2.41086 | 0.999877 | 0.999877 | -2.41086 | -2.41086 | -2.41086 | -2.41086 | 0.999877 | 0.999877 | -2.41086 | 0 | 0.999877 | 0 | 0.999877 | 0.999877 | 0.999877 | -2.41086 | 0.999877 | -2.41086 | -2.41086 | -1.41099 | -1.41099 | 0.999877 | -1.41099 | -2.41086 | -2.41086 | 0 | -2.41086 | 0 | -2.41086 | 0.999877 | -2.41086 | -2.41086 | -2.41086 | 0.999877 | -2.41086 | -2.41086 | -2.41086 | 0.999877 | -2.41086 | 0.999877 | -1.41099 | 0.999877 | -1.41099 | -1.41099 | -1.41099 | 0.999877 | 0.999877 | -2.41086 | -2.41086 | 0.999877 | -2.41086 | -2.41086 | -1.41099 | 0 | -2.41086 | 0.999877 | 0.999877 | -2.41086 | -1.41099 | -2.41086 | -1.41099 | -2.41086 | -1.41099 | -2.41086 | -2.41086 | -2.41086 | 0.999877 | 0.999877 | 0.999877 | 0.999877 | -2.41086 | -1.41099 | 0.999877 | -2.41086 | -2.41086 | -2.41086 | 0.999877 | 0.999877 | -2.41086 | -1.41099 | -1.41099 | 0.999877 | -1.41099 | 0.999877 | -2.41086 | -1.41099 | -2.41086 | -2.41086 | -1.41099 | -2.41086 | -1.41099 | -2.41086 | -1.41099 | -2.41086 | -2.41086 | -1.41099 | -2.41086 | 0.999877 | 0.999877 | -1.41099 | -1.41099 | 0 | -2.41086 | -2.41086 | 0.999877 | -2.41086 | -1.41099 | 0.999877 | 0.999877 | -2.41086 | -2.41086 | 0.999877 | -2.41086 | 0 |
```

```
0.999877 | 0 | -1.41099 | -2.41086 | -2.41086 | -2.41086 | -2.41086 | -
2.41086 | 0.999877 | -2.41086 | -2.41086 | -2.41086 | 0 | -2.41086 | -2.41086
| -2.41086 |

 Probabilities for Test Matrix:
0.5 | 0.196078 | 0.082348 | 0.731034 | 0.731034 | 0.731034 | 0.082348 |
0.082348 | 0.082348 | 0.082348 | 0.196078 | 0.082348 | 0.731034 | 0.731034 |
0.731034 | 0.731034 | 0.731034 | 0.731034 | 0.731034 | 0.082348 | 0.082348 |
0.5 | 0.731034 | 0.082348 | 0.082348 | 0.731034 | 0.082348 | 0.082348 |
0.082348 | 0.5 | 0.082348 | 0.082348 | 0.082348 | 0.082348 | 0.196078 |
0.196078 | 0.082348 | 0.731034 | 0.731034 | 0.082348 | 0.082348 | 0.196078 |
0.731034 | 0.731034 | 0.731034 | 0.731034 | 0.082348 | 0.731034 | 0.731034 |
0.082348 | 0.082348 | 0.082348 | 0.731034 | 0.082348 | 0.731034 | 0.196078 |
0.082348 | 0.731034 | 0.196078 | 0.082348 | 0.5 | 0.731034 | 0.082348 |
0.082348 | 0.082348 | 0.5 | 0.082348 | 0.731034 | 0.082348 | 0.5 | 0.082348 |
0.731034 | 0.082348 | 0.731034 | 0.731034 | 0.196078 | 0.196078 | 0.731034 |
0.731034 | 0.731034 | 0.731034 | 0.5 | 0.731034 | 0.082348 | 0.082348 |
0.196078 | 0.5 | 0.731034 | 0.196078 | 0.082348 | 0.082348 | 0.731034 |
0.731034 | 0.082348 | 0.731034 | 0.082348 | 0.082348 | 0.082348 | 0.731034 |
0.082348 | 0.731034 | 0.082348 | 0.082348 | 0.082348 | 0.196078 | 0.082348 |
0.731034 | 0.082348 | 0.731034 | 0.731034 | 0.082348 | 0.082348 | 0.731034 |
0.5 | 0.082348 | 0.082348 | 0.731034 | 0.731034 | 0.082348 | 0.082348 |
0.082348 | 0.082348 | 0.731034 | 0.731034 | 0.082348 | 0.5 | 0.731034 | 0.5 |
0.731034 | 0.731034 | 0.731034 | 0.082348 | 0.731034 | 0.082348 | 0.082348 |
0.196078 | 0.196078 | 0.731034 | 0.196078 | 0.082348 | 0.082348 | 0.5 |
0.082348 | 0.5 | 0.082348 | 0.731034 | 0.082348 | 0.082348 | 0.082348 |
0.731034 | 0.082348 | 0.082348 | 0.082348 | 0.731034 | 0.082348 | 0.731034 |
0.196078 | 0.731034 | 0.196078 | 0.196078 | 0.196078 | 0.731034 | 0.731034 |
0.082348 | 0.082348 | 0.731034 | 0.082348 | 0.082348 | 0.196078 | 0.5 |
0.082348 | 0.731034 | 0.731034 | 0.082348 | 0.196078 | 0.082348 | 0.196078 |
0.082348 | 0.196078 | 0.082348 | 0.082348 | 0.082348 | 0.731034 | 0.731034 |
0.731034 | 0.731034 | 0.082348 | 0.196078 | 0.731034 | 0.082348 | 0.082348 |
0.082348 | 0.731034 | 0.731034 | 0.082348 | 0.196078 | 0.196078 | 0.731034 |
0.196078 | 0.731034 | 0.082348 | 0.196078 | 0.082348 | 0.082348 | 0.196078 |
0.082348 | 0.196078 | 0.082348 | 0.196078 | 0.082348 | 0.082348 | 0.196078 |
0.082348 | 0.731034 | 0.731034 | 0.196078 | 0.196078 | 0.5 | 0.082348 |
0.082348 | 0.731034 | 0.082348 | 0.196078 | 0.731034 | 0.731034 | 0.082348 |
0.082348 | 0.731034 | 0.082348 | 0.5 | 0.731034 | 0.5 | 0.196078 | 0.082348 |
0.082348 | 0.082348 | 0.082348 | 0.082348 | 0.731034 | 0.082348 | 0.082348 |
0.082348 | 0.5 | 0.082348 | 0.082348 | 0.082348 |

 Prediction for Test Matrix:
0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0
| 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0
```

```
| 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 |
1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0
| 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 |
0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0
| 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 |
0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1
| 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 |
0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Elapsed time of the training algorithm: 53.3994s

Accuracy: 0.857724

Sensitivity: 0.695652

Specificity: 1

Probabilities in Naive Bayes:
0.35158 0.64842 | 0.725577 0.274423 | 0.878594 0.121406 | 0.219298 0.780702 |
0.100185 0.899815 | 0.0870726 0.912927 | 0.857998 0.142002 | 0.881538
0.118462 | 0.86616 0.13384 | 0.732418 0.267582 | 0.532521 0.467479 | 0.85712
0.14288 | 0.11319 0.88681 | 0.387272 0.612728 | 0.197192 0.802808 | 0.389397
0.610603 | 0.205582 0.794418 | 0.199492 0.800508 | 0.197192 0.802808 |
0.555296 0.444704 | 0.748541 0.251459 | 0.346738 0.653262 | 0.421811 0.578189
| 0.540428 0.459572 | 0.867093 0.132907 | 0.0985147 0.901485 | 0.870867
0.129133 | 0.845554 0.154446 | 0.871821 0.128179 | 0.373093 0.626907 |
0.75473 0.24527 | 0.859773 0.140227 | 0.745508 0.254492 | 0.747019 0.252981 |
0.537707 0.462293 | 0.869917 0.130083 | 0.869444 0.130556 | 0.23021 0.76979 |
0.21783 0.78217 | 0.869917 0.130083 | 0.847908 0.152092 | 0.863391 0.136609 |
0.117394 0.882606 | 0.103822 0.896178 | 0.412042 0.587958 | 0.116524 0.883476
| 0.887479 0.112521 | 0.381046 0.618954 | 0.09112 0.90888 | 0.757883 0.242117
| 0.738118 0.261882 | 0.87374 0.12626 | 0.21783 0.78217 | 0.875672 0.124328 |
0.123907 0.876093 | 0.568675 0.431325 | 0.874704 0.125296 | 0.360122 0.639878
| 0.867093 0.132907 | 0.530057 0.469943 | 0.375044 0.624956 | 0.10788 0.89212
| 0.535071 0.464929 | 0.871821 0.128179 | 0.521064 0.478936 | 0.360122
0.639878 | 0.859773 0.140227 | 0.225396 0.774604 | 0.856248 0.143752 |
0.37702 0.62298 | 0.872779 0.127221 | 0.375044 0.624956 | 0.558519 0.441481 |
0.414446 0.585554 | 0.371167 0.628833 | 0.860669 0.139331 | 0.788229 0.211771
| 0.414446 0.585554 | 0.121968 0.878032 | 0.421811 0.578189 | 0.206859
0.793141 | 0.358366 0.641634 | 0.244112 0.755888 | 0.851988 0.148012 |
0.872779 0.127221 | 0.540428 0.459572 | 0.383096 0.616904 | 0.10788 0.89212 |
0.544668 0.455332 | 0.74104 0.25896 | 0.86803 0.13197 | 0.0954531 0.904547 |
0.0931711 0.906829 | 0.574022 0.425978 | 0.184921 0.815079 | 0.86616 0.13384
```

```
| 0.853248 0.146752 | 0.853672 0.146328 | 0.0927428 0.907257 | 0.865232
0.134768 | 0.393724 0.606276 | 0.864309 0.135691 | 0.552156 0.447844 |
0.865232 0.134768 | 0.5476 0.4524 | 0.605416 0.394584 | 0.106478 0.893522 |
0.86616 0.13384 | 0.119178 0.880822 | 0.367388 0.632612 | 0.748541 0.251459 |
0.579544 0.420456 | 0.37702 0.62298 | 0.365535 0.634465 | 0.527679 0.472321 |
0.867093 0.132907 | 0.103188 0.896812 | 0.106478 0.893522 | 0.731024 0.268976
| 0.762679 0.237321 | 0.722933 0.277067 | 0.854524 0.145476 | 0.203087
0.796913 | 0.109331 0.890669 | 0.868972 0.131028 | 0.208157 0.791843 |
0.204324 0.795676 | 0.205582 0.794418 | 0.103822 0.896178 | 0.10788 0.89212 |
0.100185 0.899815 | 0.859773 0.140227 | 0.120093 0.879907 | 0.738118 0.261882
| 0.737395 0.262605 | 0.890467 0.109533 | 0.613975 0.386025 | 0.120093
0.879907 | 0.583322 0.416678 | 0.867561 0.132439 | 0.85712 0.14288 | 0.349943
0.650057 | 0.716569 0.283431 | 0.379021 0.620979 | 0.880063 0.119937 |
0.249901 0.750099 | 0.863391 0.136609 | 0.747019 0.252981 | 0.74104 0.25896 |
0.191767 0.808233 | 0.706045 0.293955 | 0.744008 0.255992 | 0.558519 0.441481
| 0.383096 0.616904 | 0.724248 0.275752 | 0.0949776 0.905022 | 0.742518
0.257482 | 0.367388 0.632612 | 0.549098 0.450902 | 0.549098 0.450902 |
0.568675 0.431325 | 0.212176 0.787824 | 0.422435 0.577565 | 0.735244 0.264756
| 0.862478 0.137522 | 0.114829 0.885171 | 0.865232 0.134768 | 0.862478
0.137522 | 0.785952 0.214048 | 0.356634 0.643366 | 0.871821 0.128179 |
0.0964334 0.903567 | 0.21496 0.78504 | 0.861571 0.138429 | 0.568675 0.431325
| 0.732418 0.267582 | 0.867093 0.132907 | 0.869917 0.130083 | 0.543234
0.456766 | 0.563506 0.436494 | 0.875672 0.124328 | 0.591102 0.408898 |
0.206859 0.793141 | 0.205582 0.794418 | 0.358366 0.641634 | 0.187751 0.812249
| 0.85844 0.14156 | 0.629572 0.370428 | 0.0936086 0.906391 | 0.719072
0.280928 | 0.860669 0.139331 | 0.86022 0.13978 | 0.33346 0.66654 | 0.105796
0.894204 | 0.870867 0.129133 | 0.756302 0.243698 | 0.872779 0.127221 |
0.212176 0.787824 | 0.882524 0.117476 | 0.0985147 0.901485 | 0.865232
0.134768 | 0.605416 0.394584 | 0.552156 0.447844 | 0.74104 0.25896 | 0.865232
0.134768 | 0.75473 0.24527 | 0.867093 0.132907 | 0.880554 0.119446 | 0.739573
0.260427 | 0.862478 0.137522 | 0.871821 0.128179 | 0.735244 0.264756 |
0.881538 0.118462 | 0.184921 0.815079 | 0.0936086 0.906391 | 0.870867
0.129133 | 0.553716 0.446284 | 0.0974537 0.902546 | 0.706045 0.293955 |
0.732418 0.267582 | 0.375044 0.624956 | 0.696802 0.303198 | 0.785952 0.214048
| 0.121023 0.878977 | 0.222302 0.777698 | 0.868972 0.131028 | 0.871821
0.128179 | 0.212176 0.787824 | 0.566932 0.433068 | 0.387272 0.612728 |
0.0969385 0.903062 | 0.361902 0.638098 | 0.784249 0.215751 | 0.863391
0.136609 | 0.560161 0.439839 | 0.748541 0.251459 | 0.735244 0.264756 |
0.530057 0.469943 | 0.110076 0.889924 | 0.742518 0.257482 | 0.849518 0.150482
| 0.726919 0.273081 | 0.383096 0.616904 | 0.869917 0.130083 | 0.744008
0.255992 | 0.874704 0.125296 |
```

b. Analyzing the results, it can be seen that we got a 0.8577 accuracy on the logistic regression model created using C++, which is quite good. Specificity is also 1, which is great, and sensitivity is also really close to 1. The Naïve Bayes also provided similar results when compared to the probabilities given in the logistic model, although they're skewed due to using other classifiers like pclass and age.

c. Generative classifiers try to learn the probability of a distribution. It focuses on how target variables occur, as Soner Yildirim (2020) discusses. It makes predictions out of the new data, once it understands how the data was generated. Naïve Bayes algorithm uses what we call generative classifiers, as we try to understand how different factors affect the data.
   Discriminative classifiers, in contrast, find boundaries that separate the data from each other. It checks all types of boundaries and the one that gives the lowest error is chosen, as Soner Yildirim (2020) states. There are some variations to the discriminative classifiers, like some allow for misclassification. Logistic regression is an example of an algorithm based on discrimination classifiers, and we can see it in how we can select which data to use. Both models are useful in machine learning but depending on the use case is the choice to use one. For example, the discriminative model is good against outliers while generative models need enough data to make the distributions accurately, as Soner Yildirim (2020) states.

d. Reproducibility research:
   a. Reproducibility is a practice that is common amongst researchers to help confirm research findings, get inspiration from other's work, and avoid doing research that already exists, as Zihao Ding states in his article *Reproducibility* on ML-CMU. He states that in theory, reproducibility occurs when a second researcher uses the raw data to build the same analysis and runs the same tests to get the same results.
   b. Aside from verifying that the research data presented is true, it is also important as a step to promote open and accessible research, allowing the scientific community to adopt these findings into practice, as Joelle Pineau et others state in the paper *Improving Reproducibility in Machine Learning Research*. They also state that an important process in reproducibility is for the findings to be repeatable and yield consistent results.
   c. To implement reproducibility on a machine learning model, it will need to be able recreate the workflow stablished in the original model and will have to reach the same conclusions, as Preeti Hemant states in her article for Towards Data Science. She also states that some of the factors that influence reproducibility include the

training data, features, model training, and the software environment it was run in.

Works Cited

Ding, Z. (2020, August 24). *5 - reproducibility*. ML@CMU. Retrieved October 2, 2022, from https://blog.ml.cmu.edu/2020/08/31/5-reproducibility/

Hemant, P. (2020, April 7). *Reproducible machine learning*. Medium. Retrieved October 2, 2022, from https://towardsdatascience.com/reproducible-machine-learning-cf1841606805

Pineau, J. (2021, May). *Improving reproducibility in machine learning research*. JMLR. Retrieved October 2, 2022, from https://jmlr.org/papers/volume22/20-303/20-303.pdf

Yildirim, S. (2020, November 14). *Generative vs Discriminative classifiers in machine learning*. Medium. Retrieved October 2, 2022, from https://towardsdatascience.com/generative-vs-discriminative-classifiers-in-machine-learning-9ee265be859e