Anthony Gray
EN.605.649.81
Intro to Machine Learning
Project 5

Abstract
The Logistic Regression and Naïve Bayes algorithms were implemented, trained, and used to classify examples from five public datasets. Both algorithms performed consistently and comparably on all five datasets. They both performed well on three of the datasets, with average classification accuracies of about 95%, perfectly on one dataset with 100% classification accuracy, and poorly on the final dataset, with only 50% classification accuracy. Improvements in the implementation of Logistic Regression, as well as in the pre-processing of the data would likely lead to an increase in classification accuracy. Tuning the learning rate and data processing to each dataset would also likely result in improved accuracy.

Problem Statement and Algorithms

This project compares the performance of two linear classification algorithms, Logistic Regression (6) and Naïve Bayes (7), on five different data sets. The algorithms operate by using a vector of weights associated with each feature to find a linear discriminant in order to separate each class.
The Naïve Bayes algorithm works by calculating the probability of each class and the conditional probability of each attribute by each class based on the frequency of the classes and attributes in the training data. It then uses these probabilities to calculate the probability of a given instance belonging to each class based on its attributes and assigns the classification with the greatest probability.
Logistic Regression uses a one-vs-all strategy to generate weight vectors associated with each class in the dataset, and gradient descent was used to optimize these weight vectors. Examples are classified by taking the dot product of the feature vector and each weight vector, and passing the result through the sigmoid, or logit function. This function returns the probability that the example belongs to the associated class, and the example is assigned the class associated with the vector yielding the greatest probability.
Both algorithms were expected to perform reasonably well, as both are known to be simple, yet powerful classifiers. However, both algorithms perform best over Boolean features, so one-hot encoding will be used to pre-process the datasets. Additionally, two of the datasets contained continuous variables that needed to be binned before encoding. It is expected that the performance of both algorithms will be lower on these datasets than the categorical or Boolean ones, simply due to the information loss incurred by the binning process. It is also expected that the Naïve Bayes algorithm will perform better than the Logistic Regression algorithm on datasets with a higher number of classes. Logistic Regression is best suited to binary classification problems, and while the one-vs-all strategy mentioned above was employed to mitigate this disadvantage, the Naïve Bayes algorithm does not require an extension to multiple classification problems.

Procedure/Tuning

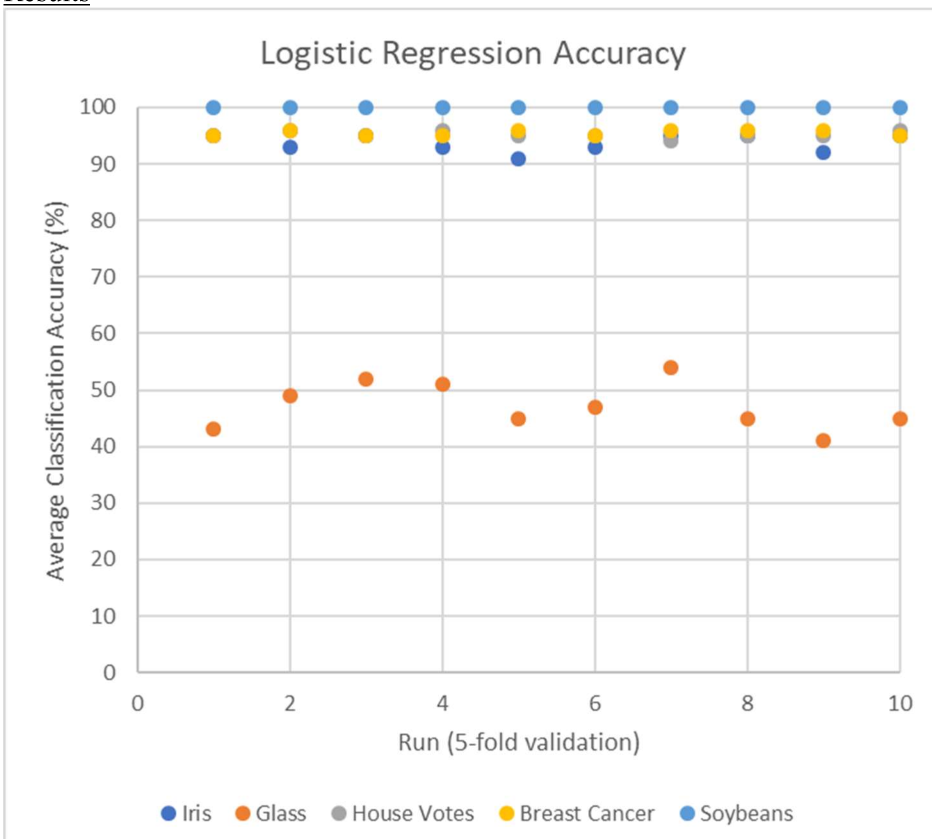
Dataset values before treatment:

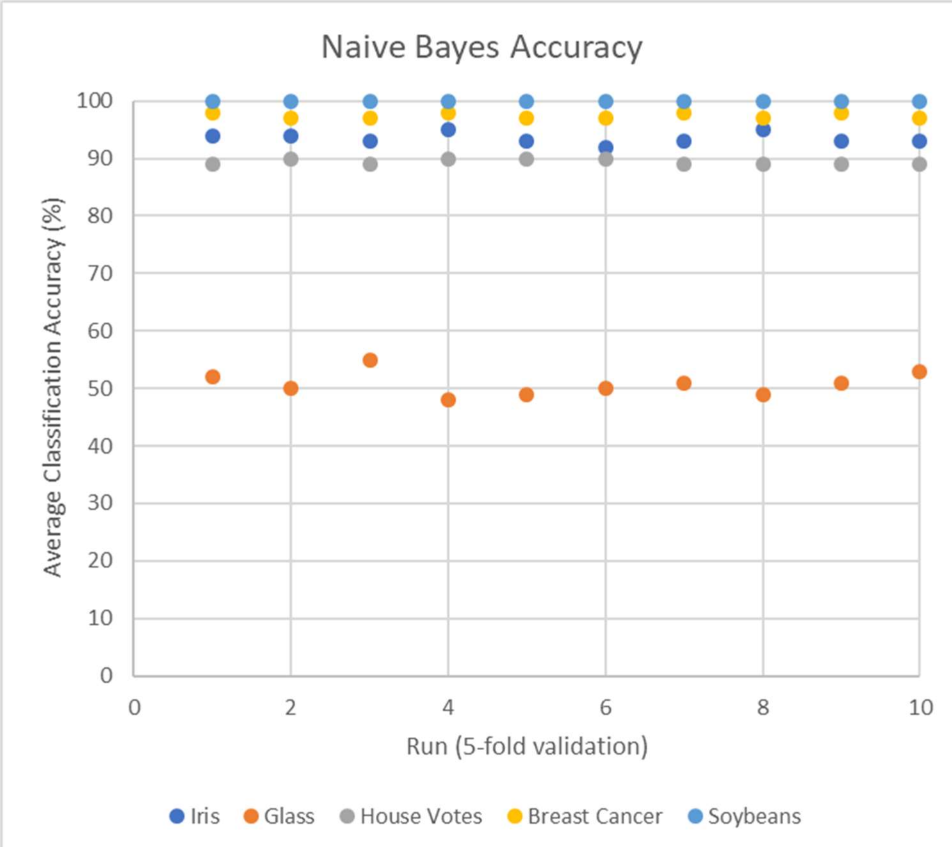
Name	Instances	Attributes	Classes	Data Type
Iris (1)	150	4	3	Continuous
Glass (2)	214	10	7	Continuous
House Votes (3)	435	16	2	Boolean
Breast Cancer (4)	699	10	2	Categorical
Soybeans (5)	47	35	4	Categorical

Each dataset was downloaded from its public repository and treated for use by the two algorithms.

The Iris dataset contained no missing values or unlearnable attributes. The ID column of the Glass dataset was removed, as it is not a learnable attribute. Missing data in the House Votes dataset was too frequent to remove, so a random Boolean value was generated for each missing data point. Instances with missing data in the Breast Cancer dataset were dropped, as there were only 16 out of almost 700 examples with missing data. The unlearnable Sample number column was also dropped from the Cancer dataset. Column order was rearranged where necessary to ensure that the Class value was always in the last column, and the data types were changed to numeric for all values except the Class. Finally, because both algorithms operate over Boolean variables, the categorical data in all datasets except the House set were converted to Boolean values using one-hot-encoding. The continuous data was binned by quintile, then converted to Boolean values using one-hot-encoding. The algorithms were evaluated on each dataset ten times using five-fold validation and the average classification accuracy was recorded as a percentage. The models produced by the learning methods were also recorded in a text file.

Results





Overall, both algorithms performed very well. They had average classification accuracies between 90% and 100% for four of the five datasets, and both algorithms achieved a 100% accuracy rate on every run on the Soybean dataset. However, both algorithms only achieved a successful classification rate of around 50% on the Glass dataset. Both algorithms had the second-best accuracy of about 95% on the Breast Cancer dataset, and while Logistic Regression had effectively the same success rate on the House Votes dataset, Naïve Bayes only achieved a success rate of about 90% on the same dataset. Both algorithms had a classification accuracy of about 93% on the Iris dataset. Across all datasets, the classification accuracy for both algorithms was remarkably similar.

Average Classification Accuracy (%) by Dataset

Dataset	Logistic Regression	Naïve Bayes
Soybeans	100	100
Breast Cancer	95.5	97.4
House Votes	95.2	89.4
Iris	93.7	93.5
Glass	47.2	50.8
Overall Mean Accuracy	86.32	86.22

<u>Discussion</u>

As expected, the two datasets comprised of continuous variables, Glass and Iris, generally had lower classification accuracies than the categorical datasets. However, that explanation is too simplistic. The algorithms produced their lowest scores on the Glass dataset by a large margin, but they performed almost identically on the Iris dataset. Additionally, the Naïve Bayes algorithm produced it's second-lowest score on the Boolean House Votes dataset, rather than the Iris dataset, so this association is not particularly clear-cut. The Iris dataset has fewer attributes with low variability, so it is possible that the quintile binning scheme used to encode the data resulted in less information loss than in the Glass dataset. It is also possible that the Glass dataset is simply less linearly separable.

There does not appear to be any relationship between the number of classes in a dataset and the relative performance of the two algorithms. The Breast Cancer and House Votes datasets had two classes each, and Logistic Regression performed better than Naïve Bayes on House Votes, but worse on Breast Cancer. Both Algorithms performed the worst on Glass, the dataset with the most classes, but they both performed the best on Soybeans, the dataset with the second highest number of classes. If the number of classes has any effect, it is not of noticeable significance.

Lastly, it is worth noting how consistent both algorithms were on each dataset. All datasets were pre-processed in a uniform way, and the same learning rate for Logistic Regression was used across all the datasets. If the Glass dataset only had two classes, the observed classification accuracy of about 50% would be the expected result if the algorithm was assigning class labels randomly. However, the Glass dataset has seven classes, so a success rate of 50% indicates that some amount of learning has occurred. It is likely that different binning schemes or a better tuned learning rate would have resulted in better accuracy in general, and on the Glass dataset in particular.

Summary

In general, both algorithms performed very well. An average overall accuracy of 86% is fairly high, and that average improves to a respectable 95% or 96% when the outlier Glass accuracy is discounted. Unfortunately, while both algorithms performed classification well, the implementation of Logistic Regression was extremely, almost unusably slow to train. This could be due to the specific implementation but is likely also due to the chosen strategy of training a separate one-vs-all weight vector for each class in the dataset. This long training time made tuning the algorithm very difficult, so binning strategies and learning rates were not optimized or tailored to each dataset. Additionally, regularization was not implemented for the Logistic Regression algorithm. As good as the overall accuracy was, a faster, regularized implementation would have allowed for better tuning and potentially even greater accuracy. Even untuned, the Logistic Regression algorithm performed comparably to the Naïve Bayes algorithm, and has the potential to perform with even greater accuracy when properly adjusted to a specific dataset.

References

1) Creator: R.A. Fisher Donor: Michael Marshall (MARSHALL%PLU@io.arc.nasa.gov) Date: July, 1988
2) Creator: B. German -- Central Research Establishment Home Office Forensic Science Service Aldermaston, Reading, Berkshire RG7 4PN Date: September, 1987
3) Title: 1984 United States Congressional Voting Records Database Source: Congressional Quarterly Almanac, 98th Congress, 2nd session 1984, Volume XL: Congressional Quarterly Inc. Washington, D.C., 1985. Donor: Jeff Schlimmer (Jeffrey.Schlimmer@a.gp.cs.cmu.edu) Date: 27 April 1987
4) O. L. Mangasarian and W. H. Wolberg: "Cancer diagnosis via linear programming", SIAM News, Volume 23, Number 5, September 1990, pp 1 & 18.
5) Michalski,R.S. Learning by being told and learning from examples: an experimental comparison of the two methodes of knowledge acquisition in the context of developing an expert system for soybean desease diagnoiss", International Journal of Policy Analysis and Information Systems, 1980, 4(2), 125-161. Donor: Doug Fisher (dfisher%vuse@uunet.uucp) Date: 1987
6) Walker, S., & Duncan, D. (1967). Estimation of the Probability of an Event as a Function of Several Independent Variables. Biometrika, 54(1/2), 167-179. doi:10.2307/2333860
7) Maron, M. E. (2002). Automatic Indexing: An Experimental Inquiry. *Journal of the ACM*, *8*(3), 404–417. https://doi.org/10.1145/321075.321084

Appendix

Models:

Iris Logistic Regression Model
Iris-versicolor [-1.2818928193068149, 0.9796627089185193, 0.3376628767941307,
0.7671236453888337, -2.039206628805601, -1.340159076889557, -2.3791298344116965, -
1.5820021850870631, 0.7592439759060476, -0.780465216830731, 2.6930494556934503, -
3.5751178521115388, -1.182054619291583, 2.9720433849627366, 2.6893091877383446, -
2.197818392924874, -1.9741695602955132, 1.0864208205212351, 2.960205513709653, -
2.1912155069460666, -1.1896411263009827]
Iris-virginica [-1.042314769260978, -1.5103285151603205, -2.4079395666115793, -
1.5942083018606934, 2.8088555758553895, 1.6741758272419767, -1.056364754551148,
1.126161462614182, 0.3567861524746573, -0.5243916838502882, -0.9520739382168001, -
2.3853708721227522, 1.938046050097104, -2.187153010707783, -0.957184803844682,
2.542075995178285, -2.0139322673654756, -1.3372902653290506, -2.177826919043512,
2.5361973086134837, 1.9304792210765014]
Iris-setosa [-1.6640235597208832, -0.1748161156057076, 0.9443452877231393, 0.127383806906081, -
1.9019607279563144, -0.6870447244999712, 2.508867390909718, -0.36555345517590937, -
2.5100952986114438, 0.8917529894920797, -2.1854570828429583, 4.538435029051705, -
1.497845255716155, -1.4263646261049483, -2.1821159557439884, -1.069352056095727,
2.9004674680635536, -0.5405794670373951, -1.4290137722342278, -1.0694826216249644, -
1.4930895329654026]

Iris Naive Bayes Model
Iris-setosa {'P(C)': 0.5833333333333334, 'P(Fi=1|C)': [0.375125, 0.375125, 0.125125, 0.000125,
0.000125, 0.500125, 0.125125, 0.000125, 0.250125, 0.000125, 0.875125, 0.000125, 0.000125, 0.000125,
0.000125, 0.750125, 0.125125, 0.000125, 0.000125, 0.000125]}
Iris-versicolor {'P(C)': 0.3333333333333333, 'P(Fi=1|C)': [0.2002, 0.2002, 0.4002, 0.0002, 0.0002,
0.0002, 0.0002, 0.6002, 0.0002, 0.2002, 0.0002, 0.0002, 0.6002, 0.2002, 0.0002, 0.0002, 0.2002, 0.6002,
0.0002, 0.0002]}
Iris-virginica {'P(C)': 0.08333333333333333, 'P(Fi=1|C)': [0.0005, 0.0005, 0.0005, 0.0005, 0.5005,
0.0005, 0.0005, 0.5005, 0.0005, 0.0005, 0.0005, 0.0005, 0.0005, 0.0005, 0.5005, 0.0005, 0.0005, 0.0005,
0.5005, 0.0005]}

House Logistic Regression Model
democrat [0.4637882026108591, 3.7230994979562433, 3.122197042528344, 2.5902352153619876, -
5.7343652221666925, 0.08561345854841822, 0.09022436333668983, -0.07786368306146002,
0.3614452092592981, 3.5882471710480246, -1.2632998970463118, 2.905900025784227, -
2.509243265898509, 0.08195801019402484, 0.07683709799518294, -4.168314969250075,
0.37685522669714455]
republican [-0.46148340871174487, -3.851219343926494, -3.2108920764073687, -
2.6924176857575493, 5.932359362689143, -0.10223148783634929, -0.09461389131970153,
0.08321829479107072, -0.3822530269324689, -3.70534569722875, 1.3048415167567664, -
2.991463143654079, 2.6070169825961815, -0.10564817453885687, -0.09768715797915441,
4.341389583675833, -0.4081168112165672]

House Naive Bayes Model
democrat {'P(C)': 0.5, 'P(Fi=1|C)': [0.5715714285714286, 0.4287142857142857, 0.5715714285714286,
0.143, 0.5715714285714286, 0.7144285714285715, 0.4287142857142857, 0.28585714285714287,

0.5715714285714286, 0.4287142857142857, 0.28585714285714287, 0.28585714285714287,
0.7144285714285715, 0.5715714285714286, 0.00014285714285714287, 0.7144285714285715]}
republican {'P(C)': 0.5, 'P(Fi=1|C)': [0.143, 0.143, 0.143, 0.8572857142857143, 0.8572857142857143,
0.8572857142857143, 0.143, 0.00014285714285714287, 0.00014285714285714287,
0.4287142857142857, 0.143, 0.8572857142857143, 0.8572857142857143, 0.8572857142857143, 0.143,
0.5715714285714286]}

Glass Logistic Regression Model
6 [-0.721899549884558, 0.0036837118491032, -2.3443306252520153, 1.6234467364098546, -
0.0012520737311523585, 0.001570489534481967, 0.8681100930018526, -3.221092960432468, -
0.0012472019120509173, -0.0023198137799575894, 1.6308594298117798, -3.1439741068894245, -
1.9477930834608341, -0.1100786697624341, 5.403170521027401, -0.9404329452423978,
2.3251147226953877, -1.0734424427855456, -0.23135380955360263, -1.1407188645987971, -
0.5859937243935762, -0.944674490366612, 0.037673085962437486, -1.4513335612863076, -
0.0030021730769252235, 1.6241647778612787, -0.7270554470991024, -0.0062118756657961 2, -
0.008795604796716561, -0.0036140875756215116, -1.4189790443838517, -0.9473297201604132,
1.6211365729675762, -0.0015975696926679196, 0.0017658779897844496, 1.0236826527903575, -
1.7410843427517533, 0.009100881160088762, -0.0071758190426410035, -0.0006334062063549153,
0.6326019625429534, -0.28695386144967766, -1.067468301551177, -0.0029483575009946472,
0.0006524037473660364]
3 [-1.668773728084417, -0.002564707677407816, -0.6764062910391229, -0.9939855446048511, -
0.0006914737148450718, -0.002863840488436134, 1.747079227883246, -2.4321904011788056,
0.0016998236398348583, -0.002122520265292525, -0.9780755460031871, 3.093448063695155, -
0.5894312420184249, -2.369345013753497, -1.0079623100324917, -0.7728406613757666, -
3.282431607730495, -0.947672012351869, 3.9492244815778346, -0.9622165111929125, -
0.4280049965838506, -0.7889142020510427, 0.9281216975897212, -0.8367129963277419,
0.00703897260447695, -0.9939220309335073, -1.6721576096372242, 0.0007085238362575779,
0.007846890130057954, 0.009261742172346576, 0.09764202518267269, -0.7823081717162419, -
0.9780094817439311, -0.0034399379170476484, -0.0026070882753820145, -0.2812805536559812, -
1.3792484171968076, -0.003719501052878371, 0.007861798375104796, -0.006717242382161717, -
0.5866061240651362, -0.12105296907521218, -0.9442919869848946, -0.003375524516497466, -
0.007828005670189622]
2 [-0.16809357939896497, 0.00327794844405609, 0.24132528944569367, -0.3951778096380456,
0.0026030209542666945, -0.0007645161437805411, -1.5796948524701069, 1.8064977134245475, -
0.006501251503646452, 0.006261419724289877, -0.3990913262575182, -4.5912767576805855,
6.02455268708158, -1.2484840258164023, -1.5381568111052195, 1.191098032238258, -
0.8150923677190831, 2.2820435439820628, -0.7828876440681976, -0.5990520569647237, -
0.25303956805754224, 1.1821694019465645, 4.038968265750579, -4.976844937361714, -
0.00029743710572123494, -0.3908460572776339, -0.1684950515624568, -0.00395267228746409,
0.009933150527025205, -0.006938081887883854, -0.9504667727229329, 1.1775445597667353, -
0.4039232395719542, 0.006143929747238231, -0.002731469973449487, 0.6855022289393036, -
0.8446297112903803, 0.009623439669311432, 0.0013015428979684836, -0.005487360343718724, -
6.2172429775316775, 3.7775013186176083, 2.280492952019321, 0.009068906465705682,
0.00016896434318977 93]
1 [-1.8560741715834772, -0.007485899734468527, -1.0602260913589439, -0.7803913493009937, -
0.002072637302835117, 0.00581843581253408, -3.9389850299941944, 2.8674811703935204,
0.001875001166474058, -0.00812441441654646, -0.7903986572437578, 5.701197334724359, -
4.201176870676652, -2.245229104431695, -0.9416639769040442, -0.1954091338186097,
3.2604469555472213, -1.5359429013184893, -2.149091168726211, -0.24414207086994213, -
1.208862210880416, -0.1946185280761603, -4.710209237206378, 3.840089779785427,
0.009188739693391576, -0.7888077429200531, -1.8618287268268463, 0.009032490029361467, -

0.009133976200180369, 0.007326169678451516, -0.8703354415336226, -0.1966046265208391, -0.7906879805227072, 0.007342045275303648, 0.004167588162225842, -0.3983405722778595, -1.4578533331491006, -0.0056984064499565395, -0.007773243379285515, -0.005351852096559651, 4.471428729431293, -4.802225853714485, -1.5358499065225313, -0.0047772488778490985, -0.0006989174183198443]
7 [-0.8211394949147868, -9.055002434479355e-05, -0.17730481182450764, -0.6521628032764295, -0.008831841292008749, -0.0025258538498864743, 1.0203791731547478, -1.1964214982778758, -0.004070006981174796, -0.0032428663072776786, -0.646953010299967, -1.021727562949045, -0.8552027113221315, 2.1842071294703285, -0.6668473347597252, -0.48463162604783283, -1.6089045334482865, -0.7162909851658795, -1.3153858196825492, 1.5858504677858711, 1.2254499603019018, -0.47959085357059833, -0.25290489238716035, 0.557220737655863, -0.007853694714721814, -0.655449622552847, -0.8307994499238711, -0.009496170227662793, 0.008726246222839925, -0.0018849808355398175, 0.2945765358655415, -0.47996721117539676, -0.6414212939368892, 0.0018758645604911427, 0.0043034152805661845, -3.6442090085663295, 2.8223815289676484, 0.00658079569788249, 0.0022975351814664997, 0.009188523086941883, 0.09566227128166754, -0.20895508417122816, -0.7220210401584071, -0.0037877056171149934, -0.00824345066486974]

Glass Naive Bayes Model

2 {'P(C)': 0.3333333333333333, 'P(Fi=1|C)': [0.0002, 0.8002, 0.0002, 0.0002, 0.0002, 0.0002, 0.8002, 0.0002, 0.0002, 0.0002, 0.0002, 0.8002, 0.0002, 0.0002, 0.0002, 0.4002, 0.4002, 0.0002, 0.0002, 0.0002, 0.0002, 0.4002, 0.4002, 0.0002, 0.0002, 0.8002, 0.0002, 0.0002, 0.0002, 0.8002, 0.0002, 0.0002, 0.0002, 0.0002, 0.8002, 0.0002, 0.0002, 0.0002, 0.0002, 0.4002, 0.2002, 0.2002, 0.0002, 0.0002]}
7 {'P(C)': 0.25, 'P(Fi=1|C)': [0.00025, 0.75025, 0.00025, 0.00025, 0.00025, 0.50025, 0.00025, 0.25025, 0.00025, 0.00025, 0.00025, 0.00025, 0.75025, 0.00025, 0.00025, 0.00025, 0.00025, 0.00025, 0.25025, 0.50025, 0.00025, 0.25025, 0.50025, 0.00025, 0.00025, 0.75025, 0.00025, 0.00025, 0.00025, 0.75025, 0.00025, 0.00025, 0.00025, 0.00025, 0.00025, 0.75025, 0.00025, 0.00025, 0.00025, 0.75025, 0.00025, 0.00025, 0.00025, 0.00025]}
1 {'P(C)': 0.25, 'P(Fi=1|C)': [0.00025, 0.75025, 0.00025, 0.00025, 0.00025, 0.00025, 0.75025, 0.00025, 0.00025, 0.00025, 0.50025, 0.25025, 0.00025, 0.00025, 0.00025, 0.75025, 0.00025, 0.00025, 0.00025, 0.00025, 0.00025, 0.75025, 0.00025, 0.00025, 0.00025, 0.75025, 0.00025, 0.00025, 0.00025, 0.75025, 0.00025, 0.00025, 0.00025, 0.00025, 0.75025, 0.00025, 0.00025, 0.00025, 0.00025, 0.75025, 0.00025, 0.00025, 0.00025, 0.00025]}
6 {'P(C)': 0.16666666666666666, 'P(Fi=1|C)': [0.0003333333333333333, 0.3336666666666666, 0.3336666666666666, 0.0003333333333333333, 0.0003333333333333333, 0.3336666666666666, 0.0003333333333333333, 0.0003333333333333333, 0.0003333333333333333, 0.3336666666666666, 0.0003333333333333333, 0.0003333333333333333, 0.3336666666666666, 0.3336666666666666, 0.0003333333333333333, 0.3336666666666666, 0.0003333333333333333, 0.3336666666666666, 0.0003333333333333333, 0.0003333333333333333, 0.0003333333333333333, 0.3336666666666666, 0.0003333333333333333, 0.0003333333333333333, 0.3336666666666666, 0.6669999999999999, 0.0003333333333333333, 0.0003333333333333333, 0.0003333333333333333, 0.3336666666666666, 0.0003333333333333333, 0.3336666666666666, 0.0003333333333333333, 0.0003333333333333333, 0.6669999999999999, 0.0003333333333333333, 0.0003333333333333333, 0.0003333333333333333, 0.0003333333333333333, 0.6669999999999999, 0.0003333333333333333, 0.0003333333333333333, 0.0003333333333333333, 0.0003333333333333333]}

Cancer Logistic Regression Model

2 [-0.5516642841055515, 0.04139488206002147, 0.005350188598646988, -0.6422204861755703, -0.008338485933007522, 0.005251472886520137, 0.989206833733217, 0.2056322622948574, -0.008294964302342108, -0.697205181883946, -0.4553895891291605, 2.051070785244048, -0.002210468262808545, -0.8059313462830627, -0.636127861652106, 0.005427435597719412, -

0.009454733423583455, -0.45776791188274923, 0.00025362351184378407, -0.6863217082109017, -
0.007509926379184812, 1.2518869296119624, -0.20260899440585253, -0.006340192349961636,
0.007657714490472705, -0.42804056042680555, -0.0009814387659174955, -0.005314974995119912, -
0.6989208505617597, -0.46980358752279994, 0.0021259662244735994, 1.3781383817296922, -
0.4568392075028046, 0.23593311994386207, -0.6406187440731479, -0.8066422242926099, -
0.0018175729007532532, 0.009659688334738152, -0.25281868680616343, 0.00021045686196813872,
0.005580858860078975, 1.2661871360882333, -0.008130238915493493, 0.22234200319201156,
0.5659788780299599, 0.00620000223889716, -0.6907075878808449, -1.2718868570555206,
0.00903378381961177, -0.6429205928339401, -0.005018908553745743, 1.612260661327596, -
1.1577726434299704, -0.6370126906892112, 0.006971030220468749, 0.4290253671846711, -
0.009373534108551683, -0.0006816092248889214, -0.0007047883805718809, -0.8041679347910007, -
0.005947045305457368, 0.5982005995017072, -1.5020368078440134, 0.823925520735419,
0.6418915290030878, 0.007371186265953967, -0.0003019543500644914, -1.0974996001020438,
0.004159777157084075, 0.006690930415622303, 0.0035256281500358096, 1.4418953792681988,
0.600149074080323, -0.692438943883336, 0.0009067950401047287, -0.006188672432639975,
0.009343469729293003, -0.8018475880902717, 0.006371195150157069, 0.0030285281276997214, -
1.09308799978426, -0.0026932846815654237, -0.004411913145842106, 0.0011412933688892644,
0.10895490361204517, -0.4652283005216046, 0.4288747752670125, -0.6427807822537255, -
0.005191445245270294, -0.006329144782705472]
4 [0.56094140479388, -0.031214125001254325, -0.003322713569820883, 0.65999252716199, -
0.0033830777396804046, 0.0076017536948826455, -1.0093241141967155, -0.2161605833342048,
0.009264635665379868, 0.7124250798028311, 0.46995010261867015, -2.098115113320617,
0.00979062201083974, 0.8293137951325542, 0.6472767635020289, -0.006129633900194789,
0.0019926423418546706, 0.47618263876629013, 0.007615123171165522, 0.7087978666071311, -
0.0029715680345679266, -1.267793788562248, 0.200614956354559, 0.008887833217923602,
0.004821724827742577, 0.4427169376950742, -0.007733570954358466, -0.003052373057409805,
0.7102738502496593, 0.47198899296293245, -0.009525313502946348, -1.396059987116267,
0.4608035297917983, -0.2380309138619032, 0.659229908247444, 0.8311917934999988, -
0.008581481752436783, -0.006585616464552097, 0.26531692775151383, 0.0055881719499842036,
0.0014289283978497318, -1.2891482549411195, -0.003210154152532123, -0.21093490286454342, -
0.5751671340075957, -9.832244818306149e-05, 0.7049375911373024, 1.2878294919477686,
0.00966816306952416, 0.6444642442799102, 0.006003003861788885, -1.6448171425908489,
1.1819555128284422, 0.6501834344343547, -0.004183439704796696, -0.44435519144681235, -
0.0005346478555862191, -0.0003268525467733166, -0.009585378702022418, 0.8244762791667686, -
0.008473436293511454, -0.6152757648171993, 1.5318845306419582, -0.8350130037091239, -
0.6377866999007461, -0.0023324773011188914, 0.004422531211986604, 1.1183392649133452, -
0.009515460192054017, 0.004766753539636563, -0.0007065299230671496, -1.4810388014947338, -
0.6234308939968233, 0.7081710545629301, -0.003852500276644433, -0.008551550824488057, -
0.0007633422363921821, 0.8240378982409235, -0.0007224279729317849, 0.0027914277247433946,
1.1282077004109907, -0.0012514766369809054, 0.0032508030547355755, 0.008368278434152615, -
0.10357242499882409, 0.4744676283122375, -0.4397629931238101, 0.6508529327532653, -
0.009616039619807597, 0.008511162260469855]

Cancer Naive Bayes Model
2 {'P(C)': 0.75, 'P(Fi=1|C)': [0.2001, 0.0001, 0.0001, 0.0001, 0.0001, 0.6001000000000001, 0.1001,
0.0001, 0.0001, 0.0001, 0.9000999999999999, 0.0001, 0.0001, 0.0001, 0.0001, 0.0001, 0.0001, 0.0001,
0.0001, 0.0001, 0.7001000000000001, 0.1001, 0.0001, 0.0001, 0.1001, 0.0001, 0.0001, 0.0001, 0.0001,
0.0001, 0.7001000000000001, 0.0001, 0.1001, 0.0001, 0.0001, 0.0001, 0.0001, 0.1001, 0.0001, 0.0001,
0.5001, 0.0001, 0.1001, 0.3001, 0.0001, 0.0001, 0.0001, 0.0001, 0.0001, 0.0001, 0.8000999999999999,
0.0001, 0.0001, 0.0001, 0.1001, 0.0001, 0.0001, 0.0001, 0.0001, 0.0001, 0.2001, 0.0001, 0.3001, 0.4001,
0.0001, 0.0001, 0.0001, 0.0001, 0.0001, 0.0001, 0.8000999999999999, 0.1001, 0.0001, 0.0001, 0.0001,

0.0001, 0.0001, 0.0001, 0.0001, 0.0001, 0.7001000000000001, 0.0001, 0.0001, 0.1001, 0.0001, 0.1001, 0.0001, 0.0001, 0.0001]}
4 {'P(C)': 0.25, 'P(Fi=1|C)': [0.00025, 0.00025, 0.25025, 0.00025, 0.00025, 0.00025, 0.00025, 0.00025, 0.25025, 0.25025, 0.00025, 0.00025, 0.00025, 0.25025, 0.00025, 0.00025, 0.25025, 0.25025, 0.00025, 0.00025, 0.00025, 0.00025, 0.25025, 0.00025, 0.25025, 0.00025, 0.00025, 0.00025, 0.25025, 0.00025, 0.00025, 0.25025, 0.00025, 0.25025, 0.25025, 0.00025, 0.00025, 0.00025, 0.00025, 0.00025, 0.00025, 0.00025, 0.25025, 0.00025, 0.00025, 0.00025, 0.25025, 0.00025, 0.25025, 0.00025, 0.00025, 0.50025, 0.25025, 0.00025, 0.00025, 0.00025, 0.00025, 0.00025, 0.00025, 0.00025, 0.00025, 0.00025, 0.00025, 0.00025, 0.00025, 0.25025, 0.50025, 0.00025, 0.00025, 0.00025, 0.25025, 0.00025, 0.00025, 0.00025, 0.00025, 0.00025, 0.00025, 0.00025, 0.00025, 0.50025, 0.00025, 0.00025, 0.00025, 0.00025, 0.50025, 0.00025, 0.25025, 0.00025, 0.00025]}

Soybean Logistic Regression Model
D3 [-0.156076820727092, 0.001390571923086479, -0.20436491300097673, -0.00547702208927797, -0.390297647495098, 0.5849721524025, 0.3308940510258177, -0.46477743500254776, -0.611280670907037, 0.44847711675390306, 0.19029807596672793, -0.1610286095961911, -0.17725966316321032, -1.219855394628688, -0.1684424963069664, 1.2179235125345504, 0.0034358992485679883, -0.15666803784799022, 0.34110697966374054, -0.588482971167198, -0.2131385126621249, 0.2919519262252664, -0.4494404445339126, 0.45106828121104064, -0.006330773969185887, -0.14935010075538246, -0.6271999409608183, 0.4560530072425212, -0.2664922340097233, 0.11233024684026667, 0.642760111306941, -0.0685492235447887, -0.7181756265820597, -0.1681749021923763, -1.6554061558221633, 1.4989686814902818, -0.16820752290835206, -0.15391635225211953, -0.15243312310750237, -0.1688817978709326, -0.1509470134138381, -0.15094322168000274, -0.16029571740808124, 0.04040049472865198, -0.2014996663302829, -0.4358644071044468, -0.1568861227493256, 0.9022435429250686, -0.47215813560067893, 1.0608099820201562, 0.0024500949153868808, -0.15668207970243353, -1.069220603357781, -0.45295112342198335, 0.28831412503482456, 1.0671325987578975, -1.220188810310343, -0.8494672865829762, 0.6712666289184915, -0.00025260322168559534, -0.16605869377992952, 0.007813984756106231, -0.1489892798216216, -0.6103749181745756, 0.44541069069562306, -0.15091477577978818, -0.15217943795162195, -0.15129213321967208, -0.1526603172353913, -0.1624154703817855, -0.15254621607292754, 0.908130920746962, -1.0659016700819997]
D4 [-0.15995452660704196, -0.0047224251656125425, -0.18644467617540078, -0.0011457061871473929, -0.5850071203983834, 0.12298680250414744, -0.3446071065056577З, 0.8250792589791931, -0.7658580946951641, 0.6074892549036226, -0.11603564118647279, -0.3711750517241062, 0.3193531124112154, 0.8738901732228583, -0.37898742817061, -0.652114818871728, 0.0021690854605491243, -0.16391910103405521, -0.5012001292086211, 0.5157596090694556, -0.18986586005586978, 0.023854335116082565, -0.39011987048866853, 0.6036048521623627, 0.00858198093528648, -0.3665954068684352, 0.0867575016666229, -0.23830987772348275, 0.5546355689455745, -0.7107090795435126, -0.13597907551704896, -0.06798322919525546, 0.05641735430311413, -0.14880106376487995, 0.9797064009084502, -1.1369754945849877, -0.16290451950114743, -0.14735509187490298, -0.1595318513454408, -0.15815910017699905, -0.16546654367254793, -0.1637951932244947, -0.1651889939512175, 0.033217661956121636, -0.18495109088619954, -0.3938238457459922, -0.37297771739747293, -0.2134938476521013, 0.8273488873234482, -1.5320440936204904, -0.0015797722759442728, -0.3828812542458146, 1.748322814653274, -0.3942139679168348, 0.23880126330600526, -1.53076555665983, 1.3679352304281633, 0.40417564721020144, -0.5524520295225492, 0.2088392453955917, -0.36996048908362733, 0.22649455293429818, -0.3833684170883863, -0.771898728478357, 0.6179203937245346, -0.1652604291030596, -0.1523412066882307, -0.1519966257552443, -0.16685002209243713, -0.15758193527144335, -0.15656328651391047, -1.891949921454326, 1.7450951732440818]

D2 [-0.18683434724034118, -0.006960321690132814, -0.19619602577031345, -0.004396999420676984, 0.6411161057242808, -0.4290189523821339, -0.048835726083366726, -0.15560948360446553, 0.4418456777601545, -0.6358464551894953, -0.874005592957194, 0.8146649513001243, -0.11328090406786354, -0.6808947265979702, 0.820140505752879, -0.31648109692667764, 0.004785490050417025, -0.16992097870007902, 0.5708860285282844, -0.43170569869917935, -0.19916662470081478, -0.13684123648456567, -0.37534344624866317, -0.6197322363447189, -0.000641891202704423, 0.8227280684534877, 0.011746301834529599, -0.18998887426343497, -0.28861640289004237, 0.11775138245909318, -0.302604415948644, -0.32489624836319003, 0.44480315609958626, -0.17981072693267577, 0.061042534551860615, -0.23834326250706306, -0.1803329070828449, -0.17468572078492187, -0.1847328115566009, -0.17421025607217602, -0.18302717920754738, -0.18501570892722793, -0.1818576827824718, 0.014113126922901544, -0.19234415099317317, -0.3637752512491436, 0.8129776255591041, -0.48705431144253997, -0.1404722281627722, -0.605352439891169, -0.002470970748044932, 0.823353832075143, -0.38710412411472866, -0.3644933522310178, 0.19233712556073423, -0.6028981323639991, 0.4362827487610889, -0.05078011169786684, -0.11606664967922962, -1.0004979492290837, 0.822150844966798, -1.0067316390530086, 0.8233987468725514, 0.4587734810226331, -0.6245929299857857, -0.17049371578371428, -0.1842065515317097, -0.17652847609562836, -0.1805881224353265, -0.1808730751849766, -0.17310921697555678, 0.20379948328365968, -0.3836869421184768]
D1 [-0.19487788662867664, -0.006677454136354371, 0.4633118031582705, 9.388795113182344e-05, 0.3041277761635972, -0.6170429117725199, -0.11361236102338718, -0.21963959051888296, 0.7624852322353038, -0.9447818705523896, 0.31828687110516896, -0.4373938989611631, -0.0690172407704618, 0.8133199927130964, -0.4343797065980767, -0.5773623678853531, 0.0038360561950733056, -0.18518733484810004, -0.7419508297859295, 0.35360683936788945, 0.4663095922571919, -0.27134467701733583, 1.1827138919288473, -0.9506892449842057, 0.005081806127334469, -0.42691272932344865, 0.15045481404955946, -0.3430288334440603, -0.3486892365188906, 0.17013093667035756, -0.502823476835448, 0.45519070353311225, -0.12965593066400974, -0.18835782596629655, 0.30073043195478827, -0.49175619124653114, -0.17779726857047792, -0.18142985007261694, -0.19057139302168993, -0.19304098308237969, -0.18693234473577633, -0.18096929879147525, -0.19559883964235572, -0.6541923414182789, 0.4593337193713762, 1.1918239379205318, -0.4340097810045702, -0.7286456591731261, -0.21693087908954586, 0.695916969948452, -0.0018342532832939587, -0.42095440225457254, -0.45859225519400504, 1.1893010215720374, -1.3724007425151143, 0.6943371845452722, -0.885908577510503, 0.08346641427438015, -0.27087196311774964, 0.24095166014576935, -0.42798306248206025, 0.24042274990430118, -0.43937634832364575, 0.7698911222175842, -0.9441156278063303, -0.19490730650231639, -0.1842508296804036, -0.19065101432002032, -0.17861914713093255, -0.19228323331345157, -0.19111559739730324, 0.26371487647924136, -0.4660376878552076]

Soybean Naive Bayes Model
D3 {'P(C)': 0.3333333333333333, 'P(Fi=1|C)': [0.0002, 0.0002, 0.0002, 0.0002, 0.6002, 0.2002, 0.0002, 0.0002, 0.8002, 0.8002, 0.0002, 0.0002, 0.0002, 0.0002, 0.8002, 0.0002, 0.8002, 0.4002, 0.2002, 0.0002, 0.2002, 0.0002, 0.8002, 0.0002, 0.0002, 0.4002, 0.4002, 0.2002, 0.6002, 0.6002, 0.2002, 0.0002, 0.8002, 0.0002, 0.8002, 0.8002, 0.8002, 0.8002, 0.8002, 0.8002, 0.8002, 0.8002, 0.8002, 0.0002, 0.0002, 0.0002, 0.8002, 0.0002, 0.8002, 0.0002, 0.0002, 0.0002, 0.0002, 0.8002, 0.8002, 0.0002, 0.4002, 0.4002, 0.8002, 0.0002, 0.8002, 0.0002, 0.0002, 0.8002, 0.8002, 0.8002, 0.8002, 0.8002, 0.8002, 0.8002, 0.8002, 0.0002]}
D4 {'P(C)': 0.3333333333333333, 'P(Fi=1|C)': [0.0002, 0.0002, 0.0002, 0.0002, 0.4002, 0.0002, 0.4002, 0.0002, 0.8002, 0.8002, 0.0002, 0.0002, 0.8002, 0.0002, 0.0002, 0.2002, 0.6002, 0.0002, 0.6002, 0.0002, 0.2002, 0.0002, 0.8002, 0.0002, 0.0002, 0.4002, 0.4002, 0.6002, 0.2002, 0.2002, 0.4002, 0.2002, 0.8002, 0.8002, 0.0002, 0.8002, 0.8002, 0.8002, 0.8002, 0.8002, 0.8002, 0.8002, 0.6002, 0.2002, 0.0002, 0.0002,

0.6002, 0.2002, 0.0002, 0.0002, 0.0002, 0.8002, 0.0002, 0.8002, 0.2002, 0.6002, 0.8002, 0.0002, 0.8002, 0.0002, 0.8002, 0.0002, 0.0002, 0.8002, 0.8002, 0.8002, 0.8002, 0.8002, 0.8002, 0.8002, 0.0002, 0.8002]}
D1 {'P(C)': 0.25, 'P(Fi=1|C)': [0.00025, 0.50025, 0.00025, 0.25025, 0.00025, 0.00025, 0.00025, 0.75025, 0.00025, 0.75025, 0.00025, 0.00025, 0.75025, 0.00025, 0.00025, 0.00025, 0.75025, 0.25025, 0.25025, 0.25025, 0.00025, 0.25025, 0.50025, 0.00025, 0.00025, 0.75025, 0.00025, 0.25025, 0.50025, 0.25025, 0.00025, 0.50025, 0.75025, 0.75025, 0.00025, 0.75025, 0.75025, 0.75025, 0.75025, 0.75025, 0.75025, 0.75025, 0.00025, 0.75025, 0.75025, 0.00025, 0.00025, 0.00025, 0.50025, 0.25025, 0.00025, 0.00025, 0.75025, 0.00025, 0.75025, 0.00025, 0.75025, 0.00025, 0.75025, 0.00025, 0.75025, 0.00025, 0.75025, 0.00025, 0.75025, 0.75025, 0.75025, 0.75025, 0.75025, 0.75025, 0.75025, 0.00025]}
D2 {'P(C)': 0.08333333333333333, 'P(Fi=1|C)': [0.0005, 0.0005, 0.0005, 0.5005, 0.0005, 0.0005, 0.0005, 0.5005, 0.0005, 0.0005, 0.5005, 0.0005, 0.0005, 0.5005, 0.0005, 0.0005, 0.5005, 0.5005, 0.0005, 0.0005, 0.0005, 0.0005, 0.0005, 0.0005, 0.5005, 0.5005, 0.0005, 0.0005, 0.5005, 0.0005, 0.0005, 0.5005, 0.5005, 0.5005, 0.0005, 0.5005, 0.5005, 0.5005, 0.5005, 0.5005, 0.5005, 0.5005, 0.5005, 0.0005, 0.0005, 0.5005, 0.0005, 0.0005, 0.0005, 0.0005, 0.5005, 0.0005, 0.0005, 0.5005, 0.0005, 0.5005, 0.5005, 0.0005, 0.0005, 0.5005, 0.0005, 0.5005, 0.5005, 0.0005, 0.5005, 0.5005, 0.5005, 0.5005, 0.5005, 0.5005, 0.5005, 0.0005]}