Name and Roll Number: 1) Priyanka Talwar (23D1387)

2) Ankit Khivasara (23D1388)

Classification using Naive Bayes:

Input: Training dataset: 30000 samples and 10 feature, 3 classes

Validation dataset: 30000 samples and 10 feature, 3 classes

10 feature: x1, x2… x10

Step 1: (on training dataset)

i) For 3 classes calculated prior probability p(y=0), p(y=1), p(y=2)
ii) For x1, x2 features (real number) calculated Gaussian distribution parameters (mean and sigma) (individually)
iii) For x3, x4 features (1/0) calculated Bernoulli distribution parameter (input 1 probability)
iv) For x5, x6 features (real number) calculated Laplace distribution parameters (median and scale for controlling spread or width of the distribution) (individually)
v) For x7, x8 features (real number) calculated Exponential distribution parameters (rate parameter which is inverse of mean) (individually)
vi) For x9, x10 features (integer value varying from 0 to k-1) calculated multinomial distribution parameter (probability of occurs of individual categories among k categories)

Step 2: Calculated Maximum Likelihood estimators (MLE) for each of these distributions and create a naive Bayes classifier with a prior to classify these points.

For calculating MLE constant parameters are ignored and taken log of individual distribution for avoiding underflow errors.

Step 3: Calculated accuracy, precision, recall and F1 score to check models performance

Results:

Model parameters: Priors: {0.0: 0.3333333333333333, 2.0: 0.3333333333333333, 1.0: 0.3333333333333333}
Gaussian: {0.0: [2.0209491, 3.9067733, 9.051951, 78.428345], 2.0: [8.02485, -0.021661418, 35.668865, 4.0075436], 1.0: [0.02138695, 0.855918, 25.160892, 230.03185]}
Bernoulli: {0.0: [0.2023, 0.104], 2.0: [0.9053, 0.1947], 1.0: [0.5984, 0.8018]}
Laplace: {0.0: [0.0766329, 0.8728088, 1.9835867, 5.9781795], 2.0: [0.79637986, 0.2125167, 3.005029, 3.061487], 1.0: [0.38284218, 0.3513043, 0.9993886, 5.998375]}
Exponential: {0.0: [1.978299710021338, 3.935424865362913], 2.0: [8.94272554396005, 14.684990355358273], 1.0: [2.984109558828107, 7.98003518599885]}
Multinomial: {0.0: [[0.2022, 0.2032, 0.2042, 0.1967, 0.1937], [0.1213, 0.1236, 0.1257, 0.1277, 0.127, 0.1271, 0.1241, 0.1235]], 2.0: [[0.2052, 0.2997, 0.1029, 0.3417, 0.0505], [0.1972, 0.0481, 0.0483, 0.1054, 0.1552, 0.153, 0.098, 0.1948]], 1.0: [[0.0977, 0.1984, 0.4047, 0.1583, 0.1409], [0.1009, 0.0506, 0.0508, 0.1998, 0.1524, 0.1487, 0.2003, 0.0965]]}

Performance:

On training dataset and validation dataset:

Training Accuracy: 0.9013666666666666

Validation Accuracy: 0.9023333333333333

Training F1 Score: [0.8562545542871024, 0.8544443897879876, 0.8657745520526194]

Validation F1 Score: [0.8579600542951329, 0.8550939663699307, 0.867312743410923]

On test dataset:

Accuracy on test dataset: 0.9023333333333333