COMP-551: Applied Machine Learning - Assignment #2

Antonios Valkanas

Student ID: 260672034

Question 1

Please see the submitted folder for this question.

Question 2

2.1 a) After having approximated the Bernoulli distribution as a Gaussian and training the GDA model using the maximum likelihood approach the following results were obtained:

```
Accuracy: 0.96375
Precision: 0.9557739557739557
Recall: 0.9725
F measure: 0.9640644361833953
```

2.1 b) The model parameters learned were the following:

```
Coefficients learned:
w0: 25.873030313005195
w: [ 13.73068895  -8.39076669  -5.75281745  -2.86687715  -9.40174385
  -3.9064478   16.32372139 -22.91392594 -27.90111292   8.83033948
 -12.50578729 -11.98214711  14.9072422   12.22579561  -5.07489982
  12.29990028  28.31986351  -6.31844394  -0.55656195  -4.84331305]
```

Question 3

3 a) The k nearest neighbor (k-NN) classifier performs significantly worse than the gaussian discriminant analysis based classifier. This result remains true for any value of k tested. In total, 18 values of k were tested between 1 and 100 and they all yielded similar results. The reason why this happens can be explained from the distribution. Since we are dealing with two gaussian distributions where the means are not very far from each other we have a lot of overlap in the data. As a result, k-NN which operates based on absolute distance is prone to a lot of errors due to said overlap. On the other hand, using GDA is a great way of classifying the data because the activation function of our model allows the decision boundary to fit the data in a non-linear manner such that most classifications are correct. It is also important to note that LDA is expected to work on the data because our data comes from Gaussian distributions with the same covariance.

Table 1: k-NN model evaluation for various k values – DS1.

k	Accuracy	Precision	Recall	F measure
1	0.509	0.508	0.525	0.517
2	0.510	0.519	0.28	0.364
3	0.534	0.533	0.532	0.533
5	0.525	0.524	0.537	0.538
9	0.544	0.543	0.548	0.545
20	0.535	0.536	0.518	0.527
30	0.549	0.550	0.538	0.544
40	0.538	0.538	0.530	0.543

50	0.540	0.539	0.555	0.547
70	0.544	0.539	0.6	0.568
90	0.551	0.545	0.623	0.581

As we can infer from the data in table 1 there does not seem to be meaningful difference in performance for the various values of k. The metrics are always near the same values for any value chosen. It is also important to note that for different runs a different optimal value of k will be chosen. This is due to the different samples that will comprise our training set generated from the multivariate Gaussians and that the model behaves roughly the same even for vastly different k values.

3 b) For the test run conducted for this report the best values that were obtained were the following:

```
For k = 90
Accuracy: 0.55125
Precision: 0.5448577680525164
Recall: 0.6225
F measure: 0.5810968494749126
```

However, as mentioned in part a of this question for different runs a different optimal k will surface based on the training data generated. Also, the best values obtained are very similar to the average value because k makes little difference on how the model behaves.

Question 4

Please see the submitted folder for this question.

Question 5

5.1 a) After having approximated the Bernoulli distribution as a Gaussian and training the GDA model using the maximum likelihood approach the following results were obtained:

```
Accuracy: 0.57125
Precision: 0.571072319201995
Recall: 0.5725
F measure: 0.571785268414482
```

5.1 b) The model parameters learned were the following:

```
Coefficients learned:
w0: 0.09192696230167563
w: [ 0.08120049  0.03562738 -0.0804858   0.02237502  0.10118552 -0.01249639
  0.08101135 -0.0393756  -0.05386362 -0.03594021 -0.13171094  0.06628183
 -0.02801044 -0.0168307  -0.02945508 -0.11210141 -0.00057707  0.0367228
  0.08563488 -0.05488231]
```

5.2) The GDA algorithm still outperforms k-NN but not by much. We notice that introducing the third distribution really deteriorated the performance of the LDA. This is expected as LDA assumes that samples are drawn from normally distributed sources with the same covariance which is not true in our case as the Gaussians that we use have different covariances. The k-NN algorithm exhibits similar behavior to the previous question and maintains similar performance for any value of k. The reasons for this are the same as explained in question 3.

Table 2: k-NN model evaluation for various k values - DS2.

k	Accuracy	Precision	Recall	F measure
1	0.503	0.502	0.520	0.511
2	0.510	0.519	0.28	0.364
3	0.514	0.514	0.500	0.507
5	0.524	0.522	0.555	0.538
6	0.521	0.527	0.413	0.463
7	0.519	0.518	0.533	0.525
8	0.528	0.538	0.435	0.479
9	0.513	0.512	0.534	0.524
20	0.506	0.507	0.425	0.484
30	0.505	0.505	0.48	0.492
40	0.499	0.500	0.475	0.487

5.3) For the test run conducted for this report the best values that were obtained were the following:

```
For k = 8
Accuracy: 0.5275
Precision: 0.5337423312883436
Recall: 0.435
F measure: 0.47933884297520657
```

Question 6

Both datasets consisted of datapoints drawn from gaussians. The LDA algorithm had excellent performance for the first dataset because the datapoints consisted of two similar distributions (with the same covariance) which meant that the assumption that the two distributions had the covariance which is necessary for LDA was correct. However, the second dataset consisted of three gaussians with different means and covariances. As a result, the assumption that the data has the same covariance was wrong and the algorithm had a poor performance. Due to the difference in the covariance matrices a QDA model would have performed much better here. On the other hand, the K-NN algorithm had similar performance for both datasets. Nevertheless, K-NN had very poor performance as it was slightly better than picking at random (accuracy near 50%). This was due to the high dimensionality of the data and the nonlinearity of the problem. Since our K-NN algorithm based its predictions on the absolute distance of the nearest neighbors and there was large overlap in the two distributions it was prone to misclassifications.