SYDE 522 A3

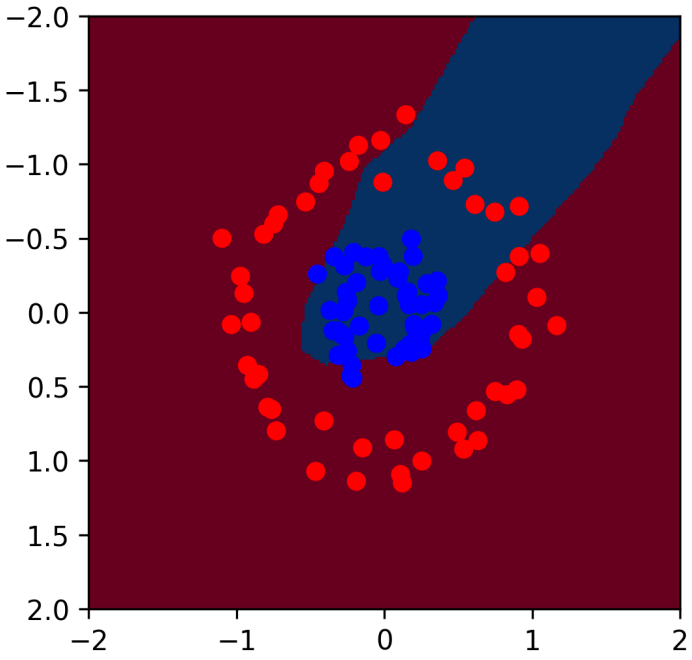
K3burke, 20838709

11/4/2023

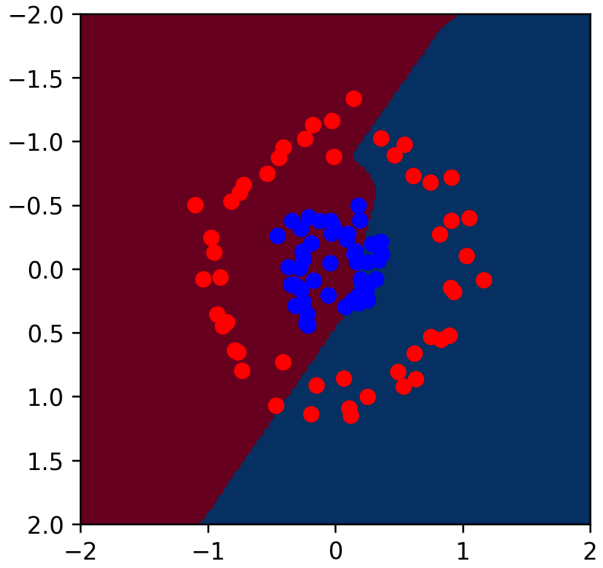
Q1A)

1A RMSE: 0.22360679774997896

Q1B)

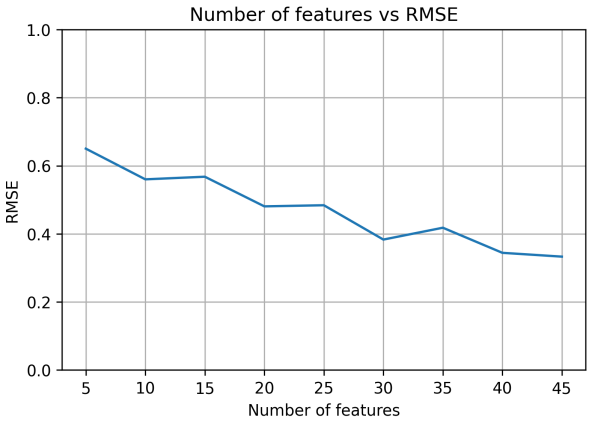


Q1C)

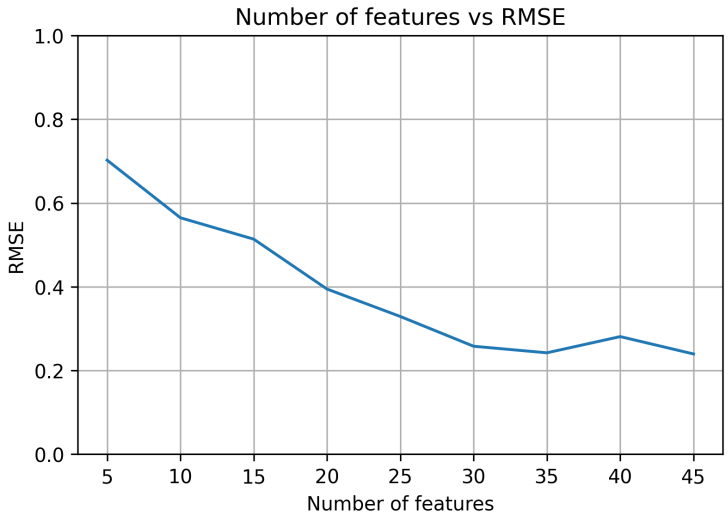


1C RMSE: 0.7416198487095663

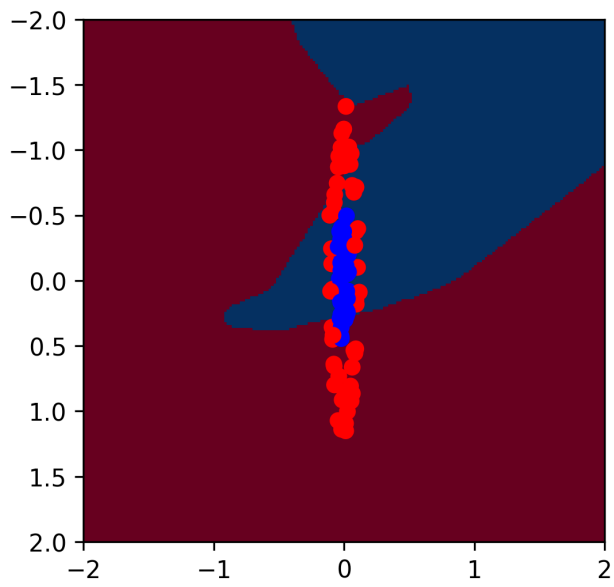
Q1D)



Q1E)

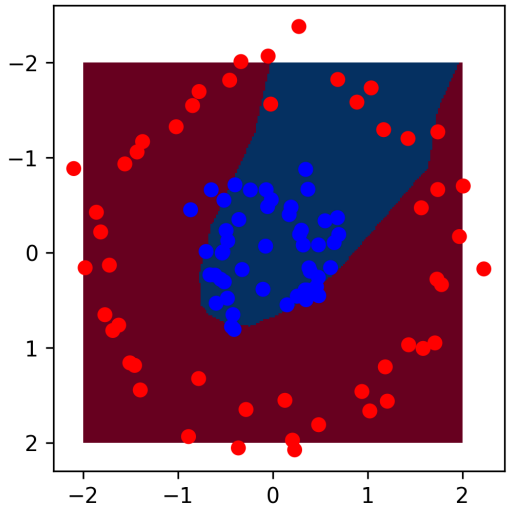


Q1F)



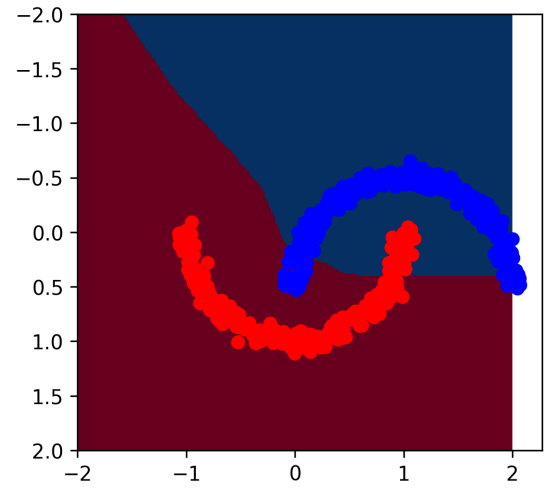
1F RMSE: 0.3872983346207417

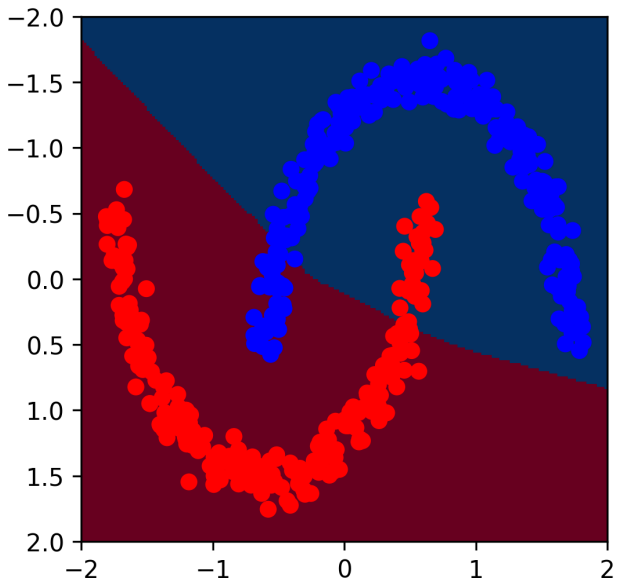
Q1G)



1G RMSE: 0.22360679774997896

Q1H)

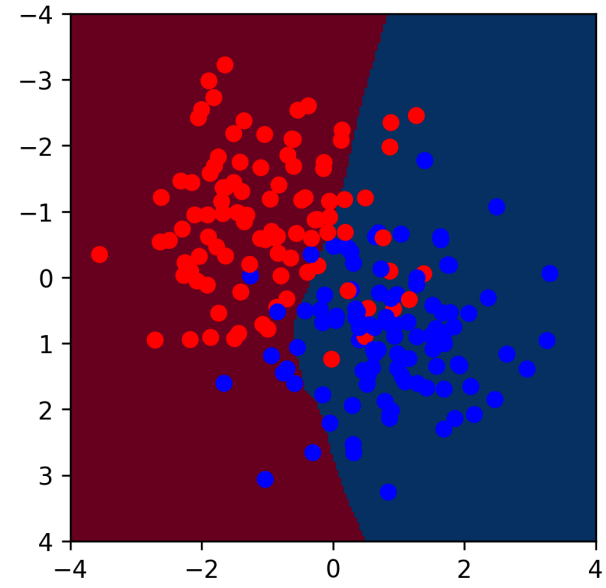


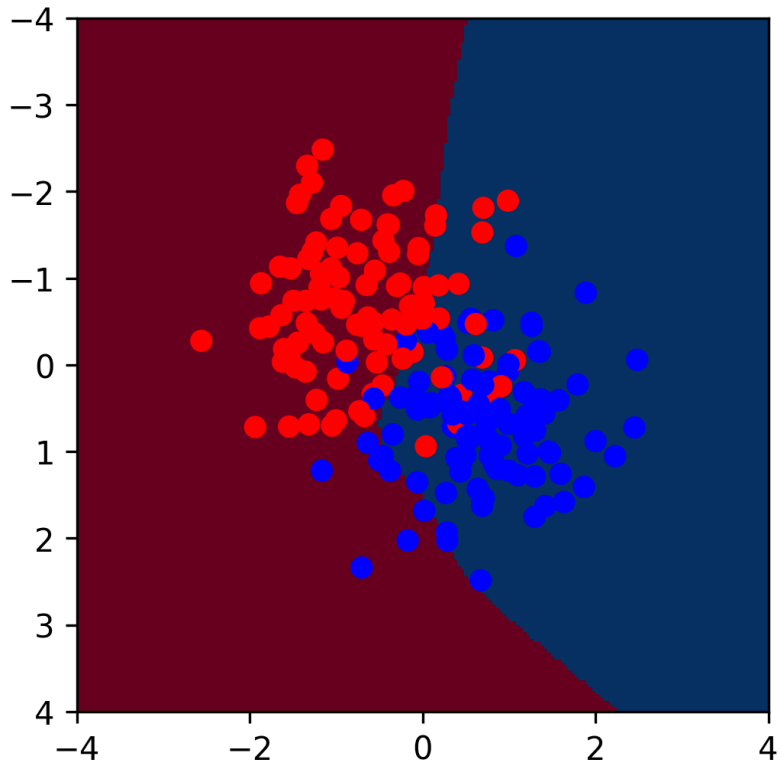


1H RMSE (no scaler): 0.37416573867739417
1H RMSE (with scaler): 0.36055512754639896

The RMSE barely changes between the scaled and unscaled version. Given that the actual data scaling method is quite nebulous to me, and therefore could produce unexpected and undesirable results when new datasets are applied, I would not use the scaler here.

Q1I)





1H RMSE (no scaler): 0.4183300132670378
1H RMSE (with scaler): 0.5

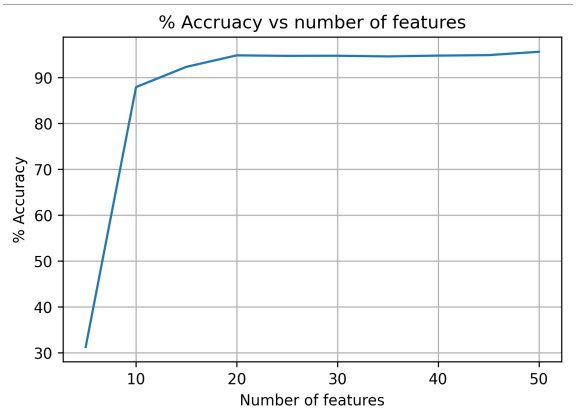
The scaling does reduce the RMSE, though perhaps artificially.
The scaler is observable to really just cluster the data more tightly, but about the origin, not about each cluster. An actually useful scaling would be scaling each dataset down about their own mean, to reduce overlap between the two datasets. Since these two datasets have lots of little 1-off random overlap, it is much harder for the model to find a linearly separable solution in high dimensional space.

2A)

Accuracy: 0.9638888888888889
Confusion matrix
[[27. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
[0. 34. 0. 0. 0. 0. 1. 0. 1. 2.]
[0. 0. 35. 0. 0. 0. 0. 0. 0. 0.]
[0. 0. 0. 29. 0. 1. 0. 0. 0. 1.]
[0. 0. 0. 0. 29. 0. 0. 1. 0. 0.]

[0. 0. 0. 0. 0. 39. 0. 0. 0. 0.]
[0. 0. 0. 0. 0. 0. 43. 0. 1. 0.]
[0. 0. 1. 0. 1. 0. 0. 38. 0. 1.]
[0. 1. 0. 0. 0. 0. 0. 0. 37. 1.]
[0. 0. 0. 0. 0. 0. 0. 0. 0. 36.]]

2B)



2C)

Accuracy:0.9638888888888889
Confusion matrix:
[[27. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
[0. 33. 0. 0. 0. 0. 1. 0. 2. 0.]
[0. 0. 33. 1. 0. 0. 0. 0. 1. 0.]
[0. 0. 2. 28. 0. 0. 0. 0. 0. 0.]
[0. 0. 0. 0. 30. 0. 0. 1. 0. 0.]
[0. 0. 0. 0. 0. 38. 0. 0. 0. 0.]
[0. 0. 0. 0. 0. 0. 43. 0. 0. 0.]
[0. 0. 0. 0. 0. 0. 0. 38. 0. 0.]
[0. 2. 1. 0. 0. 0. 0. 0. 36. 0.]
[0. 0. 0. 0. 0. 2. 0. 0. 0. 41.]]

Since there are no hidden layers, I would expect the MLP to just be a single perceptron. However, the network performs quite well for a single perceptron..

Since having an MLP with no hidden layers doesn't make a ton of sense, and since the accuracy is so high, I suspect that there may actually be some default number of layers going on in the initializer for the MLP.

2D)