EE559-Homework#8-Lei Lei

1. **Universal Function Approximation**

a)

How many parameters are in the hidden layer? How many total parameters in the ANN? Specify how each of the weights and biases in both layers should be chosen to obtain the approximation.

For the hidden layer, there would be 2(M-1) weights.

For the second layer, there would be M weights. So, for the ANN there would be 3M-2 parameters.

Text, letter

Description automatically generated

(b)

Chart, line chart

Description automatically generatedChart, line chart

Description automatically generated

Chart, line chart

Description automatically generatedChart, line chart

Description automatically generated

(c)

Chart

Description automatically generated Chart, line chart

Description automatically generated

(d) Interpretation of the NMSE results:

(i) Yes. This means that the approximation function is higher or lower than the original function. This is caused by an approximation function with a constant output regardless the input value.

(ii) No.

Chart, line chart

Description automatically generated Chart, line chart

Description automatically generated

This is because the approximation function fit the original function too well but there exits a bump in the original function. When the approximation function increases after training the left hand side points of the bump, the original function decreases immediately and increases the NMSE. We are using the mean squared error (MSE) and normalized mean squared error (NMSE).

(iii)

M = 39

Total = 3 \* M – 2 = 72 parameters.

(e)

Chart, line chart

Description automatically generated Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated Chart, histogram

Description automatically generated

(f)

Chart

Description automatically generated Chart

Description automatically generated

(g) It’s not. Because the two points in the middle are on the two sides of a bump. The left one should increase but it doesn’t because it’s lower than the initial value. And the right hand side point should decrease. However, it is lower than the last points, the derivative increased instead.

2. **Whitening and Simulation of Gaussian Random Vectors**

(a)

Chart, scatter chart

Description automatically generated

(b) (i)

Chart, scatter chart

Description automatically generated

(ii)

Chart, scatter chart

Description automatically generated

Text, letter

Description automatically generated

The sample covariance matrix is [[ 2.00033385 -5.99667272]

[-5.99667272 19.96204433]]

Comparing this to the covariance matrix of *y* as derived : [[2 -6],[-6 20]], we can found that they have very approximate value.

(iii)

The covariance of v is [[ 1.82310308e-01 -7.85113202e-03]

[-7.85113202e-03 2.17800679e+01]]

Chart, scatter chart

Description automatically generated

(iv)

The covariance of z is [[ 0.99435022 -0.00392557]

[-0.00392557 0.99832303]]

Chart, scatter chart

Description automatically generated

(v) What is the pdf of 𝑤 and the pdf of 𝑧? Is 𝑧 = 𝑤 -- i.e., since we colored 𝑤 to get 𝑦, then we whitened 𝑦 to get 𝑧? If not explain why.

The pdf of w is a Gaussian distribution and same as the pdf of z. As w and z have the same pdf, z = w because we colored w to get y, the mean of x equals to zero and the covariance of z is also an identity matrix.

3. **Comparison of PCA and MDA on wine dataset.**

(a) Baseline for comparison.

(i)

Chart, scatter chart

Description automatically generated Chart, scatter chart

Description automatically generated

(ii) Feature 1 and 2:Diagram

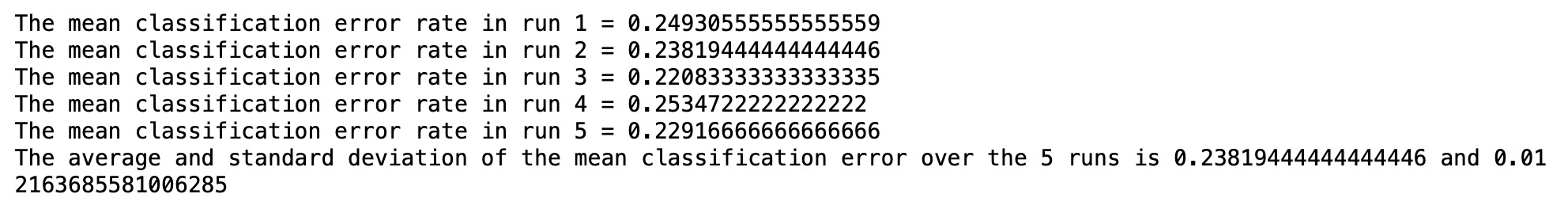
Description automatically generated with low confidence

Chart, scatter chart

Description automatically generated Chart

Description automatically generated

(iii)Feature 1 and 6:



Chart, scatter chart

Description automatically generated Chart, scatter chart

Description automatically generated

(b) PCA based on unnormalized dataset.

(i)

Chart, scatter chart

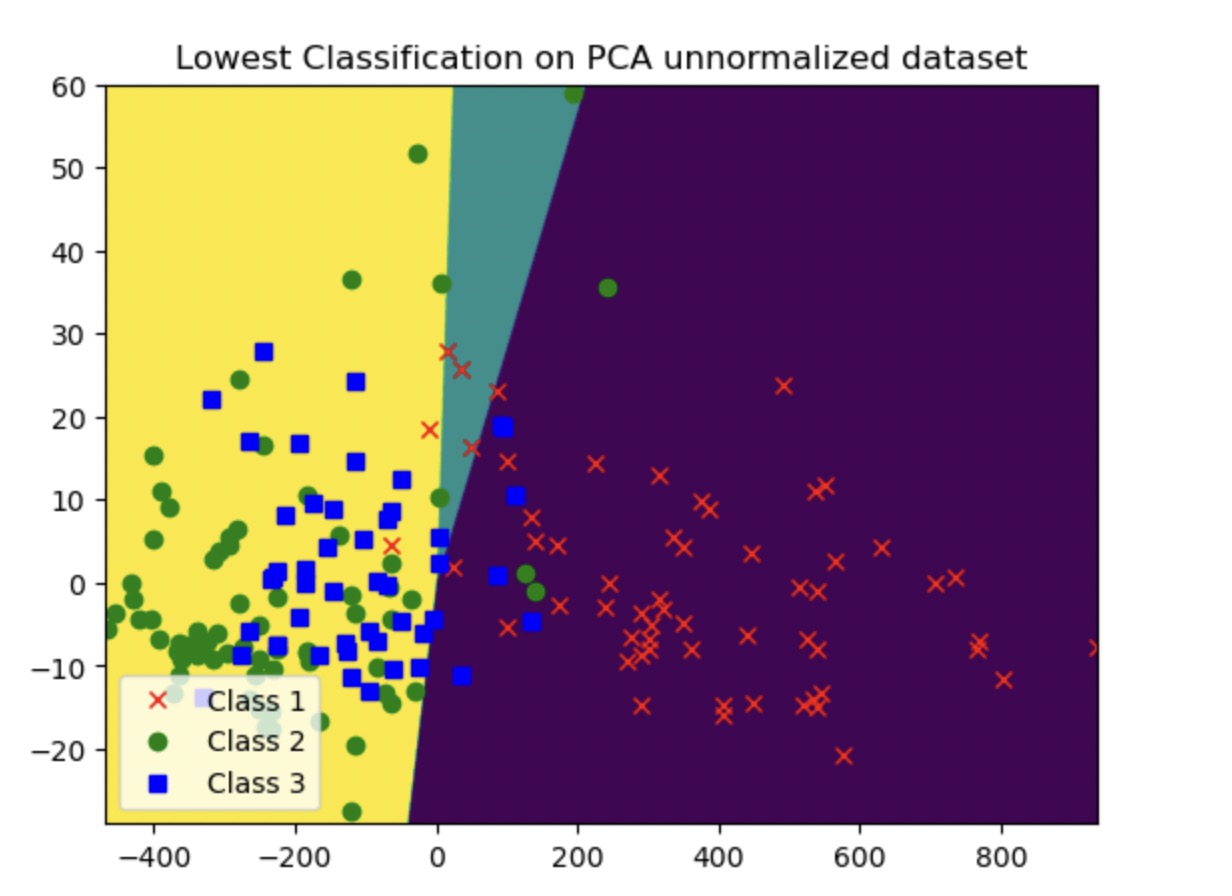
Description automatically generated

Class 2 and Class 3 almost overlap, I don’t think there would be a better classification result with PCA.

(ii)

Text

Description automatically generated

 Chart, scatter chart

Description automatically generated

(iii) How does it compare with the baselines in (a)(ii)?

The shape of decision regions seem to like the lowest error result from baselines.

(c) PCA based on standardized dataset.

(i)

Chart, scatter chart

Description automatically generated

It will have a better classification result.

(ii)

Text

Description automatically generated

Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generated

(iii)

The PCA on normalized data make a better performance on classification.

(d) MDA (using LDA as an approximation to MDA).

(i)

Chart, scatter chart

Description automatically generated

(ii)

Text

Description automatically generated

Chart, scatter chart

Description automatically generated Chart, scatter chart

Description automatically generated

(iii) According to the mean classification error, the LDA perfomance the best compare to baseline and PCA because the centers of each classes much farer after LDA.

4. **Mahalanobis distance and Bayes classification for minimum error**

(a)

Chart

Description automatically generated

(b) A close-up of a document

Description automatically generated with medium confidence

(c)

Chart

Description automatically generated

Chart

Description automatically generated

Chart

Description automatically generated