**MACHINE LEARNING FROM DATA**

**Report: Lab Session 2 – Feature selection – PCA and MDA**

**Names: Ricard**

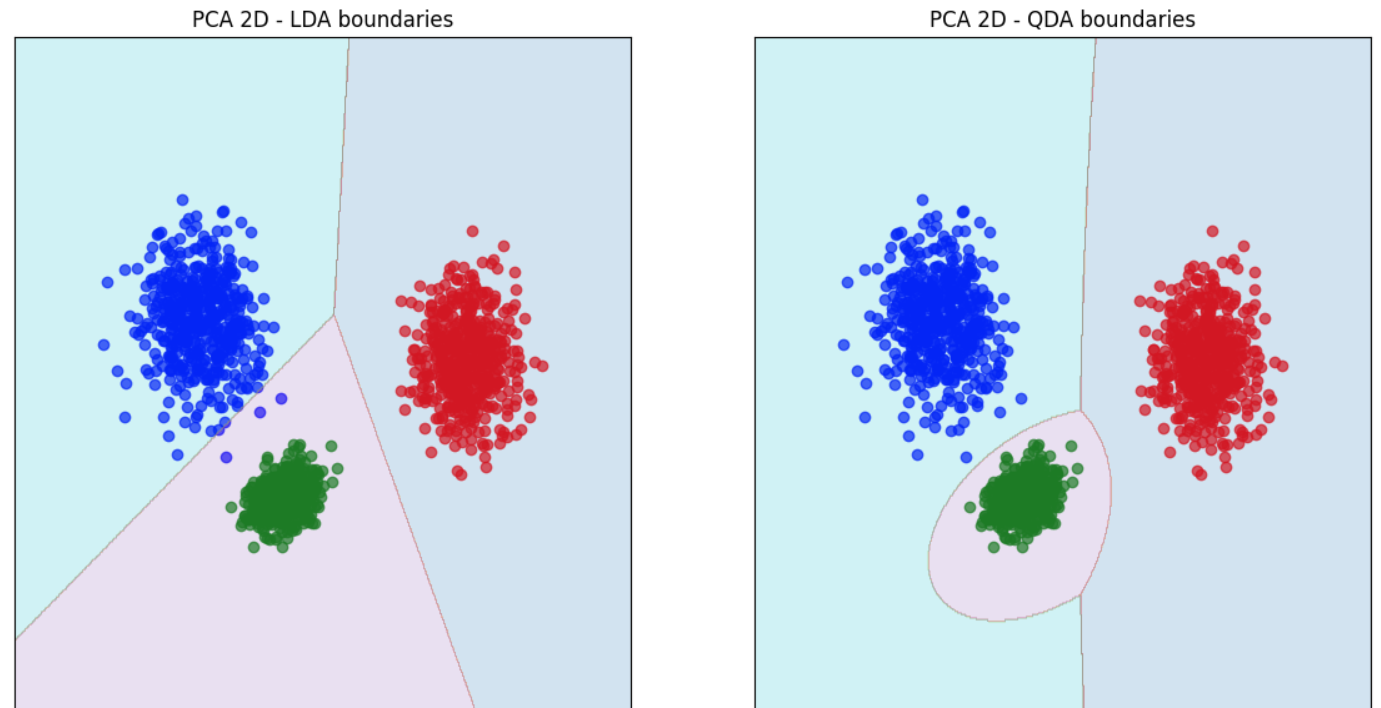
**Instructions**

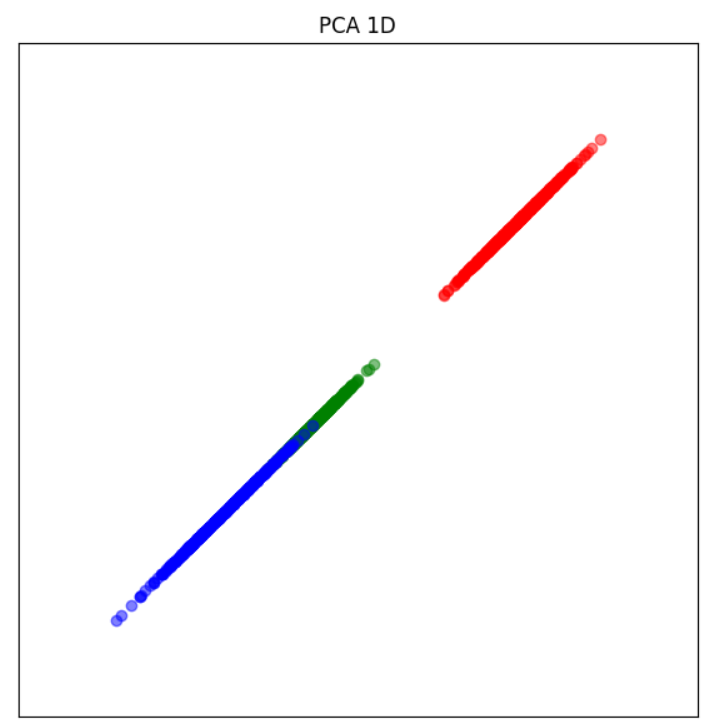
* Download and uncompress the file **Mlearn\_Lab2.zip**
* Answer the questions in the document **Mlearn\_Lab2\_report\_surnames.doc**

**Questions**

Q1: Complete the table with the training and test errors for the linear (LC) and the quadratic (QC) classifiers when using three, two and one feature, and SNR=10dB. In this case PCA is used for feature selection. Discuss the results. Analyze the scatter plots in two dimensions and in one dimension.

|  | 3 features | | 2 features | | 1 feature | |
| --- | --- | --- | --- | --- | --- | --- |
| Test | Train | Test | Train | Test | Train |
| LC | 0.000667 | 0.000000 | 0.002667 | 0.002000 | 0.039333 | 0.024667 |
| QC | 0.000000 | 0.000000 | 0.000667 | 0.000000 | 0.032000 | 0.022000 |





With 3 classifiers, we can see that for LC the test error is not 0, so we have a bit of error over there, but is really low, so we can say that the data is lineary separable.

For 2 features, we can se that the QC performs better than the LC this is suggesting that QC performs better because it can model non linear boundaries

The errors increase for both classifiers, with the LC having a test error of 0.039333 and the QC 0.032. This reduction in performance is likely due to significant information loss in the data when only one feature is used, making classification less accurate.

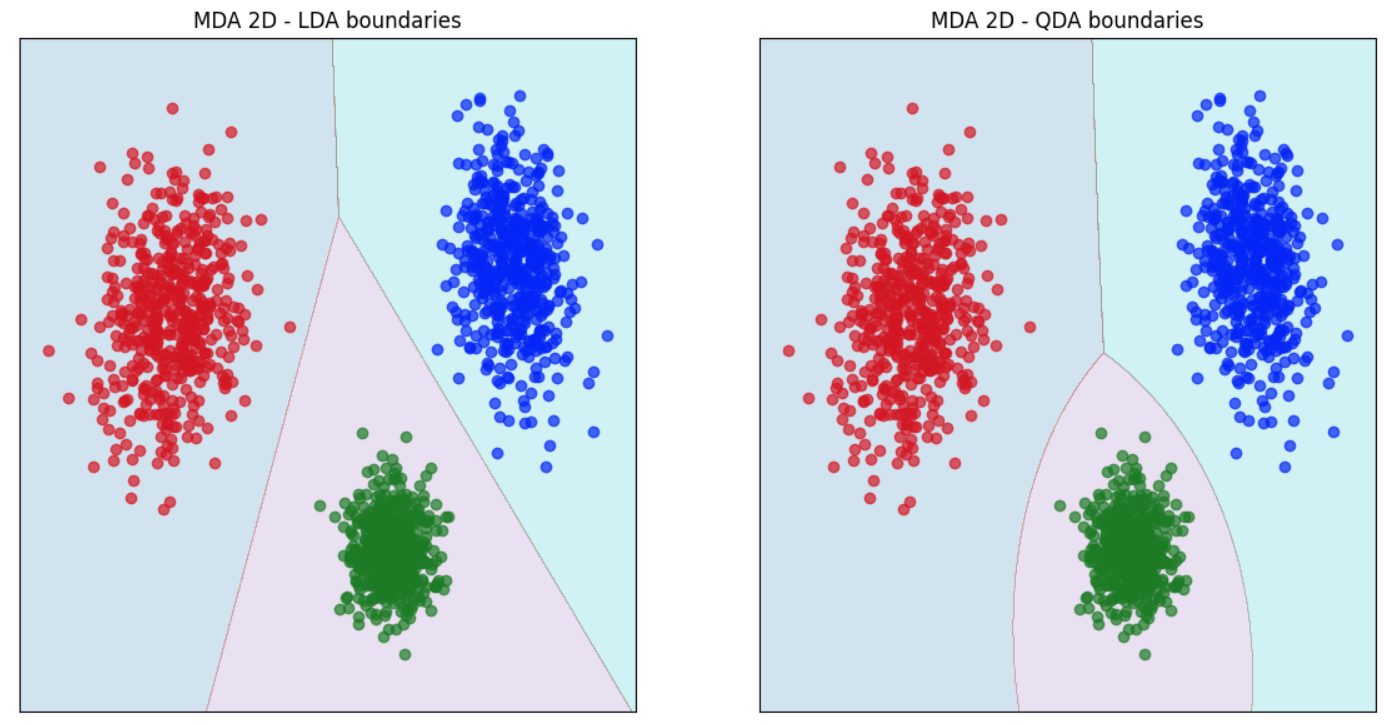
For the first scatter plot we can see that we have region overlapping, thats why we have a bit of error at the table above for 3 features.

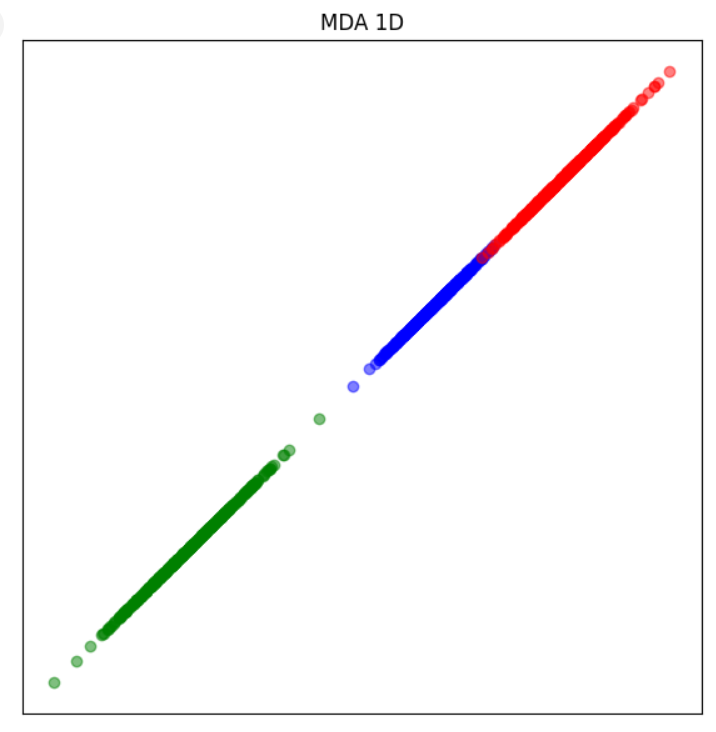
For the QDC scatter plot, be can see that the classifier dosen’t have any overlapping, so the errors will be 0

The data points are projected along a single direction, resulting in a linearly separable structure. However, this reduction simplifies the data significantly, which likely leads to an increase in classification error, as noted in the table.

Q2: Complete the table with the training and test errors for the linear (LC) and the quadratic (QC) classifiers when using three, two and one feature, and SNR=10dB. In this case MDA is used for feature selection. Discuss the results. Analyse the scatter plots in two dimensions and in one dimension.

|  | 3 features | | 2 features | | 1 feature | |
| --- | --- | --- | --- | --- | --- | --- |
| Test | Train | Test | Train | Test | Train |
| LC | 0.000667 | 0.000000 | 0.000667 | 0.000000 | 0.011333 | 0.004000 |
| QC | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.006667 | 0.003333 |





Again, it does not make sense to reduce to 3 features, as it is already the amount of features we have at the start. The errors for both classifiers are the same when using PCA and MDA, because we are not reducing the features.

With the current noise, which is quite low, we can see that the data is quite concentrated or densely clustered.

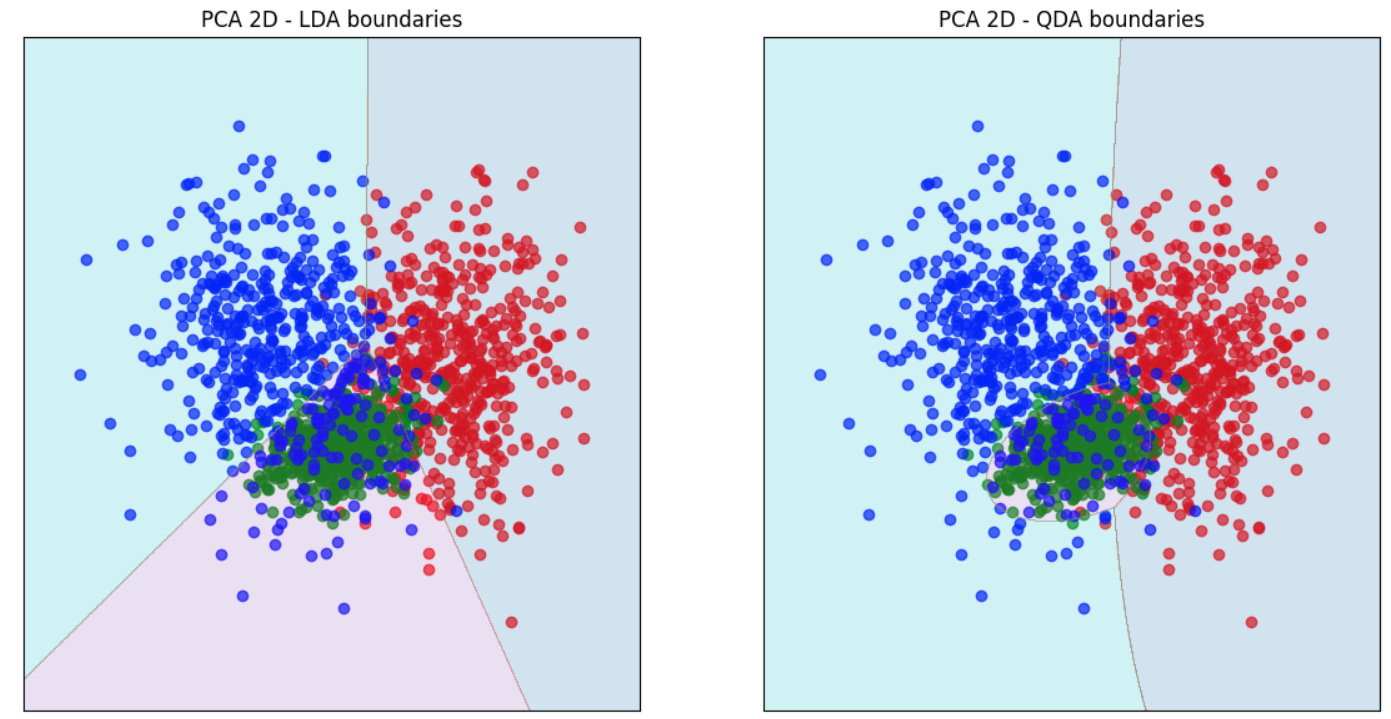
When reducing dimensionality to 2 features with MDA we see that both classifiers, LC and QC do a very good job classifying the dataset. The QC is a bit better than the LC. IN fact, QC es a perfect classifier in this case.

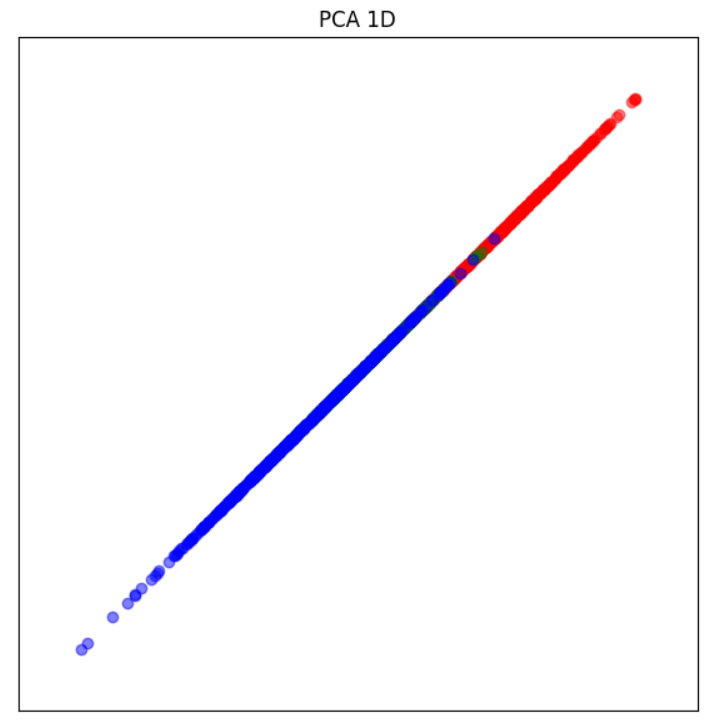
Finally we see that even when reducing to 1 feature there is still a good class separability because of the low noise. Consequently, the error is still very low. The QC outperformed the LC.

When comparing MDA and PCA, we see that in all cases, when using MDA, the projections to 2 and 1 dimensions result in a better separability (which implies better classifiers).

Q3: Use PCA for feature selection. Complete the table with the training and test errors for the linear (LC) and the quadratic (QC) classifiers when using three, two and one feature, and SNR= 0 dB. Discuss the results. Analyse the scatter plots in two dimensions and in one dimension.

|  | 3 features | | 2 features | | 1 feature | |
| --- | --- | --- | --- | --- | --- | --- |
| Test | Train | Test | Train | Test | Train |
| LC | 0.133333 | 0.115333 | 0.175333 | 0.147333 | 0.244667 | 0.250667 |
| QC | 0.062000 | 0.070667 | 0.144000 | 0.122000 | 0.239333 | 0.244667 |





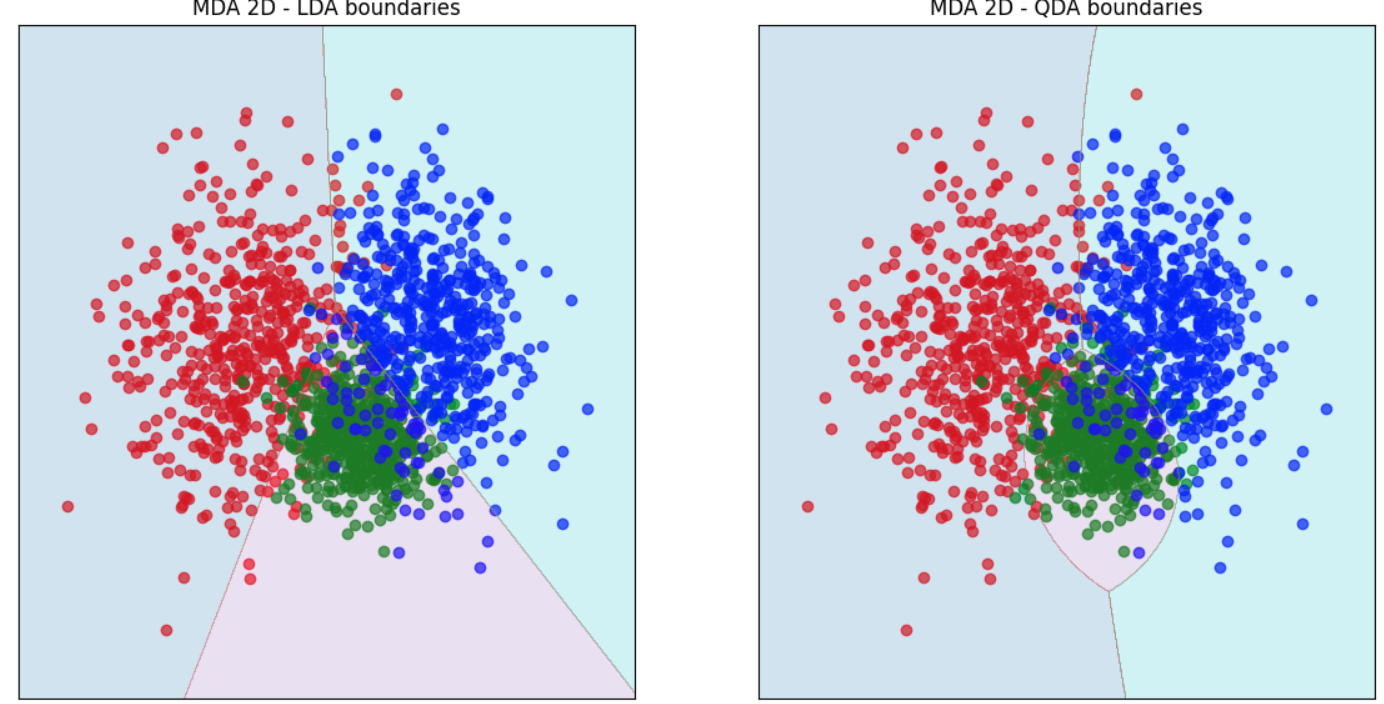
With this amount of noise, we know that the classes will be highly mixed. With 3 features the LC is not very good as it has a 13% error, the QC is a bit better with a 6% error.

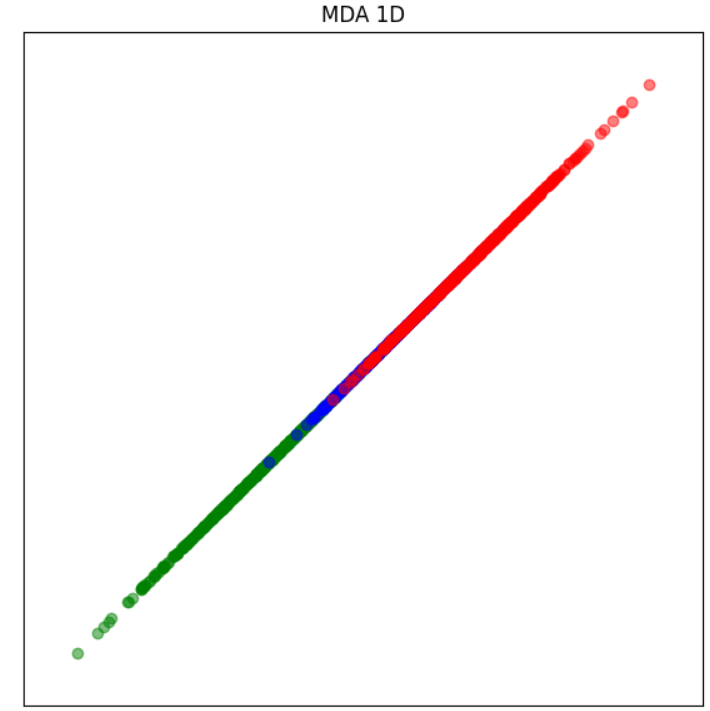
When projecting to 2 features or dimensions the high noise causes a very low separability between classes. In this case, the two classifiers are not good with an error of 17% and 14%, with the QC slightly outperforming the LC.

Finally when projecting to 1 feature we can see a big overlap between classes, resulting in both classifiers performing really bad.

Q4: Use MDA for feature selection. Complete the table with the training and test errors for the linear (LC) and the quadratic (QC) classifiers when using three, two and one feature, and SNR= 0 dB. Discuss the results. Analyse the scatter plots in two dimensions and in one dimension.

|  | 3 features | | 2 features | | 1 feature | |
| --- | --- | --- | --- | --- | --- | --- |
| Test | Train | Test | Train | Test | Train |
| LC | 0.133333 | 0.115333 | 0.133333 | 0.115333 | 0.220667 | 0.211333 |
| QC | 0.062000 | 0.070667 | 0.125333 | 0.110667 | 0.215333 | 0.208000 |





With this amount of noise, we know that the classes will be highly mixed. With 3 features the LC is not very good as it has a 13% error, the QC is a bit better with a 6% error.

When projecting to 2 features or dimensions the high noise causes a very low separability between classes. In this case, the two classifiers are not good with an error of 17% and 14%, with the QC slightly outperforming the LC.

Finally when projecting to 1 feature we can see a big overlap between classes, resulting in both classifiers performing really bad.

When comparing MDA and PCA, we see that in all cases, when using MDA, the projections to 2 and 1 dimensions result in a better separability (which implies better classifiers).

Q5. Include the plots of the phoneme spectra.

Q6. Include the error probabilities for the training and test sets obtained with the linear classifier (LC) and the quadratic classifier (QC), using all the features. Discuss the results.

Q7. Include the confusion matrices for the test set obtained with the linear classifier (LC) and the quadratic classifier (QC), using all the features. Discuss the results.

Q8. Which features would you choose? Show the error probabilities for the training and test sets obtained with the linear and the quadratic classifier. Compare with the previous case (using all features) and discuss the results.

Q9. Include the scatter plot and decision boundaries obtained. Discuss the results.

Q10. Show the error curves for the linear and the quadratic classifier on the training and on the test set.

Q11. Discuss which dimension is the most adequate for the linear classifier and which is the best one for the quadratic classifier. Remember that it is important not to overfit on the training data (the test error should not be much larger than the training error).

Q12. Which is the maximum number of features *dmax*? Show the error curves for the linear and the quadratic classifier on the training and on the test set.

Q13. Compare results and discuss the use of PCA and MDA for the Phoneme dataset using between 1 and *dmax* features