**MACHINE LEARNING FROM DATA**

**Report: Lab Session 3 – Feature selection: PCA and MDA**

**Names:Abdullah Qureshi, S M Rakib Hasan and Niina Hietamäki**

**Instructions**

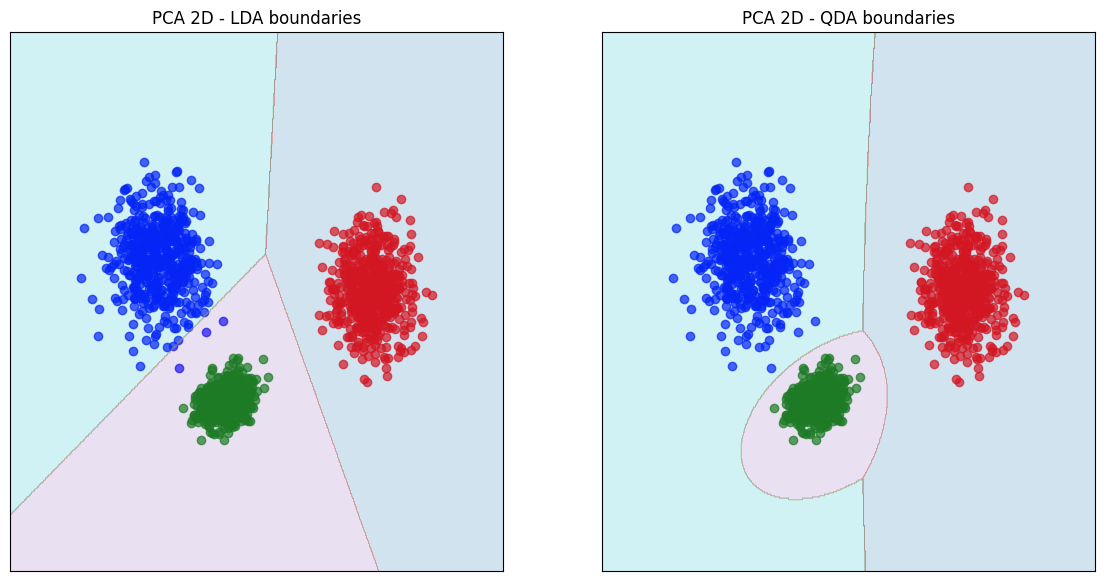
* Answer the questions in the document **Mlearn\_Lab3\_report\_surname.doc**
* Provide complete and concise answers, **maximum 5 pages**.
* Save the report, convert to pdf
* Write the new code in a Colab Notebook **Mlearn\_lab3\_3\_Hasan\_Qureshi\_Hietamäki.ipynb**.
* Zip and upload to Atenea the pdf report and the notebook in **a single file**.

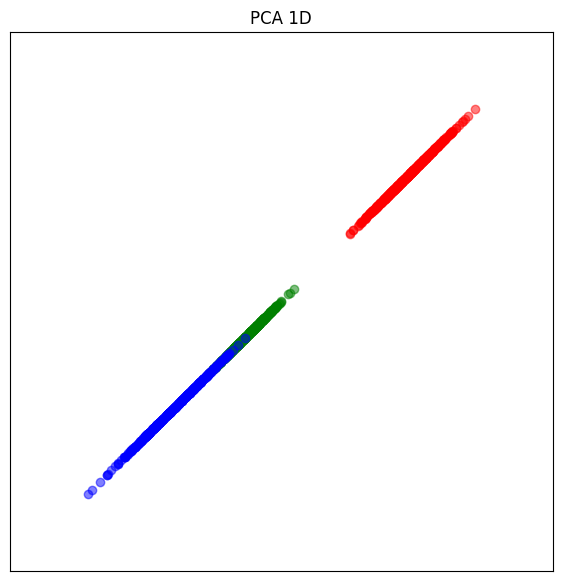
**Questions**

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 |

The Linear Classifier (LC) works well with three features but experiences a small increase in error when reduced to two features, indicating some loss of information as dimensionality decreases. The Quadratic Classifier (QC) performs well with both three and two features, as it can handle more complex, non-linear boundaries effectively.





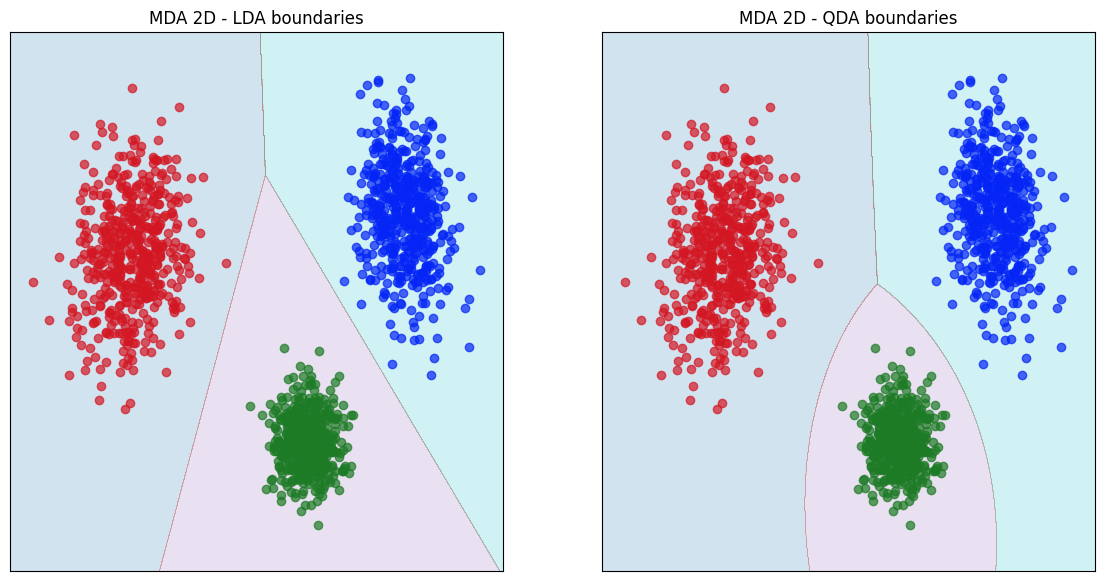
Scatter Plots Analysis:

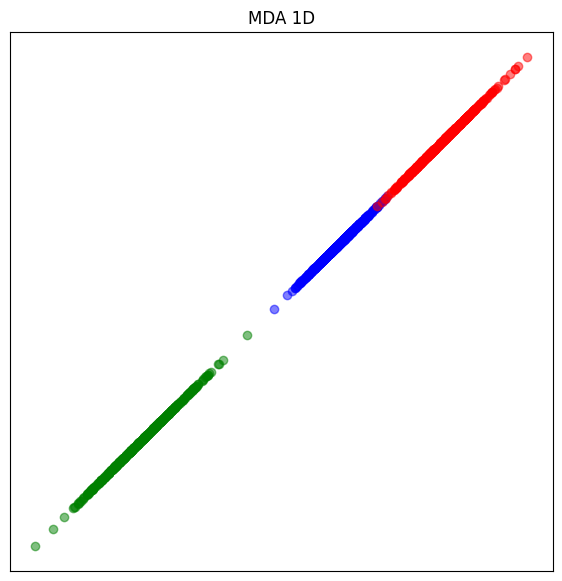
In two dimensions, the LDA plot shows straight boundaries between the classes, while QDA has curved boundaries that separate the classes more precisely. In one dimension, the separation between classes is expected to be less clear because reducing to one feature results in information loss, making class boundaries overlap more.

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 |

The MDA results show that the Linear Classifier (LC) performs well with three and two features, but the error increases significantly with one feature, indicating information loss. The Quadratic Classifier (QC) performs perfectly with three and two features and still outperforms LC with one feature, despite a slight increase in error.





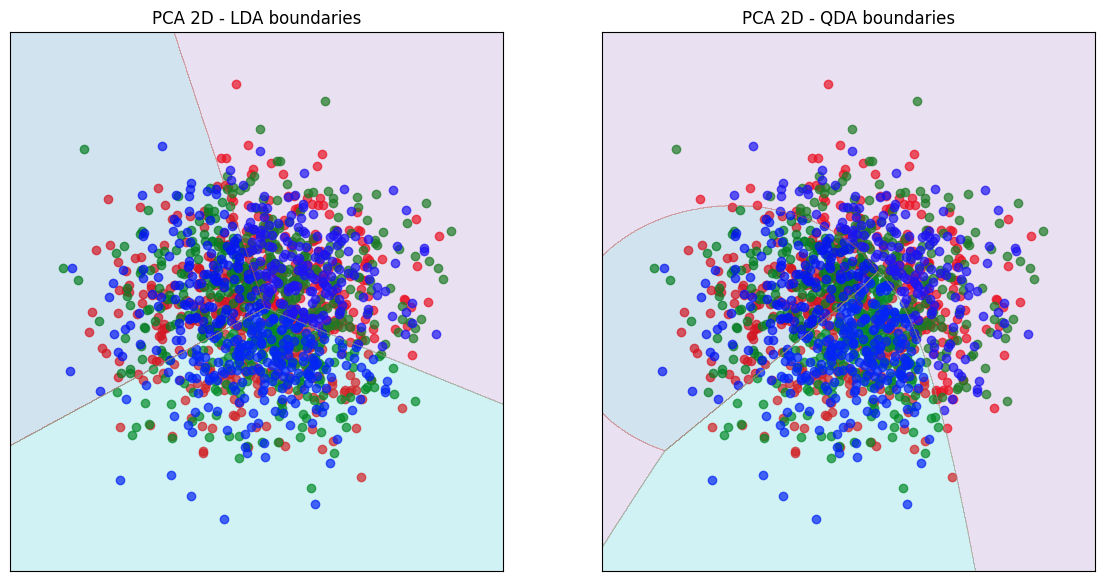
Scatter Plots Analysis:

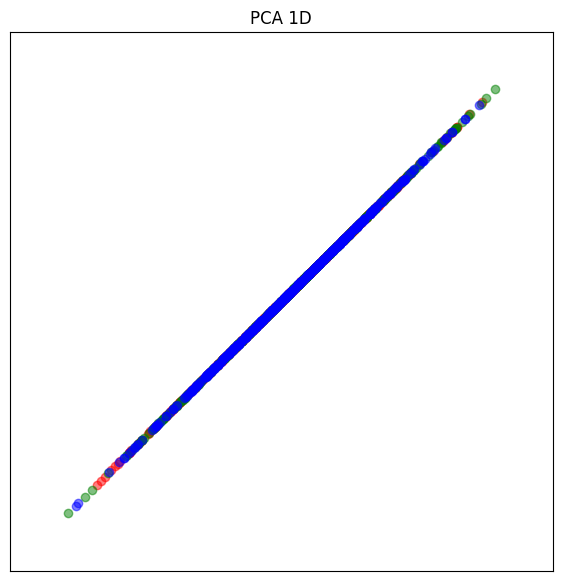
In two dimensions, MDA provides clear class separation, with LDA showing linear boundaries and QDA using curved ones. However, in one dimension, the separation is less clear due to information loss, leading to overlaps, especially for LC.

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 | 0.658667 | 0.640000 | 0.658667 | 0.640000 | 0.657333 | 0.653333 |
| QC | 0.664667 | 0.635333 | 0.664667 | 0.635333 | 0.658667 | 0.658000 |

LC and QC show a similar trend across 3, 2, and 1 features. The errors remain high due to the low SNR (0 dB), indicating a lot of noise in the data, which makes it difficult for both classifiers to distinguish the classes.





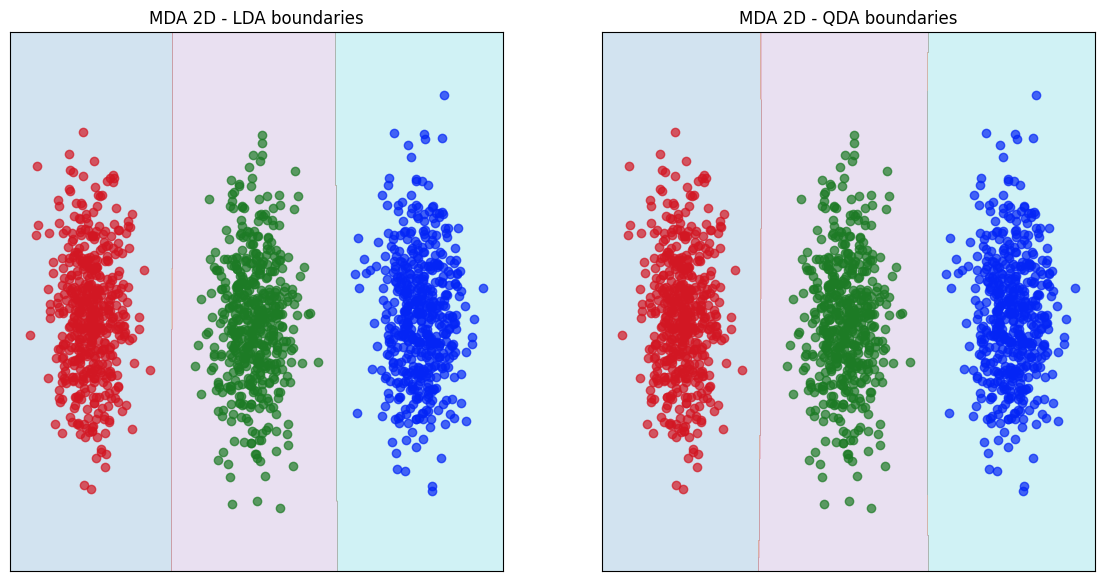
Scatter Plots Analysis

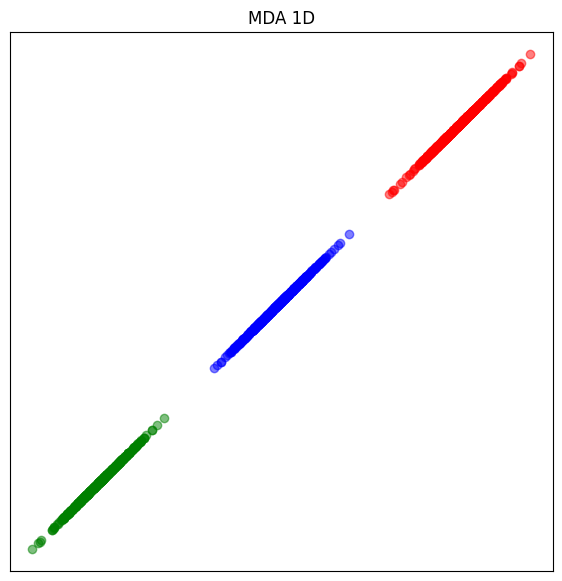
2D PCA plots for both LDA and QDA show overlapping classes, which explains the high classification error. The decision boundaries in both LDA and QDA are unable to clearly separate the classes due to the noise. 1D PCA plot further shows that class separability is minimal, and both LC and QC struggle to classify correctly with only one feature.

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 | 0 | 0 | 0 | 0 | 0 |
| QC | 0 | 0 | 0 | 0 | 0 | 0 |

Both LC and QC show perfect performance across all feature sets with zero test and train errors. This indicates that with MDA feature selection and the given dataset, both classifiers can perfectly distinguish the classes, even with high noise (SNR = 0 dB).



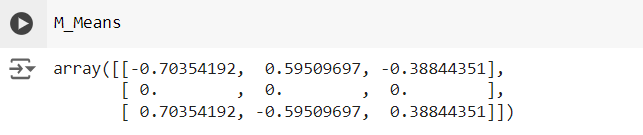


Scatter Plots Analysis:

2D Scatter Plots: Both LDA and QDA boundaries are perfectly aligned with the class separations, resulting in clear and distinct boundaries between the red, green, and blue classes.1D Scatter Plot: Even in one dimension, MDA manages to separate the classes perfectly, as the classes remain distinct along the single axis. This shows that MDA is highly effective in maintaining class separability even in noisy environments, as evidenced by the perfect classification performance.

Q5. Find and write the three vectors corresponding to the class means. Give also the value of the seed used in your experiments (if you changed it). How many features can we use with MDA?

Analysis for PCA and MDA and SNR = -5 dB. Used default seed corresponding to value of 5.



For MDA in sci-kit learn the number of components <= min(n\_classes - 1, n\_features) for dimensionality reduction. n\_components cannot be larger than min(n\_features, n\_classes - 1)

Q6. Complete a 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= -5 dB. Use PCA and MDA for feature selection. Discuss the results. In which cases is MDA clearly better than PCA?

Using PCA

|  | 3 features | | 2 features | | 1 feature | |
| --- | --- | --- | --- | --- | --- | --- |
| Test | Train | Test | Train | Test | Train |
| LC | 0.00 | 0.00 | 0.658 | 0.640 | 0.657 | 0.653 |
| QC | 0.00 | 0.00 | 0.664 | 0.635 | 0.658 | 0.658 |

Using MDA

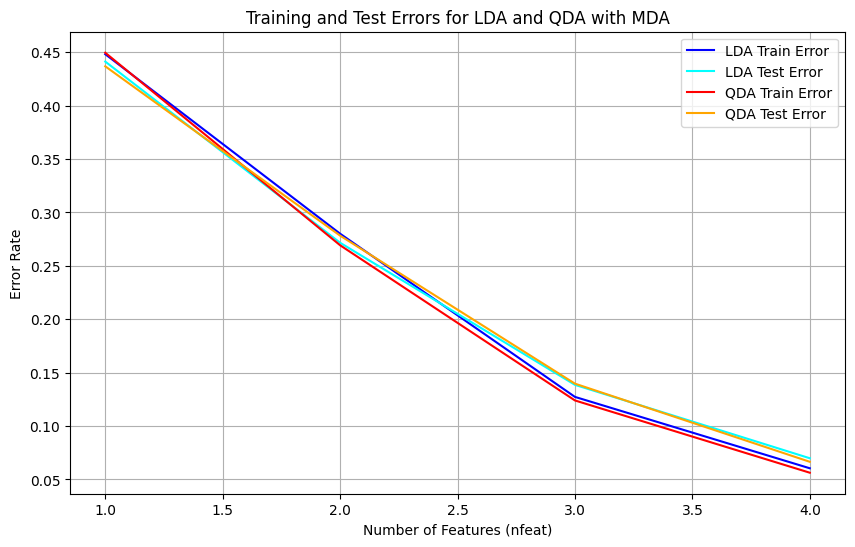
|  | 3 features | | 2 features | | 1 feature | |
| --- | --- | --- | --- | --- | --- | --- |
| Test | Train | Test | Train | Test | Train |
| LC | - | - | 0.00 | 0.00 | 0.00 | 0.00 |
| QC | - | - | 0.00 | 0.00 | 0.00 | 0.00 |

The results show that MDA clearly outperforms PCA when the number of features is reduced, especially in a noisy environment (SNR = -5 dB). While both PCA and MDA perform perfectly with 3 features, PCA's performance significantly degrades when reduced to 2 or 1 feature, with test errors rising to around 0.66 for both linear (LC) and quadratic classifiers (QC). In contrast, MDA maintains 0 test and training errors for both classifiers with 2 and 1 feature indicating that MDA is more effective at selecting discriminative features and preserving class separability under noisy conditions.

Q7. 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.

Answer: The maximum features from MDA are given by the maximum number of unique classes in the label. It is 1 less than the total unique features. As we have 5 unique classes, After MDA, 4 features will be selected.

Error curves for the linear and the quadratic classifier on the training and on the test set with 80/20 split:



Q8. Compare results and discuss the use of PCA and MDA for the Phoneme dataset

Best results from PCA (at the most optimum features selected) and MDA (with 4 features) are tabulated below:

|  | LDA Train | LDA Test | QDA Train | QDA Test |
| --- | --- | --- | --- | --- |
| PCA | 0.0798 | 0.0732 | 0.0738 | 0.0797 |
| MDA | 0.0604 | 0.0698 | 0.0562 | 0.0665 |

From the comparison table, we can see that for both classifiers, **MDA** balances the training and test performance more effectively than PCA.

LDA Performance:

* PCA shows a higher training error (0.0798) compared to MDA (0.0604), indicating that MDA is better at capturing the class-separating structure in the training data.
* Test errors for both methods are similar, with PCA slightly better (0.0732 vs. 0.0698). However, MDA maintains a lower training error, suggesting it balances class separability more effectively.

QDA Performance:

* MDA performs significantly better than PCA in both training (0.0562 vs. 0.0738) and test errors (0.0665 vs. 0.0797), indicating that MDA’s class-based feature selection benefits the quadratic decision boundaries of QDA.
* PCA, while useful for variance maximization, doesn’t capture the class separability as well, which is crucial for QDA’s non-linear boundaries.