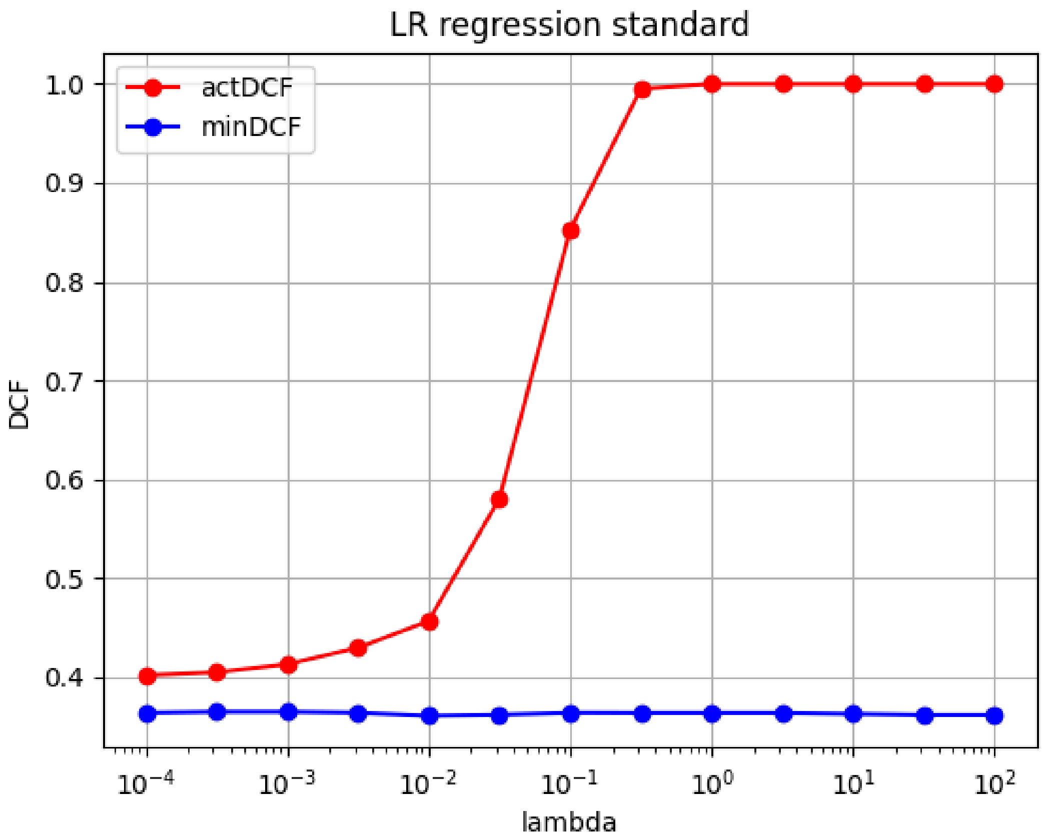
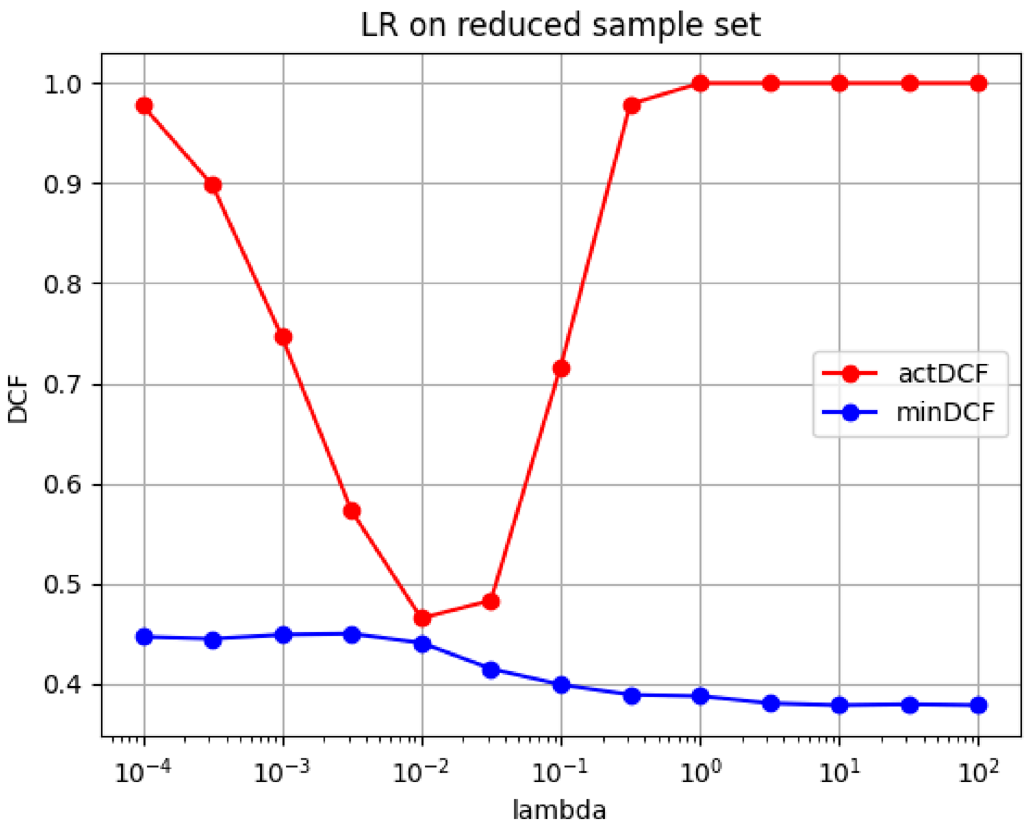
LAB08 REPORT

Let's now focus our attention on the logistic regression classifier. The standard version of this classifier aims at reducing the logistic lossJ(W,b), trying all possible combinations of W and b. We encounter a problem, if the classes are linearly separable (there is a hyperplane that correctly separates the two classes) we could increase the norm of W and the logistic loss function would keep improving, J(W,b) has an infimum at W->∞.
To avoid this issue we introduce a penalty in the form of regularization term, which contains the squared norm of $||W^2||$ multiplicated by a λ and divided by 2. Notice that λ allows us to associate a weight to the regularization term. If λ is too large, we may incur in a solution with a small norm, but not able to well separate the classes, while if λ is too small, we may overfit the training set.



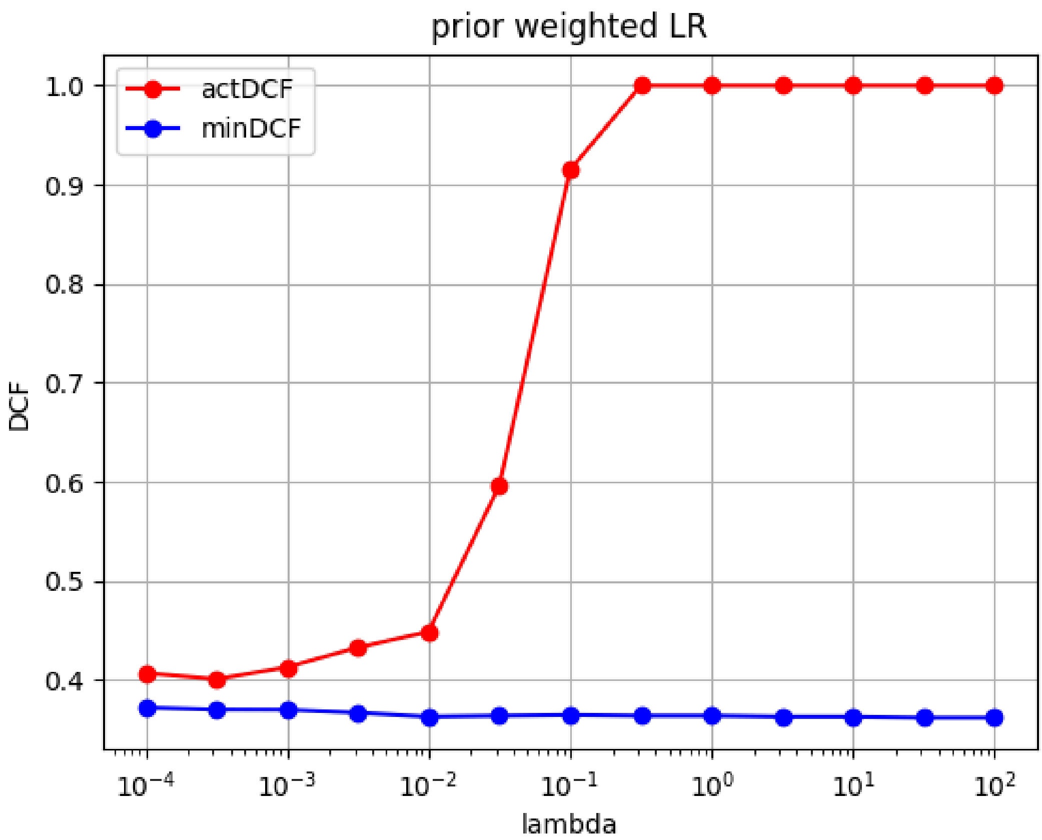
As we see from here, as the λ increase, the DCF goes up by a lot and very quickly, this is probably because the separation hyperplane is heavily influenced by the norm of W, so the LR is not able to separate well the classes, on the other hand we obtain better result with a small λ, although a ~0.4 DCF is not the best.

Now we analyze LR with the reduced dataset.



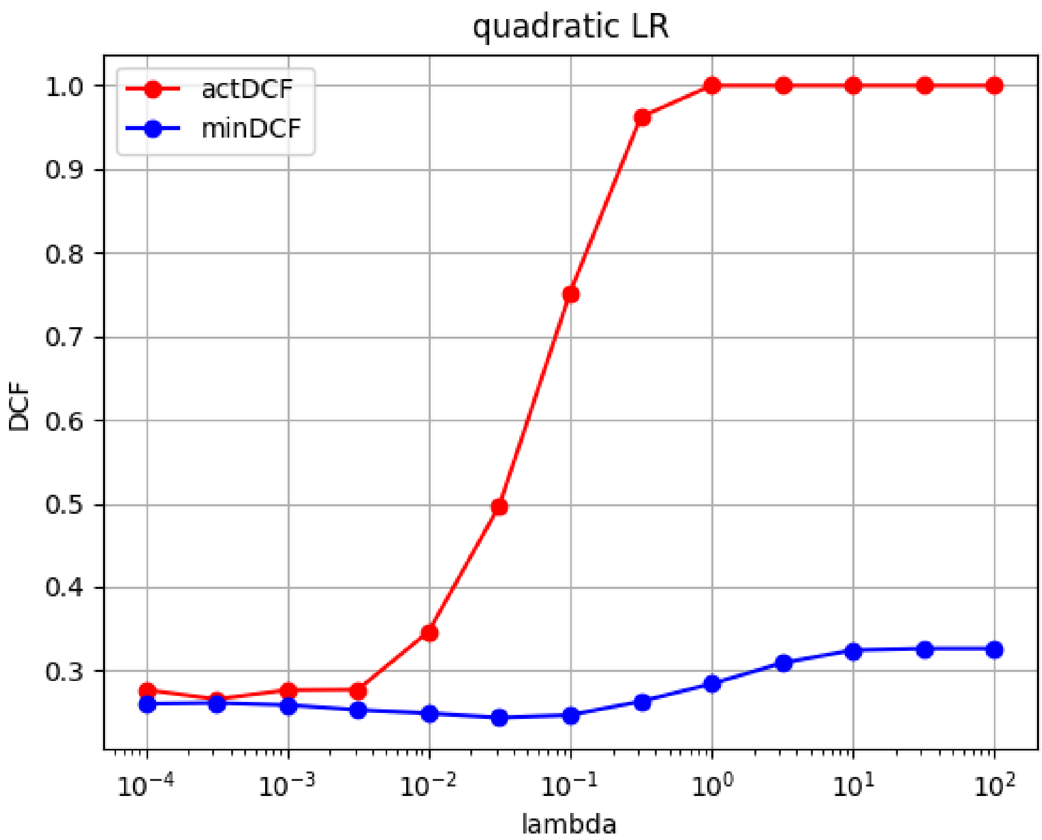
Reducing the dataset to 1/50th of the original gives us a better minDCF with greater λ values, it goes almost to 0.2. We start seeing an optimal range of values for λ, where the actDCF is near the minDCF. Due to the reduced dataset overfitting and underfitting are more severe.

Now we analyze LR with prior-weighted dataset



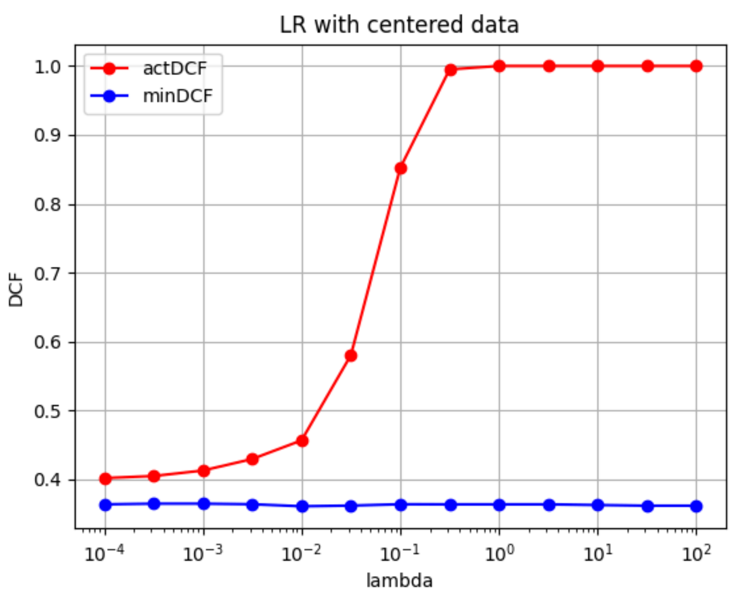
The prior weighted LR model does not give improvements, with respect to the standard model. The only advantage we could get from prior weighted LR is that the classifier is optimized for the application prior instead of the training prior.

Now we analyze quadratic LR



The quadratic LR model does not differ too much from the standard one. It has slightly better minDCF and actDCF, this is because the feature expansion can help with the classification, but the effect of the regularization term remains the same, it is good at lower values, but the underfitting, which comes with greater values of λ is more severe.

Now we analyze a preprocessing technique: data centering

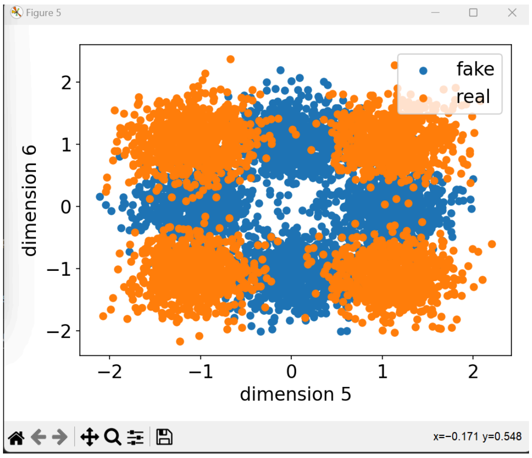


Centering the data can help reduce the computational cost of the model training and give us solutions with smaller W. In this case it doesn't seem to give significant improvement. Another approach could be Z normalization, other than subtracting the mean we could divide the samples by the standard deviance. This could unrelate the sample from their unit of measurement and increment the classification efficiency.

Final comments about the minDCF of the various models

Overall, the LR quadratic model gives us the best minDCF for various kinds of applications, there is still a significant score miss calibration on big values of λ but this can be avoided with calibration. This model is characterized by curved separation rules, and this can help with making better decisions. Another good result was for the standard MVG, also characterized by curve separation rules.

The LR still gives us better overall minDCF rather than MVG approaches.



If we take a look at the scatter plots for the last 2 features we can see why a quadraticrules classifier is better than a linear one.