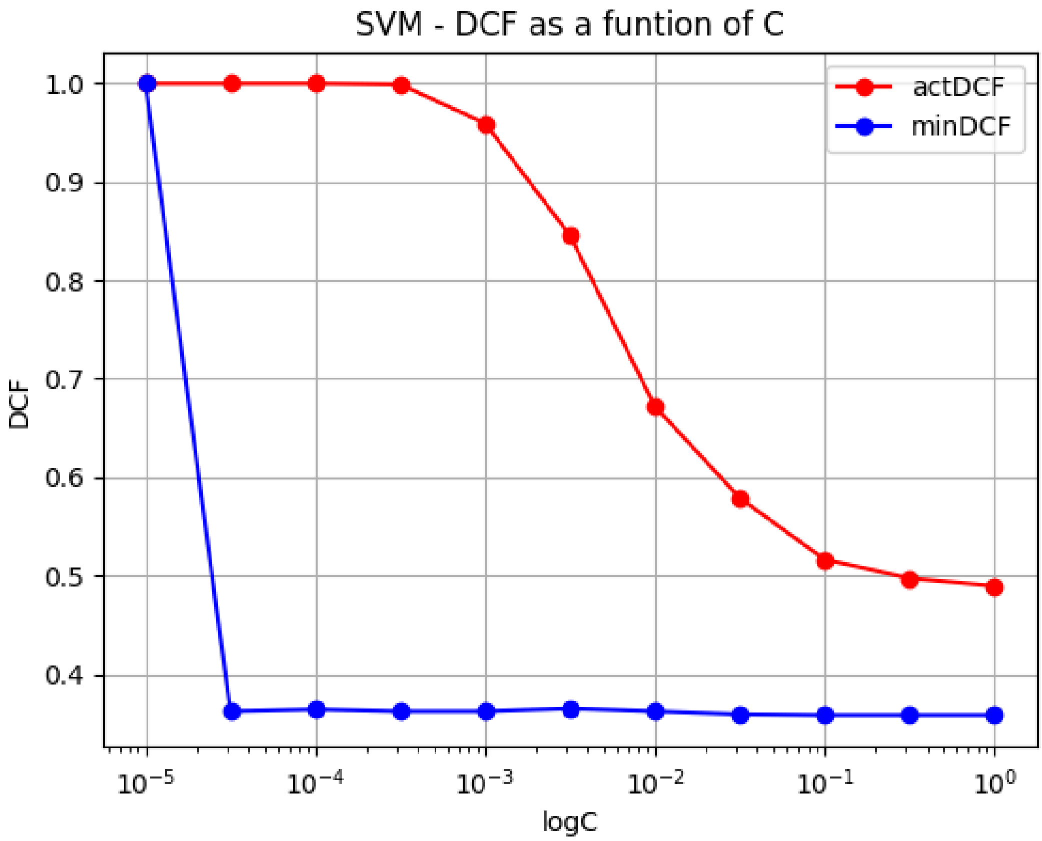
LAB09 REPORT

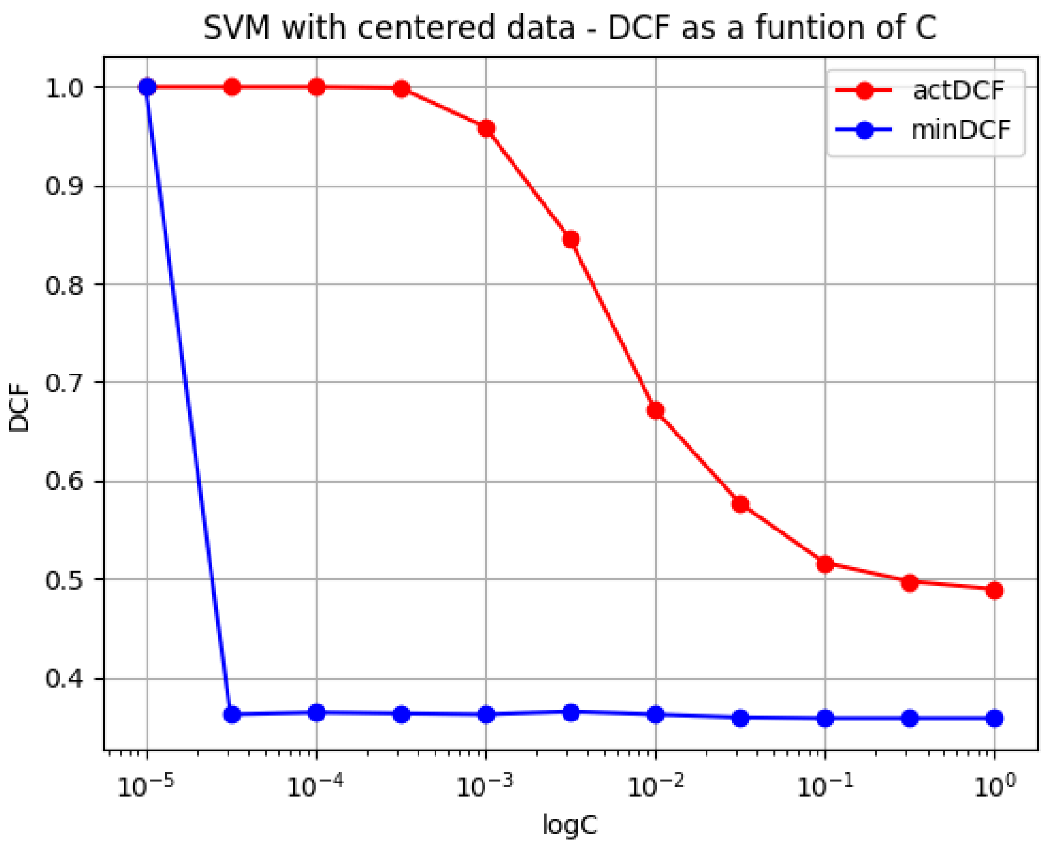
In this laboratory we need to know how the SVM performs on our data set. First of all, we try the linearSVM, where C (that indicates the weight of a misclassification) is varied.

LINEAR SVM



The linear model does not behave well, first of all the actDCF is very high, especially with strong normalization (low values of C). This is because the SVM has a geometric approach to the classification problem and tends to lose the probabilistic interpretation. Strong regularization greatly affects this. The model has a good minDCF, so that means with some calibration on the scores we should be able to obtain a good classification. This is because the calibration gives back to the llr their probabilistic interpretation.

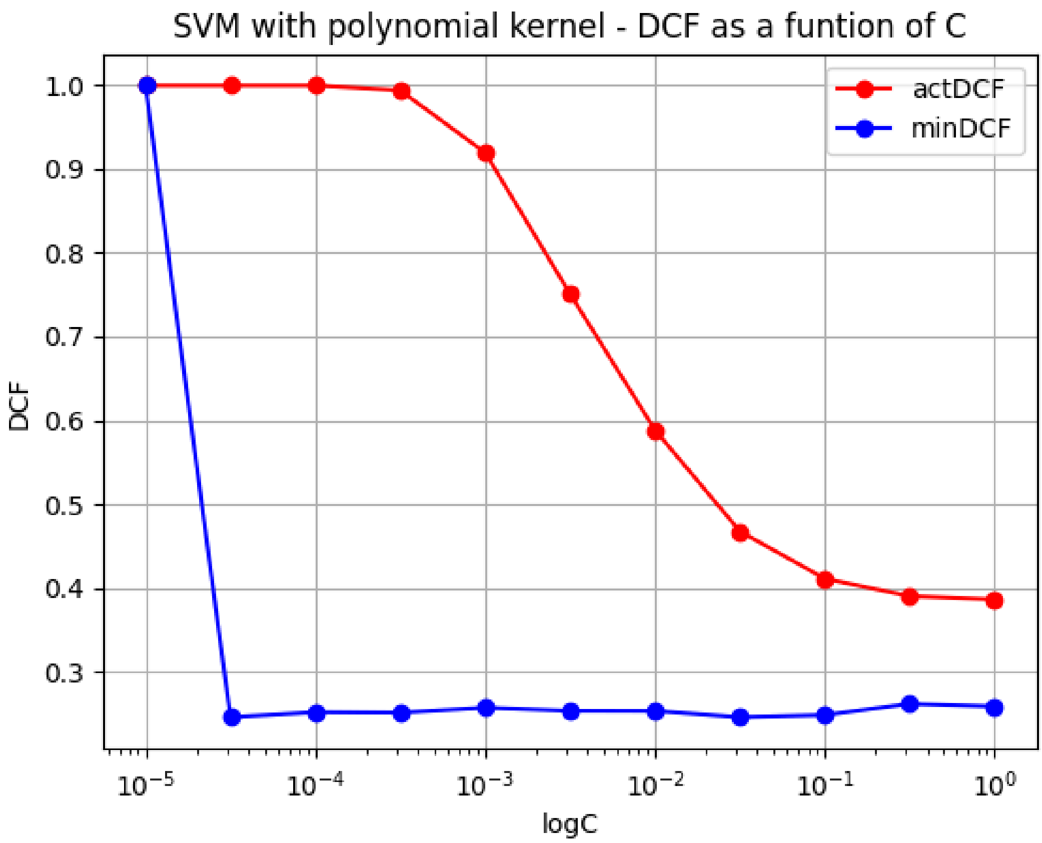
LINEAR SVM WITH CENTERED DATA



Similar to the logistic regression, centering the data does not improve the classification DCF. This is especially because the centered data is only translated, so the geometric interpretation the SVM gives to the llr does not change.

SVM WITH POLYNOMIAL KERNEL

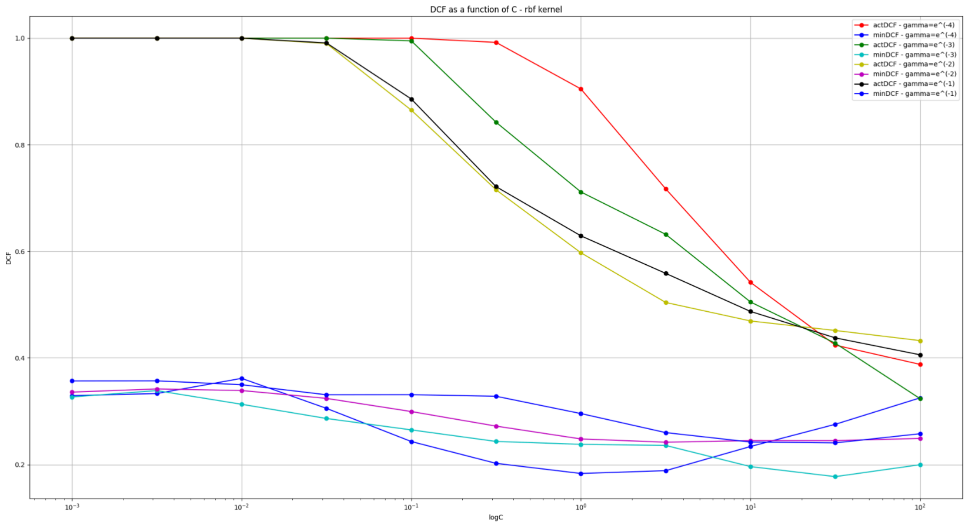
We have seen that from the solution of the dual SVM formula we have to do the dot product of the transposed sample set with the sample set. if we are able to identify a function K that efficiently does this we don't need to explicitly expand the features space and we can use that instead. This is called the polynomial kernel trick and we first try with a polynomial kernel function K. this of course can speed up computations



The results are still quite inaccurate, lower than actDCF but the model is still not calibrated well enough: The loss of the probabilistic interpretation still has a big impact on the classifications. Since now we have a quadratic separation rule, the minDCF is very similar to the LR and MVG models, even lower for certain applications.

SVM WITH RBF KERNEL

Now we try a different mapping function, the rbf (Gaussian Radial Basis Function) kernel. We can use cross validation to compute the best gamma value, this terms represent how much the points influence each other. small gamma means small influence, whereas bigger gamma values mean more influence.



As we can see, the best results have a small regularization (bigger C) and we obtain the best overall solution from bigger gamma values. From the histogram we plotted from lab 2 we could see clusters forming from the two classes so probablythe rbf approach works well on the dataset