1. **Multivariate Gaussian Classifiers**

The first models analyzed are the generative Gaussian classifiers:

-MVG (Full-Covariance)

-NBG (Diagonal-Covariance)

-TCG (Tied Full-Covariance)

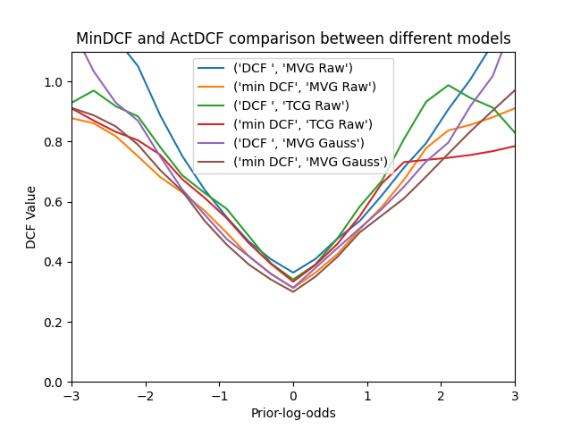
-TCNB (Tied Diagonal-Covariance).

We start doing a selection of the best models in terms of minimum DCF, which is the cost we would pay if we made optimal decision on the test set (validation set) using recognizer scores.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| **Raw Features** | | | |
| Full-Cov | **0.312** | 0.778 | 0.842 |
| Diag-Cov | 0.420 | 0.845 | 0.921 |
| Tied Full-Cov | **0.333** | 0.812 | 0.748 |
| Tied Diag-Cov | 0.403 | 0.866 | 0.932 |
| **Gauss Features** | | | |
| Full-Cov | **0.299** | 0.811 | 0.789 |
| Diag-Cov | 0.446 | 0.820 | 0.882 |
| Tied Full-Cov | 0.347 | 0.788 | 0.848 |
| Tied Diag-Cov | 0.452 | 0.866 | 0.930 |
| **Gauss Features – PCA = 10** | | | |
| Full-Cov | 0.729 | 1.007 | 0.998 |
| Tied-Full-Cov | 0.788 | 1.005 | 0.998 |

MVG Raw, TCG Raw and MVG Gaussianized are the best 3 models. Diagonal models are in general worse. PCA worsen the minDCF in all case, in fact from the Pearson Correlation Coefficient Heatmap we can notice how no pair of features are highly correlated between each other. Quite all models are useless for imbalanced tasks.

Now we analyze the models in terms of actual DCF. The plot below compares actDCF and minDCF for the best 3 models chosen earlier. The scores are quite calibrated for different applications:

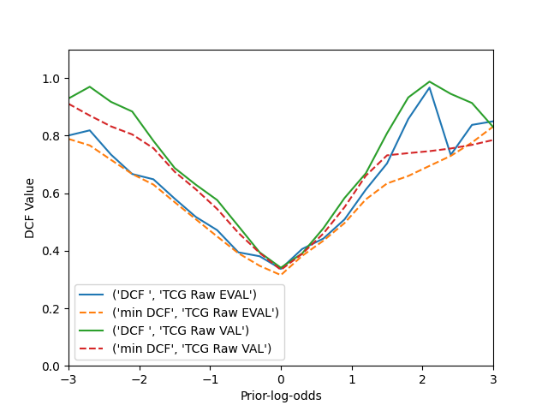


We can now analyze how the chosen models behave in the classification of evaluation set's samples. In the table are reported minDCF, actDCF and accuracy on the validation set (train set) and evaluation set (π̃=0.5)

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| **Validation Set (train set)** | | | |
| Full-Cov Raw | 0.312 | 0.364 | 16.8 |
| Tied Full-Cov Raw | 0.333 | 0.341 | 16.47 |
| Full-Cov Gauss | 0.299 | 0.313 | 14.89 |
| **Evaluation set (test set)** | | | |
| Full-Cov Raw | 0.336 | 0.353 | 18.44 |
| Tied Full-Cov Raw | 0.315 | 0.336 | 14.98 |
| Full-Cov Gauss | 0.342 | 0.364 | 17.17 |

As we can see the models perform quite well also on the evaluation set. We can deduce that train and test sets have a similar population.

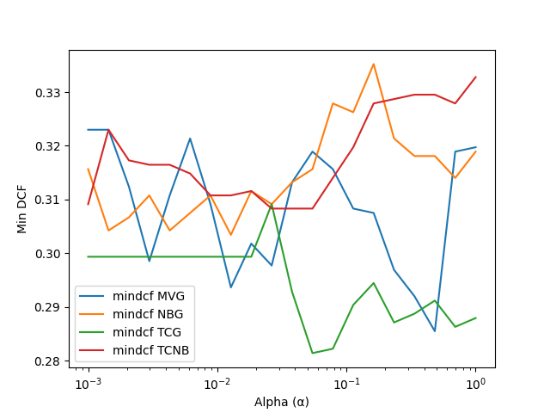
The plot below shows how evaluation and validation sets start to have a higher gap only on very imbalanced applications. (The training is done with TCG (Tied Covariance) model with raw features).



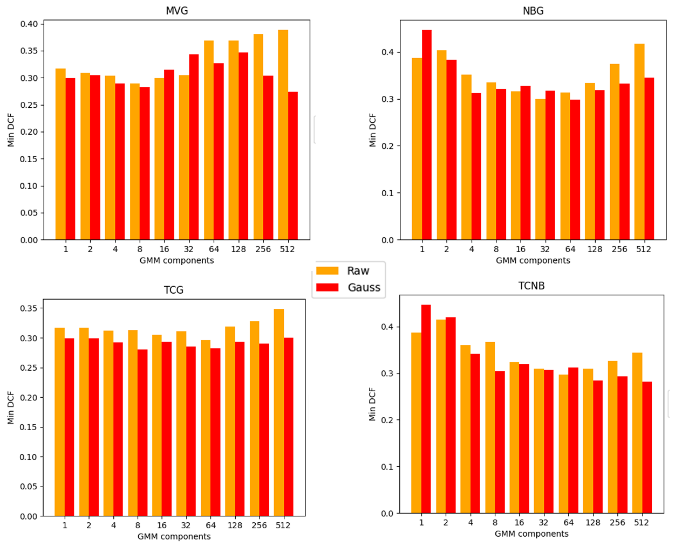
1. **Gaussian Mixture Models**

Gaussian Mixture Model are generative models that can approximate generic distributions. We expect in fact to obtain better results than Gaussian Models. We analyze GMM with diagonal and full covariance, with and without tied covariance (tying takes place at (sub-)class level).

For all models the threshold for the stopping criteria of the EM algorithm is set at 1e-6, and Ψ=0.01 to avoid degenerate cases. Also, hyperparameter α (used to split GMMs in LBG algorithm) is set to 0.1. The plot below shows that the minimum DCF doesn’t change too much for low values of α (<0.05). [π̃=0.5]



We now turn our attention to find the best number of Gaussian components (G) for all 4 Gaussian models. We try to plot the minimum DCF with different values of G (from 1 to 512) and with raw and gaussianized data, obtained from K-Fold Protocol on validation set.



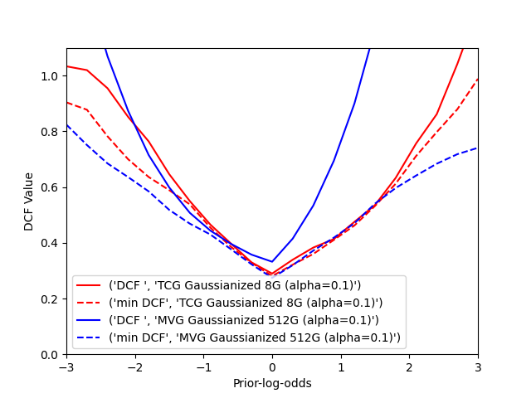
Diagonal covariance models perform worse, except for middle values of G. Tied models perform better than non-tied, especially on raw data. Gaussianization seems effective in most of the case. The best models are the full-covariance 512G and tied full-covariance 8G (both gaussianized).

We will analyze those 2 cases.

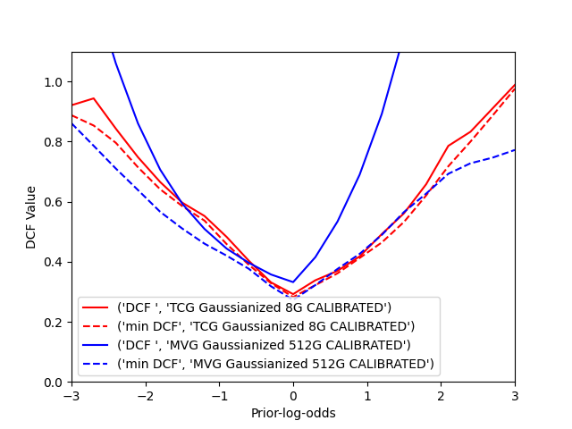
The following table shows the minDCF and (actual)DCF for different applications (π̃ = 0.5, 0.1, 0.9). The scores are uncalibrated.

|  |  |  |
| --- | --- | --- |
|  | ***minDCF*** | ***actDCF*** |
|  | | |
| Full-Cov Gau 512G | 0.274 | 0.332 |
| Tied Full-Cov Gau 8G | 0.280 | 0.289 |
|  | | |
| Full-Cov Gau 512G | 0.650 | 0.937 |
| Tied Full-Cov Gau 8G | 0.725 | 0.895 |
|  | | |
| Full-Cov Gau 512G | 0.657 | 2.254 |
| Tied Full-Cov Gau 8G | 0.751 | 0.774 |

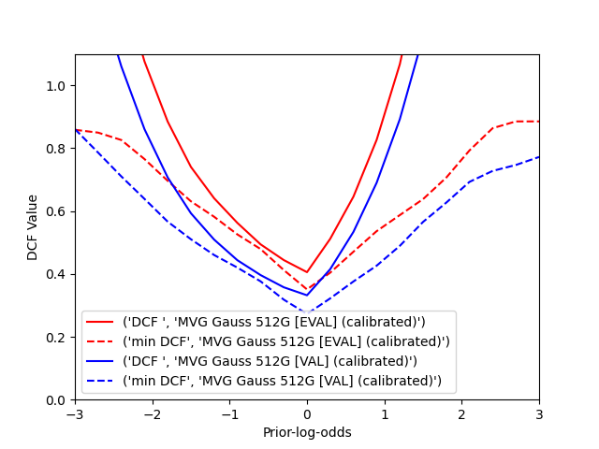
The calibration is quite poor, especially for =0.9. This is confirmed from the Bayes error plot which shows different application priors:

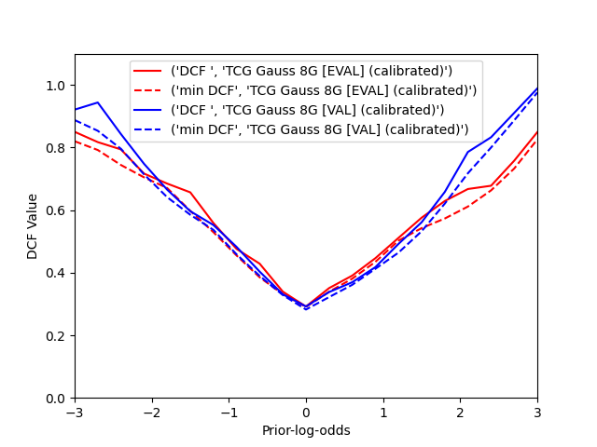


If we calibrate the scores with the logistic regression approach, we can obtain better results, although the MVG model with 512 G doesn’t calibrate well enough for higher (>0.7). (note: the choice of in the logistic regression approach doesn’t influence too much the calibration).



Now we evaluate the models on the evaluation set. The plots below show a comparison of the K-Fold protocol results on validation set and the results on evaluation set (trained with 100% of train set).





We can observe that the results on validation and evaluation sets are consistent for the TCG 8G model (lower gap). The MVG 512G model has a higher gap, so performs worse on new data, probably due to the overfitting of the 512 components.

We can conclude that the best model is indeed the Tied Covariance with 8 Gaussian components and features’ gaussianization, because it performs better in more applications (π̃) and compute better decisions on new data.