

|  |
| --- |
| GENDER SPEECH RECOGNITION  REPORT |
|  |
| 2022-07-13  S300733: Atanasio Giuseppe  S296138: Carachino Alessio |



Sommario

[Premises 3](#_Toc108647430)

[Task 3](#_Toc108647431)

[Dataset 3](#_Toc108647432)

[Features 4](#_Toc108647433)

[Histograms 4](#_Toc108647434)

[Dimensionality Reduction Techniques 7](#_Toc108647435)

[PCA 7](#_Toc108647436)

[LDA 7](#_Toc108647437)

[Classification 9](#_Toc108647438)

[What we use… 9](#_Toc108647439)

[Validation 9](#_Toc108647440)

[Generative models - Multivariate Gaussian Classifier 10](#_Toc108647441)

[Logistic Regression 11](#_Toc108647442)

[Support Vector Machine 14](#_Toc108647443)

[Gaussian Mixture Models 16](#_Toc108647444)

[The wrap up 18](#_Toc108647445)

[Final picks 20](#_Toc108647446)

[Experimental Results 21](#_Toc108647447)

[Generative models – Multivariate Gaussian Classifiers 21](#_Toc108647448)

[Logistic Regression 21](#_Toc108647449)

[Support Vector Machine 23](#_Toc108647450)

[Gaussian Mixture Model 24](#_Toc108647451)

[Final considerations 25](#_Toc108647452)

[SVM RBF vs 4-D GMM tied 25](#_Toc108647453)

[Weighted Logistic Regression vs 4-D GMM tied 26](#_Toc108647454)

Premises

# Task

The goal of the application is to build a model that best fits for the gender classification. We will discuss how they perform for the problem we have chosen, explaining pros and cons.

# Dataset

The dataset consists of synthetic speaker embeddings that represent the acoustic characteristics of a spoken utterance. Each row corresponds to a different speaker and contains 12 features followed by the gender label (1 for female, 0 for male). The features do not have any particular interpretation. Speakers belong to four different age groups. The age information, however, is not available.

The training set consists of 3000 samples for each class, whereas the test set contains 2000 samples for each class.

Features

# Histograms

Here are histograms for each of the 12 features. We can see that the raw features have an approximated gaussian distribution, but, in the next page, the features’ distributions are plotted after they have been pre-processed (Gaussianized).

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**RAW FEATURES CORRELATION HEATMAP**

Immagine che contiene testo, interni, piastrellato

Descrizione generata automaticamente

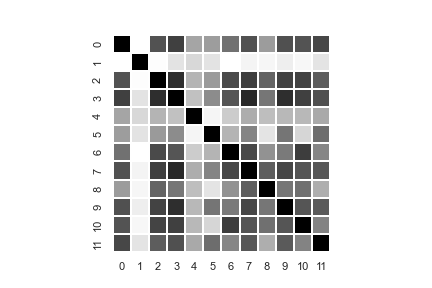
Feature 3 is highly correlated to the 0, 2, 7 and 9 ones.

This suggests we may benefit from using PCA to map data to less correlated features, but we will check later this assumption.

Here are the histograms of the gaussianized features. As we expected, the gaussianization has not brought so much improvement at all. Instead, gaussians in feature 3 and 9 were better in the previous un-pre-processed dataset.

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**GAUSSIANIZED FEATURES CORRELATION HEATMAP**



Note that, with respect to the non-gaussianized features, the correlation between 7-9 and 3 ones has decreased.

Dimensionality Reduction Techniques

# PCA

PCA is the process of computing the principal components and using them to perform a change of basis on the data. This means that it makes it possible to map in another feature space the analysis and even to reduce the number of features of the dataset, reducing for example the training overlay and the risk introduced by the curse of dimensionality.

The following images, which are for demonstration purposes only, show the training set after that PCA has been applied.



Figure 1 - PCA (12 -> 2) Figure 2 - PCA (12 -> 3)

# LDA

While PCA does not provide any guarantee about obtaining discriminant directions, Linear Discriminant Analysis is defined as a criterion of optimality finds a direction that has a large separation between the classes and small spread inside each class.

Also in this case, The following images, which are for demonstration purposes only, show the training set after that LDA has been applied, or PCA followed by LDA, as these two techniques can be used simultaneously (in this strict order).



Figure 3 – PCA (12 -> 2) + LDA (green line is the best direction) Figure 4 - PCA (12 -> 2) + LDA

(histograms show how points are projected along the direction)



Figure 5 - PCA (12 ->3) + LDA (blue line is the best direction) Figure 6 - PCA (12 -> 3) + LDA (histograms show how points are projected along the direction)



Figure 7 - LDA (12 -> 1)

# Classification

## What we use…

* Generative models – Linear and Quadratic Classifiers
  + Multivariate Gaussian Classifier (**MVG**)
  + MVG + Diagonal Covariance
  + MVG + Tied Covariance
  + MVG + Diagonal Covariance with Tied Covariance
* Logistic regression
  + Quadratic Logistic Regression
  + Prior Weighted Logistic Regression
* Support Vector Machines (SVM)
  + Primal implementation
  + Kernel SVM with polynomial kernel with degree=4
  + Kernel SVM with RFB
* Gaussian Mixture Models (GMM)
  + Gaussian Mixture Models (**GMM**)
  + GMM + Diagonal Covariance
  + GMM + Tied Covariance
  + GMM + Diagonal Covariance with Tied Covariance

# Validation

* To understand which model is most promising, and to assess the effects of using PCA, we have employed K-Fold cross validation. In facts, all of the following results have been obtained with K-Fold Validation with K = 5.
* Inside each cell of the following tables, we have reported the minDCF. We do not care about actDCF in this initial phase.
* ‘MinDCF’ has been computed with Cfp and Cfn both equal to one, as we do not have any specific requirements regarding the miss-classification costs.

## Generative models - Multivariate Gaussian Classifier

***EXPECTATIONS***

Since histograms have shown that features approximately have a gaussian distribution, it is expected that the Generative Models work well for this dataset. Furthermore, heatmaps showed that correlation is significantly spreaded along the features. Therefore, it is expected that the models based on the Naïve Bayes assumptions will perform badly.

***RESULTS***

Here are the results of Multivariate Gaussian Classifiers in three different applications (ours has π=0.5).

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | π = 0.5 | π = 0.9 | π = 0.1 |  | π = 0.5 | π = 0.9 | π = 0.1 |  | π = 0.5 | π = 0.9 | π = 0.1 |
|  | **RAW features** | | |  | **Gaussianization** | | |  | **Z NORM** | | |
|  | **no PCA** | | | | | | | | | | |
| Full-cov | 0.048 | 0.125 | 0.128 |  | 0.062 | 0.171 | 0.181 |  | 0.048 | 0.123 | 0.126 |
| DIAG-COV | 0.563 | 0.856 | 0.825 |  | 0.541 | 0.824 | 0.81 |  | 0.565 | 0.848 | 0.818 |
| Tied full-cov | 0.047 | 0.128 | 0.118 |  | 0.06 | 0.167 | 0.18 |  | 0.048 | 0.127 | 0.125 |
| tied diag-cov | 0.564 | 0.85 | 0.829 |  | 0.538 | 0.816 | 0.804 |  | 0.566 | 0.845 | 0.821 |
|  | **PCA (m=10)** | | | | | | | | | | |
| Full-cov | 0.047 | 0.124 | 0.14 |  | 0.071 | 0.204 | 0.206 |  | 0.112 | 0.263 | 0.304 |
| DIAG-COV | 0.067 | 0.156 | 0.162 |  | 0.085 | 0.223 | 0.228 |  | 0.119 | 0.275 | 0.294 |
| Tied full-cov | 0.047 | 0.125 | 0.13 |  | 0.071 | 0.206 | 0.199 |  | 0.111 | 0.259 | 0.298 |
| tied diag-cov | 0.063 | 0.149 | 0.153 |  | 0.083 | 0.225 | 0.227 |  | 0.118 | 0.272 | 0.293 |
|  | **PCA (m=9)** | | | | | | | | | | |
| Full-cov | 0.047 | 0.125 | 0.139 |  | 0.091 | 0.238 | 0.242 |  | 0.158 | 0.376 | 0.403 |
| DIAG-COV | 0.065 | 0.153 | 0.164 |  | 0.095 | 0.258 | 0.26 |  | 0.161 | 0.377 | 0.417 |
| Tied full-cov | 0.047 | 0.123 | 0.131 |  | 0.09 | 0.233 | 0.236 |  | 0.155 | 0.367 | 0.398 |
| tied diag-cov | 0.062 | 0.146 | 0.153 |  | 0.096 | 0.261 | 0.26 |  | 0.161 | 0.376 | 0.415 |

***CONSIDERATIONS***

As it was expected, PCA brings good results even with 9 features and improves models based on Naïve assumption, but they aren’t remarkable with respect to the full and tied RAW features. So, for the discriminative approaches, it’s not expected that PCA works well; however, we reported the results anyway.

Tied Models works well, and this suggests that models which exploits linear separation rules should work well. This will be confirmed later.

Lastly, Z-Normalized features does not bring remarkable results.

We can notice that the ones with diagonal covariance matrices perform worse than full covariance ones, and this is due to the highly correlation between features, as we have assumed above with the expectations.

## Logistic Regression

***OBJECTIVE FUNCTION***

Since we want to consider different applications (with different priors), we have implemented the **prior-weighted** version of the Logistic Regression. In other words, the implemented objective function is:

Immagine che contiene testo

Descrizione generata automaticamente

***EXPECTATIONS***

Since ‘TIED FULL-COV’ introduces linear separation rules and have good results, it is expected that Logistic Regression will perform well. Furthermore, we do not expect that with ‘Gaussianization’ or ‘Z-score Normalization’ we will obtain significantly better results as Logistic Regression does not require specific assumptions on the data distribution.

DA SCRIVERE CHE NON CI ASPETTIAMO CHE LA REGULARIZATION ABBIA UN IMPATTO SIGNIFICATIVO NELLA QUADRATIC

***RESULTS***

The plots show how minDCF is affected by different values of λ. They are exploited to calibrate λ, which is the regularization term.

**LR Z-NORMALIZED FEATURES – Results LR GAUSSIANIZED FEATURES – Results**

|  |  |
| --- | --- |
| **SVG SALVATA IN LOCALE** | **SVG SALVATA IN LOCALE** |
| |  |  |  |  | | --- | --- | --- | --- | |  | π = 0.5 | π = 0.9 | π = 0.1 | |  | **NO PCA** | | | | lr, lambda=1e-5 | 0.047 | 0.124 | 0.132 | | lr, lambda=1e-3 | 0.053 | 0.141 | 0.151 | | lr, lambda=1E-1 | 0.157 | 0.366 | 0.402 | | lr, lambda=1E-0 | 0.278 | 0.626 | 0.62 | |  | **PCA m=10** | | | | lr, lambda=1e-5 | 0.112 | 0.26 | 0.294 | | lr, lambda=1e-3 | 0.113 | 0.261 | 0.303 | | lr, lambda=1E-1 | 0.17 | 0.408 | 0.432 | | lr, lambda=1E-0 | 0.285 0.634 0.634 | | | |  |  | **PCA m=9** |  | | lr, lambda=1e-5 | 0.157 | 0.359 | 0.405 | | lr, lambda=1e-3 | 0.157 | 0.364 | 0.407 | | lr, lambda=1E-1 | 0.196 | 0.462 | 0.496 | | lr, lambda=1E-0 | 0.296 | 0.656 | 0.655 | | |  |  |  |  | | --- | --- | --- | --- | |  | π = 0.5 | π = 0.9 | π = 0.1 | |  | **NO PCA** | | | | lr, lambda=1e-5 | 0.056 | 0.158 | 0.162 | | lr, lambda=1e-3 | 0.056 | 0.168 | 0.168 | | lr, lambda=1E-1 | 0.122 | 0.279 | 0.313 | | lr, lambda=1E-0 | 0.228 | 0.542 | 0.552 | |  | **PCA m=10** | | | | lr, lambda=1e-5 | 0.068 | 0.197 | 0.193 | | lr, lambda=1e-3 | 0.069 | 0.203 | 0.197 | | lr, lambda=1E-1 | 0.125 | 0.292 | 0.323 | | lr, lambda=1E-0 | 0.23 | 0.545 | 0.555 | |  | PCA m=9 | | | | lr, lambda=1e-5 | 0.089 | 0.238 | 0.231 | | lr, lambda=1e-3 | 0.089 | 0.24 | 0.238 | | lr, lambda=1E-1 | 0.133 | 0.307 | 0.35 | | lr, lambda=1E-0 | 0.233 | 0.556 | 0.568 | |

**LR RAW FEATURES – Results QUADRATIC LR RAW FEATURES – Results**

|  |  |
| --- | --- |
| **SVG SALVATA IN LOCALE** | **SVG SALVATA IN LOCALE** |
| |  |  |  |  | | --- | --- | --- | --- | |  | π = 0.5 | π = 0.9 | π = 0.1 | |  | **NO PCA** | | | | lr, lambda=1e-5 | 0.047 | 0.126 | 0.131 | | lr, lambda=1e-3 | 0.047 | 0.125 | 0.132 | | lr, lambda=1E-1 | 0.048 | 0.13 | 0.127 | | lr, lambda=1E-0 | 0.059 | 0.14 | 0.154 | |  | **PCA m=10** | | | | lr, lambda=1e-5 | 0.048 | 0.122 | 0.139 | | lr, lambda=1e-3 | 0.048 | 0.122 | 0.139 | | lr, lambda=1E-1 | 0.048 | 0.125 | 0.134 | | lr, lambda=1E-0 | 0.06 | 0.155 | 0.14 | |  | **PCA m=9** | | | | lr, lambda=1e-5 | 0.048 | 0.124 | 0.137 | | lr, lambda=1e-3 | 0.048 | 0.124 | 0.136 | | lr, lambda=1E-1 | 0.048 | 0.124 | 0.135 | | lr, lambda=1E-0 | 0.059 | 0.137 | 0.156 | | |  |  |  |  | | --- | --- | --- | --- | |  | π = 0.5 | π = 0.9 | π = 0.1 | |  | **NO PCA** | | | | lr, lambda=1e-5 | 0.122 | 0.327 | 0.372 | | lr, lambda=1e-3 | 0.121 | 0.319 | 0.381 | | lr, lambda=1E-1 | 0.123 | 0.317 | 0.369 | | lr, lambda=1E-0 | 0.122 | 0.316 | 0.369 | |  | **PCA m=10** | | | | lr, lambda=1e-5 | 0.048 | 0.122 | 0.139 | | lr, lambda=1e-3 | 0.048 | 0.122 | 0.139 | | lr, lambda=1E-1 | 0.048 | 0.125 | 0.134 | | lr, lambda=1E-0 | 0.06 | 0.14 | 0.155 | |  | **PCA m=9** | | | | lr, lambda=1e-5 | 0.048 | 0.124 | 0.137 | | lr, lambda=1e-3 | 0.048 | 0.124 | 0.136 | | lr, lambda=1E-1 | 0.048 | 0.124 | 0.135 | | lr, lambda=1E-0 | 0.059 | 0.137 | 0.156 | |

***CONSIDERATIONS***

The calibration of the hyperparameter λ shows that regularization is not required in none of the above cases. However, we decided to show the minDCF value for different λ values. Results for λ=0 are not reported since they are similar to λ=1e-6 or λ=1e-4. PCA does not bring remarkable improvements, except for the Quadratic Logistic Regression.

The best performing model of Log-Reg family on the validation set is the Weighted Logistic Regression with no regularization (or λ close to zero).

Up to now, the top 2 performing models are Weighted Log-Reg and MVG with same covariance matrixes (Tied).

DA SCRIVERE CHE LA REG NON HA UN IMPATTO NELLA QUAD COME PREVISTO

## Support Vector Machine

***EXPECTATIONS***

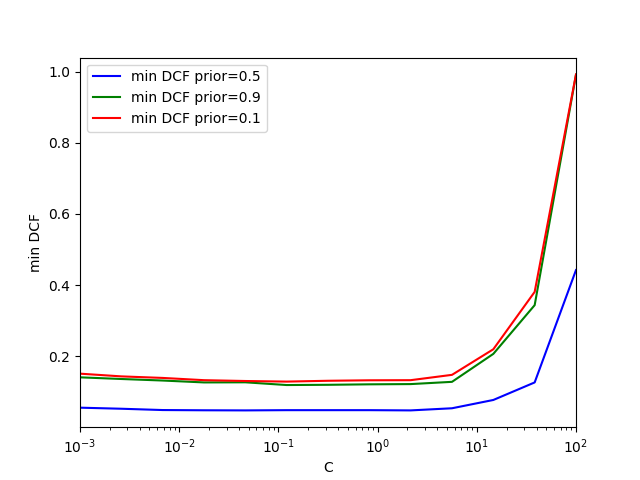
Up to now, model which exploit linear separation rules have performed very well. So it is expected that Linear SVM (the primal implementation) will provide good results as well.

***RESULTS***

The plots show how minDCF is affected by different values of the hyper-parameter of C. They are exploited to tune C.

**Primal, RAW FEATURES Implementation – Results**

**C TUNING WITH K=1.0**

****

|  |  |  |  |
| --- | --- | --- | --- |
| π = 0.5 | K = 0.1 | K = 1.0 | K = 10.0 |
| SVM, C=0.01 | 0.06 | 0.049 | 0.048 |
| SVM, C=0.1 | 0.057 | 0.048 | 0.048 |
| SVM, C=1.0 | 0.051 | 0.048 | 0.047 |
| SVM, C=10.0 | 0.093 | 0.096 | 0.065 |

|  |  |  |  |
| --- | --- | --- | --- |
| π = 0.1 | K = 0.1 | K = 1.0 | K = 10.0 |
| SVM, C=0.01 | 0.163 | 0.135 | 0.132 |
| SVM, C=0.1 | 0.161 | 0.13 | 0.13 |
| SVM, C=1.0 | 0.14 | 0.133 | 0.138 |
| SVM, C=10.0 | 0.252 | 0.232 | 0.165 |

|  |  |  |  |
| --- | --- | --- | --- |
| π = 0.9 | K = 0.1 | K = 1.0 | K = 10.0 |
| SVM, C=0.01 | 0.151 | 0.13 | 0.124 |
| SVM, C=0.1 | 0.146 | 0.12 | 0.118 |
| SVM, C=1.0 | 0.129 | 0.119 | 0.127 |
| SVM, C=10.0 | 0.291 | 0.291 | 0.157 |

**Dual + Polynomial, RAW FEATURES Implementation – Results**

|  |  |  |
| --- | --- | --- |
| C = 1.0, D=4, con=1000 | K = 1.0 | K = 10.0 |
| SVM, π=0.5 | 0.11 | 0.166 |
| SVM, π=0.1 | 0.334 | 0.543 |
| SVM, π=0.9 | 0.322 | 0.44 |

**Dual + RBF, RAW FEATURES Implementation – Results**

|  |  |  |  |
| --- | --- | --- | --- |
| gamma=0.001 | C=1.0 | K = 0.1 | K = 1.0 | K = 10.0 |
| SVM , π = 0.5, | 0.039 | 0.039 | 0.04 |
| SVM, π = 0.1 | 0.122 | 0.123 | 0.12 |
| SVM, π = 0.9 | 0.12 | 0.12 | 0.12 |

***CONSIDERATIONS***

We can compare linear models in terms of min DCF:

|  |  |  |  |
| --- | --- | --- | --- |
|  | π = 0.5 | π = 0.9 | π = 0.1 |
| MVG (Tied Full-Cov) | 0.047 | 0.128 | 0.118 |
| Log Reg (λ = 10−6, πT = 0.5) | 0.048 | 0.126 | 0.131 |
| wLog Reg (λ = 10−4, πT = 0.5) | 0.047 | 0.126 | 0.131 |
| SVM, C=1.0 | 0.048 | 0.119 | 0.133 |

Linear SVM performs similarly to other linear approaches, as expected.

Since Linear models perform similarly to non-linear models, we decided to report the results for the two SVM non-linear formulations

## Gaussian Mixture Models

***EXPECTATIONS***

GMMs can approximate generic distributions, so we expect to obtain better results than with the Gaussian model. We are going to train GMM up to 128 components for full and diagonal covariance, and tied/non tied.  
Please note: raw and z-normalization retrieve the same results.

***RESULTS***

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **cOMPONENTS** | **2** | **4** | **8** | **16** | **32** | **64** | **128** |
|  | Z-Normalization | | | | | | |
| **Full-cov** | ﻿0.043 | 0.033 | 0.032 | 0.044 | 0.057 | 0.087 | 0.151 |
| **Diag-cov** | 0.4 | 0.108 | 0.087 | 0.081 | 0.088 | 0.086 | 0.095 |
| **Tied full-cov** | 0.048 | 0.03 | 0.03 | 0.032 | 0.035 | 0.035 | 0.039 |
| **Tied Diag-cov** | 0.402 | 0.105 | 0.082 | 0.083 | 0.072 | 0.068 | 0.069 |
|  | Gaussianized | | | | | | |
| **Full-cov** | 0.053 | 0.047 | 0.048 | 0.055 | 0.064 | 0.089 | 0.106 |
| **Diag-cov** | 0.346 | 0.127 | 0.106 | 0.107 | 0.093 | 0.096 | 0.101 |
| **Tied full-cov** | 0.062 | 0.057 | 0.053 | 0.046 | 0.043 | 0.042 | 0.049 |
| **Tied Diag-cov** | 0.358 | 0.148 | 0.113 | 0.103 | 0.098 | 0.086 | 0.08 |

|  |  |
| --- | --- |
|  |  |
|  |  |

4-D GMM

|  |  |
| --- | --- |
|  |  |
|  |  |

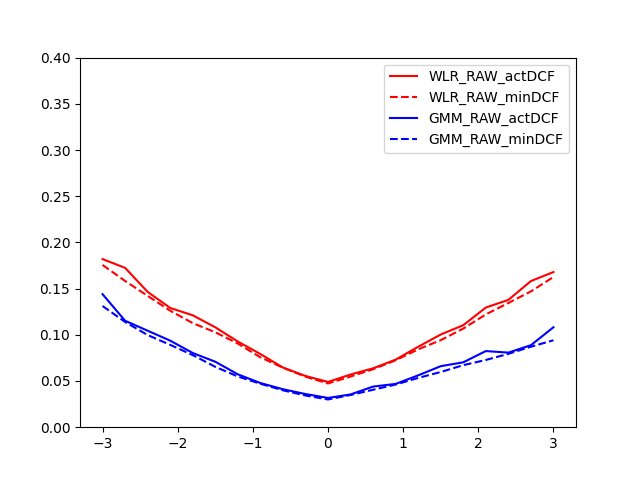
***CONSIDERATIONS***

We can check that the best result that it could be obtained is for full-covariance tied GMM, especially the 4-D one.   
Taking a zoom for the 4-D models, we can analyse bayes error plots: the best is again the full-covariance tied.

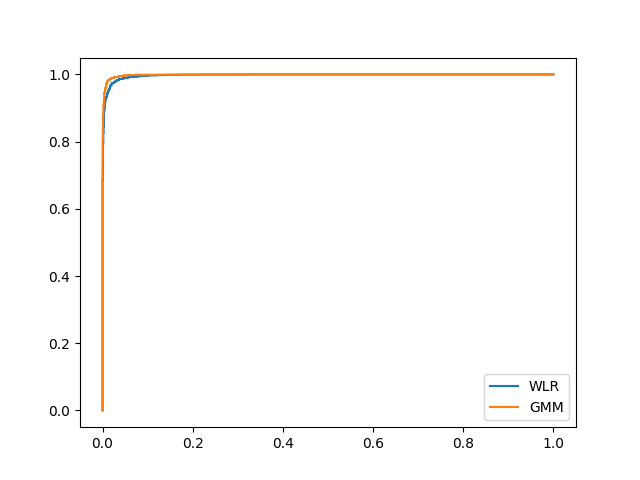
# The wrap up

We still should compare the models. It’s time to exploit ROC curves and Bayes Error Plots.

Firstly, here is the comparison (on RAW features) between the GMM with 4 components and Tied Full Covariance Matrix and Weighted Logistic Regression.

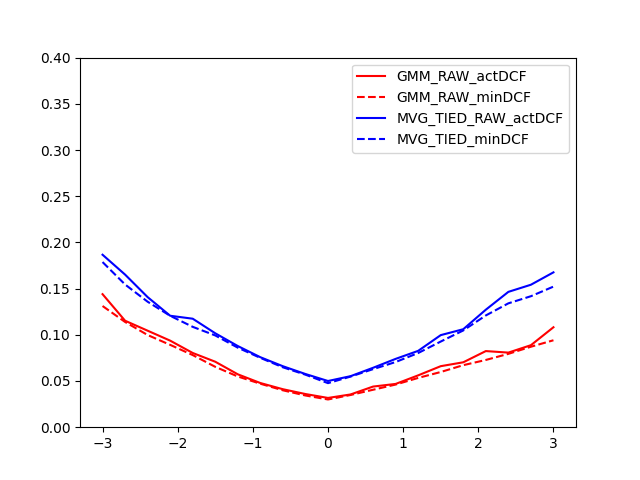


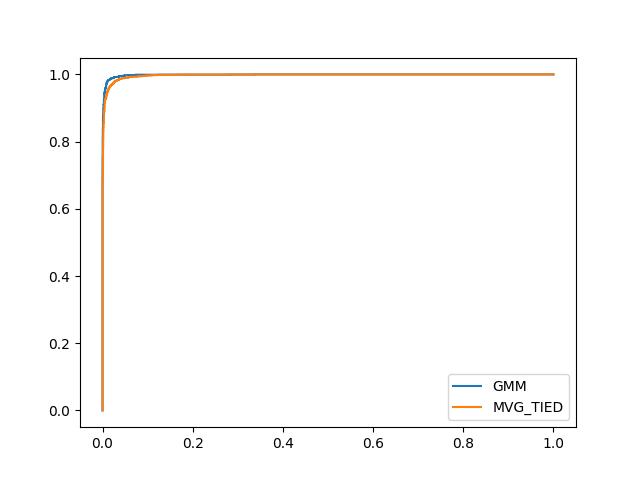
It is clear that GMM outperforms Weighted Logistic Regression. However, calibration is not required as the actDCF curve is already quite close to the minDCF curve.



The ROC Curve shows the better performance on the validation set of the GMM too.

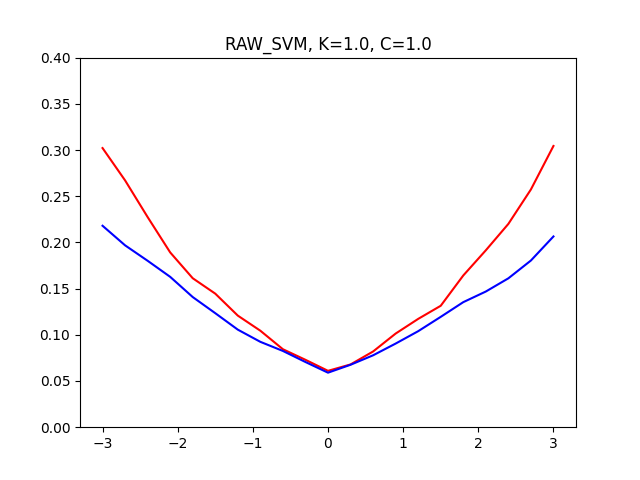
Now, let’s turn the attention on the comparison (RAW features in this case too) between MGC with Tied Full Covariance and the same GMM as above.





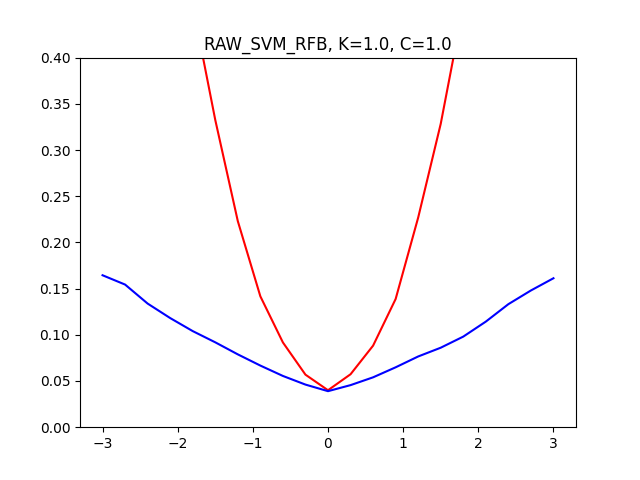
Where the exact same considerations of the previous comparison are still valid.

For what regards SVM, the following image shows the Bayes Error Plots for the case of the linear SVM:



The above results are good but not better than the previously compared models.

While the next one, shows the Bayes Error Plot for the RFB Kernel case.



In this case, scores are highly uncalibrated. They can be improved by basically applying Logistic Regression on the scores returned by the model.

# Final picks

We expect that the **4D-GMM Tied Full Covariance Model** outperforms the others.

Experimental Results

We now turn the attention to the evaluation part. It’s useful to choose the minimum DCF to evaluate because it’s an optimistic estimation for the actual DCF in the hypothetical case in which it could be chosen the optimal threshold for the evaluation set.

Since validation has performed with 5-fold cross-validation, the evaluation part will be executed using the whole dataset (all training data for training, all test data for test).

## Generative models – Multivariate Gaussian Classifiers

Here are reported the evaluation results for MVG.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | π = 0.5 | π = 0.9 | π = 0.1 |  | π = 0.5 | π = 0.9 | π = 0.1 |
|  | **RAW features** | | |  | **Gaussianization** | | |
|  | **no PCA** | | | | | | |
| Full-cov | 0.053 | 0.138 | 0.134 |  | 0.073 | 0.182 | 0.202 |
| DIAG-COV | 0.57 | 0.882 | 0.81 |  | 0.547 | 0.846 | 0.792 |
| Tied full-cov | 0.05 | 0.135 | 0.132 |  | 0.069 | 0.177 | 0.184 |
| tied diag-cov | 0.57 | 0.88 | 0.808 |  | 0.545 | 0.846 | 0.793 |
|  | **PCA (m=10)** | | | | | | |
| Full-cov | 0.052 | 0.139 | 0.136 |  | 0.08 | 0.206 | 0.203 |
| DIAG-COV | 0.068 | 0.18 | 0.198 |  | 0.088 | 0.229 | 0.243 |
| Tied full-cov | 0.054 | 0.134 | 0.138 |  | 0.078 | 0.205 | 0.199 |
| tied diag-cov | 0.068 | 0.181 | 0.196 |  | 0.085 | 0.225 | 0.234 |
|  | **PCA (m=9)** | | | | | | |
| Full-cov | 0.054 | 0.14 | 0.144 |  | 0.098 | 0.23 | 0.255 |
| DIAG-COV | 0.068 | 0.18 | 0.204 |  | 0.102 | 0.248 | 0.29 |
| Tied full-cov | 0.054 | 0.136 | 0.142 |  | 0.098 | 0.232 | 0.258 |
| tied diag-cov | 0.068 | 0.18 | 0.2 |  | 0.104 | 0.258 | 0.286 |

The results are consistent with those obtained in the validation part. We can confirm that the MVG model with tied covariance is the best option to choose, and this also confirm that linear separation rules are the best choose.  
PCA isn’t outperforming with respect to the previous, even if it improves the MVG models with diagonal covariance.  
Gaussianization worsen the data.

## Logistic Regression ( DA AGGIORNARE IN SEGUITO A WLR IN LR )

We now move our attention to the Logistic Regression.

|  |  |  |  |
| --- | --- | --- | --- |
|  | π = 0.5 | π = 0.9 | π = 0.1 |
| LR raw lambda=1e-6 | 0.053 | 0.133 | 0.135 |
| lr gauss. lambda=1e-6 | 0.059 | 0.165 | 0.162 |
| lr ZNORM. lambda=1e-6 | 0.052 | 0.135 | 0.133 |
| WLR RAW, lambda=1e-4 | 0.053 | 0.133 | 0.135 |
| QLR RAW, lambda=1e-4 | 0.114 | 0.285 | 0.353 |

Even these results are consistent with our expectations.  
The results are quite similar to the MVG tied, and it can be confirmed, also in this case, that linear separation rules are the best road to follow.

Although we do not expect significant deviations, we can also verify whether the chosen λ provides close to optimal results.

|  |  |  |
| --- | --- | --- |
|  | VALIDATION RESULTS | EVALUATION RESULTS |
| LR  RAW |  |  |
| WLR  RAW |  |  |
| QLR  RAW |  |  |

## Support Vector Machine

We report the results for all types of SVM, in particular for the hyperparameters C=1.0 and K=1.0

|  |  |  |  |
| --- | --- | --- | --- |
|  | π = 0.5 | π = 0.9 | π = 0.1 |
| SVM primal | 0.051 | 0.129 | 0.138 |
| SVM DUAL, PolyNomial  D=4 | 0.156 | 0.392 | 0.486 |
| SVM dual, rfb  gAMMA=1e-3 | 0.044 | 0.112 | 0.133 |

We reported the comparison on the evaluation set of SVM in the next pages.

## Gaussian Mixture Model

We have chosen 4 components results to analyze with other applications.

Please note: raw and z-normalization retrieve the same results.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **π = 0.5** | **π = 0.1** | **π = 0.9** |
|  | RAW | | |
| **Full-cov** | 0.041 | 0.114 | 0.119 |
| **DIAG-COV** | 0.112 | 0.258 | 0.29 |
| **Tied full-cov** | 0.03 | 0.08 | 0.092 |
| **tied diag-cov** | 0.111 | 0.256 | 0.295 |
|  | Gaussianization | | |
| **Full-cov** | 0.05 | 0.149 | 0.147 |
| **DIAG-COV** | 0.132 | 0.328 | 0.357 |
| **Tied full-cov** | 0.06 | 0.16 | 0.176 |
| **tied diag-cov** | 0.161 | 0.411 | 0.399 |

|  |  |
| --- | --- |
|  |  |
|  |  |

The outcomes are consistent with those in the validation part.   
We can confirm that linear models outperform, especially for the 4-D GMM tied full covariance case.  
It is remarkable that full-covariance models over the 8-D tends to have more problems of overfitting than the diagonal ones.

## Final considerations

As we expected, the evaluation confirms every result obtained during the validation phase. We can conclude that our results are reliable.

4-D GMM Tied outperforms with respect to the other models.

### SVM RBF vs 4-D GMM tied

We now turn our attention to the comparison between SVM RBF and GMM tied. We have chosen 4-G Tied full-cov and SVM dual RBF (C=1.0, K=1.0). From now on we are going to call them simply SVM and GMM. Following: bayes error (left), ROC (right), comparison (bottom).

|  |  |
| --- | --- |
|  |  |
| |  |  |  |  | | --- | --- | --- | --- | |  | **π=0.5** | **π=0.1** | **π=0.9** | | **GMM** | 0.03 | 0.08 | 0.092 | | SVM | 0.044 | 0.133 | 0.112 | | |

As we can see, GMM performs better than SVM. We could calibrate the scores of the latter exploiting logistic regression and check if it performs better, even though GMM remains the best model with the lowest minDCF and actDCF curves. We can conclude that calibrating SVM isn’t useful to do at all.

### Weighted Logistic Regression vs 4-D GMM tied

We now turn our attention to the comparison between Weighted Logistic Regression (lambda=1e-4) and 4-D GMM tied. For simplicity we will call them WLR and GMM. Following: bayes error (left), ROC (right), comparison (bottom).

|  |  |
| --- | --- |
|  |  |
| |  |  |  |  | | --- | --- | --- | --- | |  | **π=0.5** | **π=0.1** | **π=0.9** | | **GMM** | 0.03 | 0.08 | 0.092 | | WLR | 0.053 | 0.135 | 0.133 | | |

Even though WLR has a lower minDCF than the above SVM, the actDCF curve suggests that it performs better than the above SVM model, but not enough to bring us to choose 4-D GMM tied full-cov as the best model for our application.