

|  |
| --- |
| GENDER SPEECH RECOGNITION  REPORT |
|  |
| 7 luglio  S300733: Atanasio Giuseppe  S296138: Carachino Alessio |



Sommario

[Premises 3](#_Toc104837004)

[Task 3](#_Toc104837005)

[Dataset 3](#_Toc104837006)

[Dimensionality Reduction Techniques 4](#_Toc104837007)

[PCA 4](#_Toc104837008)

[LDA 4](#_Toc104837009)

[Considerations regarding DRT for Gender Speech Recognition 6](#_Toc104837010)

[Classification 7](#_Toc104837011)

[What we use… 7](#_Toc104837012)

[Expectation 7](#_Toc104837013)

[Results 7](#_Toc104837014)

[Summary and Considerations 7](#_Toc104837015)

[Validation 9](#_Toc104837016)

[K-Fold 9](#_Toc104837017)

[Leave-One-Out 9](#_Toc104837018)

[Holdout 9](#_Toc104837019)

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 per class, whereas the test set contains 2000 samples per class.

Features

# Histograms

Here are histograms for each of the 12 features. We can see that features are already pretty well gaussianized, but we’re going to apply a gaussianization as well.

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

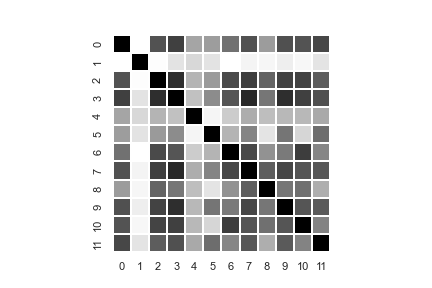
Immagine che contiene testo, interni, piastrellato

Descrizione generata automaticamente

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

Here are the gaussianized features. As we expected, the gaussianization hasn’t brought so much improvement at all. Instead, gaussians in feature 3 and 9 were better in the previous unmodified dataset.

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



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

Here are the results of Multivariate Gaussian Classifiers in three different applications (ours is π=0.5). We can notice that the ones with diagonal covariance matrix perform worst than full covariance matrix, and this is due to the highly correlation between features, as we can see above in the heatmap.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | 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 | 0.048 | 0.125 | 0.128 |  | 0.061 | 0.178 | 0.189 |
| DIAG-COV | 0.563 | 0.856 | 0.825 |  | 0.538 | 0.825 | 0.819 |
| Tied full-cov | 0.047 | 0.128 | 0.118 |  | 0.06 | 0.166 | 0.184 |
| tied diag-cov | 0.564 | 0.85 | 0.829 |  | 0.535 | 0.822 | 0.808 |
|  | **PCA (m=10)** | | | | | | |
| Full-cov | 0.047 | 0.124 | 0.14 |  | 0.073 | 0.213 | 0.21 |
| DIAG-COV | 0.067 | 0.156 | 0.162 |  | 0.083 | 0.225 | 0.239 |
| Tied full-cov | 0.047 | 0.125 | 0.13 |  | 0.071 | 0.199 | 0.207 |
| tied diag-cov | 0.063 | 0.149 | 0.153 |  | 0.082 | 0.218 | 0.224 |
|  | **PCA (m=9)** | | | | | | |
| Full-cov | 0.047 | 0.125 | 0.139 |  | 0.092 | 0.239 | 0.24 |
| DIAG-COV | 0.065 | 0.153 | 0.164 |  | 0.096 | 0.254 | 0.266 |
| Tied full-cov | 0.047 | 0.123 | 0.131 |  | 0.091 | 0.233 | 0.249 |
| tied diag-cov | 0.062 | 0.146 | 0.153 |  | 0.096 | 0.248 | 0.261 |

Since ‘TIED FULL-COV’ introduces linear separation rules, it is expected that Linear Regression will perform well. Here are the results.

|  |  |  |  |
| --- | --- | --- | --- |
|  | π = 0.5 | π = 0.9 | π = 0.1 |
| LR, lambda=1e-6 | 0.06 | 0.158 | 0.171 |
| lr, lambda=1e-4 | 0.06 | 0.158 | 0.171 |
| lr, lambda=1e-2 | 0.061 | 0.161 | 0.173 |
| lr, lambda=1.0 | 0.119 | 0.291 | 0.309 |

|  |  |  |  |
| --- | --- | --- | --- |
| π = 0.5 | K = | K = | K = |
| SVM, C= | 0. | 0 | 0. |
| SVM, C= | 0. | 0. | 0. |
| SVM, C= | 0. | 0. | 0. |
| SVM, C= | 0. | 0. | 0. |

#### Gaussianization

|  |  |  |  |
| --- | --- | --- | --- |
|  | π = 0.5 | π = 0.9 | π = 0.1 |
| Full-cov | 0.048 | 0.125 | 0.128 |
| DIAG-COV | 0.563 | 0.856 | 0.825 |
| Tied full-cov | 0.047 | 0.128 | 0.118 |
| tied diag-covm | 0.564 | 0.85 | 0.829 |
|  |  |  |  |
| LR, lambda=1e-6 | 0.06 | 0.158 | 0.171 |
| lr, lambda=1e-4 | 0.06 | 0.158 | 0.171 |
| lr, lambda=1e-2 | 0.061 | 0.161 | 0.173 |

Dimensionality Reduction Techniques

We have performed feature extraction within dimensionality reduction using PCA.

# PCA

In this part we’ve tested PCA for reducing the features, and we’ve empirically concluded that the best number of those to save is m=8. In fact, we can check that the error rate in the classification methods is quite similar for increasing values of m.

When we generate a plot scatter with this value, we can check that there are four different areas, one for each group age:



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

# LDA

Text text …



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)

# Considerations regarding DRT for Gender Speech Recognition

As we can see with m=2 and m=3 -> bad separation -> bad accuracy -> we keep results in the tables just to describe…

Classification

# What we use…

* Generative models – Linear and Quadratic Classifiers
  + Multivariate Gaussian Classifier (**MGC**)
  + MGC + Naïve Bayes Classifier
  + MGC + Tied Covariance
  + MGC + Naïve Bayes Classifier with Tied Covariance
* Linear regression

# Expectation

Since this is the …, this should be the best for our goal..

# Results

Tables with results and combinations with pca and lda

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Classifier | Raw | PCA  (m=2) | PCA (m=8) | LDA | PCA + LDA | PCA |
| MGC | 2.75% |  | 2.675 % | 2.8 % | 2.7 % |  |
| MGC - Naïve | 29.5% |  | 3.525 % | 2.8 % | 2.7 % |  |
| MGC - TC | 2.775% |  | 2.7 % | 2.775 % | 2.7 % |  |
| MGC - Naïve + TC | 29.4% |  | 3.5 % | 2.775 % | 2.7 % |  |

Table 1 - Error rates for Generative Models Classifiers and Dimensionality Reduction Techniques

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| λ | 0 | 0.000001 | 0.001 | 0.1 | 1 | 3 | Tied Gau |
| Raw | 2.75% | 2.75% | 2.725% | 2.8% | 3.7% | 5.325% | 2.775% |
| PCA m=8 | 2.7% | 2.7% | ﻿2.725% | 2.825% | 3.775% | 5.3% |  |
| PCA |  |  |  |  |  |  |  |
| PCA[m=8] + LDA | 2.675% | 2.675% | 2.675% | 2.7% | 2.7% | 2.7% |  |

Table 2 - Error rates for Linear Regression and Dimensionality Reduction Techniques and Comparison with Tied Cov

# Summary and Considerations

Dato che i risultati dei Naive sono scarsi, allora le features devono essere molto correlate tra di loro.

Validation

# K-Fold

Text Text

# Leave-One-Out

Text Text

# Holdout

Text Text