Logistic Regression - Gender Recognition by Voice

João Brito, M9984
Department of Computer Science
University of Beira Interior
Covilhã, Portugal
joao.pedro.brito@ubi.pt

***Abstract*—In this work, a binary classifier was implemented to classify human voice recordings as coming from a male subject or a female counterpart. Several measures will be highlighted, to evaluate the capabilities of our model, as well as, some experiments aiming to find the optimal configuration for this problem.**

I. Introduction

Given the nature of this project, a binary classification dataset was in order. To that end, the ”Gender Recognition by Voice” dataset [1] (freely available on Kaggle [2]) was chosen for its binary target classes and possibly interlaced features.

As it will become increasingly clear, the following sections are devoted to providing a complete overview of the dataset, the implementation details regarding the classifier and the conducted experiments.

II. Dataset Overview

A. Base Configuration

The dataset used for the purposes of this project contains 20 acoustic features of human voices. The following list describes those independent variables in more detail:

- **MeanFreq**: mean frequency of the recorded voice (in kHz);
- **sd**: standard deviation of the recorded frequency;
- **Median**: median frequency (in kHz);
- **Q25**: first quantile (in kHz);
- **Q75**: third quantile (in kHz);
- **IQR**: interquantile range (in kHz);
- **Skew**: how skewed the recording was;
- **Kurt**: the kurtosis value for the recording;
- **Sp_ent**: the spectral entropy;
- **Sfm**: the spectral flatness;
- **Mode**: the mode frequency;
- **Centroid**: the frequency centroid;
- **Meanfun**: the average of fundamental frequency measured across acoustic signal;
- **Minfun**: the minimum fundamental frequency measured across acoustic signal;
- **Maxfun**: the maximum fundamental frequency measured across acoustic signal;
- **Meandom**: the average of dominant frequency measured across acoustic signal;
- **Mindom**: the minimum of dominant frequency measured across acoustic signal;
- **Maxdom**: the maximum of dominant frequency measured across acoustic signal;
- **Dfrange**: the range of dominant frequency measured across acoustic signal;
- **Modindx**: the modulation index, calculated as the accumulated absolute difference between adjacent measurements of fundamental frequencies divided by the frequency range.

Finally, the independent variable (denoted as ”**Label**”), is either 1 (when the instance came from a male subject) or 0 (when the instance came from a female subject).

B. Pre-processing Steps

In terms of pre-processing steps, the chosen dataset didn’t have missing values. Given this, the steps taken consisted in shuffling the dataset (to make sure that the instances were pseudo-randomly distributed, with no real order to it all) and refactoring some aspectes of the original file.

On a later stage, a new file is responsible for passing the data through the K-Fold Cross Validation algorithm. This procedure takes as input a number ”k” (the number of folds) and splits the data into a set of ”k” training subsets and ”k” test subsets. With that, each instances gets to be in the test set and the training set at some point, promoting an unbiased distribution. For the purposes of this work, a value of 5 was given to ”k”.

To finalise the pre-processing stage, a final file is called with the responsibility of normalize the data. It has three modes: [0,1] normalization; [-1,1] normalization or standardization (a mean of 0 and standard deviation of 1). The experiments described here were performed with a standardization process in place.

III. Classifier Architecture

The model used is a logistic regression model, whose technical formula (a composition of 2 formulas) is as follows:

$$f = g \circ h$$

with

$$g(x) = \frac{1}{1 + e^{-x}}$$

$$h(x) = \theta_0 + \theta_1 x_1 + ... + \theta_n x_n$$

.

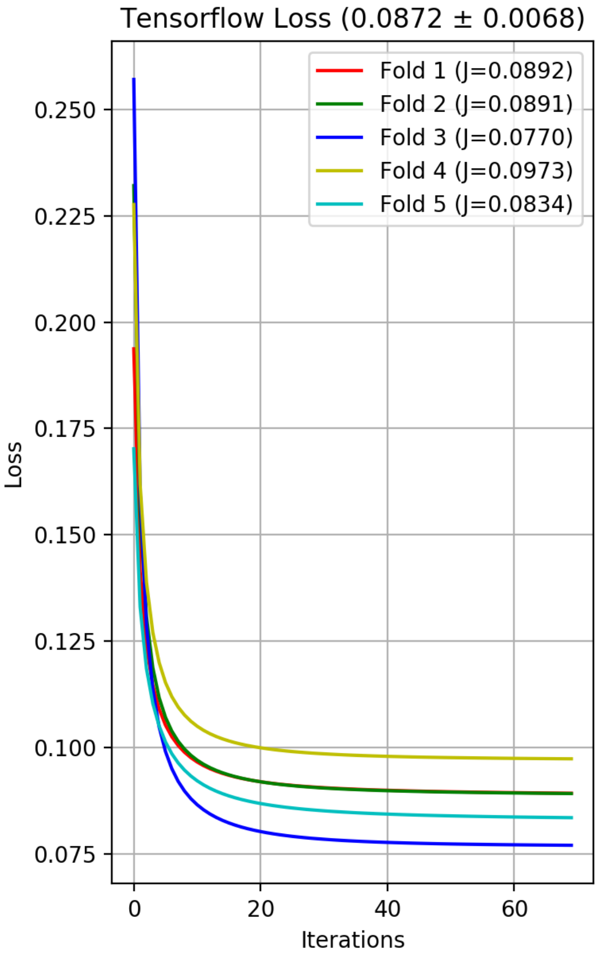

Fig. 1: Performance recorded for the original classifier with the following parameters: **70000 epochs**, **learning rate of 0.01** and **no regularization**

The above formulas show us that a logistic regression model is an extension of the linear regression model. It computes a weighted sum of all of X's features and passes it to the sigmoid function. The later outputs a value ranging from 0 to 1, with both ends being exclusive.

IV. Experiments

A. Reference performance

The reference classifier achieved really impressive results in this dataset. Figure 1 shows how training was handled on each fold given by the K-Fold algorithm, with the corresponding final loss value on display. Figure 2 highlights the confusion matrix generated by the test evaluation. Generally it performed really well, and was able to distinguish between classes. That behaviour is more evident in figure 3, where the plot closely reaches the desired (0,1) corner.

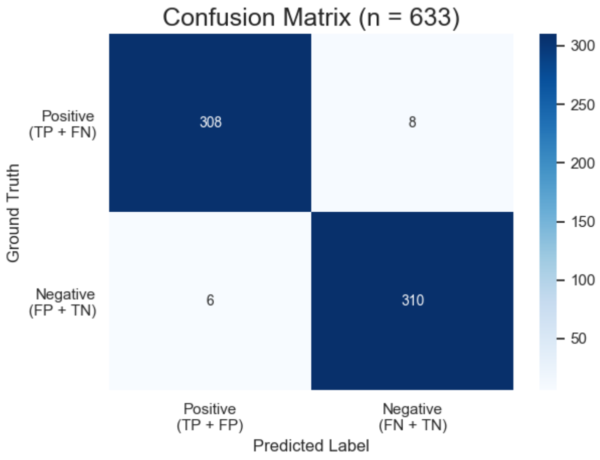

Fig. 2: Confusion matrix on the test set

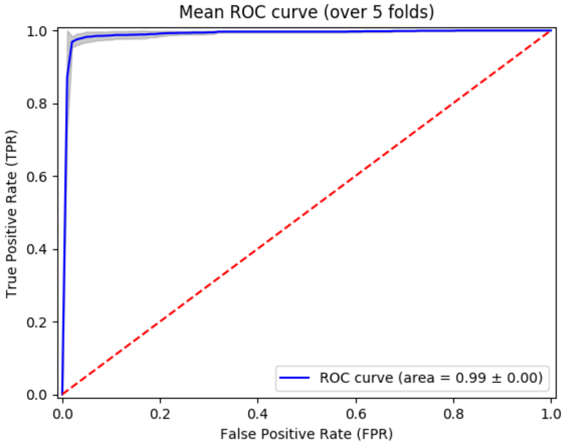

Fig. 3: ROC curve given by the base classifier: **AUC** = 0.99 ± 0.00; **Accuracy** = 0.98 ± 0.01; **Precision** = 0.98 ± 0.00 and **Recall** = 0.97 ± 0.02

B. Principal Component Analysis

As a mere experimental exercise, the Principal Component Analysis (PCA) algorithm was applied to the original data. We gave it a 95% variance input (meaning that we wanted to keep 95% of the variance in the data, whilst the number o variables needed for that goal was decided by the algorithm). The implementation available in the Scikit-learn [4] library decided that 10 variables (i.e. dimensions) were enough to capture 95% of the original variance. The source data (the same as in the previous experiment) was projected to the aforementioned 10-dimensional space. Then, the same steps were taken: training with the same hyper-parameters and evaluation on the test subsets (figures 4, 5 and 6). It is really impressive how a reduction from 20 dimensions to just 10 yielded roughly the same results. Such results show how the PCA algorithm was able to use the core features and properties of the original data, while discarding what wasn't vital for the final prediction.

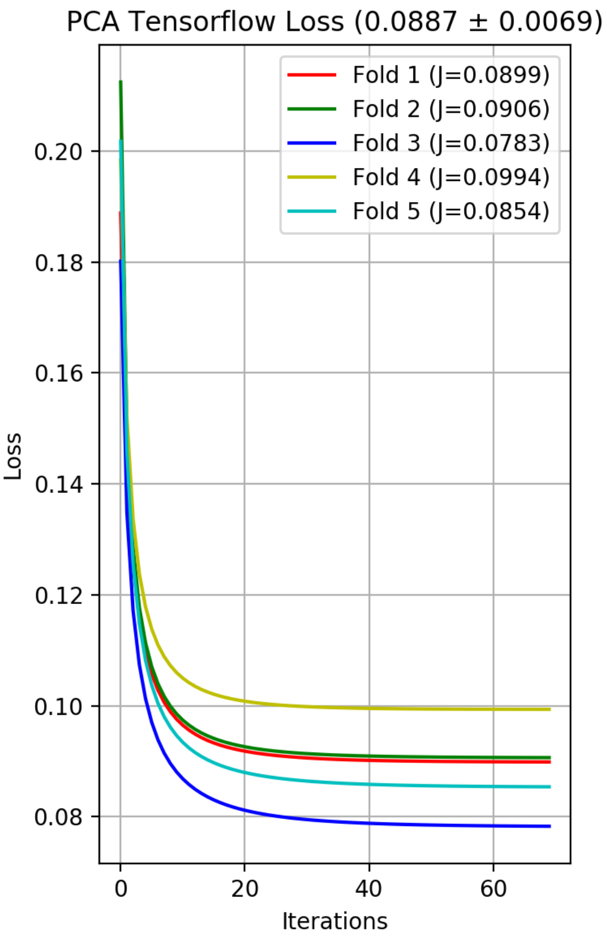

Fig. 4: Performance recorded for the classifier on the projected space (using PCA with **95% variance**)

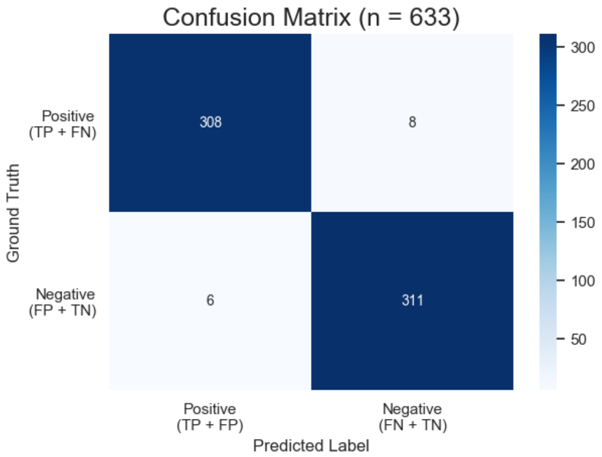

Fig. 5: Confusion matrix on the test set (of the newly projected test data)

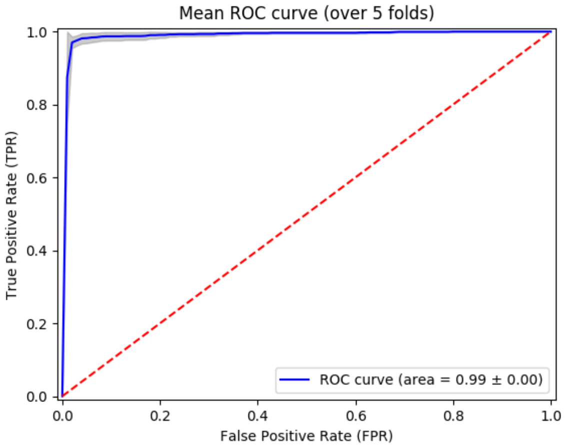

Fig. 6: ROC curve given by the improved classifier: **AUC** = 0.99 ± 0.00; **Accuracy** = 0.98 ± 0.01; **Precision** = 0.98 ± 0.01 and **Recall** = 0.97 ± 0.02

V. Conclusion

The report that now comes to an end, described the approach taken to solve a binary classification problem. With the help of a logistic regression model and then a 10-dimensional version, it was possible to achieve a really solid performance on the test set. Finally, it was demonstrated the viability of logistic regression models for such tasks and how we are able to free ourselves from the dimensionality curse with feature reduction tools (like PCA).

References

[1] Gender Recognition by Voice Dataset. [Online] https://www.kaggle.com/primaryobjects/voicegender

[2] Kaggle. [Online] https://www.kaggle.com/

[3] Tensorflow. [Online] https://www.tensorflow.org/

[4] Scikit-learn. [Online] https://scikit-learn.org/stable/