ARN - Laboratory 03

Anthony Coke, Guilain Mbayo, Mehdi Salhi

HE IG VD

April 12, 2022

Contents

Introduction

Man vs Woman

The purpose of this experiment was to train our model to be able to differentiate between men and women.

> The associated notebook for this experiment is 'MaleFemale-model_selection.ipynb'

Number of observations

Our dataset consists of audio samples of vowels pronounced by men and women. For this part, we have treated 36 values for each class (men and women). We then computed 13 MFCCs for each sample.
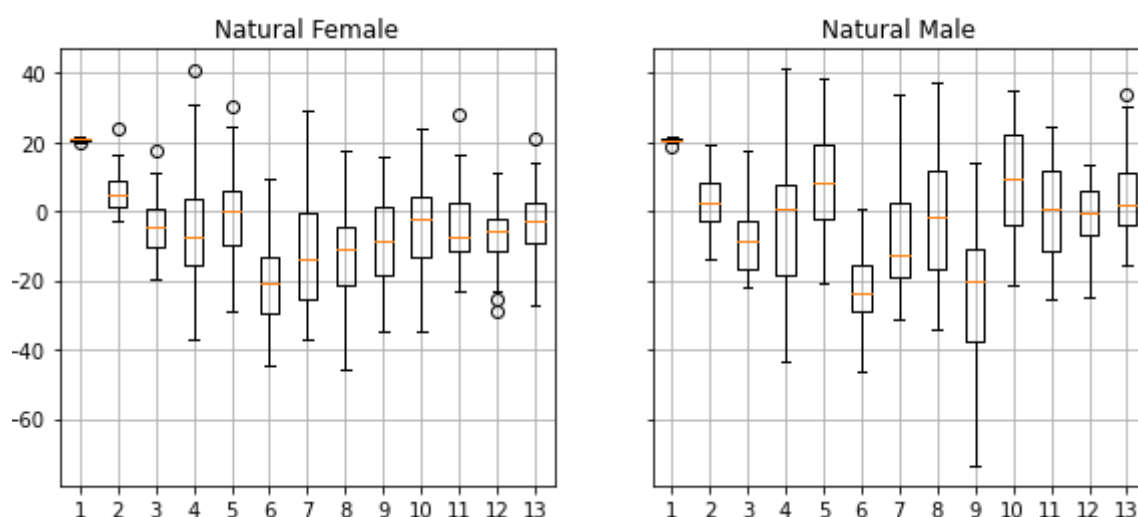
Figure 1: Boxplot of the MFCCs for men and women

Features to train the model

As we can see on the precedent box plot, there are some mfccs that are better than others to classify our dataset. For example, the mfcc number 9 is quiet different between male and female voices. However, we can see as well that there is still an overlap. As none of the mfccs seemed sufficient in itself or in combination with some others, we chose to use the 13 mfccs in order to get classification as accurate as possible.

Procedure for model selection

The very first step after getting the dataset is to normalize and label it. We chose to use the tanh function for this part, because after running some tests the sigmoid was kind of stuck at an error of 0.25, which was significantly higer than the hyperbolic tangent. We didn't really see a learning phase with the logistic function. As tanh outputs value between minus one and one, we thought it would make sense to output 1 if the observation is classified as a man or -1 if the model thinks it's a female. Then, in order to choose the best parameters for our model, we tried at first the basis value of 0.001 for the learning rate, 0.5 for the momentum and 50 epochs. After observing the results, we adjusted our parameters. Those steps where repeated several times in order to narrow our results and reduce the oscillation. When we obtained a satisfying curve for the training and test sets, we generated the confusion matrix to verify that our data were indeed well classified.
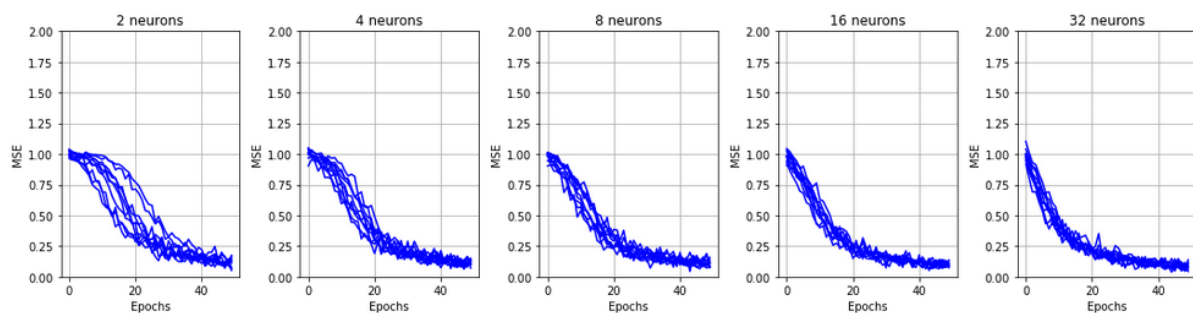
Figure 2: Exploring the number of neurons

Description of the final model and performance evaluation

Our final model used the following hyper-parameters:

- tanh activation function
- learning rate of 0.0009
- momentum of 0.85
- 2 hidden neurons
- one output neuron
- number of epochs: 100
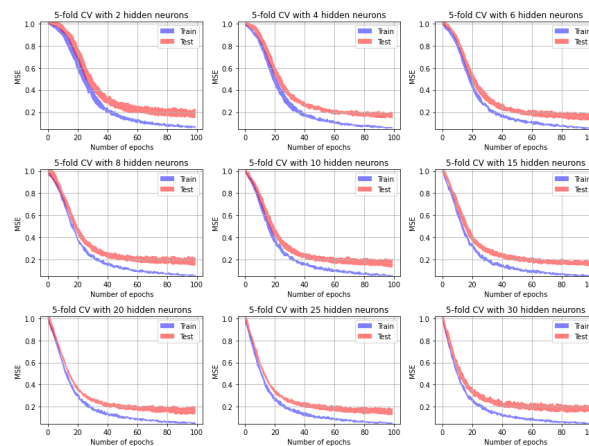- threshold at 0.0.

Results :

- MSE on train set : 0.059
- MSE on test set : 0.178.

Our confusion matrix was

[35. 1.]

[1. 35.]

We measured the performance of our model by using a 5-fold cross-validation.

Figure 3: Final Model Evaluation

We can see that our model has a good ability to generalize. Our final choice would be to use 2 or 4 neurons because the MSE result for the test set does not spread too much and it follows the train set accurately. Also, having a small number of neurons avoids the risk of overfitting and is a simple enough model. It's not a perfect model, there's still a significant gap between the train error and test error, but it does the job.

We also computed the following scores to confirm the performance of our model:

Male: - Accuracy : 0.972 - F1-Score : 0.972 **Female:** - Accuracy : 0.972 - F1-Score : 0.972

Comments

We had a problem with data normalization. At first, we normalized the female and male dataset separately, which produced a curious error. We needed to give output value between 0 and 1 instead of -1 and 1 for the tanh validation function in order to get acceptables MSE curves for both training and test sets. This problem has been fixed by merging both dataset before the normalization, which is of course the correct way to normalize a dataset. Now we have a working model which generalize two different classes.
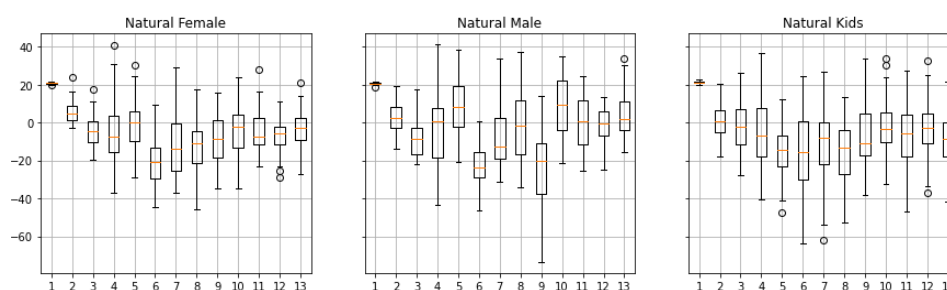
Man vs Woman vs Children

The purpose of this experiment was to train our model to be able to differentiate between men, women and kids.

> The associated notebook for this experiment is 'MaleFemaleKid-model_selection.ipynb'

Number of observations

The dataset was composed of 180 values with 13 mfccs each. This represents all the male, female and kids voices.


Figure 4: MFCCs Men Women Kids

Features to train the model

For this second part, we had the same observation as for the first. The mfccs of the three classes were too close to one another to be taken independently, but by taking the whole dataset, we were able to separate each class from another due to some little differences on several of the mfccs.

Procedure for model selection

This part required a different approach than the first one, as our goal was to classify the data into three classes instead of two. We labeled those data with three distinct columns taking the values (1,-1,-1) for male, (-1,1,-1) for female or (-1,-1,1) for kid. With this, we could use the activation function tanh in order to train and test our dataset.

Other than that, the procedure that we used to select the model was the same as for the first part, except that we specified the last three columns as classes labels to the "fit" function.
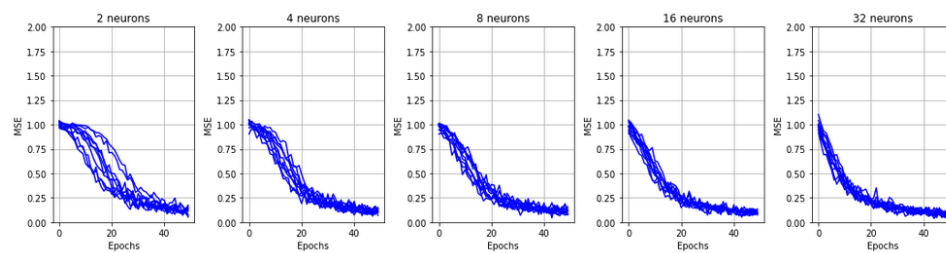
Figure 5: Exploring Number of Neurons

Description of the final model and Performance evaluation

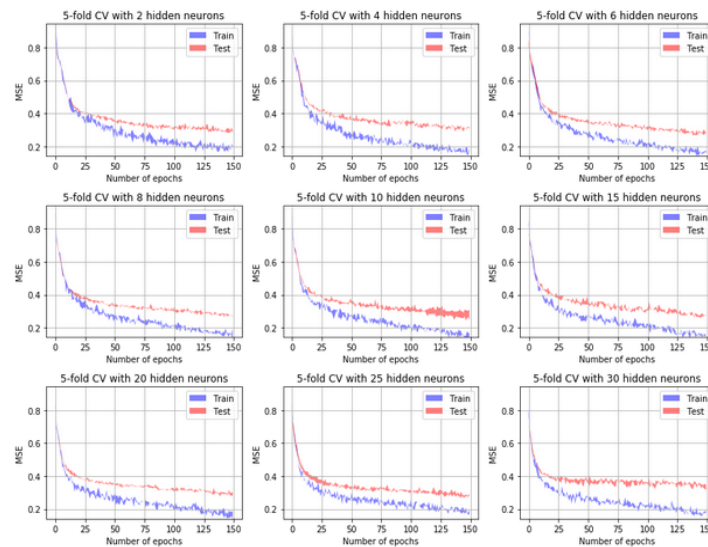Our final model uses the following hyper-parameters:

- tanh activation function
- learning rate : 0.0008
- momentum : 0.9
- 2 hidden neurons
- 3 output neurons
- Number of epochs : 150
- threshold : 0.0

Results

- MSE for train set: 0.17

- MSE for test of 0.32

- Our confusion matrix was

  [[33. 3. 0.]]

We measured the performances of our model by using a 5-fold cross-validation.

Figure 6: Final Model Test

We can directly see that this problem was harder to generalize. The result are not as good as with only men and women, which is expected. Our model still has a good ability to generalize. Our final choice would be to use 2 neurons because the MSE result for the test set does not spread too much and it follows the train set accurately. Also, having only 2 neurons avoid the risk of overfitting and is a simple enough model.

We also computed the following scores to confirm the performances of our model:

Male: - Accuracy : ~0.96 - F1-Score : ~0.96

Female: - Accuracy : ~0.97 - F1-Score : ~0.61

Kid: - Accuracy : ~0.86 - F1-Score : ~0.89

Comments

The results of our confusion matrix are inaccurate. It seems to be due to the implementation of the cross-validation function, which seems to act oddly when in presence of three classes. As a result, our F1-score and accuracies are also biased, but the different diagrams and calculation are correct. A solution to this problem would be to use sklearn instead of the given functions, but we saw this problem too late and didn't want to modify our code the day of the deadline. We can also see that our model has a hard time generalizing women as the f1-score for this particular class is significantly lower than the others. This problem may be fixed by choosing specific mfccs instead of taking the whole package.

Final experiment

The purpose of this experiment was to train our model to be able to differentiate between natural and synthetic voices.

> The associated notebook for this experiment is 'MaleFemaleKidSynthetic-model_selection.ipynb'

Number of observations

Our dataset was composed of 360 values of 13 mfccs each. We used all the natural voices values as well as all the synthetic voices. Our objective for this experiment was to classify values as either human or synthetic.
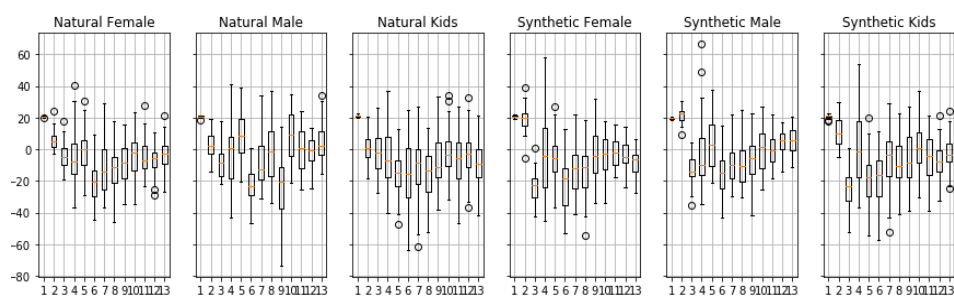
Figure 7: MFCCs Men Women Kids Synthetic

Features to train the model

In this experiment, we could see that the columns 1 and 2 of ours mfccs seemed to be sufficient to classify our dataset. We tried to apply the same method as for precedents parts, but using only those two colums and got very good curves on our graphs. But in the end, the confusion matrix wasn't good at all. We decided to go back to taking all our mfccs in order to correct the problem and to have a more generic model, and it did correct the problem.

Procedure for model selection

As our goal was to separate two classes (synthetic or human), we chose to use the same method as for the first part. We chose again to use the tanh function for this part. This problem is basically the same as the first experiment. It uses 13 inputs and outputs 2 possible values, 1 for natural and -1 for synthetic. Of course, the exploration of hyper-parameters was different as the dataset was bigger and composed of different values.

Description of the final model

The final model is similar to the first one (man and woman only) but has even better scores.
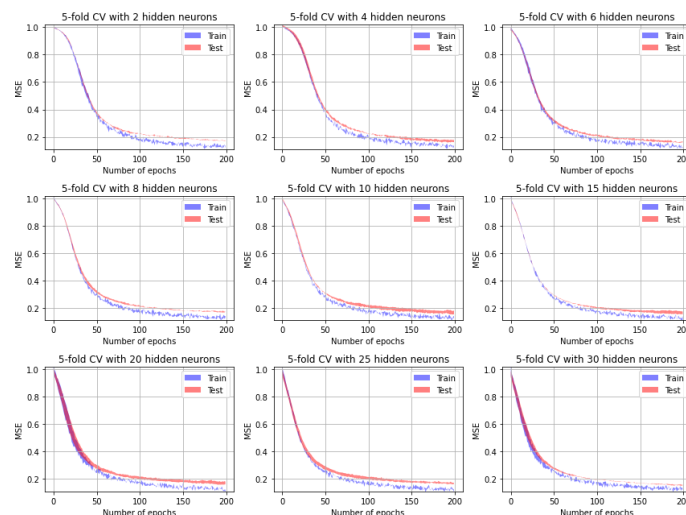
Figure 8: Final Model Men Women Kids Synthetic

Performance evaluation

Our final model uses the following hyper-parameters:

- tanh activation function
- learning rate : 0.0008
- momentum : 0.5
- 2 hidden neurons
- 1 output neurons
- Number of epochs : 250
- threshold : 0.0

We came out with the following values for the evaluation of our final model:

- MSE training : 0.117

- MSE test : 0.159

- Confusion matrix:

  [[174. 6.]]

Human: - Accuracy : ~0.95 - F1-Score : ~0.95

Synthetic: - Accuracy : ~0.95 - F1-Score : ~0.95

The results shows that this problem is not hard to generalize for our model, even with only 2 neurons.

Comments

This part was the easiest as we widely took advantage of our past experience with the two first parts of this lab.