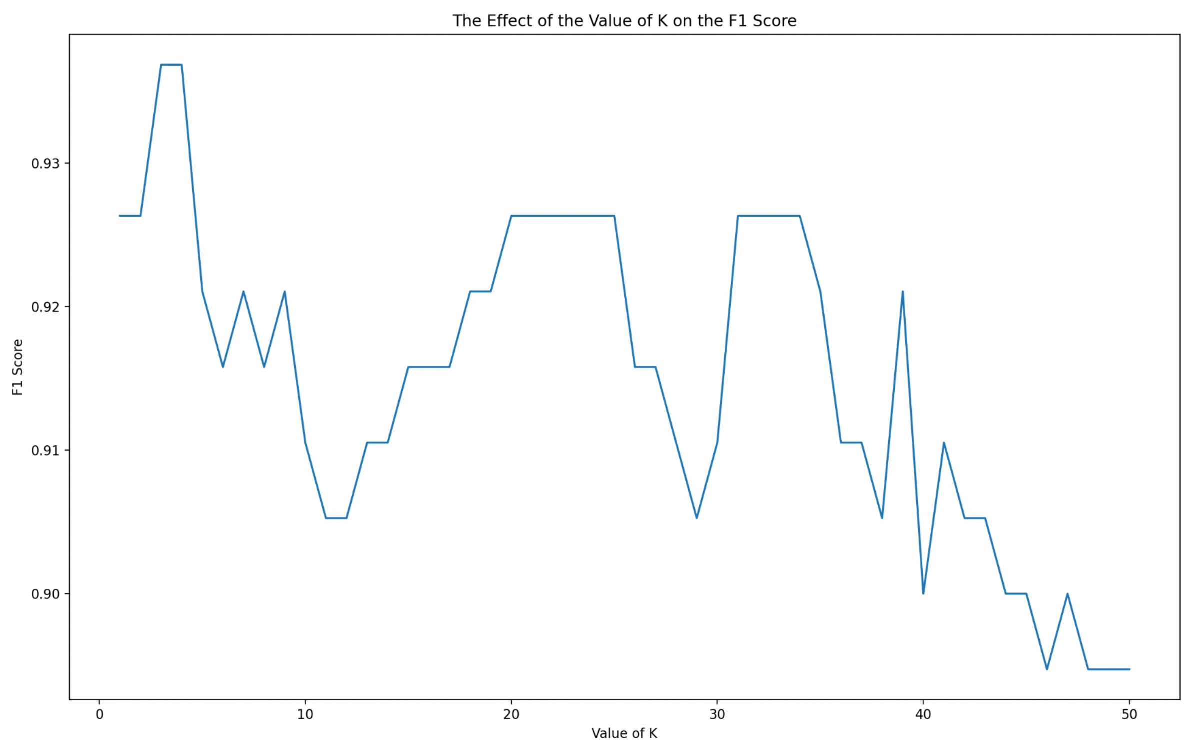
ICS2203 & ARI2203 - Speech Phoneme Analysis and Classification Assignment - Documentation

1) How Performance Changes with Different Values of K

In order to properly analyse the difference that K is having on the accuracy of the classifier, I made a method which I named **run_classifier_changing_k**.

This method is similar to the **run_classifier** method, however instead of leaving k as its default value (5), the method performs the classification algorithm 50 times, each time with a different value of k, starting from 1 and incrementing its value by 1 each iteration. While doing this, the mean F1 score of each iteration is recorded in a list which is used in order to plot a graph depicting the mean F1 score corresponding to each value of K.

The resulting graph produced can be seen below:



As can be seen from this graph, the F1 score peaks at K = 3 and K = 4, producing an F1 score of approximately 0.935. The classifier continues producing similar but lower F1 scores as K increases up till around K = 40, after which the F1 score starts to drop off quite significantly. I took note of the optimal value of K and started using K = 3 from here on out.

2) Distance Metrics Used

Prior to answering this question, I was using euclidean distance for my classifier, which is the default metric used for the **KNeighborsClassifier** function.

I later ran the classifier using Manhattan distance and Chebyshev distance and the results can be seen below.

Euclidean Distance:

```
F1 Score for Iteration 1: 0.9210526315789473
F1 Score for Iteration 2: 0.868421052631579
F1 Score for Iteration 3: 0.9736842105263158
F1 Score for Iteration 4: 0.9736842105263158
F1 Score for Iteration 5: 0.9473684210526315
Average F1 Score: 0.9368421052631579
```

Manhattan Distance:

```
F1 Score for Iteration 1: 0.9210526315789473
F1 Score for Iteration 2: 0.8947368421052632
F1 Score for Iteration 3: 0.9736842105263158
F1 Score for Iteration 4: 0.9736842105263158
F1 Score for Iteration 5: 0.9473684210526315
Average F1 Score: 0.9421052631578949
```

Chebyshev Distance:

```
F1 Score for Iteration 1: 0.9473684210526315
F1 Score for Iteration 2: 0.868421052631579
F1 Score for Iteration 3: 0.9736842105263158
F1 Score for Iteration 4: 0.9736842105263158
F1 Score for Iteration 5: 0.9473684210526315
Average F1 Score: 0.9421052631578949
```

From these results I concluded that Manhattan distance and Chebyshev produced the best results, both producing the same F1 score, that being slightly higher than the one produced when using Euclidean distance.

I also attempted to use various other distance metrics which produced some strange results. From the confusion matrices produced I could notice that for each formant value, the classifier was only classifying phonemes into 1 class, rather than into all 3 as was being done prior. This led me to believe that these metrics are not meant to be used for this specific scenario. These metrics included hamming distance, cosine similarity and jaccard distance.

3) Both Genders vs One Gender

In order to analyse the performance of the classifier when taking into consideration data from only one gender, I created a new function **extract_data_male_female** that creates two new data frames, one for the formant values produced by the males and another for those produced by the females. It also creates another two data frames which hold the phoneme labels produced by each respective gender.

These data frames are then passed into the **run_classifier** function where the classifier is trained and tested with data gathered from only one of the genders.

Below are the results which I obtained (K = 3 and Manhattan Distance):

Both Genders:

```
F1 Score for Iteration 1: 0.9210526315789473
F1 Score for Iteration 2: 0.8947368421052632
F1 Score for Iteration 3: 0.9736842105263158
F1 Score for Iteration 4: 0.9736842105263158
F1 Score for Iteration 5: 0.9473684210526315
Average F1 Score: 0.9421052631578949
```

Female:

```
F1 Score for Iteration 1: 1.0
F1 Score for Iteration 2: 0.9473684210526315
F1 Score for Iteration 3: 1.0
F1 Score for Iteration 4: 1.0
F1 Score for Iteration 5: 0.9473684210526315
Average F1 Score: 0.9789473684210526
```

Male:

```
F1 Score for Iteration 1: 1.0
F1 Score for Iteration 2: 0.8421052631578947
F1 Score for Iteration 3: 0.8947368421052632
F1 Score for Iteration 4: 1.0
F1 Score for Iteration 5: 1.0
Average F1 Score: 0.9473684210526315
```

As can be seen from the results, the predictions are more accurate when only data from one gender is considered, which holds up to certain studies such as [1] which have concluded that certain phoneme formant values vary depending on gender.

4) Phonemes that Produced the most Confusion

From the 5 iterations which the **run_classifier** function performs I was unable to determine any phonemes which produced confusion as the classifier performed quite well in these scenarios. Because of this, I increased the number of iterations to 20 in the hopes that the new iterations would show some of these phonemes.

Overall, I still did not notice a standout confusing phoneme, however the most common confusion which I noticed was that IY phonemes were being classified as EH phonemes and EH phonemes were being classified as IY phonemes, as well as AA phonemes being classified as EH. This slight confusion could be due to various reasons, however I suspect that it is due to some unorthodox pronunciations of the words “heed” and “head” and “hod” caused by the accents spoken by the people in the dataset which causes a vowel to sound similar to another.

References

[1] E. Pépiot, "Male-female acoustic differences and cross-language variation in English and French speakers," Corela, June 2012. doi: 10.4000/corela.3783