GENDER RECOGNITION BY VOICE

Project for the course

MATH-412 Statistical Machine Learning,
Ecole Polytechnique Federale de Lausanne,
Fall 2017

Authors: Besson Adrien, Hippolyte Lefebvre, Greg
Professor: Dr. Emeric Thibaud

December 15, 2017

Contents

List of Figures

List of Tables

1 Objective of the project

2 The dataset

2.1 General considerations

The voice gender dataset[1] consists of features extracted from 3168 recorded voice samples, collected from male and female speakers. The features have been computed using tuneR[2] and seewave[3], two acoustic analysis packages of R.

The dataset takes the form of a csv files where each row is composed of the following acoustical features of each voice:

- **meanfreq:** mean frequency (in kHz)
- **sd:** standard deviation of frequency
- **median:** median frequency (in kHz)
- **Q25:** first quantile (in kHz)
- **Q75:** third quantile (in kHz)
- **IQR:** interquantile range (in kHz)
- **skew:** skewness of the spectrum
- **kurt:** kurtosis
- **sp.ent:** spectral entropy
- **sfm:** spectral flatness
- **mode:** mode frequency
- **centroid:** frequency centroid
- **peakf:** peak frequency (frequency with highest energy)
- **meanfun:** average of fundamental frequency measured across acoustic signal
- **minfun:** minimum fundamental frequency measured across acoustic signal
- **maxfun:** maximum fundamental frequency measured across acoustic signal
- **meandom:** average of dominant frequency measured across acoustic signal
- **mindom:** minimum of dominant frequency measured across acoustic signal
- **maxdom:** maximum of dominant frequency measured across acoustic signal
- **dfrange:** range of dominant frequency measured across acoustic signal
- **modindx:** modulation index. Calculated as the accumulated absolute difference between adjacent measurements of fundamental frequencies divided by the frequency range
- **label:** male or female

The features are all quantitative and represents frequency characteristics of the voices.

[1] `https://www.kaggle.com/primaryobjects/voicegender`
[2] `https://cran.r-project.org/web/packages/tuneR/tuneR.pdf`
[3] `https://cran.r-project.org/web/packages/seewave/seewave.pdf`

2.2 Description of the features

Before starting the data analysis, it is important to perfectly understand the features involved in the exercise. This will be very useful in a preprocessing step, since it will allow us to remove collinear features. It will also be a great asset when it will come to the analysis of the most important features in the gender recognition.

As already pointed out in Section 2.1, the extracted features are all related to the spectrum.

Frequency-related features The mean frequency corresponds to a weighted average of the frequency by the amplitude of the spectral components:

$$\mu_f = \sum_{i=1}^{N} f_i y_i, \tag{2.1}$$

where N is the number of frequency components of the spectrum, f_i is the i-th frequency and y_i is the relative amplitude of the i-th component of the spectrum. As described in p.163 of the seewave documentation, it is equal to the feature 'centroid'. The standard deviation is calculated as:

$$\sigma_f = \sqrt{\sum_{i=1}^{N} y_i \left(f_i - \mu_f\right)^2} \tag{2.2}$$

The median frequency is calculated as the frequency where the spectrum is divided into frequency intervals of same energy. The calculation of the quartiles are based on the same criterion. The interquartile range is calculated as the difference between the third and the first quartile.

The feature 'mode' characterizes the dominant frequency of the spectrum, *i.e.*the one with the highest amplitude. It is very similar to the peak frequency which corresponds to the frequency with the highest energy. The fundamental frequency is the lowest frequency of the spectrum.

The features 'meanfun', 'minfun', 'maxfun', 'meandom', 'maxdom', 'mindom', 'dfrange' and 'modindx' are based on short-time Fourier transform applied on segments of fixed durations, small compared to the duration of the whole signal. This permits to have features more localized in time.

In addition to the frequency-related features, we can find measures on the shape of the spectrum which may give very interesting additional information.

Skewness of the spectrum The skewness of the spectrum is a measure of its asymmetry around the mean frequency. It is calculated as follows:

$$S = \frac{1}{\sigma_f^3} \frac{\sum_{i=1}^{N} \left(f_i - \mu_f\right)^3}{N-1}. \tag{2.3}$$

From (2.3), it is clear that the sign of S gives information of the left or right asymmetry of the spectrum while the absolute value of S gives the strength of the asymmetry.

Kurtosis The Kurtosis is a measure of the "tailedness" of a probability distribution. It is calculated as the fourth order moment of the frequency distribution, described below:

$$K = \frac{1}{\sigma_f^4} \frac{\sum_{i=1}^{N} \left(f_i - \mu_f\right)^4}{N-1}. \tag{2.4}$$

When $K = 3$, the frequency distribution is normal. When $K < 3$, the frequency distribution is said to be *platikurtic*, it has fewer items around the means than in the tails, compared to a normal distribution. When $K > 3$, the distribution is said to be *leptokurtic* and has more frequency around the mean than in the tails, compared to a normal distribution.

Shannon spectral entropy The Shannon entropy is used to discriminate whether the voice signal is noisy or pure **Nunes2004** it is calculated as follows:

$$H = \frac{-\sum_{i=1}^{N} y_i \log_2 (y_i)}{\log_2 (N)} \tag{2.5}$$

If the signal is pure, then all the energy is concentrated in one frequency component, let us say the j-th component for which $y_j = 1$. In this case, $H = 0$. If the signal is a white noise, then $y_i = 1/N$, $\forall i \in \{1, ..., N\}$ and $H = 1$.

Spectral flatness The spectral flatness is rather similar to the spectral entropy. It is measured as the ratio between the geometric mean and the arithmetic mean:

$$F = N \frac{\sqrt[N]{\prod_{i=1}^{N} y_i}}{\sum_{i=1}^{N} y_i}. \tag{2.6}$$

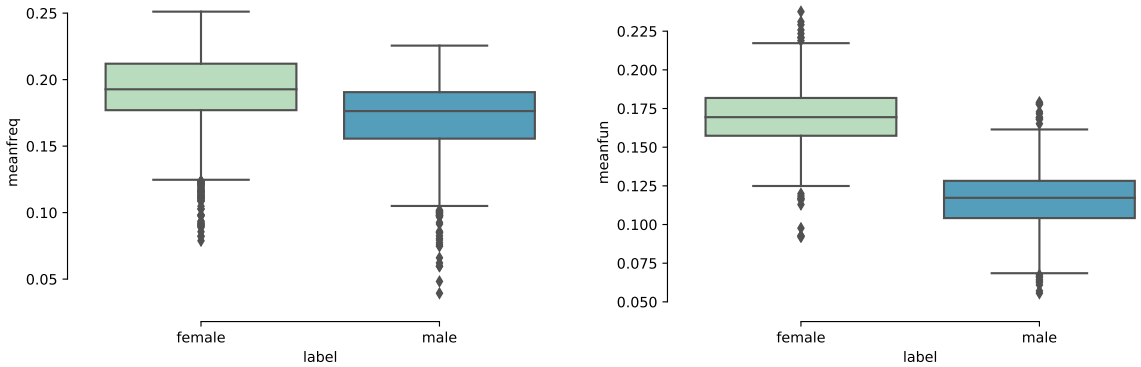
In case of a white noise, the spectrum is flat and $H = 1$. In case of a pure tone, the geometrical mean is equal to zero and $H = 0$.

2.3 Cleaning the dataset

From the description of the features given in Section 2.2, a first cleaning of the dataset may be achieved before starting the analysis. Indeed, several features are exactly the same of collinear:

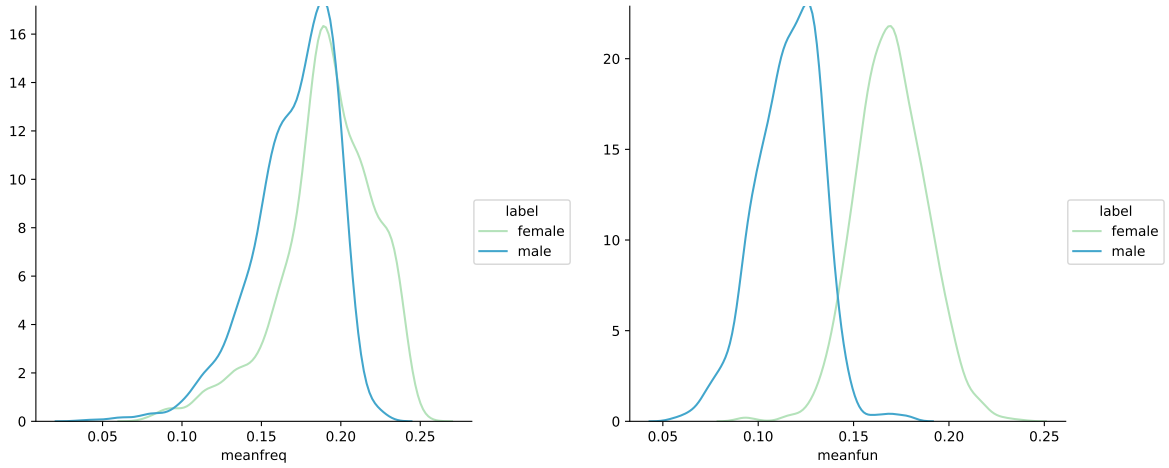
- The features 'meanfreq' and 'centroid' are exactly similar. So 'centroid' has been removed;
- The following relationship holds: $'IQR' =' Q75' -' Q25'$. 'IQR' has been removed.
- The following relationship holds: $'dfrange' =' maxdom' -' mindom'$. 'dfrange' has been removed.

3 Exploratory Data Analysis



(a) (b)

Figure 3.1



(a) (b)

Figure 3.2

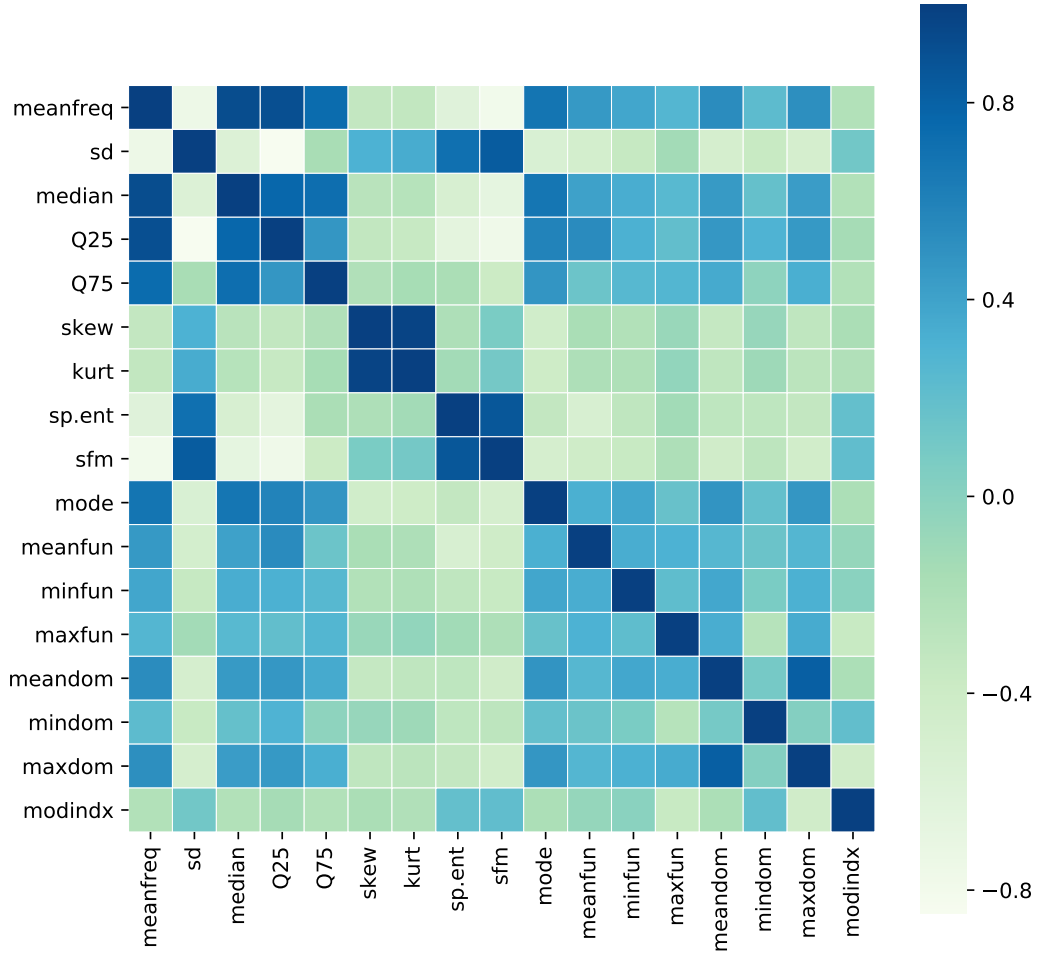

meanfreq
sd
median
Q25
Q75
skew
kurt
sp.ent
sfm
mode
meanfun
minfun
maxfun
meandom
mindom
maxdom
modindx
0.8
0.4
0.0
−0.4
−0.8

Figure 3.3

4 Evaluation of the Best Classification Method

4.1 Considered classification methods

4.2 The naive strategy

4.2.1 Description

4.2.2 Results

Table 4.1 Classification Error of the Methods for Different Seed Numbers

Type	Methods	Seed number				
		1	2	3	4	5
Max. Likelihood	Logistic reg.	0.0158	0.0347	0.0315	0.0237	0.0221
	Logistic reg. - Ridge	0.0158	0.0315	0.0315	0.0189	0.0284
	Logistic reg. - Lasso	0.0315	0.0363	0.0379	0.0284	0.0300
	LDA	0.0315	0.0410	0.0379	0.0284	0.0268
	QDA	0.0347	0.0347	0.0347	0.0268	0.0363
Trees	Tree	0.0379	0.0426	0.0315	0.0284	0.0300
	Pruned Tree	0.0394	0.0473	0.0347	0.0363	0.0300
	Bagging	0.0237	0.0410	0.0142	**0.0174**	0.0284
	Random Forest	0.0189	0.0347	**0.0126**	0.0205	0.0205
	XGBoost	0.0189	**0.0268**	**0.0126**	0.0205	**0.0189**
SVM	Linear	**0.0142**	0.0315	0.0284	0.0189	0.0252
	Gaussian	0.0158	0.0315	0.0284	0.0189	0.0221
x	kNN	0.0300	0.0347	0.0252	0.0379	0.0300

A Heading on Level 0 (chapter)

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

A.1 Heading on Level 1 (section)

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

A.1.1 Heading on Level 2 (subsection)

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Heading on Level 3 (subsubsection)

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Heading on Level 4 (paragraph) Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how

the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

A.2 Lists

A.2.1 Example for list (itemize)

- First item in a list
- Second item in a list
- Third item in a list
- Fourth item in a list
- Fifth item in a list

Example for list (4*itemize)

- First item in a list
  - First item in a list
    - First item in a list
      - First item in a list
      - Second item in a list
    - Second item in a list
  - Second item in a list
- Second item in a list

A.2.2 Example for list (enumerate)

1. First item in a list
2. Second item in a list
3. Third item in a list
4. Fourth item in a list
5. Fifth item in a list

Example for list (4*enumerate)

1. First item in a list
   1.1. First item in a list
      1.1.1. First item in a list
         1.1.1.1. First item in a list
         1.1.1.2. Second item in a list
      1.1.2. Second item in a list
   1.2. Second item in a list
2. Second item in a list

A.2.3 Example for list (description)

First item in a list

Second item in a list

Third item in a list

Fourth item in a list

Fifth item in a list

Example for list (4*description)

First item in a list

First item in a list

First item in a list

First item in a list

Second item in a list

Second item in a list

Second item in a list

Second item in a list