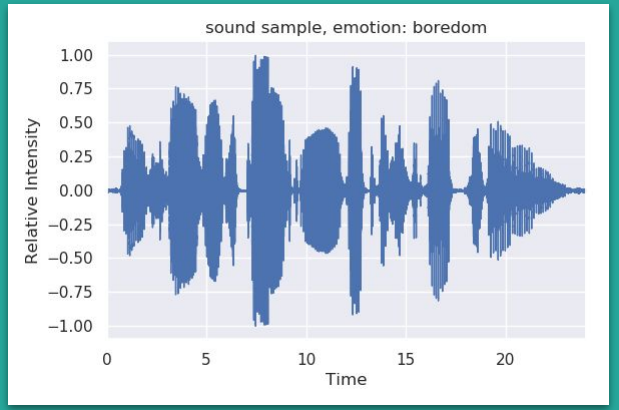
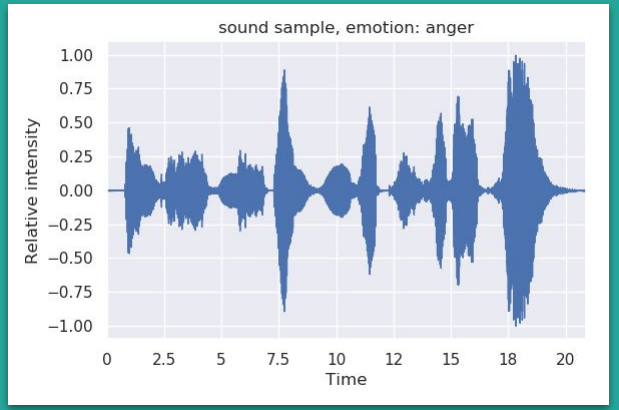
Speech emotion recognition

Charles Dufour

Given an audio recording, can we identify the emotion of the speaker ?



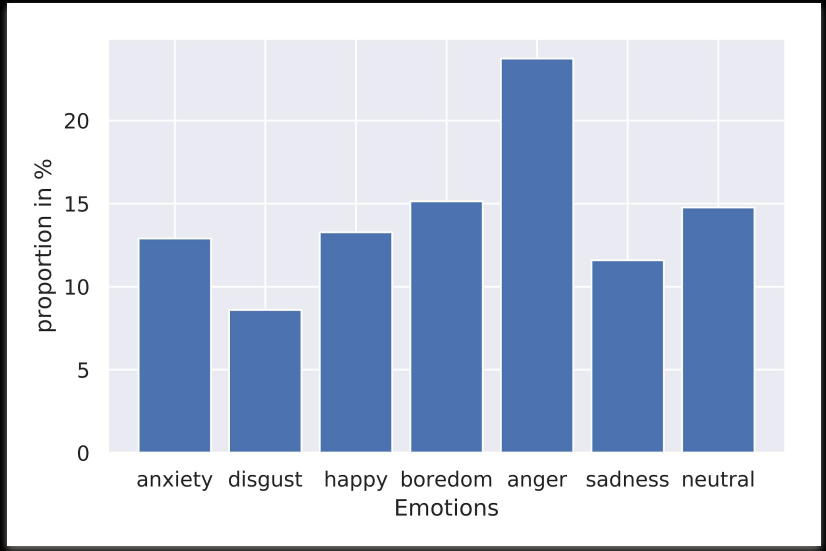


The Data

Source

Emo-DB

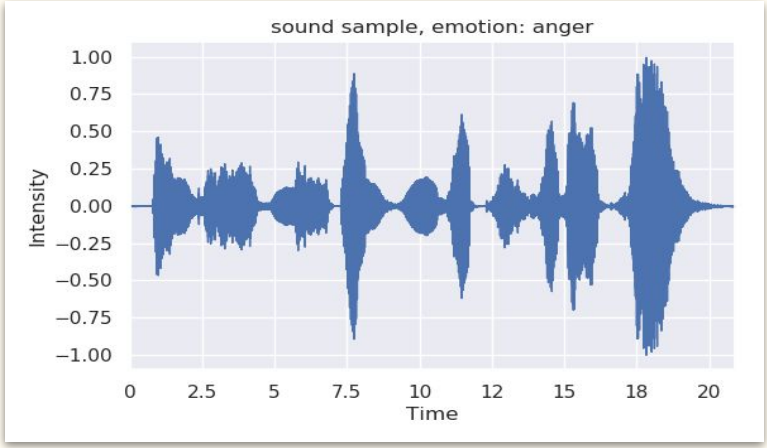
- Samples spoken by actors
- 535 audio files, 7 emotions
- Great quality



Building the features

1. Mel-frequency cepstral coefficients (MFCC)
2. Spectral components (steepness changes, filter banks,...)

   Only report statistics of these components (mean, standard deviation, minimum, maximum,...)
3. Ratio spoken time vs unspoken time

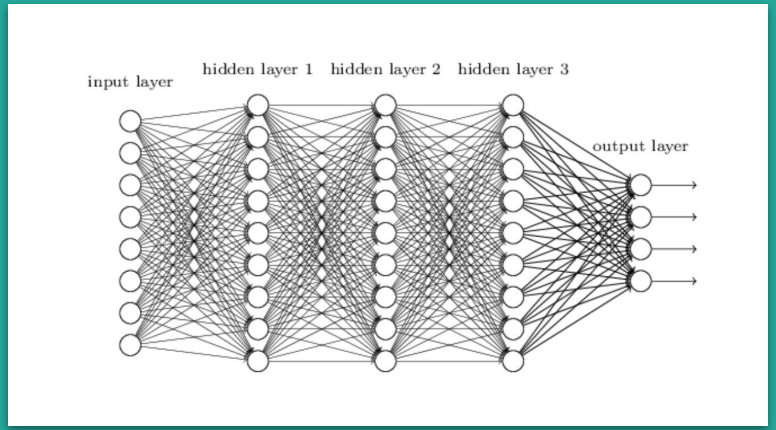


Libraries used:

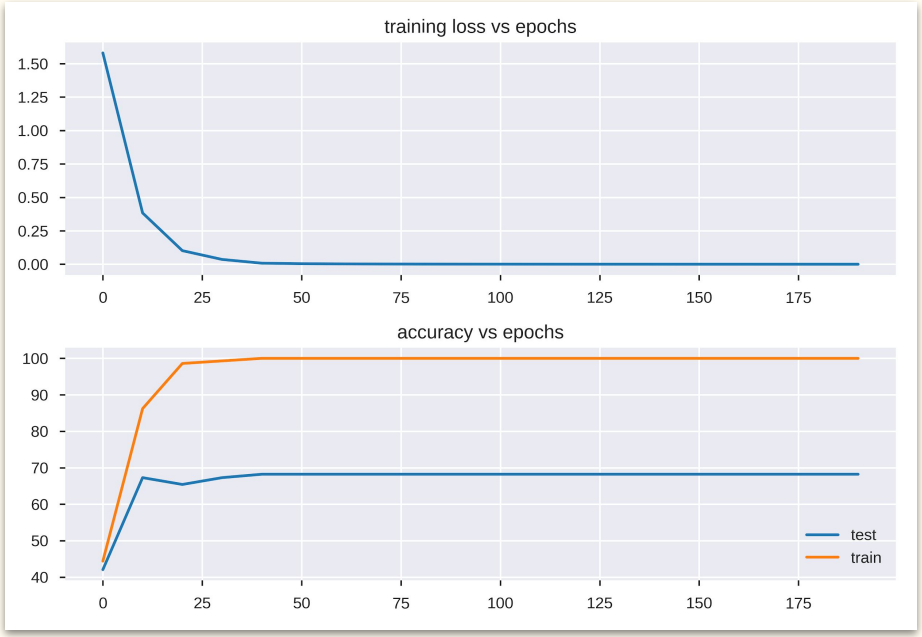
Librosa, python_speech_features, ...
Sklearn, Torch, seaborn,..

Model

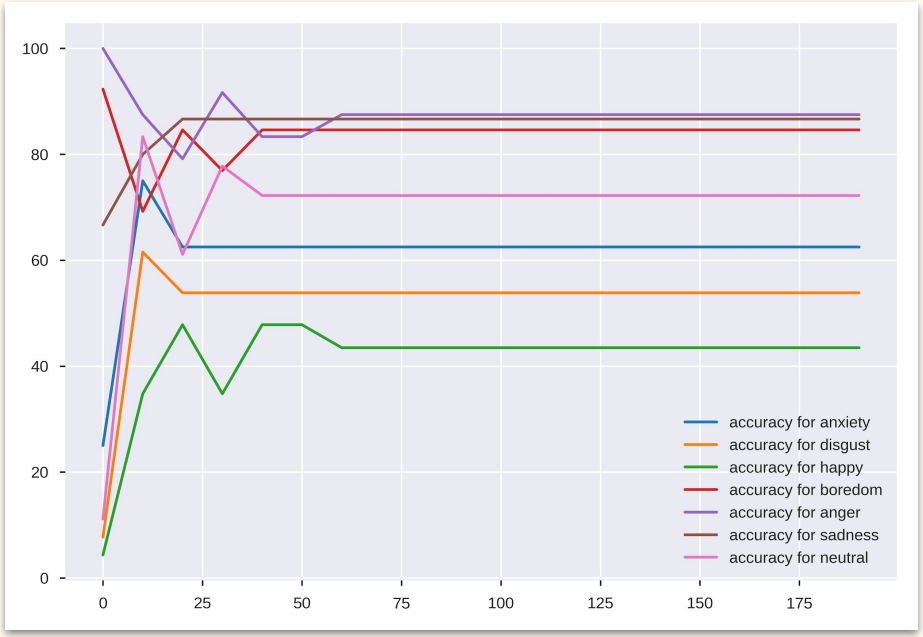
Fully connected neural network with 4 hidden layers and Relu activation



Results



global accuracy vs epochs for one split of the data

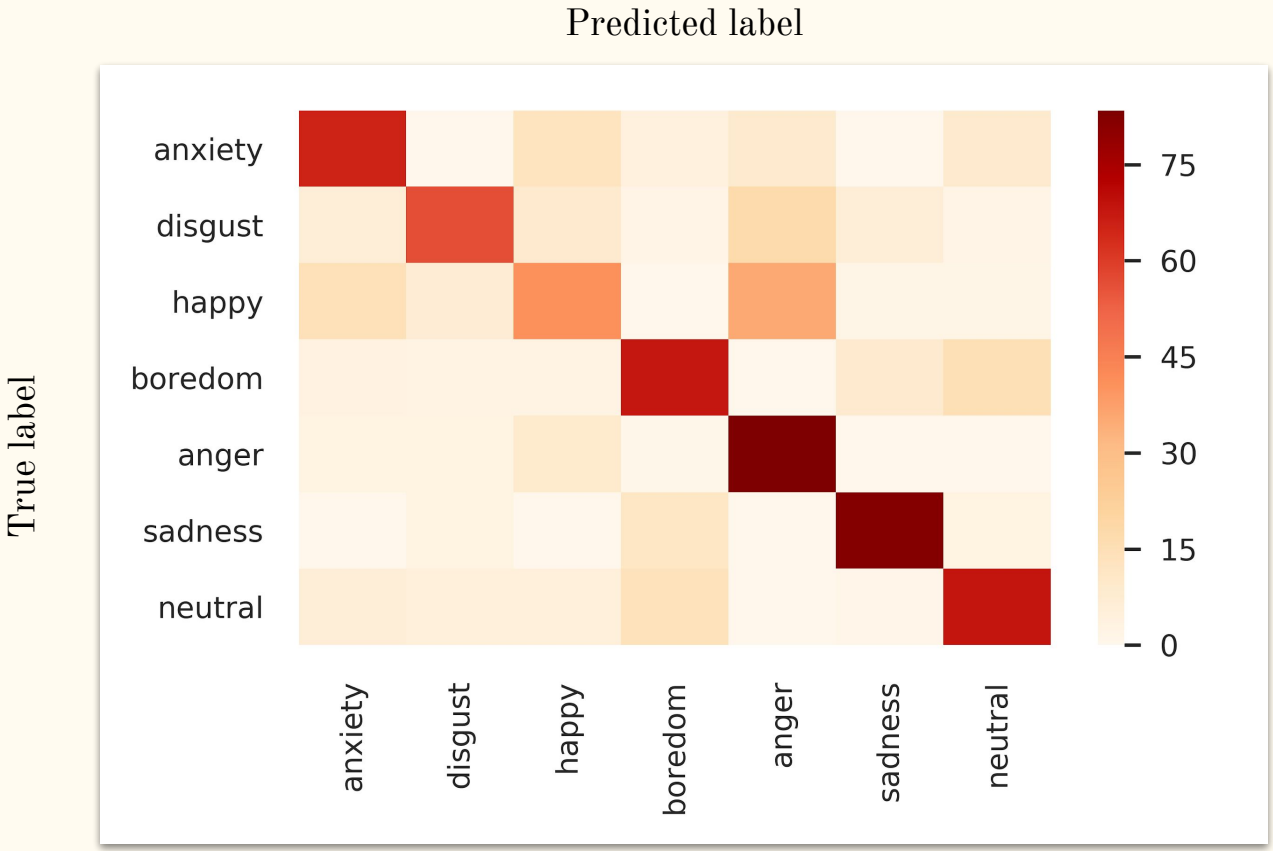


accuracy per class vs epochs for one split of the data

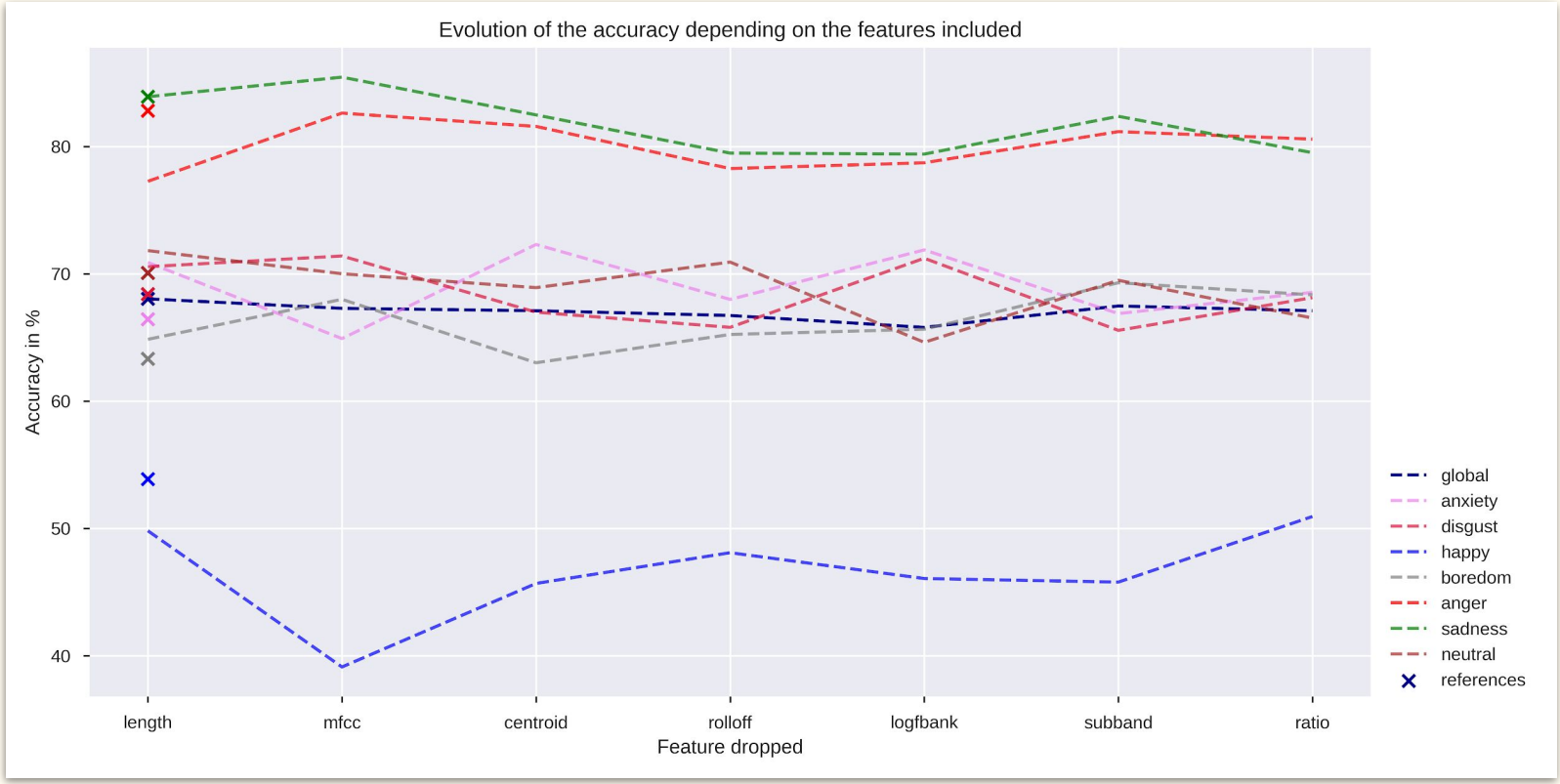
Accuracy
Obtained with 5-fold CV

global	anxiety	disgust	happy	boredom	anger	sadness	neutral
68.41	68.97	64.02	46.45	69.68	84.13	83.82	70.35

Confusion matrix (in %)



Influence of the features



Sources

- **Emotion recognition from the human voice**. Parlak, Cevahir & Diri, Banu. (2013). *21st Signal Processing and Communications Applications Conference, SIU 2013. 1-4. 0.1109/SIU.2013.6531196*
- **Speech Emotion Recognition: Methods and Cases Study**. Kerkeni, Leila & Serrestou, Youssef & Mbarki, Mohamed & Raoof, Kosai & Mahjoub, Mohamed. (2018). *175-182. 10.5220/0006611601750182*
- **Emotion Recognition from Speech using Discriminative Features**. Chandrasekar, Purnima & Chapaneri, Santosh & Jayaswal, Deepak. (2014). *International Journal of Computer Applications. 101. 31-36. 10.5120/17775-8913*
- Berlin emotional speech database
- Librosa: Audio and music signal analysis in python.McFee, Brian, Colin Raffel, Dawen Liang, Daniel PW Ellis, Matt McVicar, Eric Battenberg, and Oriol Nieto
- **The Elements of Statistical Learning** ,Hastie, T.; Tibshirani, R. & Friedman, J. (2001), *Springer New York Inc. , New York, NY, USA* .
- **Feature extraction Mel frequency cepstral coefficients (MFCC)**, Mustafa Yankayış, *pdf presentation link*