UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Sound classification of the UrbanSound8K

Subject: Pattern Recognition and Machine Learning

Professors: Andreu Català
Cecilio Angulo
Raul Benítez

Students: Fàtima El Baghdadi
Gerard Franco Panadés
David Pérez Ruiz

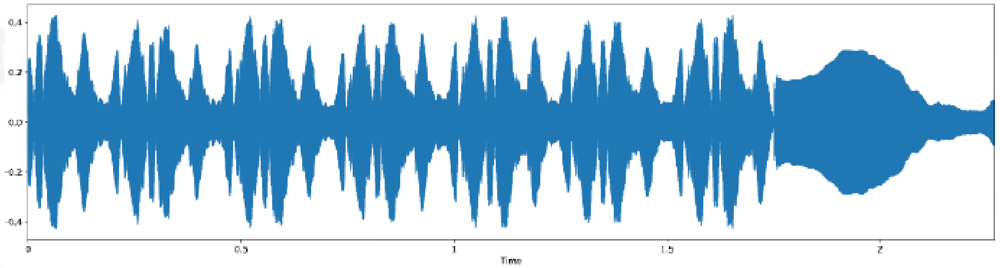
Objectives

Develop an audio classifier

Capable of distinguishing between the several sound classes

- 0 = air conditioner
- 1 = car horn
- 2 = children playing
- 3 = dog bark
- 4 = drilling
- 5 = engine idling
- 6 = gun shot
- 7 = jackhammer
- 8 = siren
- 9 = street music

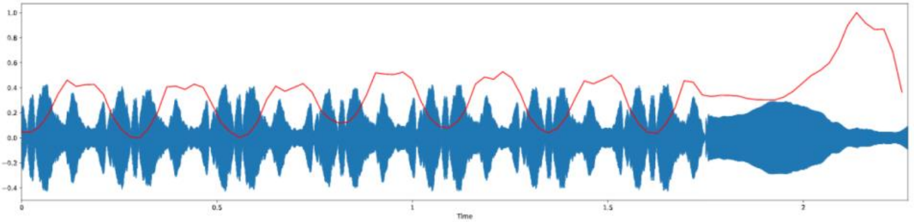
Different labeled classes from the Urban Sound Dataset



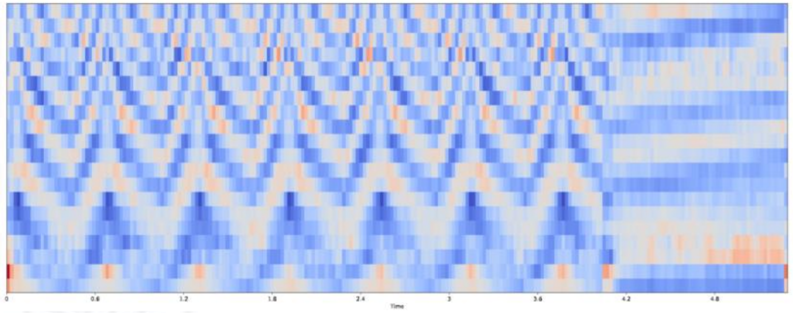
Plot of the audio array in waveform manner

Candidate features

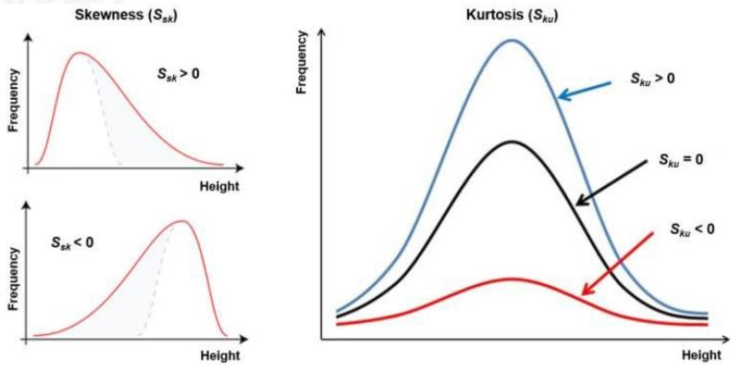
1. Zero Crossing Rate (1x1 dimension)
2. Spectral Centroid (1x1 dimension)
3. Spectral Rollof (1x1 dimension)
4. Mel Frequency Cepstrum Coefficient (MFCC) (13x1 dimension)
5. Spectral Flatness (1x1 dimension)
6. Skewness (1x1 dimension)
7. Kurtosis (1x1 dimension)
8. Spectral Entroypy (1x1 dimension)
9. Tempo (1x1 dimension)
10. Tonal centroids (6x1 dimension)



Plot Spectral Centroid (red) with the waveform (blue)

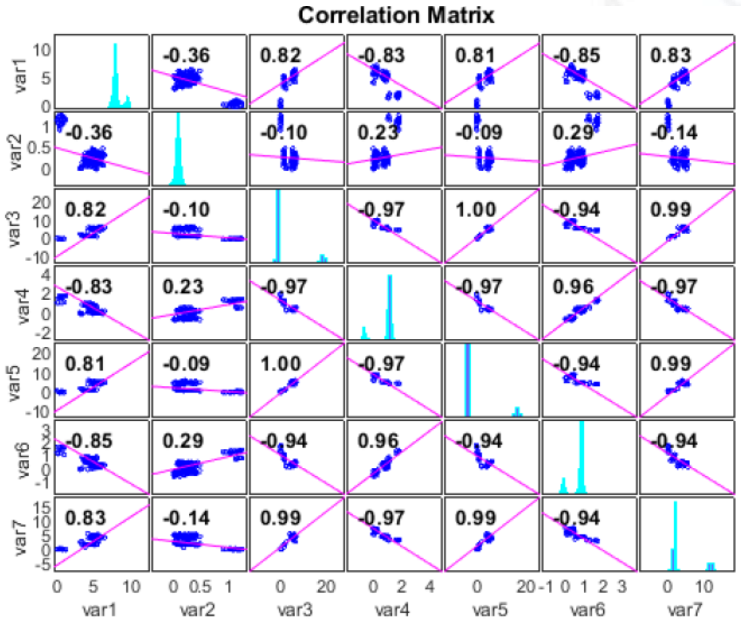


MFCCs scaled such that each coefficient dimension has zero mean and unit variance

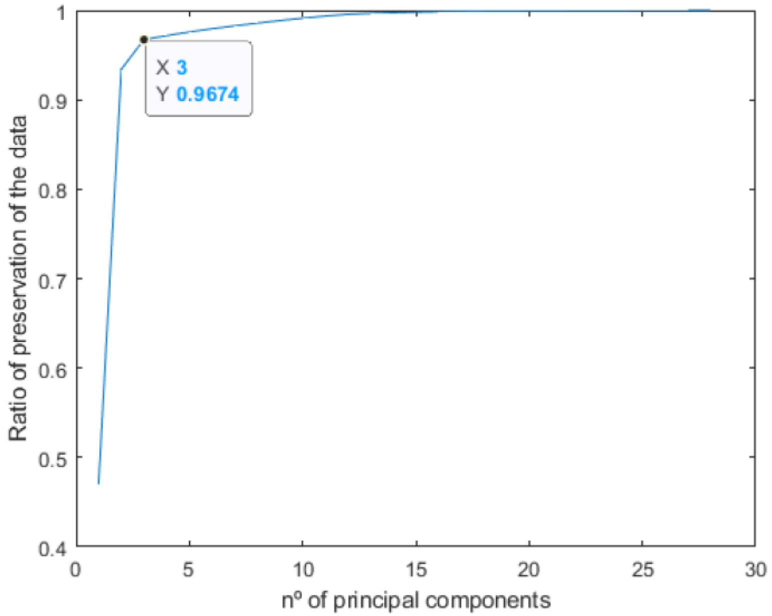


Skewness and Kurtosis graphically explained

Reduction of the features



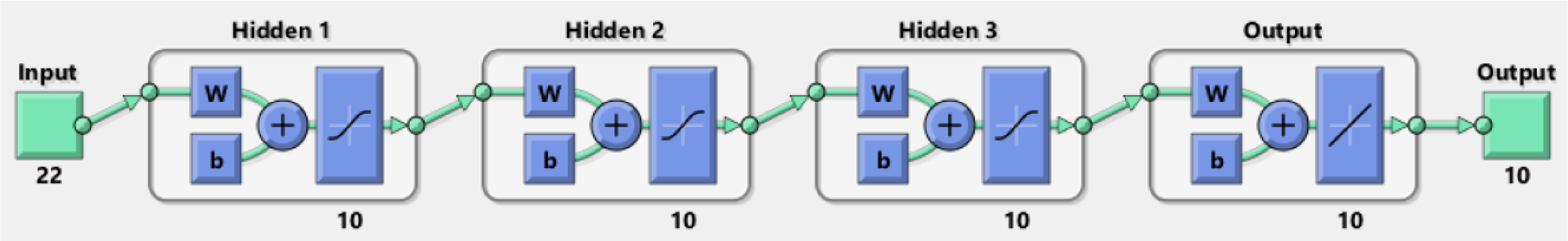
Correlation matrix with 7 features of the original dataset



Number of components that explain the data

As it can be perceived, it is needed 3 components to explain the 97% of the data.

First classifier: NN



Configuration of the network.

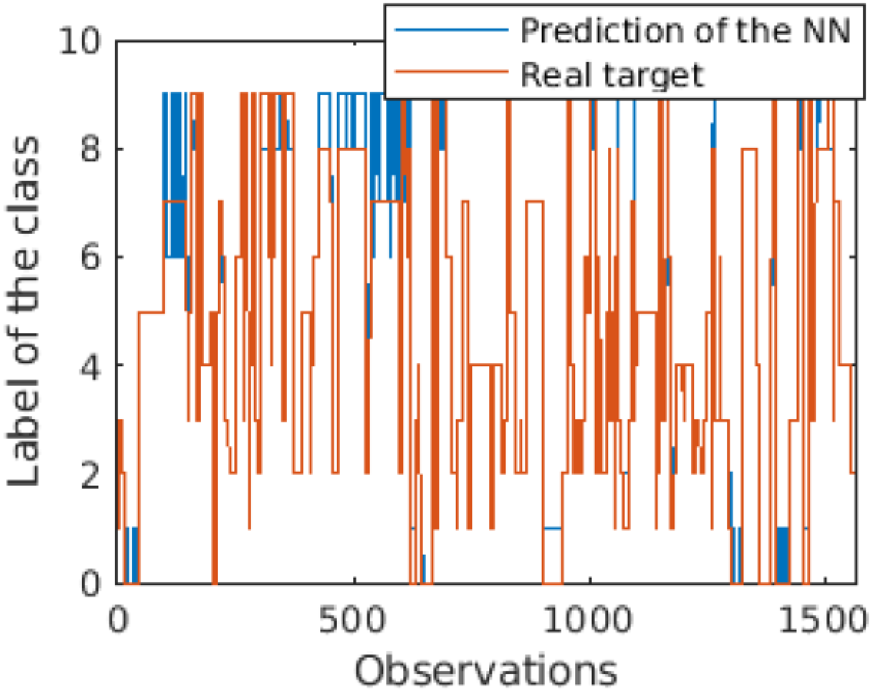
3 methods to explain the different types of training functions

1. Levenberg-Marquardt
2. Bayesian Regularization
3. BFGS Quasi-Newton

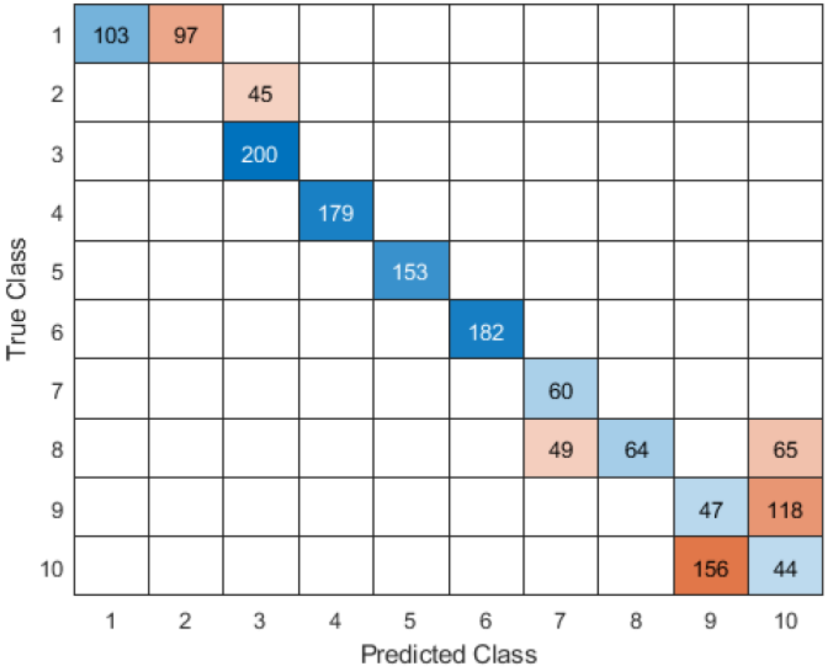
Training functions	1	2	3
Accuracy(%) all features	100	100	24.9
Accuracy(%) applying PCA	66.1	100	77.8

UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

First classifier: NN



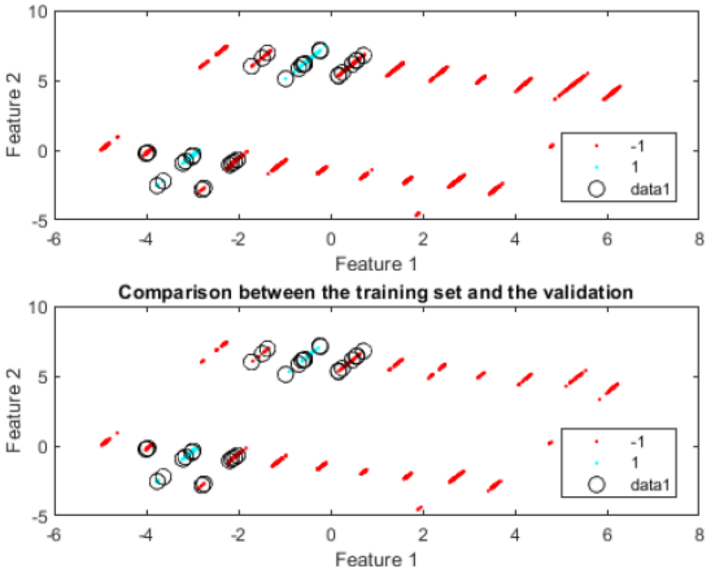
Results between the prediction and real target using Levenberg-Marquardt



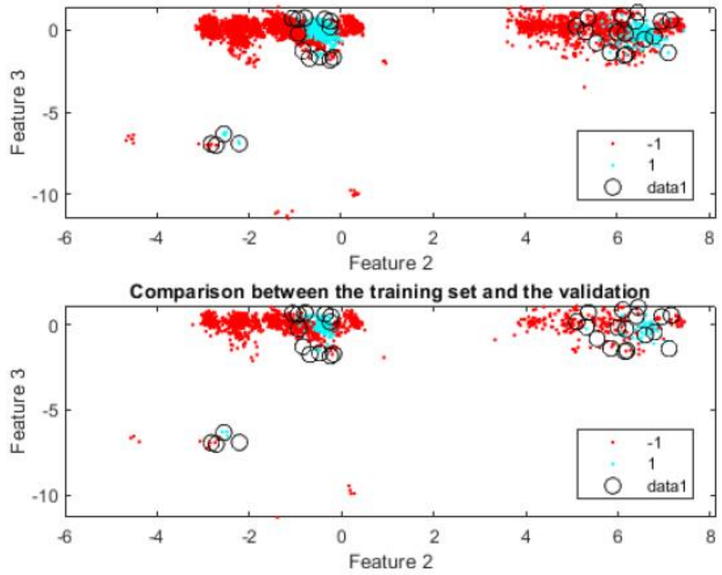
Plot of the confusion matrix of NN applied to the validation data.

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

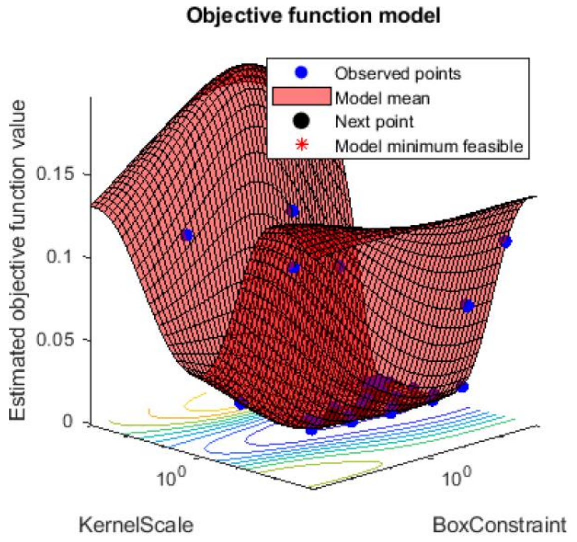
Second classifier: SVM



First and second features and support vectors for the training (above) and testing (below)



Second and third features and support vectors for the training (above) and testing (below)

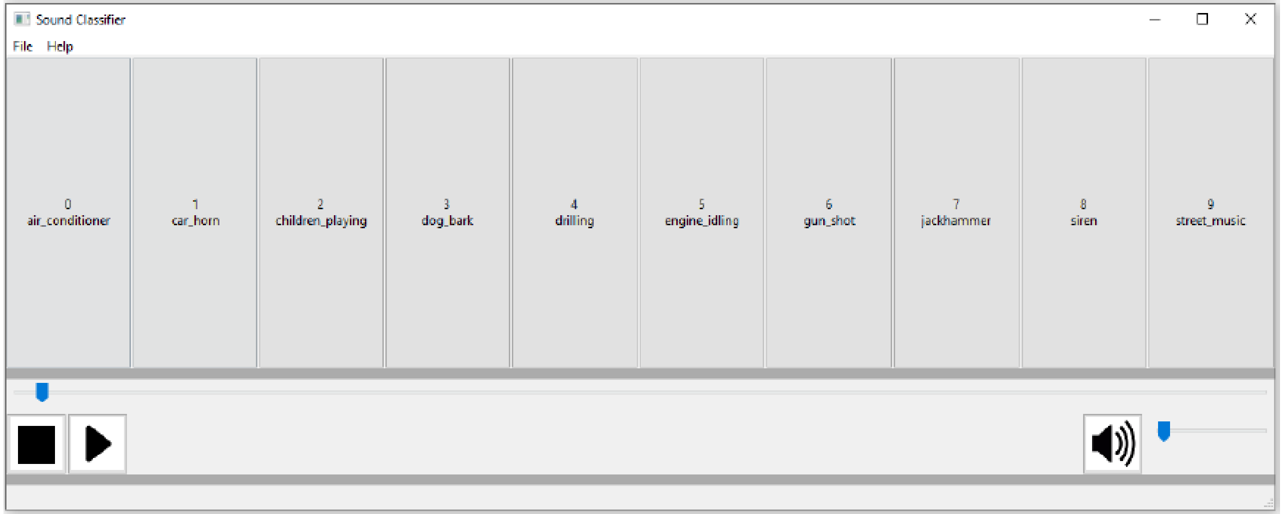


Plot of the objective function when the class is 2

10 different models with its optimized parameters → Observing results in training and predicting → Accuracy obtained with SVM is 99,88%, ✓

Interface to classify new sounds

- To annotate new incoming data that were to make its way to the existing dataset.
- Make it easier and less time consuming for the user to store a label.
- Created with the framework wxPython, a cross-platform GUI toolkit for the Python language.



Snapshot of the interface created to ease the data labelling process

Conclusions

- Analyzing a dataset and extracting the features.
- Applying methods studied during the course and prove how it changes the performance tuning the different parameters of these methods.
- Best results using the Levenberg-Marquardt and Bayesian Regularization functions with 1000 epochs.
- Respect the SVM classification, a very good performance has been proven.
- The computational cost is lower using SVM respect the NN, however the accuracy using NN is 100 respect the 99,8% of the SVM.
- Finally, an interface has been created to obtain new sounds.

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Thank you