**INSTITUTO TECNOLOGICO Y DE ESTUDIOS SUPERIORES DE OCCIDENTE**

Imagen que contiene Logotipo

Descripción generada automáticamente

**MASTER’S IN DATA SCIENCE**

**REPORT #2:**

**Descriptive Analysis of Data and Comparative Analysis of the Model**

Directors: PhD. Jorge Arturo Pardiñas Mir, PhD. Riemann Ruiz Cruz

Presents: Ing. Adrián Ramos Pérez

adrian.ramos@iteso.mx

April 25th, 2022

CONTENT TABLE

[1. Descriptive Analysis of the Data (Signals) 3](#_Toc102022212)

[1.1 Obtención de la Respuesta al Impulso 5](#_Toc102022213)

[2. Potential models to solve the problem 5](#_Toc102022214)

[2.1 Baseline techniques 5](#_Toc102022215)

[2.1.1 Support Vector Machine (SVM) 5](#_Toc102022216)

[2.2 Convolutional Neural Network (CNN) 6](#_Toc102022217)

[2.2.1 Arquitectura de la Red 6](#_Toc102022218)

[2.2.2 Activation Function 6](#_Toc102022219)

[2.2.3 Regularización 7](#_Toc102022220)

[2.2.4 Cálculo del Error y Eficiencia del Modelo 7](#_Toc102022221)

[Particle Swarm Optimization 7](#_Toc102022222)

[Bibliografía 7](#_Toc102022223)

# Descriptive Analysis of the Data (Signals)

A signal per se is a representation of real world, and thus information or data.

Audio signals are considered the data for the proposed problem.

As digital signals are a representation of a real analog signal, amplitude of the digital signal is quantized and limited by the Analog-to-Digital converter (ADC) used to convert the signal, which is part of the sound card of a computing system and thus lies in a range from -1 to 1 Volts.

Sampling frequency shall follow Nyquist theorem to ensure there is no aliasing, which states that sampling frequency should be at least double the maximum frequential content of the signal.

A monoaural speech signal, with a duration of 2 seconds and sampled at 22.05 kHz is being used for experimental purposes, the signal wave can be seen in Figure 1:

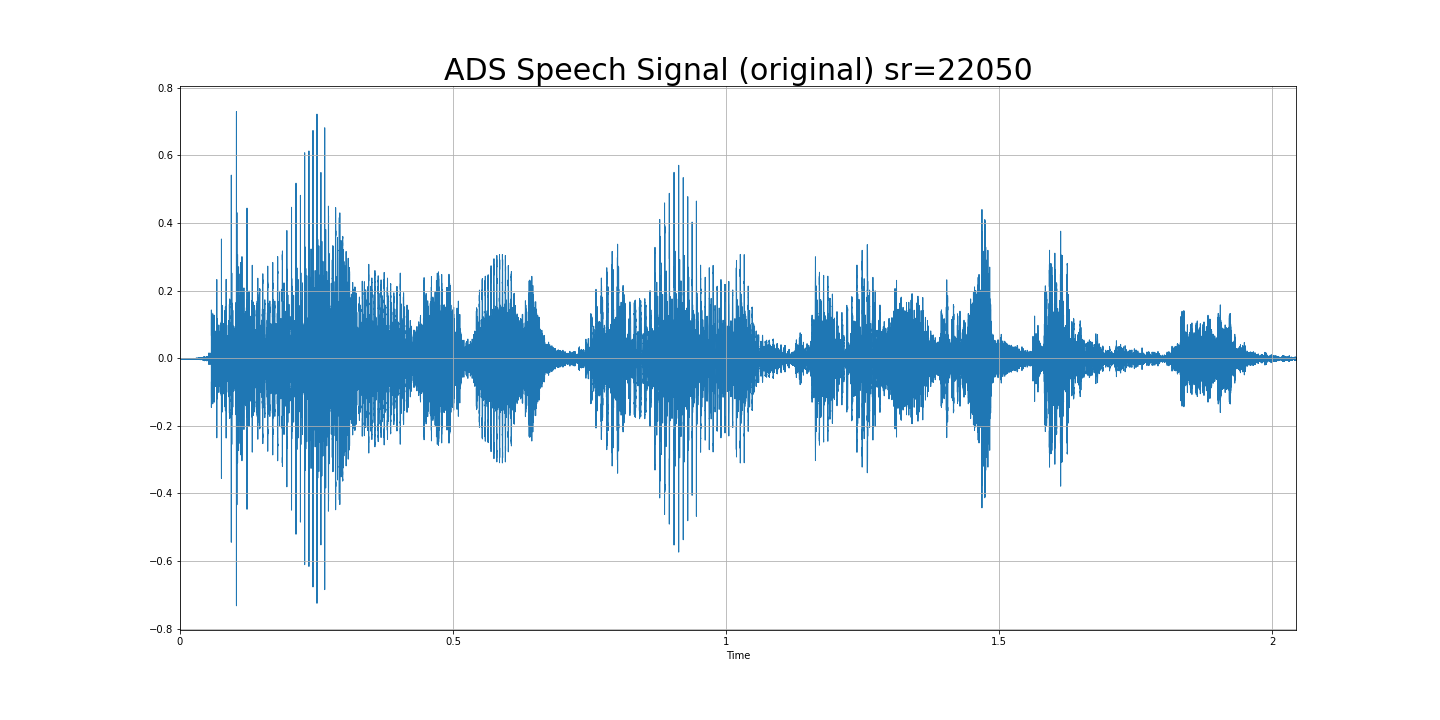


Figure 1. Original Speech signal of 2 seconds lenght at a sampling rate of 22.05 kHz

To simulate a room’s impulse response the proposed signal is one coming from a room with echo.

Gráfico, Gráfico de cajas y bigotes

Descripción generada automáticamente

Figure 2. Impulse Response of a Room with Echo, sampled at 22.050 kHz

To simulate original speech signal passing through the the room, the convolution operation is performed with original signal and impulse response, resulting in the signal of Figure 3:

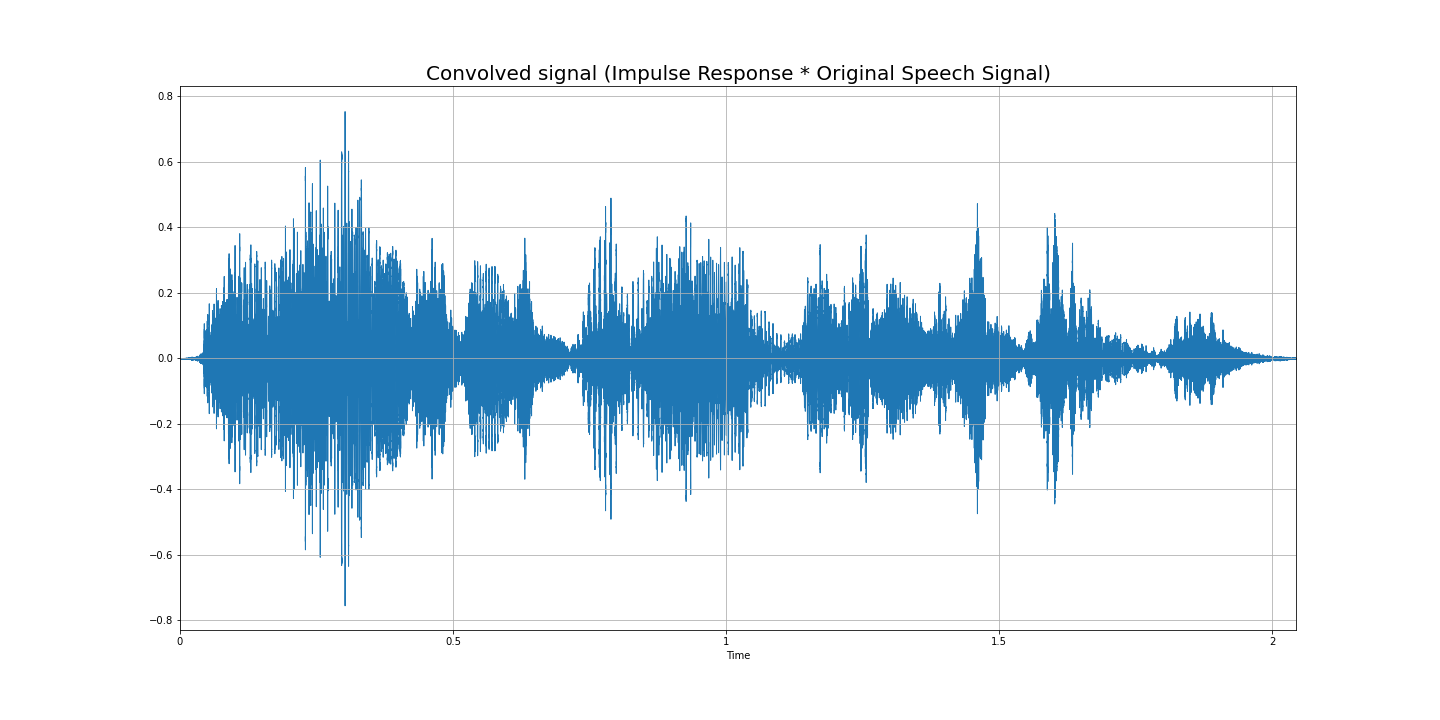


Figure 3. Resultant signal from convolution of Original Speech Signal and the Impulse Response

It is important to note that to be able to convolve two signals, it is necessary that both of them had been digitalized with the same sampling frequency.

To reduce the amount of data it was preferred to use a sampling frequency of 22.050 kHz instead of typical 44.1 kHz, which for this experiments does not represent a problem as the frequential component of speech lies between 250 Hz and 4 kHz, according to Nyquist theorem.

# Potential models that can solve the problem

Within *machine learning* there are two main focuses or approaches: ***supervised learning*** and ***not supervised learning***. It is important to note this because we start from the assumption that the current problem to solve requires **supervised learning**, which involves having the output signal for the model.

In ***supervised learning*** in addition to having the input data ,we have an output vector too, which feedbacks the model and tell us the “correct” answer so we can define the error, the objective is to train the model with features and training labels to later use a new set of data (signals) and predict those labels (classification) or data (regression).

Possible models to solve the problems are those who are supervised learning and specifically those for regression, which can be:

* Support Vector Regressors
* Convolutional Neural Networks
* Time-Delay Neural Networks
* Gravitational Search Algorithm
* Particle Swarm Optimization

## Neural Networks

Neural networks are the traditional deep learning approach, which are a structure of interconnected neurons (basic processing unit) and vary depending on its architecture, activation functions

Forma

Descripción generada automáticamente con confianza media

Figure 4. Graphical representation of a single Neuron

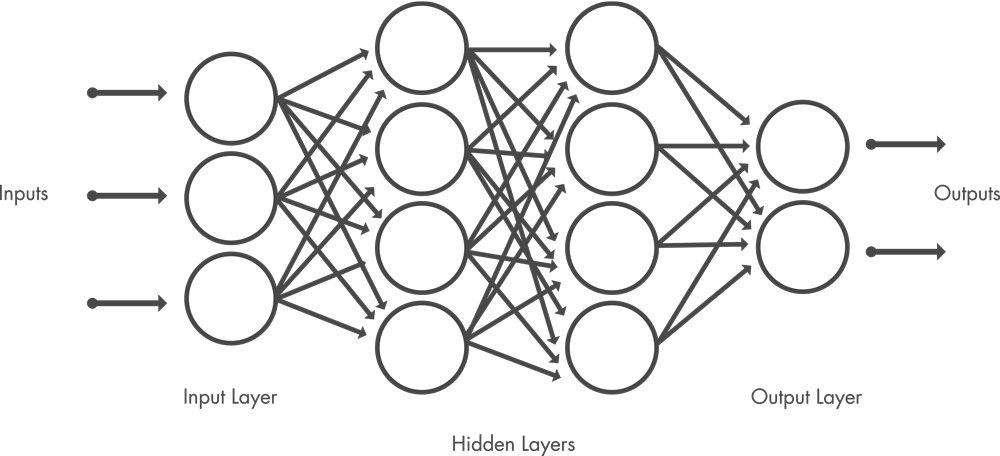


Ilustración . Arquitectura de una Red Neuronal densa

Entrenar a una red neuronal es un problema de optimización donde una función de pérdida es minimizada por propagación hacia atrás del error a través de la red neuronal.

**Existe una curva deseada en la respuesta al impulso**, la que se toma como la salida ideal y contra la que se compara la diferencia de la curva de respuesta en frecuencia alcanzada, **lo que representa la función de coste a minimizar**.

### Network arquitecture

Diagrama

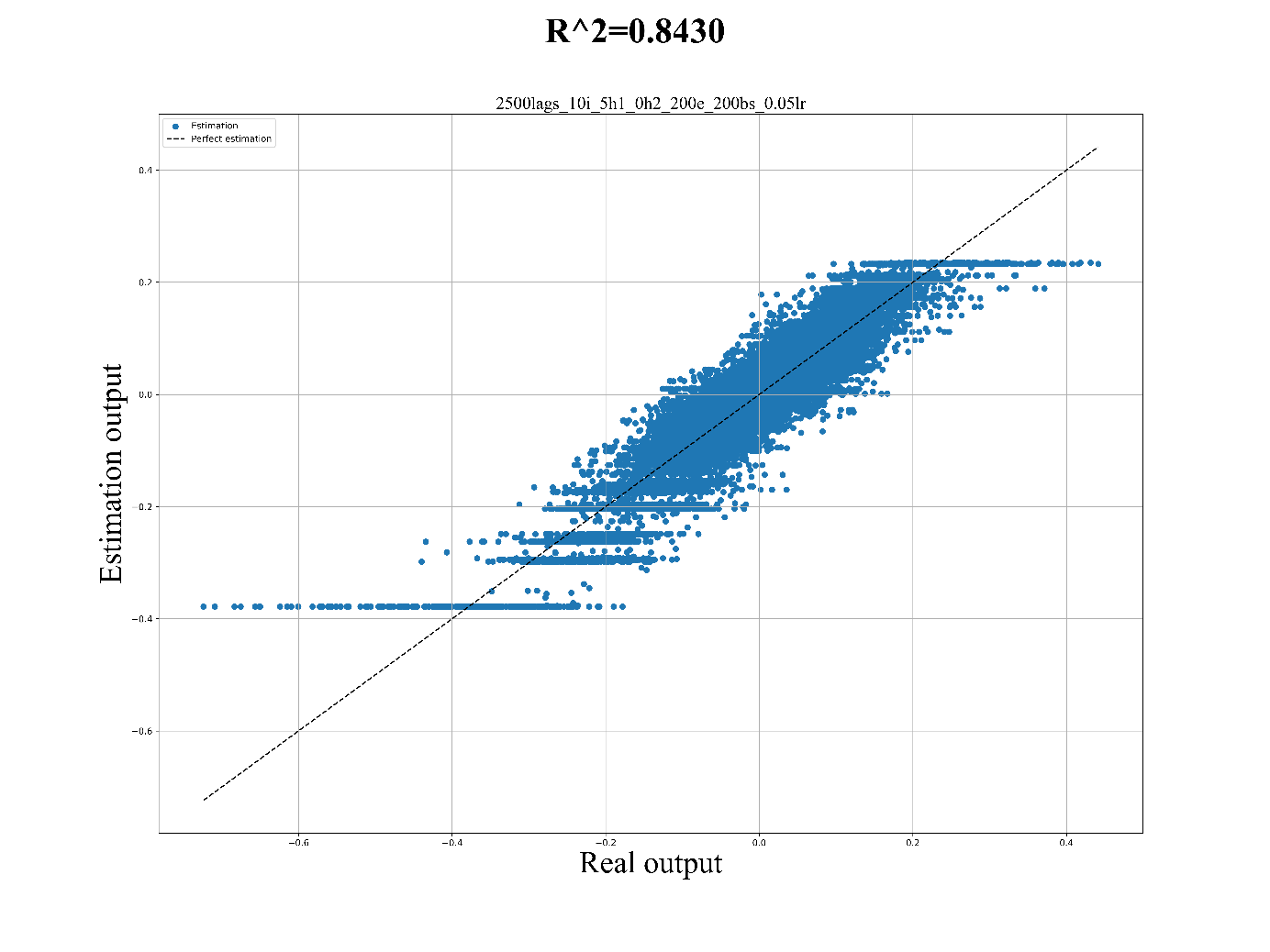
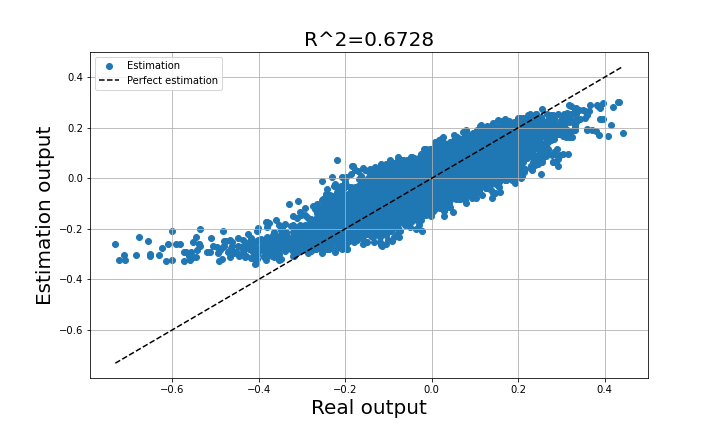
Descripción generada automáticamente

Figure 5. Neural Network’s Layers Structure

## 2.2.4 Error calculation and model efficiency

Typical **numerical** metrics for a regression problem is R2 which is a statistical measure, that describes quality of the model and the proportion of variation that can be described by it. Equation is:

However, as we are working with audio signals, it is equally important to listen to resultant audio files to identify effects as distortion and a more subjective validation of the prediction.



Cost function to minimize is Mean-squared error for a typical regression problem.

The equation is:

# Bibliography

[1] Cecchi, S; Carini, A.; Spors, S. Room Response Equalization-A Review. Appl. Sci. 2018, 8,16.

[2] Audio Equalization – Deep Neural Networks. Appl. Sci. 2020, 10, 2843.