

# Tarea 1

Entrenando una red neuronal

Alumna: Romina Romero Oropesa

Profesor: Alexandre Bergel

Auxiliares: Juan Pablo Silva

Ayudantes: Alonso Reyes Feris  
Gabriel Chandía

Fecha de entrega: 6 de noviembre de 2018  
Santiago, Chile

# Índice de Contenidos

<b>1. Implementación</b>	<b>1</b>
1.1. Instalación de dependencias . . . . .	1
<b>2. Spambase Data Set</b>	<b>1</b>
<b>3. Tests</b>	<b>1</b>
<b>4. Resultados</b>	<b>1</b>

## Lista de Códigos

1. Ejemplo en Python. . . . .	1
-------------------------------	---

# Implementación

El código de la implementación se encuentra en el repositorio de github [https://github.com/romina-romero/redes\\_neuronales\\_2018\\_2](https://github.com/romina-romero/redes_neuronales_2018_2). El lenguaje utilizado es python. La red neuronal se encuentra implementada en la carpeta trabajo\_incremental. Aquí se incluye además una serie de tests que muestran gráficas de ejemplo.

Las clases usadas para procesar el dataset elegido son **SigmoidNeuron**, **NeuronLayer** y **Neuralnetwork**. NeuralNetwork y SigmoidNeuron incluyen unittest que validan su funcionamiento.

## Instalación de dependencias

Código 1: Instalación de dependencias.

```
1 pip install x
```

## Spambase Data Set

El dataset elegido es el Spambase Data Set (<https://archive.ics.uci.edu/ml/datasets/Spambase>). Es una colección de 4601 emails, clasificados como spam o no spam, caracterizados en un vector de largo 57.

Cada email se describe de la siguiente forma:

- 48 porcentajes de aparición de palabras claves sobre el total de palabras del email. Una "palabra.<sup>en</sup> este caso es cualquier conjunto de caracteres alfanuméricos delimitados por caracteres no alfanuméricos.
- 6 porcentajes de aparición de caracteres claves sobre el total de caracteres del email.
- Promedio de los largos de las secuencias ininterrumpidas de mayúsculas, en el email.
- Largo de la secuencia ininterrumpida de mayúsculas más larga.
- Total de letras mayúsculas del email.

El vector incluye además un ítem número 58, donde se indica si es (1) o no (0) es spam.

## Tests

## Resultados