

# Homework 1

August 7, 2025

- You can install all the required dependencies by running the following command in your terminal: **pip install -r requirements.txt**
1. Run and carefully review the Jupyter notebook on Linear Regression.
  2. Use the label Male and Female as labels (+1 and -1) for a linear classification of the following data set
    - Context  
The age of abalone is determined by cutting the shell through the cone, staining it, and counting the number of rings through a microscope, a boring and time-consuming task. Other measurements, which are easier to obtain, are used to predict the age. Further information, such as weather patterns and location (hence food availability) may be required to solve the problem.
    - Original Dataset  
The original dataset can be accessed at <https://archive.ics.uci.edu/ml/datasets/abalone>.
    - The data set is included in the files.
  3. Split the data in training, validation and testing as necessary.
  4. Given that the labels used for classification are +1 and -1 the natural threshold to obtain a classification in 0. How is this

$$class(\mathbf{x}) = \begin{cases} C_1 & \text{if } f(\mathbf{x}) > 0 \\ C_2 & \text{if } f(\mathbf{x}) < 0 \end{cases}$$

5. Please solve the problem of linear classification by the use of three different methods
  - (a) Canonical solution (Split only training and testing)
  - (b) Gradient descent solution (Split training, validation and testing) here you have a **learning rate** that needs to be adjusted to obtain a good performance. Please use validation to avoid over-fitting.
  - (c) Model by Sklearn (Split training, validation and testing), but you still need to use a threshold.
6. Report the following metric
  - (a)  $Accuracy = \frac{\text{Correct Number of classifications}}{N}$ 
    - i. Where N is the total number of samples
    - ii.  $\text{Correct Number of classifications} = \text{Correct number of generated labels } C_{male} + \text{Correct number of generated labels } C_{female}$
  - (b) Whenever possible report Loss graph for validation and training and tell me when cut training.
7. Report everything in a Jupyter Notebook.