hugomcp@di.ubi.pt, 2024/25

Machine Learning

Practical Sheet 4: Experimental Setup

Consider the "banknote.csv" dataset, available at the course web page, and taken from the "UCI: Machine Learning Repository", of the University of California web page.

Suppose that we are interested in developing a machine learning model able to distinguish between **genuine** and **forged** bank notes. To do that, experts informed that we would have to measure four features in each note:

1. variance of the wavelet transformed image
2. skewness of wavelet transformed image
3. curtosis of wavelet transformed image
4. entropy of the image

That is exactly the information provided (in tabular form) in the "csv" file. The fifth column provides the class information, i.e., whether the note is **genuine (1)** or a **fake (0)**.

1. Start from the "logistic_regression.py" script, and implement the regularized version of logistic regression.
2. See how different values of λ lead to different models

3. Implement the “K-fold” cross validation and bootstrapping performance evaluation strategies.
    a. Report the corresponding “mean $\mp$ standard deviation”, and AUC values.