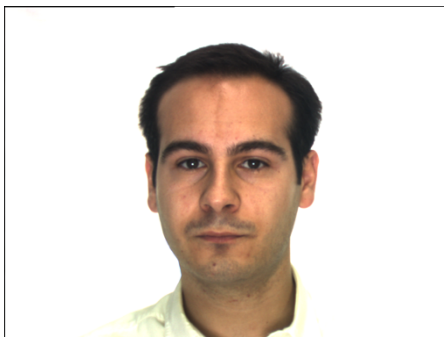
hugomcp@di.ubi.pt, 2023/24

Machine Learning

Practical Sheet 5: Dimensionality Reduction



Consider the "AR.csv" dataset, available at the course web page. It contains a "csv" representation of [48 x 64] face images. Each value provides a grayscale representation of the pixel at a corresponding position, for a total of 3,072 values. We will use it to distinguish between "**Male**" and "**Female**" samples. In each line, the last column gives the corresponding class (1="**Male**"; 0="**Female**")

1. Implement a "Python" script that performs PCA dimensionality reduction, on the selected data set

    a. Programming it from scratch

    b. Using the "sklearn" library

2. Divide the available data into learn, validation and test sets.

3. Use the "logistic_regression.py" script previously done to compare the effectiveness of this classifier in the:

    a. Original data set (100%)

    b. Dimensionality reduced, keeping 99% of the amount of variability.

    c. Dimensionality reduced, keeping 95% of the amount of variability.

    d. Dimensionality reduced, keeping 90% of the amount of variability.

    e. Dimensionality reduced, keeping 50% of the amount of variability.