Practical exercise

Regression

Linear regression allows predicting the value of one variable using another variable.

1. Load the *diabetes* dataset using datasets().
   Ten baseline variables, age, sex, body mass index, average blood pressure, and six blood serum measurements were obtained for each of n = 442 diabetes patients, as well as the response of interest, a quantitative measure of disease progression one year after baseline.
   The goal here is to predict the disease progression for patients (last variable).

   Compute the correlation matrix between all the variables and conclude which are the most correlated to the target variable.

   Use a projection method (as PCA) and plot the data in 2 dimensions.

2. Import *linear_model* from *sklearn*.
   Split the dataset in two subsets: training and test subset. Create the linear regression model (object in python) using:
   linear_model.LinearRegression()

   Train the regression model using *fit* function on the training subset and *predict* on the test subset.

3. Import mean_squared_error from sklearn. This function allows computing the MSE error of the prediction.
   Use *scatter* and *plot* functions to plot the data and the regression line on the same figure.
   Analyse the obtained image and the MSE error. Conclude.

4. To improve the results, it is possible to test other regression methods, and also to ensemble the obtained results.
   Use others regression methods and ensemble learning. Some information can be found here:
   https://scikit-learn.org/stable/auto_examples/ensemble/plot_voting_regressor.html?highlight=diabetes#sphx-glr-download-auto-examples-ensemble-plot-voting-regressor-py