Course: Introduction to materials informatics

Tutorial 3

Title: Supervised machine learning

Questions

1. Calculate the first three principal components (PC) from the autocorrelations data. Prepare the input and output data.
2. Perform data splitting into training and testing data
    a. Use random splitting
    b. Use specific class-wise splitting
3. Perform linear regression on the data and calculate the R square value on the test data prediction. Plot the actual vs predicted stress plot.
4. Perform random forest regression on the data and calculate the R square value on the test data prediction. Plot the actual vs predicted stress plot.
5. Perform multi-layer perceptron based regression on the data and calculate the R square value on the test data prediction. Plot the actual vs predicted stress plot.

****

Additional references:

PCA

https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html

Random train/ test data splitting

https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html

Linear regression

https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html

Random forest regression

https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html

Multilayer perceptron regression

https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPRegressor.html

R square metric

https://scikit-learn.org/stable/modules/generated/sklearn.metrics.r2_score.html