

Applied Data Science with Python

Labwork #2

Follow the guide from this link to install the necessary packages (use skcikit-learn) https://github.com/justmarkham/scikit-learn-videos/blob/master/02_machine_learning_setup.ipynb

Linear Regression:

- Import Pandas and select one dataset available from it.
- Study the number of data instances and features.
- Select two features which are the most correlated (provided by the correlation matrix)
- Compute and visualize the linear regression.
- Compute the mean square error (MSE is equal to the total sum of distances from data instances to the regression line).

Classify with k-nn :

- Charge a dataset with the label being available.
- Divide the dataset into two subsets: training and testing.
- Perform different experimental runs of k-nn with different values for k.
- Comment on the result.