18 Jan 2022, Wednesday

Machine Learning Lab
Assignment 1
Regression

Dataset: Car Price Prediction

1. Import the libraries:
    a. "**pandas**" for the dataset management,
    b. "**matplotlib**" or "**seaborn**" for the plots, and
    c. "**sklearn**" for the Linear Regression algorithm.
        i. "**train_test_split**" for the training set and test set to split
        ii. "**LinearRegression**" for the algorithm itself
        iii. "**MinMaxScaler**" for data normalization
        iv. "**r2_score**" for the evaluation
    d. "**numpy**"
        i. "**corrcoef**" for Pearson's correlation coefficient
2. Read the dataset. Observe #observations, #features, and target variable.
3. Remove all the columns/features that are unusable from the algorithm. Linear Regression can only use **numerical features**, so to make this dataset usable, **remove all the other features**.
4. Plot and compute **correlations** between the remaining cars' features and the target variable 'price.' **Drop the unnecessary columns** (uncorrelated) for the Linear Regression algorithm.
5. Normalize all the numerical features.
6. Divide dataset into a **training set** and a **test set**. Use the "train_test_split" method of the sklearn library to create the training set and the test set.
7. Create the linear regression model with the training set.
8. Make predictions using the test set.
9. Find the accuracy of the model, and use the most popular metric for linear regression called "**R-squared**" (the closer the value is to 1, the more accurate the model is).