**LINEAR REGRESSION**

The goal of this part is to show how to use Machine Learning techniques, measurements, and show some general basic knowledge about the model construction (Linear regression in our case).

Linear regression has many practical uses. In our case it is:

* Prediction, or forecasting, where linear regression is used to fit a predictive model to an observed data set of **y** and **X** values. After developing such a model, if an additional value of **X** is then given without its accompanying value of **y**, the fitted model can be used to make a prediction of the value of **y**

**N.B.** For linear regression task we will use Python libraries: Sklearn, Pandas & Numpy (so please include it in your Python scripts)

**LINEAR REGRESSION TASK**

1.      Create Python scripts for linear regression model construction:

-        read data from PREPARED/YYYY/

- Transform data into 2D for using in Linear Regression model construction.

Where **X value** will be “value index for houses”, the formula is: (*(“livingArea” column \* 1.6) + (“basement” column \* 0.6)) \* (“quality” column \* 3) \* (“condition” condition \* 0.8)*

Where **y value** will be: *“Sale Price” column (the values which we will predict)*

- Split the data on 2 parts:

**Training set** – the set which will be used for model training

**Test set** – the set which will be our actual data, where we will try to predict “Sale Price”.

- Create linear regression object (using “Sklearn” library) and train the model using **training set**.

- Use calculated model to predict values of **y values** using **X values** from **test set**.

- Visualize the results using Residual plots where you will compare **X** and **y values** from **test set** and predicted values from previous action point.

- Transform arrays for visualization to 1D arrays for dataframes

- Write Index values to new data frame, as “HouseIndex” column

- Write Predicted price values to same data frame, as “PredictedPrice” column

- Write Real sell price values to same data frame as “SellPrice” column

- Write prediction error values to same data frame as “PredictionError” column:

Calculate as: absolute value from (1 - ("PredictedPrice" / "SellPrice")

- print mean values of “PredictionError” column in console.

- write resulting data frame results to PROCESSED/YYYY/Houses\_predicted.csv