REGRESSION:we use an explanatory variable inorder to predict the output.

We use X to predict Y.

Normally you need to know that data is in form of table.And the features are in form of matrix.

For example:in iris data set if you we have features as sepal and petal length and width which is nothing but the feature and they all mentioned in column part.usually at he end of the column we have our target variable which is nothing but “species” here.

For any basic machine learning algorithm what you need to do is,

1.choose the model

2.choose model hyperparameters

3.arrange data into features matrix and target array.

4.fit the model

5.apply trained model to new data.

Some of the basic packages have been used by ml in sckit learn.go through that.

Such as model.fit(x,y),fit\_intercept,model.coef\_,model.intercept\_

x,y 🡪represent values plotted

x\_fit and y\_fit 🡪represent the straight line.

Np.linspace 🡪used to fit sline.

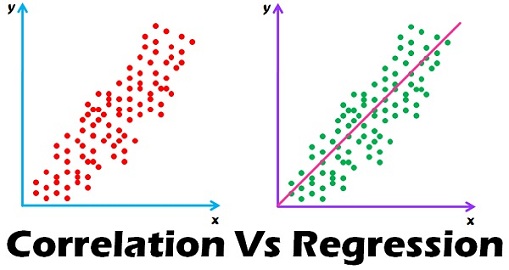
Since after importing the dataset if the rowa and columns are dissimilar or the columns acts as a row we use the parameter

Header=none🡪which makes the correct ordering.

CORRELATION ANALYSIS AND FEATURE SELECTION:

Correlation🡪it determines co relation or association of two variables.

Regression🡪it describes how independent variable is numerically related to dependent variable.



Inorder to understand the dataset.we may use of seaborn for viewing it.by use of the method

Sns.heatmap(…….)🡪it gives a type of correlation matrix with the values.

There can be negative values of an dataset.if the dataset is negative it means that people are not intrested on something related to the attribute

Example:housing rediction.

Most people don’t preffer houses near industry in that cases the values will be negative.

According to the choosen y and x value the pearson coefficient varies

Example:housing prediction

First case:we predict the price of the house by keeping

No of rooms as ‘x’ value

For this it gives one correlation.

Second case:we predict the price of house according to the status of peoples income.

For this it gives different correlation value.

