Hands on Tutorial on Linear Regression

Agenda For this Video:

1. Understanding the Linear Regression Algorithm Briefly.
2. Real Life Examples and Use Cases.
3. Taking an Example (Toy data set) for House Price Prediction
4. Coding the Linear Regression in Python
5. Checking the answers using Sklearn
6. Task For you.....

Linear Regression

It is a Procedure which enables us to find a line (for 2-d data) which is best in among all the possible sets of lines.

What about data which is 3-d,4-d or n-d?
For those data, we don't say a line but rather Hyperplane which is just fancy way of saying line in higher dimension.

But once we understand the theory behind the 2-d Linear Regression , we can easily generalize it to more dimensions.

So, let us start to talk about an example….

For a 2-d line, the equation will be of form:

$$Y = mx + C \text{ or}$$

$$\text{price} = m * \text{number\_of\_rooms} + C$$

Once we obtain the best values of m and c, we can just plug the value of X (number of rooms) and we can get the value of price.

So, how do we do this with computer?

1). Take some random values of m and c.
2). Decide the number of iterations.
3). Repeat till number of iterations:
   a) Take random data (X,y)
   b) Obtain Prediction using the above random values of m and c
   c) Update this values by a very small margin called learning rate.

The procedure that we just discussed is what is known as Linear Regression Algorithm. This is what we are going to implement in Python From Scratch.

And also, we will be checking the validity of our procedure with a Library in Python called Sklearn.

Use Cases of Linear Regression

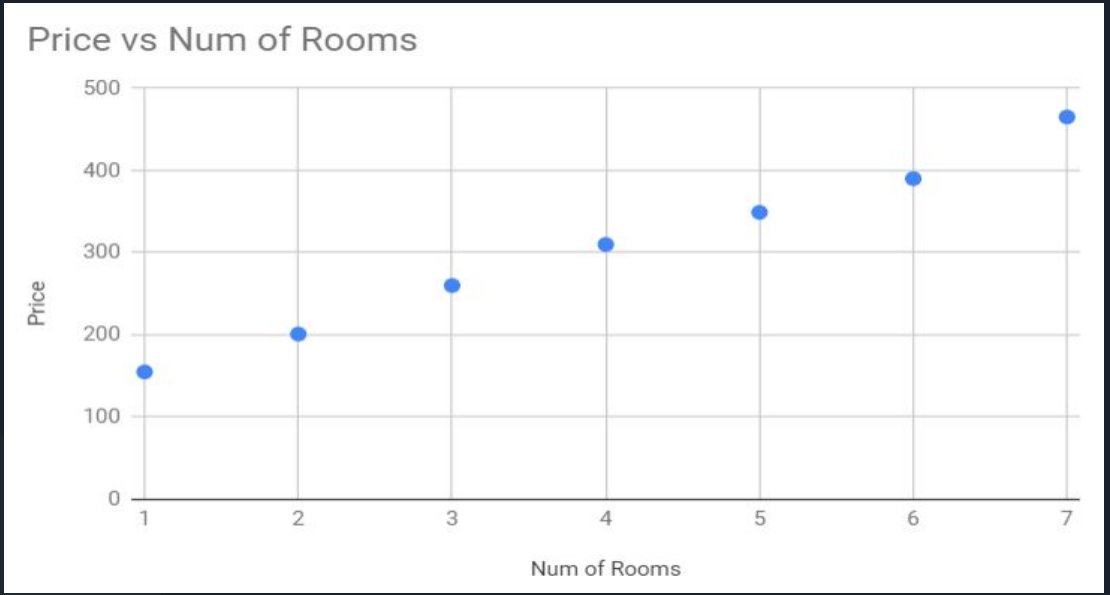
1. Prediction of future revenues or sales of a company based on some features like advertisement.
2. Predicting of Electricity consumption by consumers
3. Prediction of Price of house based on its various features.
4. Prediction of Market Effectiveness
5. Predicting the risk value of a customer.
6. Prediction of Stock Prices (Time Series Regression).

Best Line for Our Data

Num of Rooms	Price
1	155
2	201
3	260
4	310
5	349
6	390
7	465

Let us say the variable that we want to predict is Price of House. So, this variable will be our y based on number of rooms as our feature X.

Let’s Plot the data



Pseudocode For Linear Regression

```
# Step.1 Take some number of iterations...
number_of_iterations = 10000
# Step. 2 Intialize some random values of m and c
m = some_random
c = some_random
# Step. 3 Loop through number_of_iterations times
for value in range(number_of_iterations):
  # Randomly choose a data point (X and y)...
  random_X = something
  random_y = something
  # For each loop we obtain the prediction for the current m and c...
  prediction = m * random_X + x
  # then we compare, prediction with actual y
  # then we adjust m and c accordingly.
  # Usually, we take the difference of prediction and actual y
  # Because if the difference is close to zero, then that means we don't
  # have to Update that much..
  m += (actual_y - prediction) * random_X * learning_rate
  c += (actual_y - prediction) * learning_rate

# Step. 4 After the loop ends, we get a value of m and c which is hopefully
# near to perfect.
```

I hope you were able to follow through and understand everything that we discussed so far.

Let us get started with Coding.....

Let us turn that Pseudocode to actual code....

Now, let us check our answers using Sklearn Library.

Task For You....

Replicate the same procedure on the following dataset...

Advertisement Expenditure	Sales
100	289
200	364
250	402
310	451
401	522
500	597
615	689

Build a linear regression model to predict sales based on expenditure on advertisement.
Also, check your answers using sklearn.

Another Task...

In this task, I want you to point to another data which is going to be a real world dataset.

Go to Kaggle.com and Search for House Price Prediction.

This is a very popular dataset and it has got multiple features.
Try to perform Linear Regression on this dataset.