

Linear Regression with Multiple Variable

Outlines

- Problem
- Import libraries
- Load the Data
- Data Preprocessing
- Linear Regression Using Multiple Variables
- Calculate The Coefficients
- Prediction
- Quiz 1
- Quiz 2
- Exercise

Problem:

Based on the table shown below, find out the price of a home that has:

3000 sq ft area, 3 bedrooms, and 40 year old.

2500 sq ft area, 4 bedrooms, and 5 year old.

area	bedrooms	age	price
2600	3	20	550000
3000	4	15	565000
3200		18	610000
3600	3	30	595000
4000	5	8	760000
4100	6	8	810000

$$price = m_1 * area + m_2 * bedrooms + m_3 * age + b$$

Dependent variable

Independent variables (**features**)

$$price = m_1 * area + m_2 * bedrooms + m_3 * age + b$$

Coefficients

Intercept

$$y = m_1 x_1 + m_2 x_2 + m_3 x_3 + b$$

Import Libraries

```
In [9]: import pandas as pd
import numpy as np
from sklearn import linear_model
```

Load the Data

```
In [10]: df=pd.read_csv('D:/Data_Science/My Github/3. Machine-Learning-with-Python/2. Line
df
```

```
Out[10]:
```

	area	bedrooms	age	price
0	2600	3.0	20	550000
1	3000	4.0	15	565000
2	3200	NaN	18	610000
3	3600	3.0	30	595000
4	4000	5.0	8	760000
5	4100	6.0	8	810000

[Tip]: Before solving any machine learning problem we need to clean the data.

Data Preprocessing: Handling Missing Data

```
In [11]: # calculate the median for NaN Data
import math
median_bedrooms = math.floor(df.bedrooms.median())
median_bedrooms
```

Out[11]: 4

```
In [12]: # import median_bedrooms into the data by using fillna
df.bedrooms = df.bedrooms.fillna(median_bedrooms)
df
```

Out[12]:

	area	bedrooms	age	price
0	2600	3.0	20	550000
1	3000	4.0	15	565000
2	3200	4.0	18	610000
3	3600	3.0	30	595000
4	4000	5.0	8	760000
5	4100	6.0	8	810000

Linear Regression Using Multiple Variables

```
In [13]: reg = linear_model.LinearRegression()
reg.fit(df[['area', 'bedrooms', 'age']], df.price)
```

Out[13]: LinearRegression()

Calculate the Coefficients

```
In [20]: print('m1 = %s m2 = %s m3 = %s b = %s' % (reg.coef_[0], reg.coef_[1], reg.coef_[2], reg.coef_[3]))

m1 = 112.06244194213451 m2 = 23388.88007793923 m3 = -3231.7179086329634 b = 221323.00186540425
```

$$y = m1 * x1 + m2 * x2 + m3 * x3 + b$$

Prediction

Predict the price of home by Linear Regression model with 3000 sqr ft area, 3 bedrooms, 40 year old

```
In [21]: reg.predict([[3000,3,40]])
```

```
Out[21]: array([498408.25158031])
```

Quiz 1:

Find the price of home by Linear Regression model with 2500 sqr ft area, 4 bedrooms, 5 year old

[Click here for the solution](#)

Quiz 2:

Find the price of home by handy calculation with 2500 sqr ft area, 4 bedrooms, 5 year old

[Click here for the solution](#)

Exercise

In exercise folder there is **hiring.csv**. This file contains hiring statics for a firm such as experience of candidate, his written test score and personal interview score. Based on these 3 factors, HR will decide the salary. Given this data, you need to build a machine learning model for HR department that can help them decide salaries for future candidates. Using this predict salaries for following candidates,

2 yr experience, 9 test score, 6 interview score

12 yr experience, 10 test score, 10 interview score

[guide: preprocess your data with these 3 steps](#)

1. In experience column two values are missed. Assumed them to be zero.
2. In the test_score one data is missed. Filled this with median of this column.
3. Ususally Linear Regression model work on numbers. So, convert the experience column to numbers.(you should use **pip install word2number** in your cmd)

[Click here for the solution](#)

Author Ehsan Zia