linear regression multivariate

- y=m1x1 + m2x2 + m3x3 + c
- m - Coefficient
- c - Intercept

Steeps

- 1) Import libraries
- 2) Import the train data file and clean the data
- 3) Separate features data
- 4) Separate Labels data
- 5) Create a linear Regression model
- 6) Fit the data into the model
- 7) Prediction result

In [1]:

```
# 1) Import libraries
import pandas as pd
import numpy as np
import math
from sklearn import linear_model
```

In [2]:

```
# 2) Import train data file
train_data = pd.read_csv('homeprices_multivariate.csv')
```

In [3]:

```
train_data
```

Out[3]:

	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

In [4]:

```
# Data Preprocessing
bedrooms_mean=math.floor(train_data['bedrooms'].mean())
train_data['bedrooms'].fillna(bedrooms_mean,inplace = True)

train_data
```

Out[4]:

	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

In [5]:

```
# 3) Separate features data
features = train_data.drop('price',axis='columns')
```

In [6]:

```
features
```

Out[6]:

	area	bedrooms	age
0	2600	3.0	20
1	3000	4.0	15
2	3200	4.0	18
3	3600	3.0	30
4	4000	5.0	8
5	4100	6.0	8

In [7]:

```
# 4) Separate Labels data
Label = train_data[['price']]
```

In [8]:

```
Label
```

Out[8]:

	price
0	550000
1	565000
2	610000
3	595000
4	760000
5	810000

In [9]:

```
# 5) Creat linear Regression model
lin_reg = linear_model.LinearRegression()
```

In [10]:

```
# 6) Fit the data into model
lin_reg.fit(features,Label)
```

Out[10]:

```
LinearRegression()
```

In [11]:

```
# 7) Prediction result
area = input('Enter area(sq_ft):')
bedrooms = input('Enter bedrooms:')
age = input('Enter age:')
lin_reg.predict([[area,bedrooms,age]])
```

```
Enter area(sq_ft):3300
Enter bedrooms:2
Enter age:25

C:\Users\Gopi\anaconda3\lib\site-packages\sklearn\base.py:450: UserWarnin
g: X does not have valid feature names, but LinearRegression was fitted wi
th feature names
  warnings.warn(
C:\Users\Gopi\anaconda3\lib\site-packages\sklearn\base.py:566: FutureWarni
ng: Arrays of bytes/strings is being converted to decimal numbers if dtype
='numeric'. This behavior is deprecated in 0.24 and will be removed in 1.1
(renaming of 0.26). Please convert your data to numeric values explicitly
instead.
  X = check_array(X, **check_params)
```

Out[11]:

```
array([[557113.8727145]])
```

In [ ]:

```
```

Test the formula

In [12]:

```
# Tack 'm' values
lin_reg.coef_
```

Out[12]:

```
array([[  112.06244194, 23388.88007794, -3231.71790863]])
```

In [13]:

```
# Tack 'c' vlue
lin_reg.intercept_
```

Out[13]:

```
array([221323.0018654])
```

In [14]:

```
# y=m1x1 + m2x2 + m3*x3 + c
y=112.06244194*float(area) + 23388.88007794*float(bedrooms) + (-3231.71790863*float(age))
```

In [15]:

```
y
```

Out[15]:

```
557113.87270753
```

In [ ]:

```
```