hiringprediction

June 14, 2022

```python
[51]: import numpy as np
      import pandas as pd
      import matplotlib.pyplot as plt
      %matplotlib inline
      from sklearn import linear_model
      from word2number import w2n
      import math
```

```python
[52]: df=pd.read_csv('https://raw.githubusercontent.com/codebasics/py/master/ML/
      ↪2_linear_reg_multivariate/Exercise/hiring.csv')
```

```python
[53]: df
```

```
[53]:   experience  test_score(out of 10)  interview_score(out of 10)  salary($)
      0        NaN                    8.0                           9      50000
      1        NaN                    8.0                           6      45000
      2       five                    6.0                           7      60000
      3        two                   10.0                          10      65000
      4      seven                    9.0                           6      70000
      5      three                    7.0                          10      62000
      6        ten                    NaN                           7      72000
      7     eleven                    7.0                           8      80000
```

```python
[54]: df.experience=df.experience.fillna('zero')
      df.experience
```

```
[54]: 0       zero
      1       zero
      2       five
      3        two
      4      seven
      5      three
      6        ten
      7     eleven
      Name: experience, dtype: object
```

```python
[55]: df
```

```
[55]:    experience  test_score(out of 10)  interview_score(out of 10)  salary($)
    0       zero                    8.0                          9      50000
    1       zero                    8.0                          6      45000
    2       five                    6.0                          7      60000
    3        two                   10.0                         10      65000
    4      seven                    9.0                          6      70000
    5      three                    7.0                         10      62000
    6        ten                    NaN                          7      72000
    7     eleven                    7.0                          8      80000
```

```
[ ]:
```

```
[56]: df.experience=df.experience.apply(w2n.word_to_num)
      df
```

```
[56]:    experience  test_score(out of 10)  interview_score(out of 10)  salary($)
    0          0                    8.0                          9      50000
    1          0                    8.0                          6      45000
    2          5                    6.0                          7      60000
    3          2                   10.0                         10      65000
    4          7                    9.0                          6      70000
    5          3                    7.0                         10      62000
    6         10                    NaN                          7      72000
    7         11                    7.0                          8      80000
```

```
[65]: k=math.floor(df[["test_score(out of 10)"]].mean())
      k
```

```
[65]: 7
```

```
[66]: df[["test_score(out of 10)"]]=df[["test_score(out of 10)"]].fillna(k)
      df[["test_score(out of 10)"]]
```

```
[66]:    test_score(out of 10)
    0              8.000000
    1              8.000000
    2              6.000000
    3             10.000000
    4              9.000000
    5              7.000000
    6              7.857143
    7              7.000000
```

```
[67]: reg=linear_model.LinearRegression()
      reg.fit(df[['experience','test_score(out of 10)','interview_score(out of
       ↪10)']],df[['salary($)']])
```

```
[67]: LinearRegression()
```

```
[48]: reg.coef_
```

```
[48]: array([[2812.95487627, 1845.70596798, 2205.24017467]])
```

```
[49]: reg.predict([[2,9,6]])
```

```
[49]: array([[53205.96797671]])
```

```
[ ]:
```