sk_multipleRegression

April 15, 2019

1 Multiple Regression

Almost all real world problems that you are going to encounter will have more than two variables.
Linear regression involving multiple variables is called "multiple linear regression". The steps
to perform multiple linear regression are almost similar to that of simple linear regression. The
difference lies in the evaluation. You can use it to find out which factor has the highest impact on
the predicted output and how different variables relate to each other.

```python
In [1]: # Multiple linear regression
        import numpy as np
        from sklearn import datasets
        from sklearn.linear_model import LinearRegression
        from sklearn import metrics
        from sklearn.model_selection import train_test_split
```

```python
In [2]: dataset = datasets.load_diabetes()
        dataset.feature_names
```

```python
Out[2]: ['age', 'sex', 'bmi', 'bp', 's1', 's2', 's3', 's4', 's5', 's6']
```

```python
In [3]: # train test split
        X_train, X_test, y_train, y_test = train_test_split(dataset.data, dataset.target,test_s
```

```python
In [4]: regressor = LinearRegression()
        regressor.fit(X_train, y_train)
```

```python
Out[4]: LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None,
                 normalize=False)
```

```python
In [5]: print(">", regressor.intercept_)
        print(">",regressor.coef_)
```

```
> 153.96242985120966
> [ -55.62047247 -259.84522708  540.30079099  344.62715568 -971.71436208
   612.43396894  155.15860756  190.63482994  830.76852218   73.86967485]
```

```python
In [6]: # predction for test data
        y_pred = regressor.predict(X_test)
```

```
In [7]:  # Error report
         print('Mean Absolute Error:', metrics.mean_absolute_error(y_test, y_pred))
         print('Mean Squared Error:', metrics.mean_squared_error(y_test, y_pred))
         print('Root Mean Squared Error:', np.sqrt(metrics.mean_squared_error(y_test, y_pred)))
         print('Variance score: %.2f' % metrics.r2_score(y_test, y_pred))

Mean Absolute Error: 42.097780223111066
Mean Squared Error: 2966.0402992842164
Root Mean Squared Error: 54.461365198498434
Variance score: 0.40
```

1.1 References:

1. https://stackabuse.com/linear-regression-in-python-with-scikit-learn/