sk_linearRegression

April 15, 2019

1 Linear regression

In statistics, linear regression is a linear approach to modelling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables). The case of one explanatory variable is called simple linear regression. For more than one explanatory variable, the process is called multiple linear regression. This term is distinct from multivariate linear regression, where multiple correlated dependent variables are predicted, rather than a single scalar variable.

In [1]:
```
# linear regression
import numpy as np
import matplotlib.pyplot as plt
from sklearn import datasets, linear_model
from sklearn.model_selection import train_test_split
from sklearn import metrics
```

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

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

In [3]:
```
# feature selection
bmi = dataset.data[:, np.newaxis ,2]
# train test split
X_train, X_test, y_train, y_test = train_test_split(bmi, dataset.target, test_size=0.2
```

In [4]:
```
# model
regressor = linear_model.LinearRegression()
regressor.fit(X_train.reshape(-1, 1) , y_train.reshape(-1, 1) )
```

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

In [5]:
```
print(regressor.intercept_)
print(regressor.coef_)
```

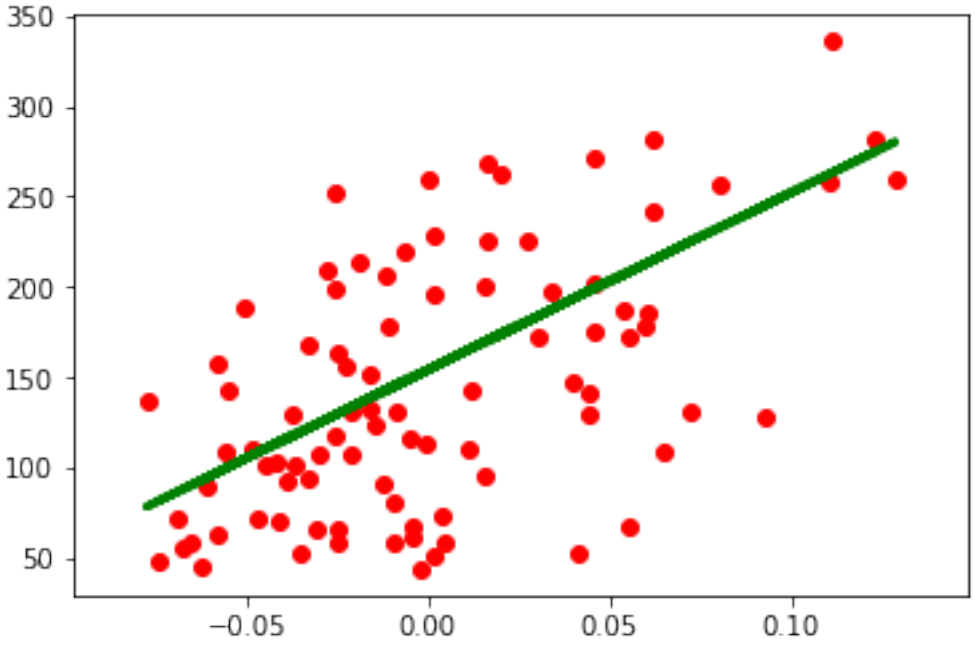
```
[154.27729208]
[[979.8299551]]
```

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

In [7]: 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: 49.4200929535062
Mean Squared Error: 3612.4239077907882
Root Mean Squared Error: 60.10344339379224
Variance score: 0.27


In [8]: plt.scatter(X_test, y_test,  color='red')
        plt.plot(X_test, y_pred, color='green', linewidth=3)
        plt.show()
```



1.1 References:

1. https://scikit-learn.org/stable/auto_examples/linear_model/plot_ols.html
2. https://stackabuse.com/linear-regression-in-python-with-scikit-learn/
3. https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html
4. https://en.wikipedia.org/wiki/Linear_regression