Linear Regression

February 15, 2022

```python
[32]: import pandas as pd
      import numpy as np
      from sklearn.linear_model import LinearRegression
      import matplotlib.pyplot as plt
      import numpy as np
      from sklearn.metrics import mean_squared_error, r2_score
      from sklearn import datasets, linear_model
```

```python
[33]: df = pd.read_csv("Advertising.csv")
      x_arr = df["radio"].to_numpy()
      y_arr = df["sales"].to_numpy()
      x_arr = x_arr.reshape(-1,1)
      y_arr = y_arr.reshape(-1,1)
```

```python
[34]: regr = linear_model.LinearRegression()
      regr.fit(x_arr,y_arr)
```

```
[34]: LinearRegression()
```

```python
[35]: print("Coefficients: ", regr.coef_,"Intercepts: ",regr.intercept_)
```

```
Coefficients:  [[0.20249578]] Intercepts:  [9.3116381]
```

```python
[36]: arr_y_pred = regr.predict(x_arr)
      print('Predicted response:', arr_y_pred[195:])
```

```
Predicted response: [[10.06087249]
 [10.30386743]
 [11.19484888]
 [17.816461  ]
 [11.05310183]]
```

```python
[37]: print("Mean squared error: %.2f" % mean_squared_error(x_arr, y_arr))
      print("Coefficient of determination: %.2f" % r2_score(x_arr, y_arr))
```

```
Mean squared error: 242.99
Coefficient of determination: -0.11
```

```
[38]: plt.scatter(x_arr, y_arr, color="black")
      plt.plot(x_arr, arr_y_pred, color="blue", linewidth=3)
      plt.title("Prediction vs data")
      plt.xlabel("radio spending")
      plt.ylabel("No of units sold")
      plt.show()
```
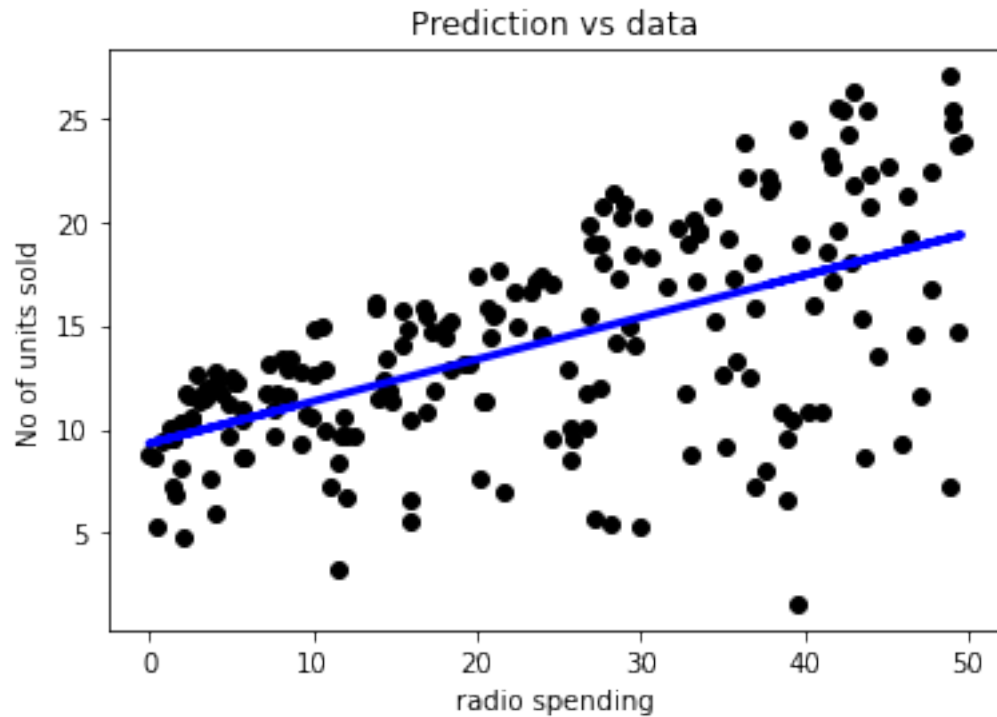


```
[39]: #to find the value of units sold for every 25M$ spent we use the y=mx+c formula
      units_sold = (regr.coef_*23) + regr.intercept_
      units_sold
```

```
[39]: array([[13.96904111]])
```