regression_101

March 9, 2023

```python
[9]: import os
     import sys
     import pandas as pd
     import numpy as np
     from sklearn.linear_model import LinearRegression
     from sklearn.model_selection import train_test_split
     from matplotlib import pyplot as plt
```

```python
[10]: %matplotlib inline
```

```python
[11]: df =pd.read_csv('test.csv')
```

```python
[12]: df.head()
```

```
[12]:     x          y
      0  77  79.775152
      1  21  23.177279
      2  22  25.609262
      3  20  17.857388
      4  36  41.849864
```
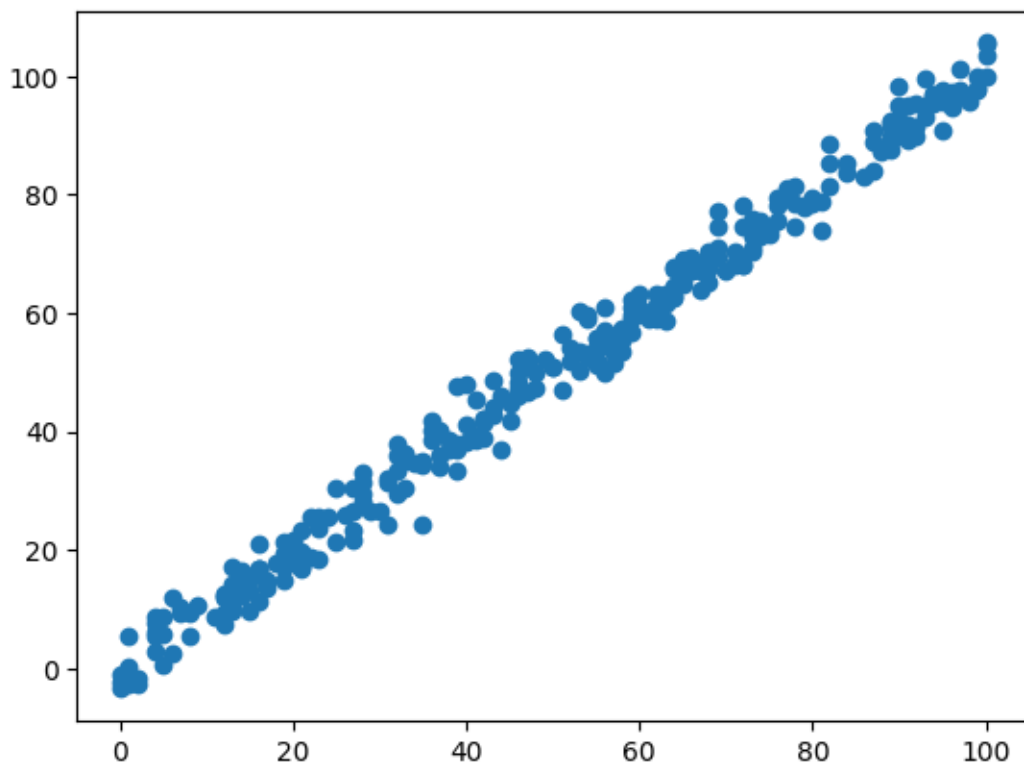
```python
[13]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 300 entries, 0 to 299
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   x       300 non-null    int64
 1   y       300 non-null    float64
dtypes: float64(1), int64(1)
memory usage: 4.8 KB
```

```python
[14]: df.fillna(-99999,inplace=True)
```
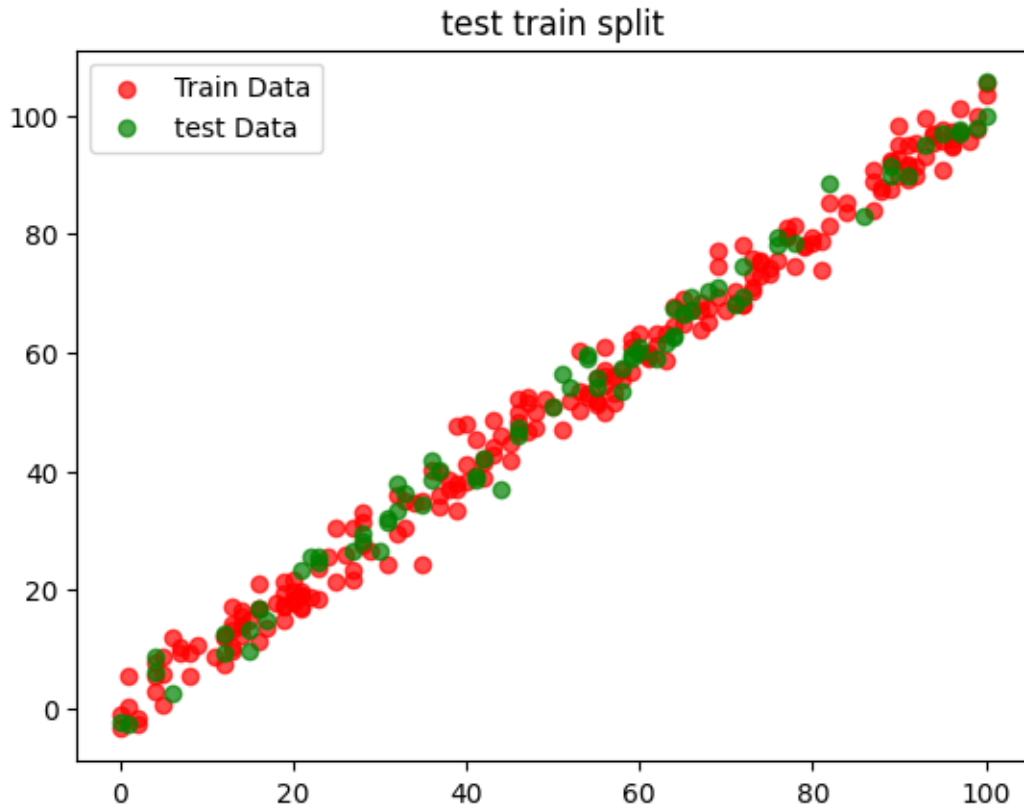
```python
[15]: xpoints=np.array(df['x'])
      ypoints=np.array(df['y'])
```

```
[20]: plt.plot(xpoints,ypoints,'o')
      plt.show()
```



```
[27]: x_train,x_test,y_train,y_test=train_test_split(df['x'],df['y'])
```
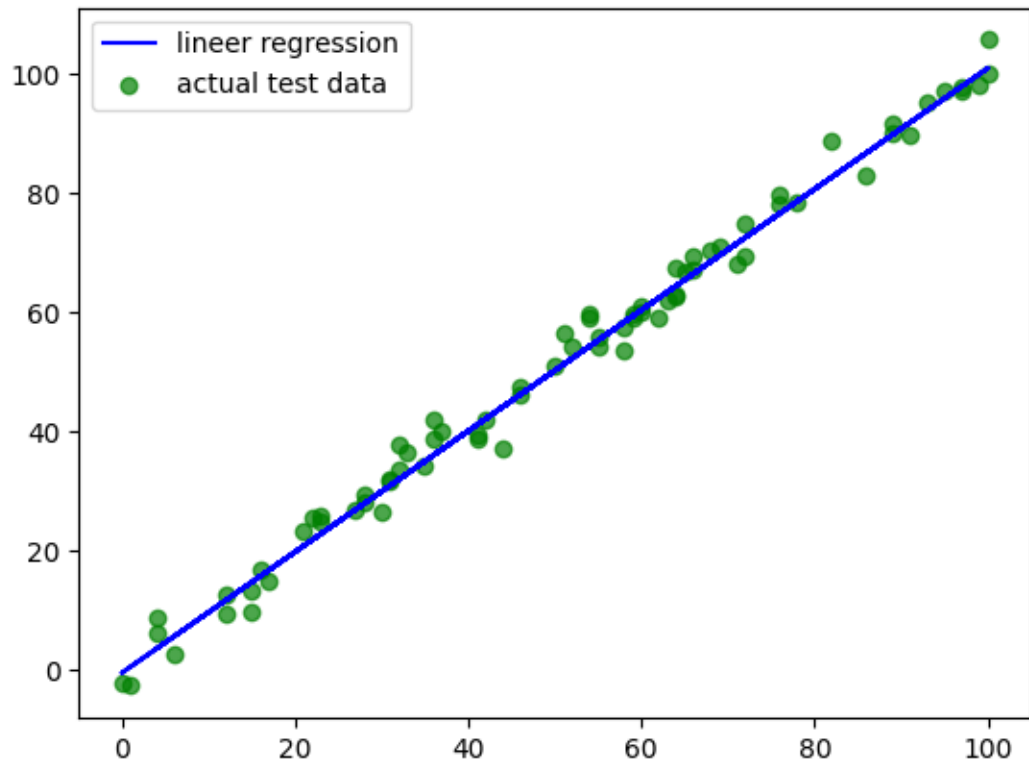
```
[31]: plt.scatter(x_train,y_train,label="Train Data",color='r',alpha=.7)
      plt.scatter(x_test,y_test,label="test Data",color='g',alpha=.7)
      plt.legend()
      plt.title('test train split')
      plt.show()
```

test train split

```
[33]: LR=LinearRegression()
      LR.fit(x_train.values.reshape(-1,1),y_train.values)
```

```
[33]: LinearRegression()
```

```
[35]: prediction=LR.predict(x_test.values.reshape(-1,1))
      plt.plot(x_test,prediction,label='lineer regression',color='b')
      plt.scatter(x_test,y_test,label='actual test data',color='g',alpha=0.7)
      plt.legend()
      plt.show()
```

[ ]: