Linear Regression

November 25, 2018

In [3]:
```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd

data = pd.read_csv('train.csv')
dataset = data.dropna()

test = pd.read_csv('test.csv')
test = test.dropna()

# check wheather to keep values or not
X = dataset.iloc[:, :-1].values
y = dataset.iloc[:, 1].values

X_test = test.iloc[:, :-1].values
y_test = test.iloc[:, 1].values

from sklearn.linear_model import LinearRegression

regressor = LinearRegression()
regressor.fit(X, y)

y_pred = regressor.predict(X_test)

plt.subplot(2,1,1)
plt.scatter(X, y, c = 'magenta')
plt.plot(X, regressor.predict(X), color = 'purple')
plt.xlabel("X datasets")
plt.ylabel("y datasets")
plt.title("Fitted regression line for the training data")
plt.show()

plt.subplot(2,1,2)
plt.scatter(X_test, y_test, c = 'orange')
plt.plot(X, regressor.predict(X), color = 'black')
plt.xlabel("X datasets")
plt.ylabel("y datasets")
```

```
plt.title("Fitted regression line for the test data")
plt.show()
```