9TH GRADE

Student Grades Prediction

การทำนายเกรดนักเรียน

Linear regression Algorithm

```
In [1]: import numpy as np
        import pandas as pd
        from sklearn.linear_model import LinearRegression
        from sklearn.model_selection import train_test_split
        from matplotlib import pyplot as plt
```

```
In [2]: data = pd.read_csv("student-mat.csv")
        data.head()
        #data.info()
```

Out[2]:

	school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	...	famrel	freetime	goout	Dalc	Walc	health	absences	G1	G2	G3
0	GP	F	18	U	GT3	A	4	4	at_home	teacher	...	4	3	4	1	1	3	6	5	6	6
1	GP	F	17	U	GT3	T	1	1	at_home	other	...	5	3	3	1	1	3	4	5	5	6
2	GP	F	15	U	LE3	T	1	1	at_home	other	...	4	3	2	2	3	3	10	7	8	10
3	GP	F	15	U	GT3	T	4	2	health	services	...	3	2	2	1	1	5	2	15	14	15
4	GP	F	16	U	GT3	T	3	3	other	other	...	4	3	2	1	2	5	4	6	10	10

โดย data set จะมีข้อมูลทั้งหมด 32 คอลัมน์

ข้อมูลเด็กนักเรียน 395 คน

DATA SET

Check out the beta version of the new UCI Machine Learning Repository we are currently testing! Contact us if you have any issues, questions, or concerns. Click here to try out the new site. ×

Student Performance Data Set

Download: Data Folder, Data Set Description

Abstract: Predict student performance in secondary education (high school).

Data Set Characteristics:	Multivariate	**Number of Instances:**	649	**Area:**	Social
Attribute Characteristics:	Integer	**Number of Attributes:**	33	**Date Donated**	2014-11-27
Associated Tasks:	Classification, Regression	**Missing Values?**	N/A	**Number of Web Hits:**	1044710

Source:

Paulo Cortez, University of Minho, GuimarÃ£es, Portugal, http://www3.dsi.uminho.pt/pcortez

DATA SET

Attribute Information:

Attributes for both student-mat.csv (Math course) and student-por.csv (Portuguese language course) datasets:
1 school - student's school (binary: 'GP' - Gabriel Pereira or 'MS' - Mousinho da Silveira)
2 sex - student's sex (binary: 'F' - female or 'M' - male)
3 age - student's age (numeric: from 15 to 22)
4 address - student's home address type (binary: 'U' - urban or 'R' - rural)
5 famsize - family size (binary: 'LE3' - less or equal to 3 or 'GT3' - greater than 3)
6 Pstatus - parent's cohabitation status (binary: 'T' - living together or 'A' - apart)
7 Medu - mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 â€" 5th to 9th grade, 3 â€" secondary education or 4 â€" higher education)
8 Fedu - father's education (numeric: 0 - none, 1 - primary education (4th grade), 2 â€" 5th to 9th grade, 3 â€" secondary education or 4 â€" higher education)
9 Mjob - mother's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other')
10 Fjob - father's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other')
11 reason - reason to choose this school (nominal: close to 'home', school 'reputation', 'course' preference or 'other')
12 guardian - student's guardian (nominal: 'mother', 'father' or 'other')
13 traveltime - home to school travel time (numeric: 1 - <15 min., 2 - 15 to 30 min., 3 - 30 min. to 1 hour, or 4 - >1 hour)
14 studytime - weekly study time (numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours)
15 failures - number of past class failures (numeric: n if 1<=n<3, else 4)
16 schoolsup - extra educational support (binary: yes or no)
17 famsup - family educational support (binary: yes or no)
18 paid - extra paid classes within the course subject (Math or Portuguese) (binary: yes or no)
19 activities - extra-curricular activities (binary: yes or no)
20 nursery - attended nursery school (binary: yes or no)
21 higher - wants to take higher education (binary: yes or no)
22 internet - Internet access at home (binary: yes or no)
23 romantic - with a romantic relationship (binary: yes or no)
24 famrel - quality of family relationships (numeric: from 1 - very bad to 5 - excellent)
25 freetime - free time after school (numeric: from 1 - very low to 5 - very high)
26 goout - going out with friends (numeric: from 1 - very low to 5 - very high)
27 Dalc - workday alcohol consumption (numeric: from 1 - very low to 5 - very high)
28 Walc - weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)
29 health - current health status (numeric: from 1 - very bad to 5 - very good)
30 absences - number of school absences (numeric: from 0 to 93)

these grades are related with the course subject, Math or Portuguese:
31 G1 - first period grade (numeric: from 0 to 20)
31 G2 - second period grade (numeric: from 0 to 20)
32 G3 - final grade (numeric: from 0 to 20, output target)

อ่านรายละเอียดเพิ่มเติม

https://archive.ics.uci.edu/ml/datasets/Student+Performance#

CODE

In [3]:
```
data = data[["G1", "G2", "G3", "studytime", "failures", "absences"]]
predict = "G3"
x = np.array(data.drop([predict], 1))
y = np.array(data[predict])

x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2)
```

In [4]:
```
linear_regression = LinearRegression()
linear_regression.fit(x_train, y_train)
result = linear_regression.score(x_test, y_test)
print(result)
```

```
0.8992967960785373
```

Test 20% และ Train 80%

ความแม่นยำ 89%

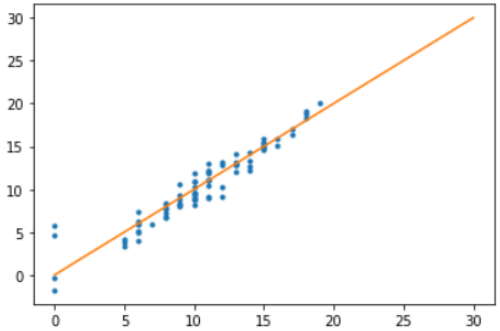
CODE

In [5]:
```
predictions = linear_regression.predict(x_test)

print(predictions[x], x_test[x], y_test[x])
```

```
[[ 8.34743983  6.31363602  4.91534485 12.86002299  6.31363602]
 [ 8.34743983  8.34743983  4.91534485 12.86002299 16.42262056]
 [ 5.81850882 19.1915428   4.91534485  9.51825514 11.0477117 ]
 ...
 [11.0477117  19.1915428   5.90724692  9.51825514  9.51825514]
 [10.86853588  8.8730898   5.90724692 12.86002299 12.86002299]
 [19.1915428  10.55761543  5.90724692 12.86002299  8.34743983]] [[[10  9  3  0  4]
  [ 8  7  2  0  4]
  [ 6  6  2  0  2]
  [11 13  1  1 10]
  [ 8  7  2  0  4]]

 [[10  9  3  0  4]
  [10  9  3  0  4]
  [ 6  6  2  0  2]
  [11 13  1  1 10]
  [16 16  2  0  2]]

 [[ 6  7  2  0  0]
  [17 18  2  0 21]
  [ 6  6  2  0  2]
  [10 10  2  0  2]
  [ 9 12  3  0  3]]

 ...

 [[ 9 12  3  0  3]
  [17 18  2  0 21]
  [ 7  6  1  0 18]
  [10 10  2  0  2]
  [10 10  2  0  2]]

 [[11 11  1  0  0]
```

CODE

In [6]:
```
y_pred = linear_regression.predict(x_test)
plt.plot(y_test,y_pred, '.')

#line
x = np.linspace(0,30, 50)
y = x
plt.plot(x, y)
plt.show()
```



REFERENCE

Paulo Cortez. (2008). *Student Performance Data Set*. สืบค้น 25 ตุลาคม 2564, จาก https://archive.ics.uci.edu/ml/datasets/Student+Performance#

AMAN KHARWAL. (2020). *Linear Regression with Python*. สืบค้น 26 ตุลาคม 2564, จาก https://thecleverprogrammer.com/2020/11/20/linear-regression-with-python/

THANK YOU

จัดทำโดย นางสาว สุนิชา เฟื่องฟู 6210450776