

4.7 แบบฝึกหัด

1. [simple linear regression] จากข้อมูลของ X และ Y ที่กำหนดให้ ให้ทำงานต่อไปนี้

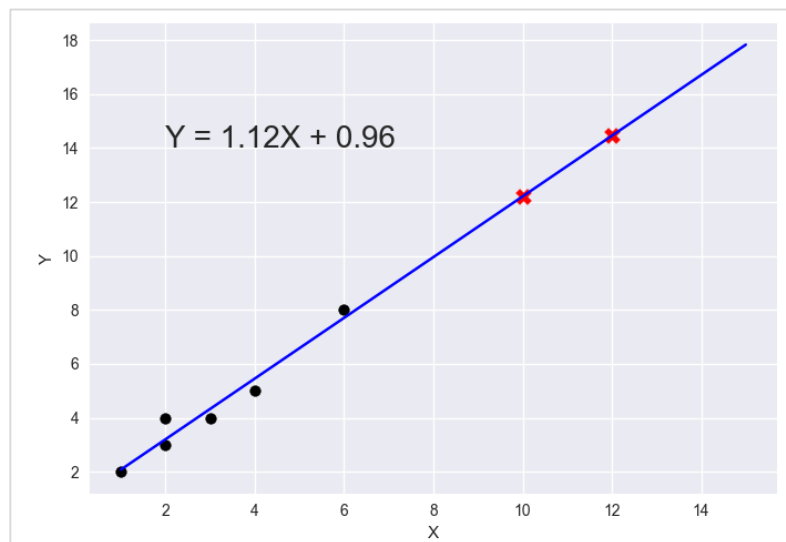
a. โดยการคำนวณ: **prymania note : can show output from the code**

- สร้างสมการ linear regression, หาค่า R^2 และ MSE
- วิเคราะห์ว่าข้อมูลชุดนี้ควรใช้การวิเคราะห์แบบ simple linear regression หรือไม่
- หาค่าของ Y เมื่อ X มีค่าเป็น 10 และ 12

b. โดย python code:

- แสดง output: สมการ linear regression, หา R^2 และ MSE
- วาดกราฟโดยแสดง
 - ข้อมูลทั้งหมดเป็นจุดสีเขียว
 - สมการ linear regression เป็นเส้นสีฟ้า จากจุด 0 – 15
 - สมการ linear regression บนกราฟ
 - จุดการทำนายค่าของ Y เมื่อ X มีค่าเป็น 10 และ 12 เป็นกากบาทสีแดง

Y	X
2	1
3	2
4	2
4	3
5	4
8	6



2. [simple linear regression]

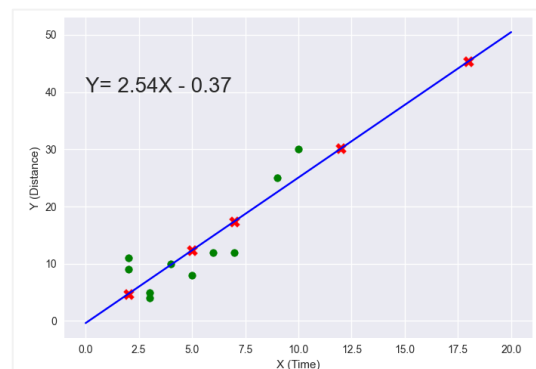
เขียน python โดยใช้ข้อมูลการวิ่งในตัวอย่าง 4.2 เพื่อวิเคราะห์ simple linear regression

- แสดง output: สมการ linear regression, หา R^2 และ MSE
- วาดกราฟโดยแสดง
 - ข้อมูลทั้งหมดเป็นจุดสีเขียว
 - สมการ linear regression เป็นเส้นสีฟ้า จากจุด 0 – 20
 - สมการ linear regression บนกราฟ
 - จุดการทำนายค่าของ Y เมื่อ X มีค่าเป็น 2, 5, 7, 12 และ 18 เป็นกากบาทสีแดง
- output และกราฟควรเท่ากับหรือใกล้เคียงกับที่คำนวณได้ในตัวอย่าง 4.2

คนที่	ระยะทาง (เมตร)	เวลา (วินาที)
1	5	3
2	9	2
3	12	7
4	11	2
5	12	6
6	10	4
7	8	5
8	4	3
9	25	9
10	30	10

```
Linear equation:   Y = 2.54X - 0.37
R_squared      : 0.75
MSE       : 16.09

Predicted response of X:
2      4.72
5      12.35
7      17.43
12     30.15
18     45.41
```



3. [simple linear regression]

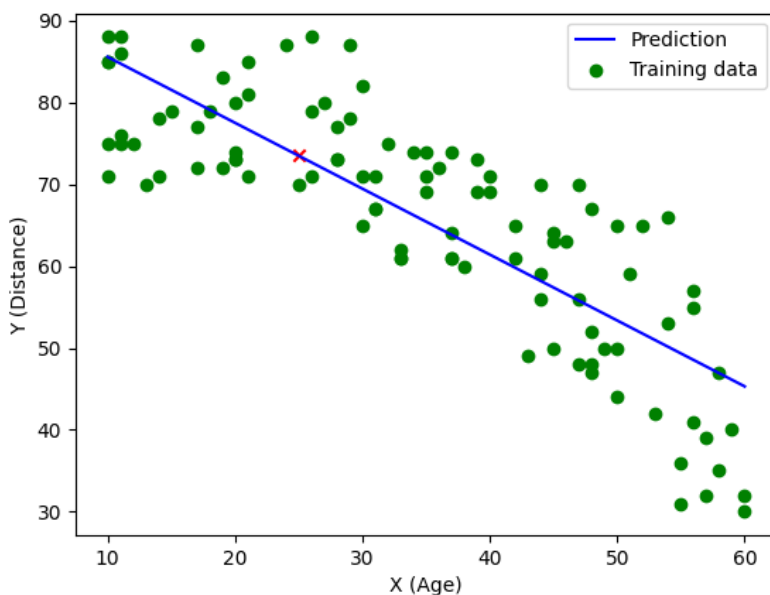
อ่านข้อมูลจากไฟล์ simple_run.csv ซึ่งเก็บข้อมูลอายุและระยะทางที่แต่ละคนวิ่งได้ใน 10 วินาที

ตัวอย่างข้อมูล

age	distance
10	71
10	88
11	88
11	75

- วาดกราฟโดยแสดง
 - จุดสีเขียว แสดงข้อมูลทั้งหมด
 - เส้นสีน้ำเงิน แสดง linear regression
 - จุด x สีแดง แสดงการประเมินว่าคนอายุ 25 วิ่งได้ระยะทางเท่าไร
- สมการ linear regression คือ ?
- ค่า R^2 และ MSE ของ Training data คือ ?
- ประเมินว่าคนอายุ 25 ปี ควรวิ่งได้ระยะทางเท่าไร
- วิเคราะห์และอธิบายความสัมพันธ์ระหว่างตัวแปร X และ Y
- ข้อมูลชุดนี้ควรใช้การวิเคราะห์แบบ linear regression หรือไม่

ตัวอย่างกราฟที่ต้องการ



4. [Multiple linear regression]

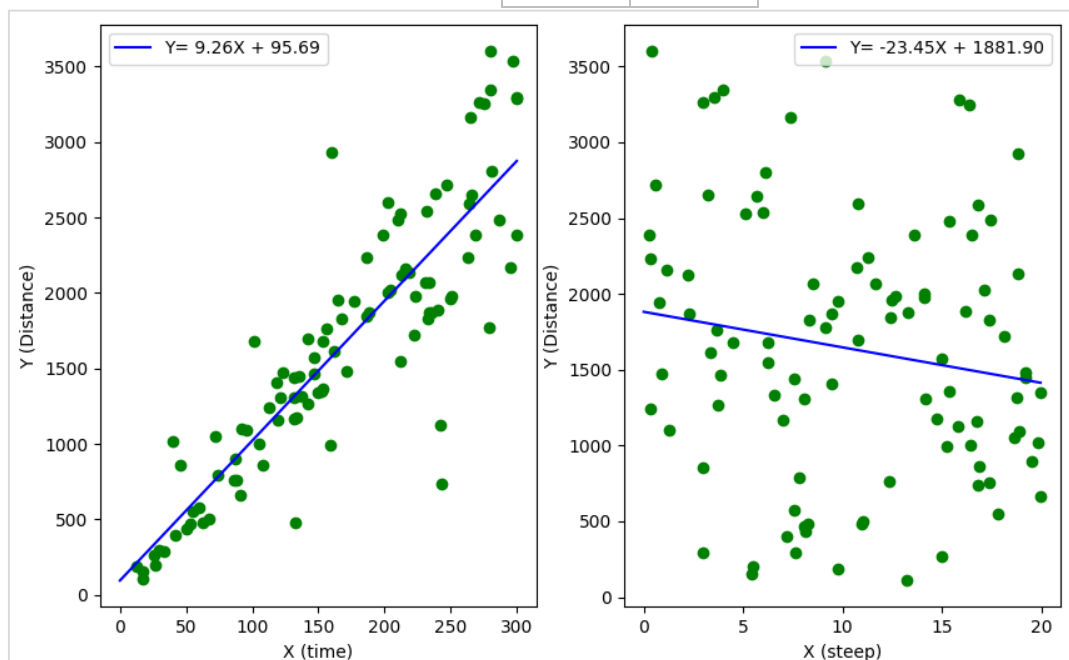
ไฟล์ multi_run.csv เป็นข้อมูลการวิ่ง เป็นความสัมพันธ์ระหว่าง เวลา และความชัน (Steep) ซึ่งส่งผลต่อระยะทางการวิ่ง ให้วิเคราะห์ความสัมพันธ์ของข้อมูลชุดนี้ โดย

ตัวอย่างข้อมูล

distance	time	steep
109.58	17	13.21
153.77	17	5.41
267.12	26	14.96
199.5	27	5.5

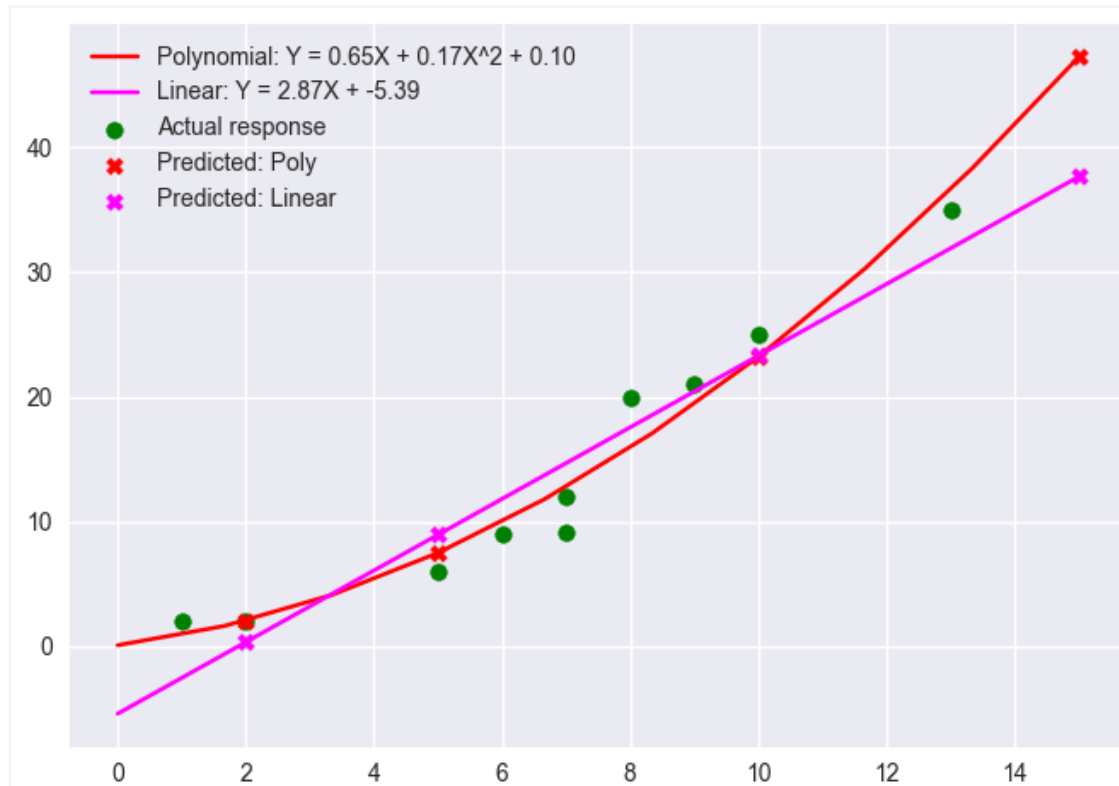
- แสดง multiple linear regression equation
- แสดงค่า R^2 และ MSE ของ Training data และ Testing data
- วิเคราะห์ค่า R^2 และ MSE ของ Training data และ Testing data ว่าเป็นเพราะอะไร และควรจัดการการเทรนข้อมูลชุดนี้อย่างไร
- แยกการวิเคราะห์เป็น simple linear regression โดยวิเคราะห์เป็น
time --> distance และ Steep --> distance และแสดงกราฟของทั้ง 2 ตัวแปร
- ทำนายว่า ถ้าวิ่งด้วยข้อมูลต่อไปนี้จะได้ระยะทางเท่าใด โดยใช้ multiple regression

time	steep
10	20
15	2
20	10



5. [เปรียบเทียบ Polynomial regression และ Linear regression]

อ่านข้อมูลจากไฟล์ polydata.csv วิเคราะห์ข้อมูลด้วย Linear regression และ Polynomial degree 2 แสดงกราฟ



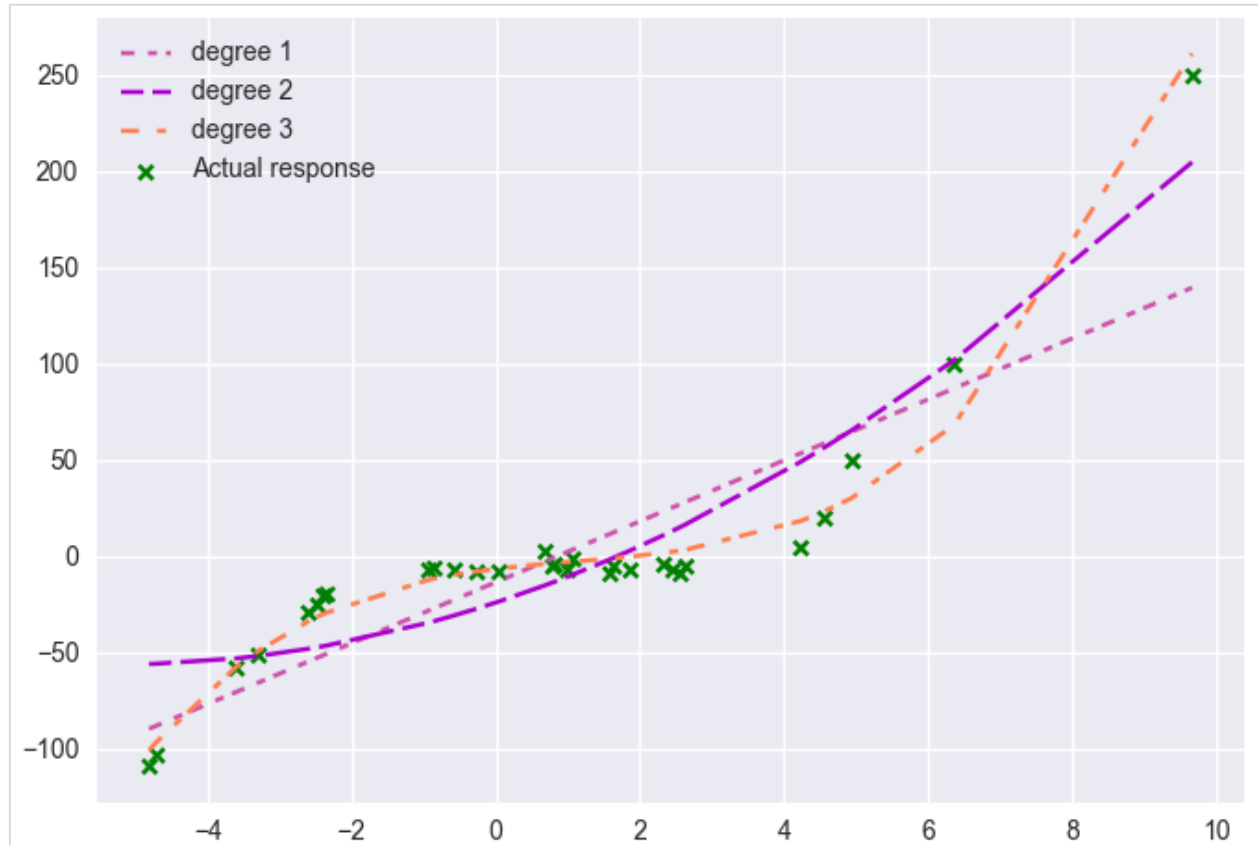
6. [Polynomial, degree 2, 3, 4]

อ่านข้อมูลจากไฟล์ rate.csv ซึ่งมีข้อมูล X และ Y วิเคราะห์ข้อมูลด้วย Polynomial degree 1, 2 และ 3

วิเคราะห์หาค่า R^2 และ MSE ว่าข้อมูลชุดนี้ควรใช้การวิเคราะห์แบบใด

สร้างกราฟแสดงข้อมูลทั้งหมด และเส้น polynomial regression ทั้ง 3 degree

```
Degree 1:  
a = -13.67  
b = ['0.00', '15.84']  
r2 = 0.74, mse= 922.21  
  
Degree 2:  
a = -24.24  
b = ['0.00', '12.30', '1.18']  
r2 = 0.83, mse= 603.75  
  
Degree 3:  
a = -6.76  
b = ['0.00', '4.51', '-1.29', '0.38']  
r2 = 0.97, mse= 90.30
```



7. [Logistic regression]

จากการเก็บข้อมูลการสอบสัมภาษณ์ของบริษัทแห่งหนึ่ง พบว่าข้อมูลที่ส่งผลต่อการตัดสินใจรับผู้สมัครเข้าทำงาน คือ คะแนนสอบภาษาอังกฤษ คะแนนเฉลี่ย และประสบการณ์การทำงานของผู้สมัคร

ตัวอย่างข้อมูลจากไฟล์ interview.csv

admitted	english	gpa	experience
Accepted	470	3.74	9
Accepted	410	2.73	2
Accepted	410	3.3	7
Rejected	270	1.82	2
Rejected	400	2.87	3

- สร้างโมเดลวิเคราะห์ข้อมูลชุดนี้ด้วย logistic regression แสดง confusion matrix และ accuracy
- ทำนายว่าบุคคลต่อไปนี้มีโอกาสสอบสัมภาษณ์ผ่านหรือไม่

English	GPA	Experience
400	2.2	1
255	3.5	5
390	2.8	6
100	4.00	10

4.8 References

<https://sites.google.com/site/mystatistics01/regression-correlation-analysis/simple-linear-regression-step>

<https://towardsdatascience.com/linear-regression-python-implementation-ae0d95348ac4>

<https://stackabuse.com/linear-regression-in-python-with-scikit-learn/>

<https://www.geeksforgeeks.org/linear-regression-python-implementation/>

<https://www.kaggle.com/sudhirnl7/linear-regression-tutorial>