

LINEAR REGRESSION

BY TAUTOLOGY

Linear Regression



Introduction

Introduction

What is Linear
Regression?

Data for Linear
Regression

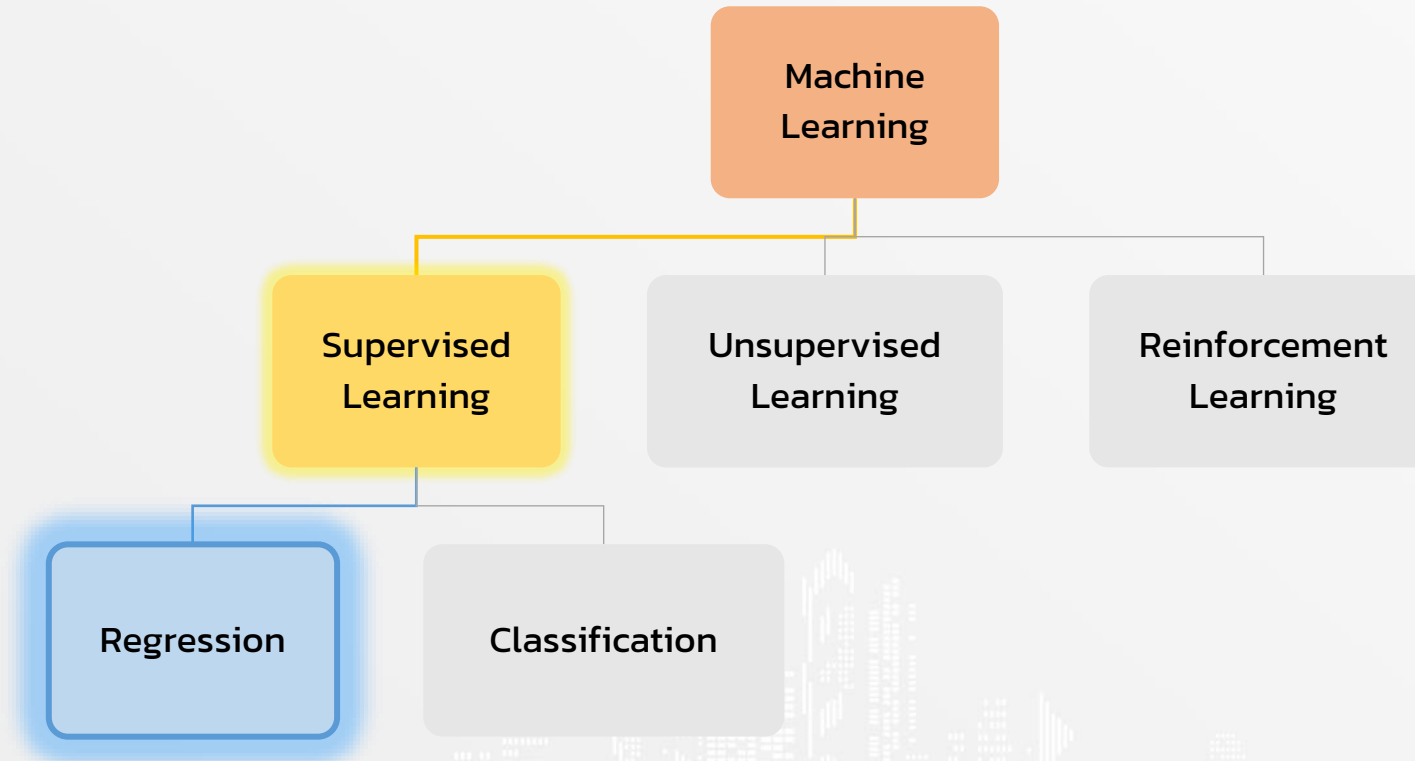
Pros & Cons

Real World
Application

What is Linear Regression?

Linear Regression เป็นหนึ่งใน algorithm ประเภท **supervised learning** ที่ใช้สำหรับแก้ปัญหา regression โดยมีหลักการทำงานคือ การสร้างสมการเชิงเส้นที่ใช้ตัวแปรต้นเพื่อพยากรณ์ตัวแปรตาม

What is Linear Regression?



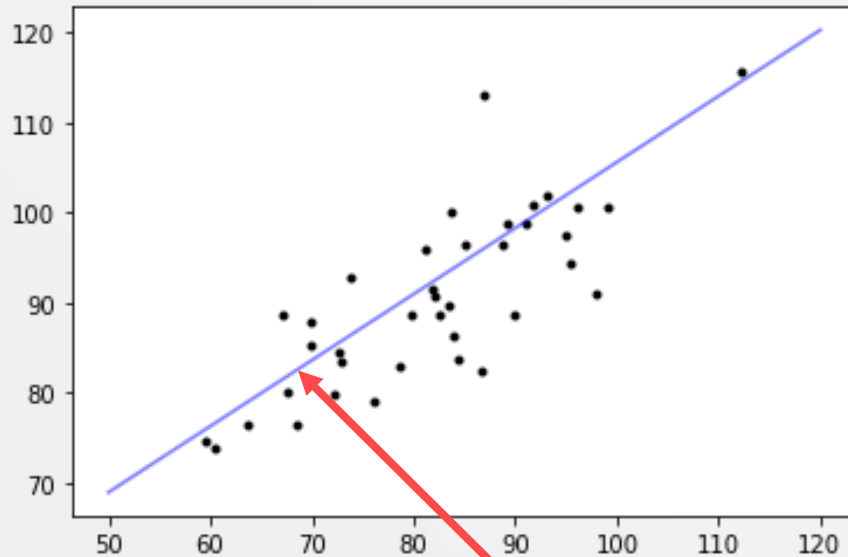
What is Linear Regression?

สมการคณิตศาสตร์ของ Linear Regression

$$\hat{y} = w_0 + w_1x_1 + w_2x_2 + \dots + w_px_p$$

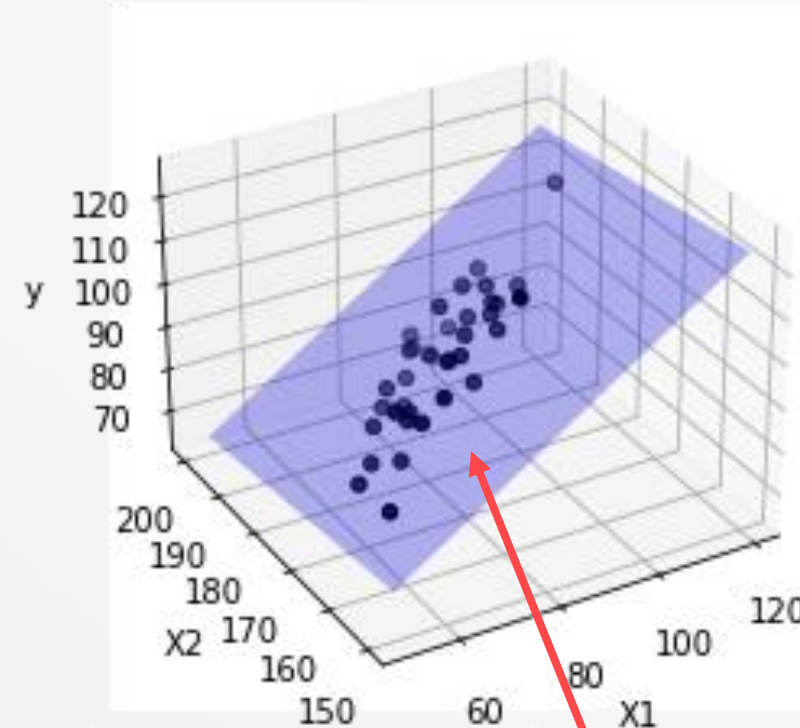
- โดย
- \hat{y} คือ ตัวแปรตาม (predicted target)
 - $x_1, x_2, x_3, \dots, x_p$ คือ ตัวแปรต้น (feature)
 - $w_0, w_1, w_2, \dots, w_p$ คือ สัมประสิทธิ์ (coefficient)

What is Linear Regression?



$$\hat{y} = w_0 + w_1x_1$$

$$(\hat{y} = a + bx)$$



$$\hat{y} = w_0 + w_1x_1 + w_2x_2$$

Introduction

What is Linear Regression?



Data for Linear Regression



Pros & Cons

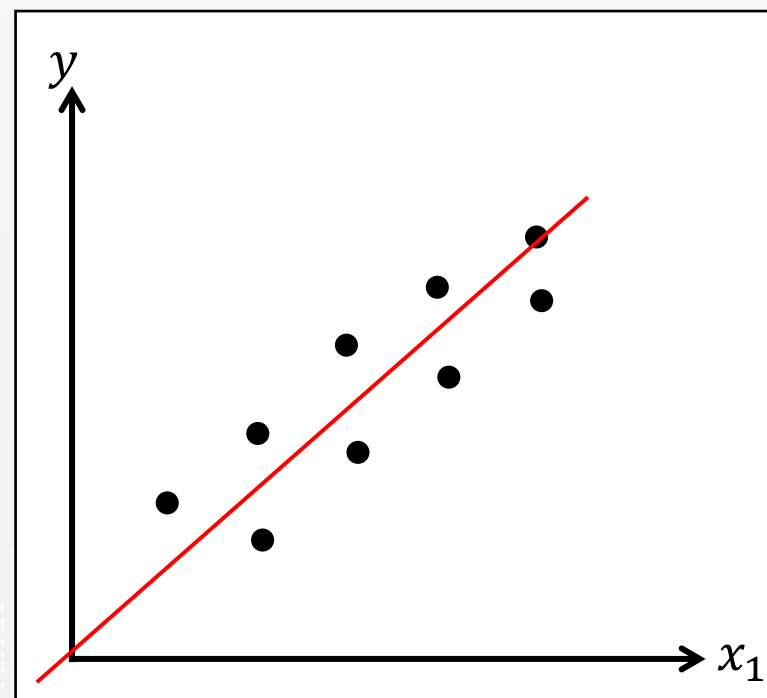
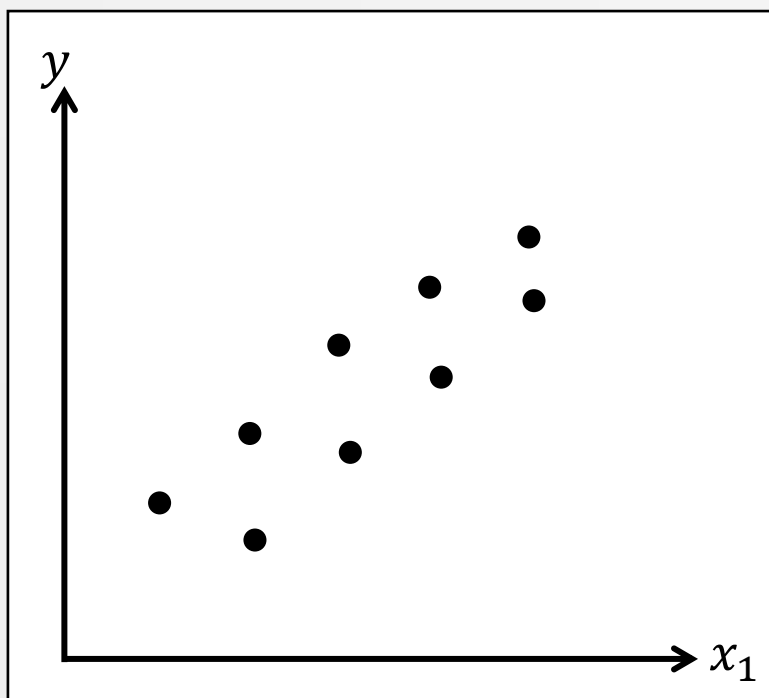


Real World Application



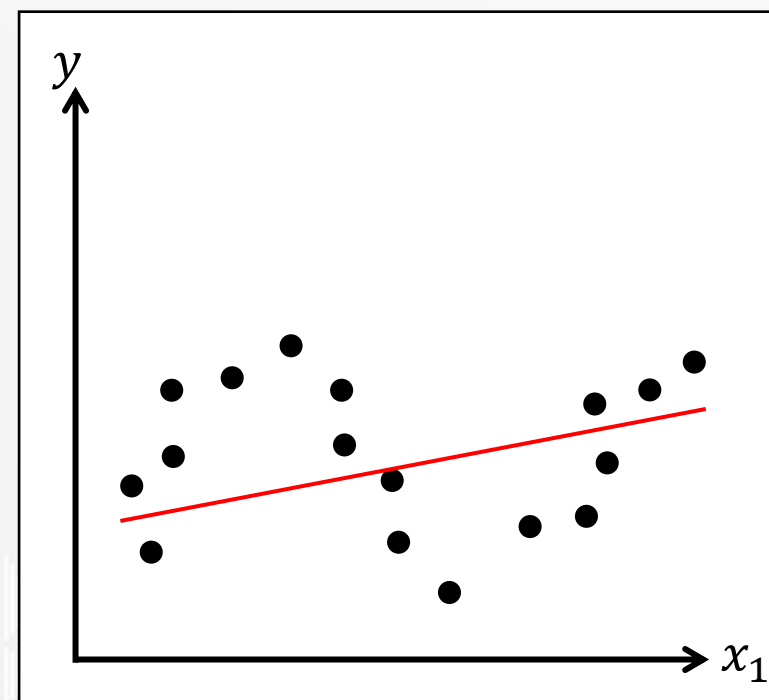
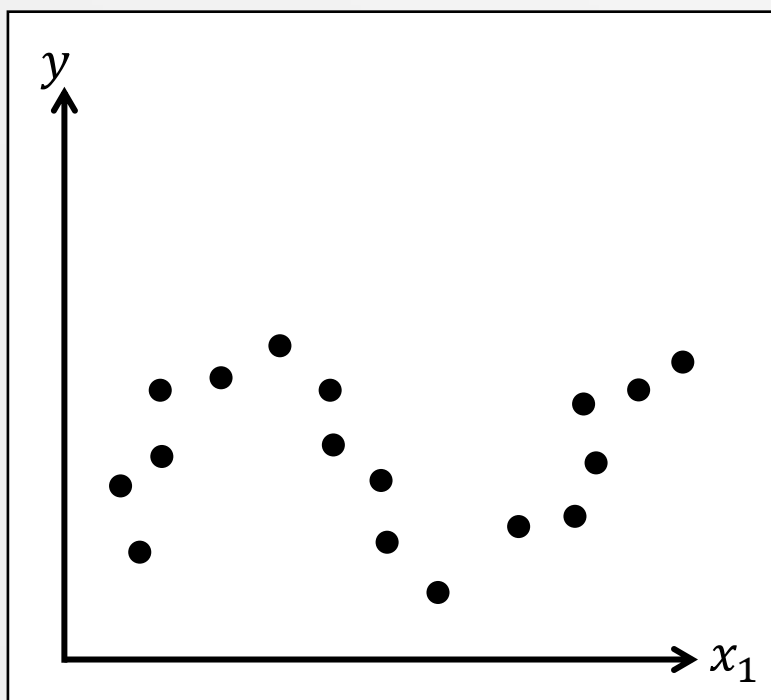
Data for Linear Regression

ตัวอย่างของข้อมูลที่เหมาะกับ Linear Regression



Data for Linear Regression

ตัวอย่างของข้อมูลที่ไม่เหมาะกับ Linear Regression



Introduction

**What is Linear
Regression?**



**Data for Linear
Regression**



Pros & Cons



**Real World
Application**



Pros & Cons

ข้อดี

- เป็น algorithm ที่เรียบง่าย
- เป็น algorithm ที่เหมาะสมที่สุดกับข้อมูลที่มีความสัมพันธ์ระหว่างตัวแปรต้นและตัวแปรตามเป็นเชิงเส้น

ข้อเสีย

- sensitive กับ outliers

ข้อจำกัด

- ใช้งานได้กับข้อมูลที่มีความสัมพันธ์ระหว่างตัวแปรต้นและตัวแปรตามเป็นเชิงเส้นเท่านั้น

Introduction

**What is Linear
Regression?**



**Data for Linear
Regression**



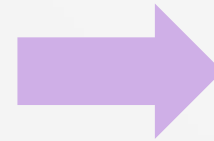
Pros & Cons



**Real World
Application**



Real World Application



**การประเมินค่าตัวของนัก
ฟุตบอลในตำแหน่งกองหน้า**
โดยพิจารณาจาก อายุ ส่วนสูง การมี
ส่วนร่วมกับประตู จำนวนเกมที่ลง
เล่น เป็นต้น

อ้างอิง : [2018, Yunus et al] Multiple Linear Regression Approach
For Estimating the Market Value of Football Players in Forward
Position

Real World Application



การคาดการณ์ค่าน้ำค่าไฟ ของฟาร์มโคนม

โดยพิจารณาจาก จำนวนโคนม
จำนวนเครื่องรีดนมวัว จำนวนเครื่อง
ขัดพื้นไฟฟ้า เป็นต้น

อ้างอิง : [2018, Shine et al] Multiple linear regression modelling of on-farm direct water and electricity consumption on pasture based dairy farms

Real World Application



**พยากรณ์ปริมาณพืชที่จะ
ปลูกได้**

โดยพิจารณาจาก สภาพดิน สภาพ
อากาศ ปริมาณน้ำฝน อุณหภูมิ เป็น
ต้น

อ้างอิง : [2017, Aditya Shastry, HA Sanjay and E. Bhanusree]
Prediction of Crop Yield Using Regression Techniques

Introduction

**What is Linear
Regression?**



**Data for Linear
Regression**



Pros & Cons



**Real World
Application**

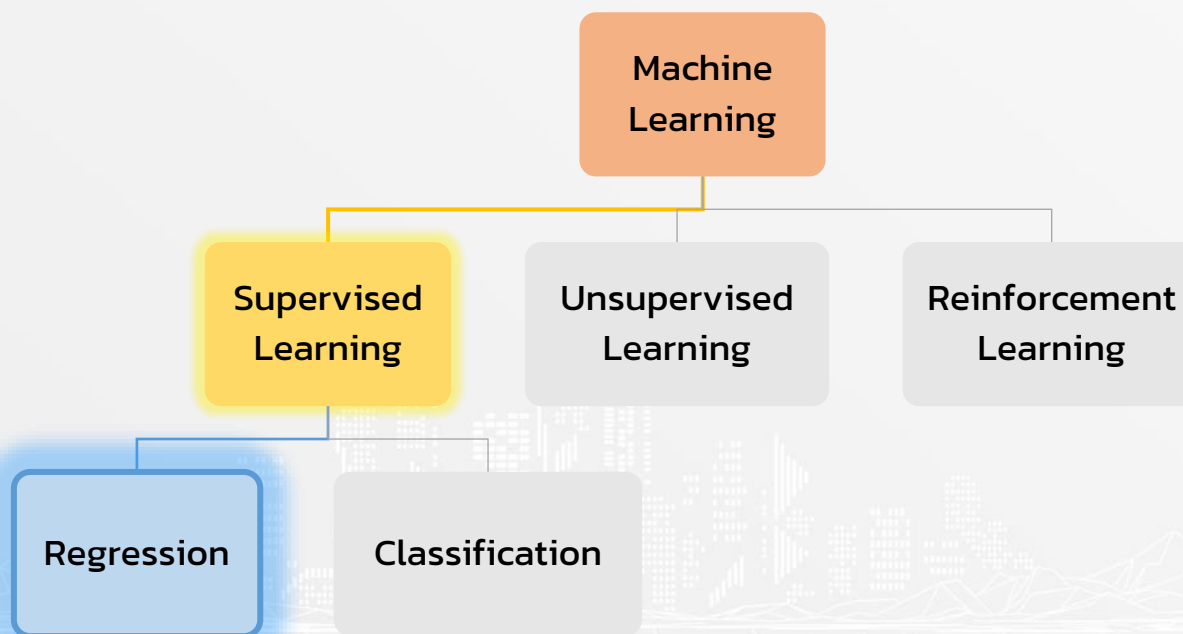


Linear Regression



Linear Regression

Linear Regression เป็นหนึ่งใน algorithm ประเภท supervised learning



Concept of Supervised Learning

Data \Rightarrow **Model** \Rightarrow **Prediction**

Model Creation

Model

Assumption

Real Face of the
Model

Cost Function and
Cost Landscape

How to Create
Model (Math)

How to Create
Model (Code)

Further Reading

Assumption

1. Linear Relationship
2. Normality of Residuals
3. Homoscedasticity
4. No Missing Features
5. No Multicollinearity

Model

Assumption



Real Face of the
Model



Cost Function and
Cost Landscape



How to Create
Model (Math)



How to Create
Model (Code)

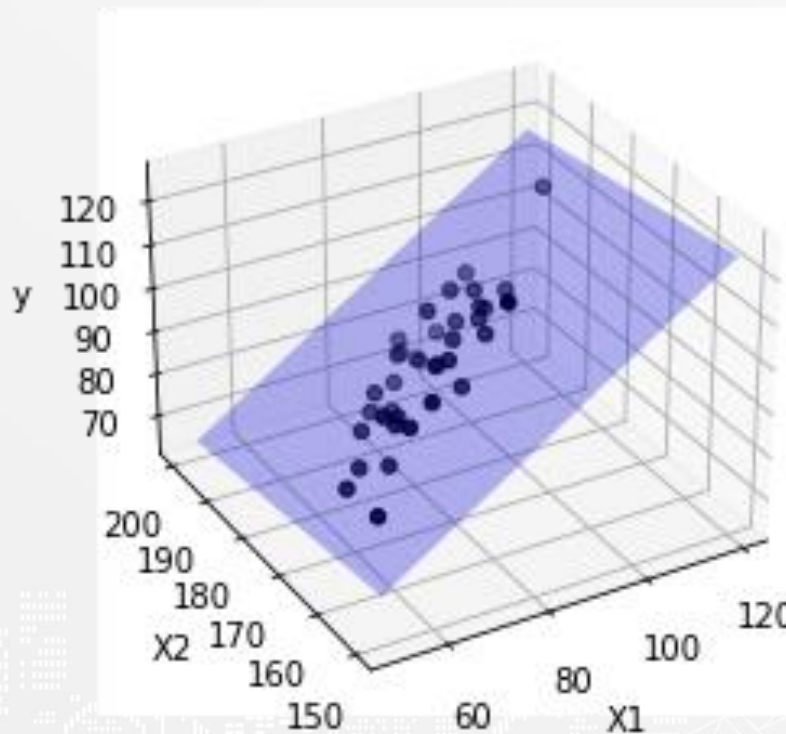


Further Reading



Real Face of the Model

Linear regression คือ สมการเชิงเส้นที่ใช้ตัวแปรต้นเพื่อพยากรณ์ตัวแปรตาม



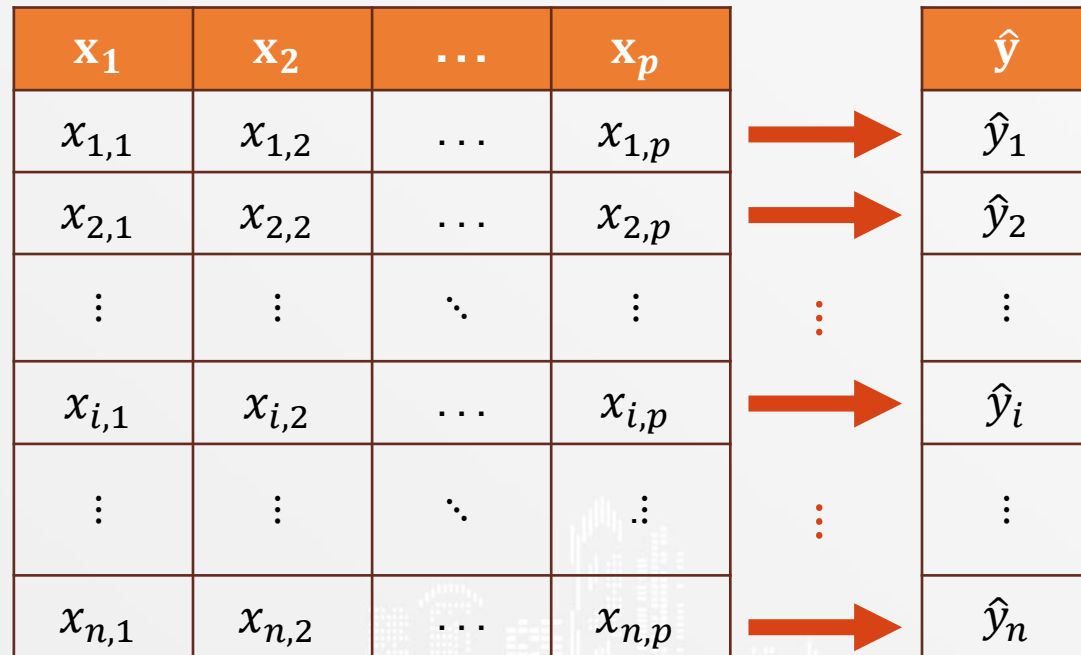
Real Face of the Model

Linear regression คือ สมการเชิงเส้นที่ใช้ตัวแปรต้นเพื่อพยากรณ์ตัวแปรตาม

$$\hat{y} = w_0 + w_1x_1 + w_2x_2 + \dots + w_px_p$$

- โดย
- ♦ \hat{y} คือ ตัวแปรตาม (predicted target)
 - ♦ x_1, x_2, \dots, x_p คือ ตัวแปรต้น (feature)
 - ♦ w_0, w_1, \dots, w_p คือ สัมประสิทธิ์ (coefficient)

Real Face of the Model



Real Face of the Model

$$\hat{y} = w_0 + w_1x_1 + w_2x_2 + \cdots + w_px_p$$



$$\hat{y}_1 = w_0 + w_1x_{1,1} + w_2x_{1,2} + \cdots + w_px_{1,p}$$

$$\hat{y}_2 = w_0 + w_1x_{2,1} + w_2x_{2,2} + \cdots + w_px_{2,p}$$

⋮

$$\hat{y}_i = w_0 + w_1x_{i,1} + w_2x_{i,2} + \cdots + w_px_{i,p}$$

⋮

$$\hat{y}_n = w_0 + w_1x_{n,1} + w_2x_{n,2} + \cdots + w_px_{n,p}$$

Real Face of the Model

“เป้าหมายของเราคือการหา $w_0, w_1, w_2, \dots, w_p$ เพื่อสร้าง model ของ linear regression :

$$\hat{y} = w_0 + w_1x_1 + w_2x_2 + \dots + w_px_p$$
”



Real Face of the Model

“ โดยที่ $w_0, w_1, w_2, \dots, w_p$ เหล่านี้ต้องทำให้ทำให้ผลรวมของ error ระหว่างค่าจริง (y_i) กับค่าพยากรณ์ (\hat{y}_i) น้อยที่สุด ”

| y | | \hat{y} |
|----------|-----------------------|-------------|
| y_1 | \longleftrightarrow | \hat{y}_1 |
| y_2 | \longleftrightarrow | \hat{y}_2 |
| \vdots | \vdots | \vdots |
| y_n | \longleftrightarrow | \hat{y}_n |



Real Face of the Model

“ เราต้องการหา $w_0, w_1, w_2, \dots, w_p$ ที่ทำให้ผลรวมของ error ระหว่าง y_i กับ \hat{y}_i น้อยที่สุด ”



Real Face of the Model

ผลรวมของ error ระหว่าง y_i กับ \hat{y}_i เรียกว่า “Cost function”

Real Face of the Model



“เราต้องการหา $w_0, w_1, w_2, \dots, w_p$ ที่ทำให้
cost function ต่ำที่สุด”

Model

Assumption



**Real Face of the
Model**



**Cost Function and
Cost Landscape**



**How to Create
Model (Math)**



**How to Create
Model (Code)**



Further Reading



Cost Function and Cost Landscape

Cost function ที่เราจะใช้ในการสร้าง model คือ

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2$$

โดยสูตรข้างต้นมีชื่อว่า Sum of Squared Errors หรือ SSE

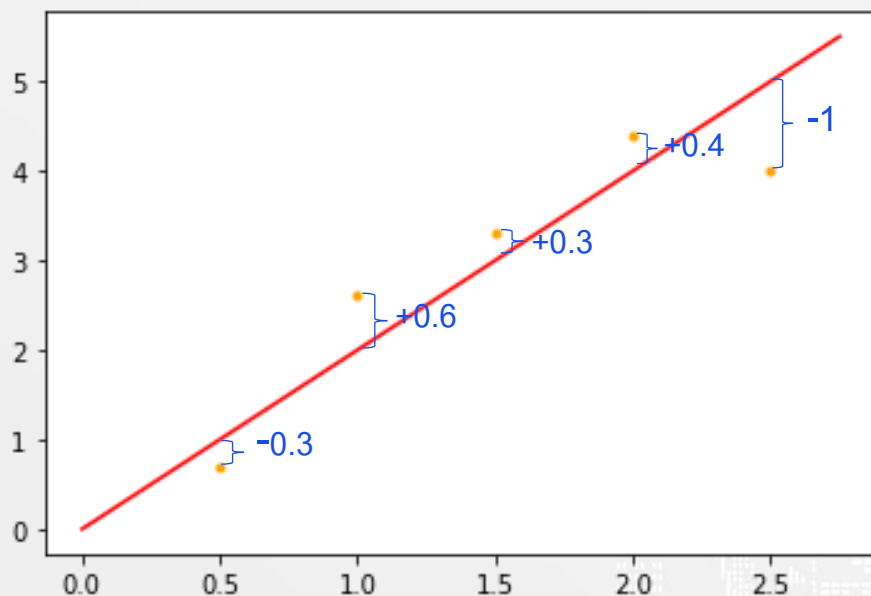
Cost Function and Cost Landscape

เหตุผลที่เลือกใช้ Sum of Squared Errors (SSE)

1. Error ของแต่ละ sample จะไม่หักล้างกัน
2. Cost function ที่นิยามแบบ SSE จะสามารถ diff ได้ และมีความต่อเนื่องทุกจุด
3. Cost function ที่นิยามแบบ SSE เป็น convex function และมีจุดต่ำสุดเพียงจุดเดียว

Cost Function and Cost Landscape

1. Error ของแต่ละ sample จะไม่หักล้างกัน



กราฟแสดงข้อมูลระหว่างค่าจริง y_i (จุด) และค่าพยากรณ์ \hat{y}_i (เส้น)

$$SE = \sum_{i=1}^n (y_i - \hat{y}_i)$$

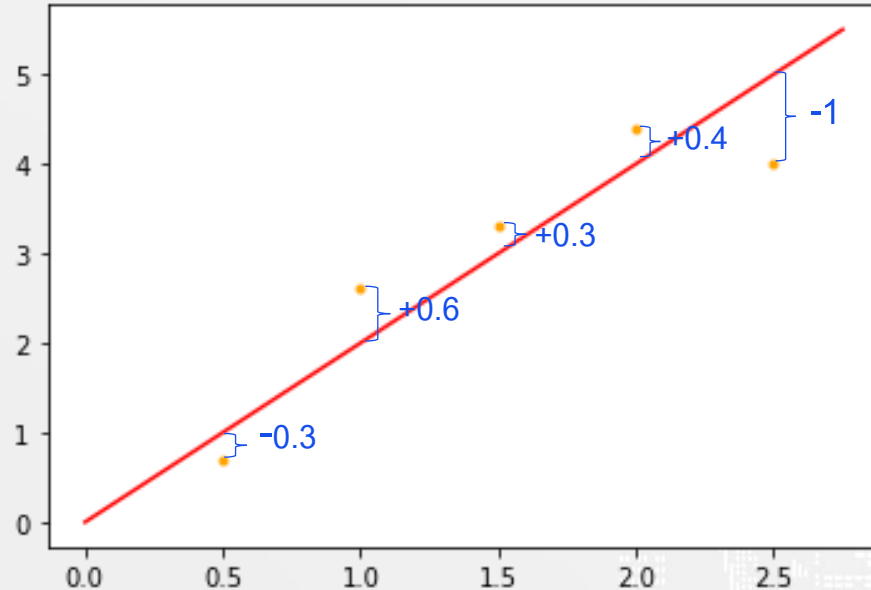
$$= (-0.3) + (0.6) + (0.3) + (0.4) + (-1)$$

$$= 0$$

 การมีค่าติดลบทำให้ error หักล้างกัน

Cost Function and Cost Landscape

1. Error ของแต่ละ sample จะไม่หักล้างกัน



กราฟแสดงข้อมูลระหว่างค่าจริง y_i (จุด) และค่าพยากรณ์ \hat{y}_i (เส้น)

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$= (-0.3)^2 + (0.7)^2 + (0.2)^2 + (0.4)^2 + (-1)^2$$

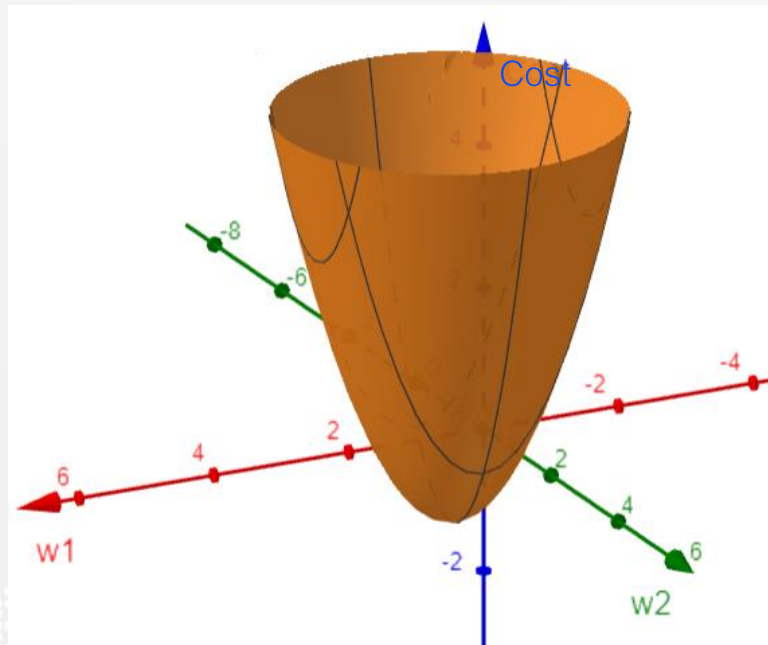
$$= 0.09 + 0.49 + 0.04 + 0.16 + 1$$

$$= 1.78$$

☑ การยกกำลังสองทำให้ error ไม่หักล้างกัน

Cost Function and Cost Landscape

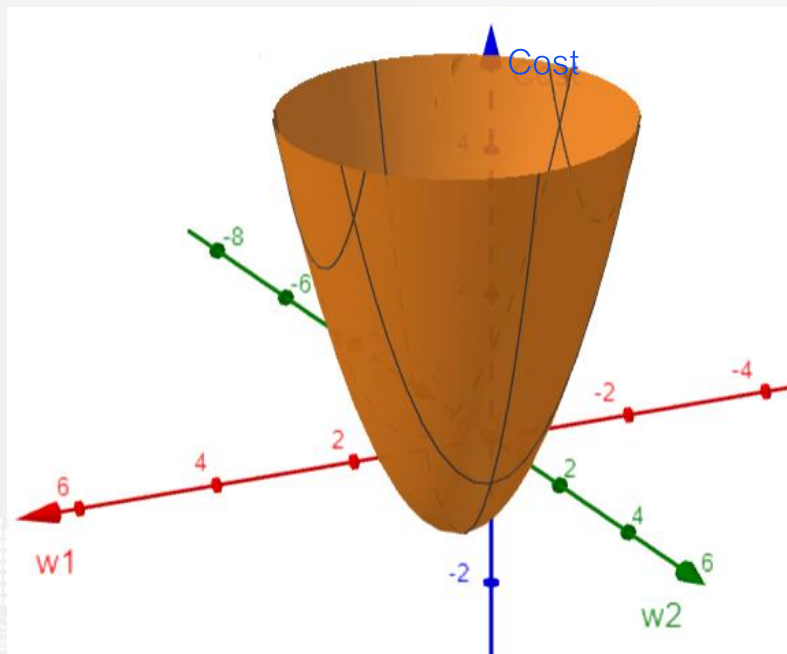
2. Cost function ที่นิยามแบบ SSE จะสามารถ diff ได้ และมีความต่อเนื่องทุกจุด



กราฟของ cost landscape โดยที่ cost function เป็น SSE

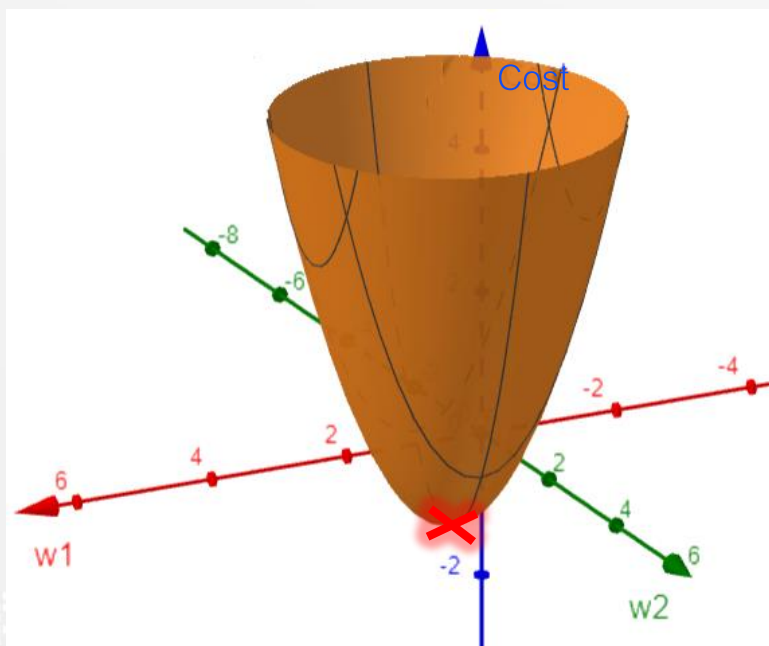
Cost Function and Cost Landscape

การที่ function **diff** ได้ และต่อเนื่องทุกจุดทำให้เราสามารถ**ใช้ calculus** ได้อย่างเต็มที่



Cost Function and Cost Landscape

3. Cost function ที่นิยามแบบ SSE เป็น convex function และมีจุดต่ำสุดเพียงจุดเดียว



กราฟของ cost landscape โดยที่ cost function เป็น SSE

Cost Function and Cost Landscape



Cost Landscape Plotting



Open File

Cost_Landscape_Plotting.pdf

Cost Function and Cost Landscape



Convexity of SSE



Open File

Convexity_SSE_LR.pdf

Model

Assumption



**Real Face of the
Model**



**Cost Function and
Cost Landscape**



**How to Create
Model (Math)**



**How to Create
Model (Code)**



Further Reading



How to Create Model (Math)



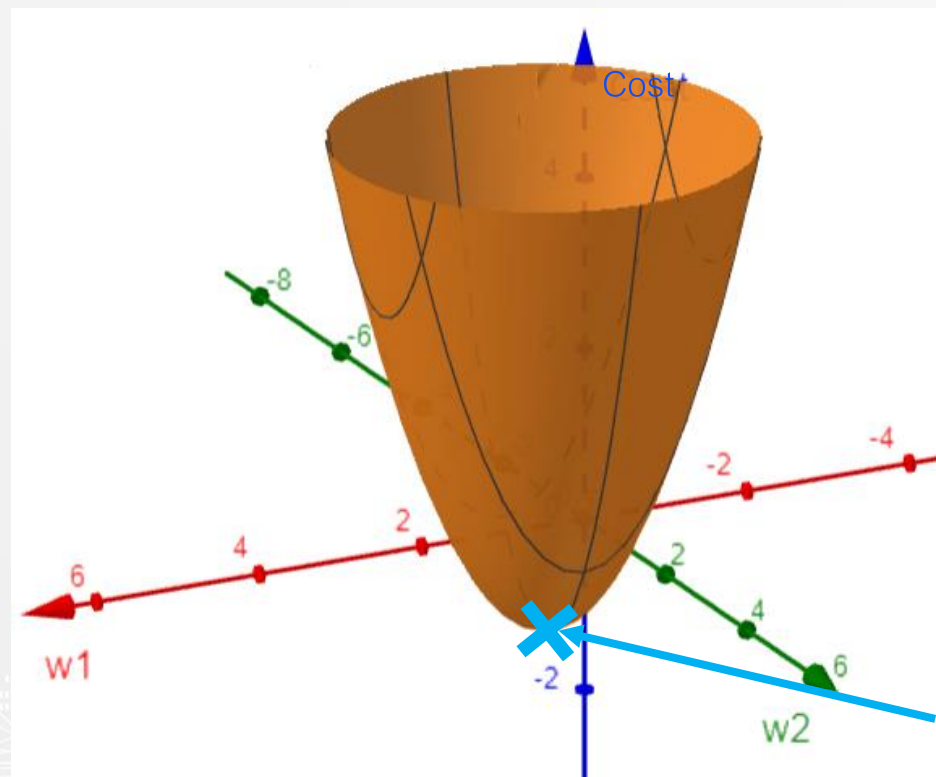
“เราต้องการหา $w_0, w_1, w_2, \dots, w_p$ ที่ทำให้
cost function น้อยที่สุด”

How to Create Model (Math)

- ☐ Least Squares Method
- ☐ Calculation Example

Least Squares Method

เป้าหมาย เราต้องการจุดต่ำสุดของ cost function ซึ่งสามารถใช้ calculus ในการหาได้



สามารถใช้ calculus
เพื่อหาจุดต่ำสุดได้เลย

Least Squares Method

วิธีการที่เราจะใช้ในการหาคำตอบมีชื่อว่า

“Least Squares Method”

Least Squares Method

จากคุณสมบัติของจุดต่ำสุดคือ

“gradient ของ cost function เท่ากับ 0 ($\nabla Cost = 0$)”

ซึ่งหมายความว่า ความชันของ cost function ในมิติใด ๆ เท่ากับ 0

$$\frac{\partial Cost}{\partial w_d} = 0 ; d = 0, \dots, p$$

Least Squares Method

การที่ความชันของ cost ในมิติใด ๆ เท่ากับ 0 หรือ $\frac{\partial Cost}{\partial w_d} = 0 ; d = 0, \dots, p$
สามารถเขียนเป็นสมการได้ดังต่อไปนี้

$$\frac{\partial Cost}{\partial w_0} = 0$$

$$\frac{\partial Cost}{\partial w_1} = 0$$

$$\vdots$$

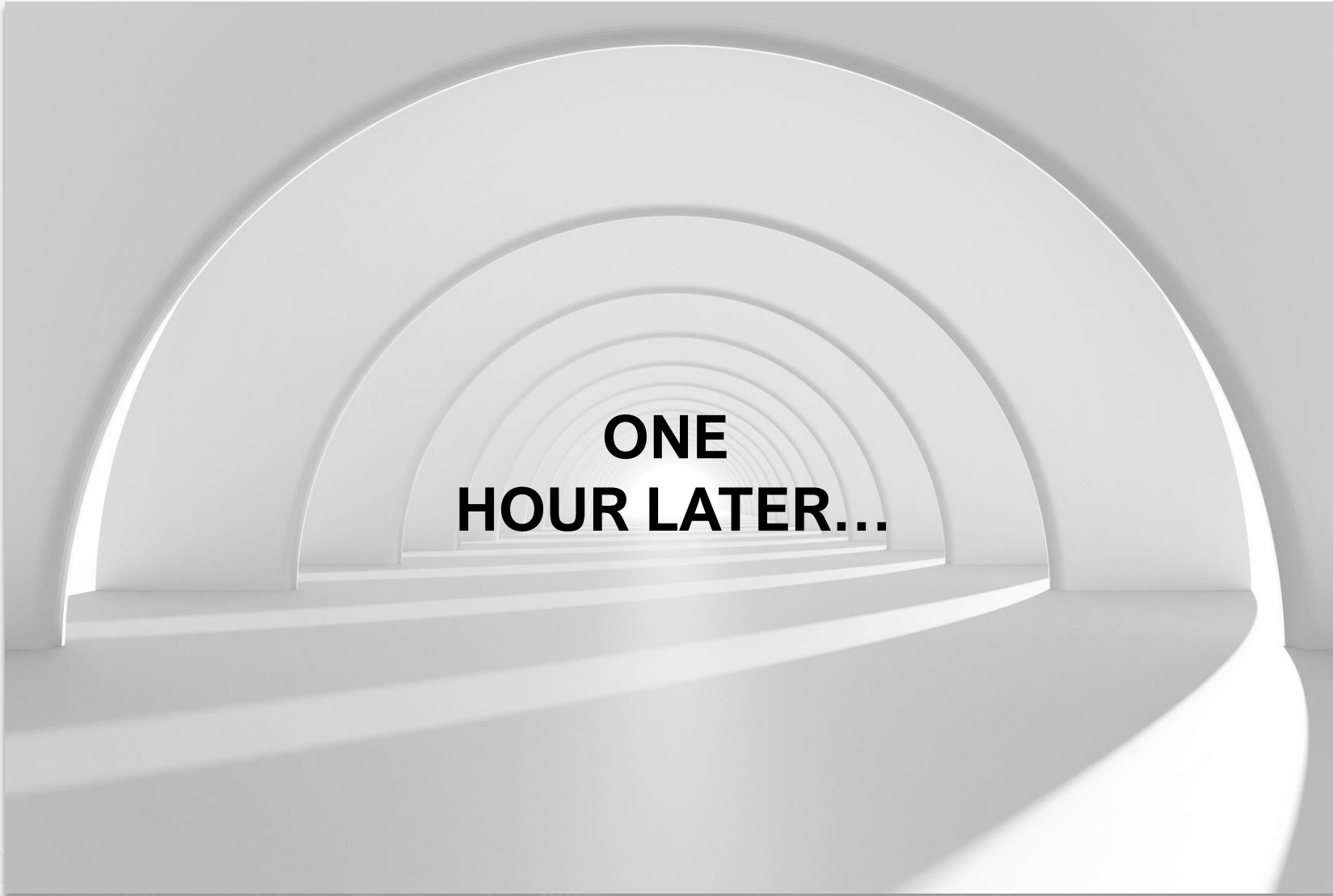
$$\frac{\partial Cost}{\partial w_p} = 0$$

Least Squares Method

ดังนั้น $\nabla Cost = 0$ สามารถเขียนใหม่ได้ดังนี้

$$\nabla Cost = \begin{bmatrix} \frac{\partial Cost}{\partial w_0} \\ \frac{\partial Cost}{\partial w_1} \\ \vdots \\ \frac{\partial Cost}{\partial w_p} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\text{โดยที่ } Cost = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$



Least Squares Method

Normal Equation คือ สมการที่ใช้ในการหา weight ของ model

$$w = (X_b^T X_b)^{-1} X_b^T y$$

โดยที่ $w = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_p \end{bmatrix}$, $X_b = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \cdots & x_{1,p} \\ 1 & x_{2,1} & x_{2,2} & \cdots & x_{2,p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & x_{n,2} & \cdots & x_{n,p} \end{bmatrix}$, $y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$

Least Squares Method



Derivation of Normal Equation



Open File

Derive_NormalEq.pdf

How to Create Model (Math)

- ☒ **Least Squares Method**
- ☐ Calculation Example

Calculation Example

ตัวอย่างการคำนวณ w ด้วย normal equation

| x_1 | x_2 | y |
|-------|-------|-----|
| 0 | 1 | 4 |
| 2 | 1 | 8 |
| 1 | 1 | 6 |
| 2 | 0 | 5 |

ตารางแสดง toy dataset

Calculation Example

- จากข้อมูลใน dataset เราสามารถเขียน X , y และ X_b ได้ดังต่อไปนี้

$$X = \begin{bmatrix} 0 & 1 \\ 2 & 1 \\ 1 & 1 \\ 2 & 0 \end{bmatrix}, y = \begin{bmatrix} 4 \\ 8 \\ 6 \\ 5 \end{bmatrix} \text{ และ } X_b = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \\ 1 & 2 & 0 \end{bmatrix}$$

Calculation Example

- จากสูตร normal equation $\mathbf{w} = (X_b^T X_b)^{-1} X_b^T \mathbf{y}$ จะได้ว่า

$$\begin{aligned}\mathbf{w} &= \left(\begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 1 & 2 \\ 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \\ 1 & 2 & 0 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 2 & 1 & 2 \\ 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 4 \\ 8 \\ 6 \\ 5 \end{bmatrix} \\ &= \begin{bmatrix} 4 & 5 & 3 \\ 5 & 9 & 3 \\ 3 & 3 & 3 \end{bmatrix}^{-1} \begin{bmatrix} 23 \\ 32 \\ 18 \end{bmatrix} \\ \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix} &= \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}\end{aligned}$$

Calculation Example

ดังนั้น เราจะได้ model ของ linear regression สำหรับข้อมูลชุดนี้คือ

$$\hat{y} = 1 + 2x_1 + 3x_2$$

Calculation Example



Exercise of Normal Equation



Open File

Exercise_NormalEq.pdf

How to Create Model (Math)

- ✓ **Least Squares Method**
- ✓ **Calculation Example**

Model

Assumption



**Real Face of the
Model**



**Cost Function and
Cost Landscape**



**How to Create
Model (Math)**



**How to Create
Model (Code)**



Further Reading



How to Create Model (Code)

ตัวอย่าง code สำหรับคำนวณ w

| x_1 | x_2 | y |
|-------|-------|-----|
| 0 | 1 | 4 |
| 2 | 1 | 8 |
| 1 | 1 | 6 |
| 2 | 0 | 5 |

ตารางแสดง toy dataset

How to Create Model (Code)

- Code สำหรับสร้าง model จากข้อมูลของเราโดยที่

$$X = \begin{bmatrix} 0 & 1 \\ 2 & 1 \\ 1 & 1 \\ 2 & 0 \end{bmatrix}, \quad y = \begin{bmatrix} 4 \\ 8 \\ 6 \\ 5 \end{bmatrix}$$

```
In [1]: ▶ 1 reg = LinearRegression()  
        2 reg.fit(X, y)
```

```
Out[1]: LinearRegression()
```

How to Create Model (Code)

- ค่า w_0 ถูกเก็บไว้ใน attribute ชื่อ `intercept_`

```
In [2]: ▶ 1 reg.intercept_
```

```
Out[2]: 1.
```

How to Create Model (Code)

- ค่า w_1, \dots, w_p ถูกเก็บไว้ใน attribute ชื่อ `coef_`

```
In [3]: 1 reg.coef_  
Out[3]: array([2., 3.])
```


How to Create Model (Code)

ดังนั้น เราจะได้ $w_0 = 1, w_1 = 2$ และ $w_2 = 3$ ซึ่งสามารถเขียนเป็น model ของ linear regression สำหรับข้อมูลชุดนี้ได้ดังนี้

$$\hat{y} = 1 + 2x_1 + 3x_2$$

How to Create Model (Code)



Code for this section



Open File
Model Creation.ipynb

How to Create Model (Code)



MATH

$$\mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_p \end{bmatrix}$$



CODE

```
intercept_ = w0
coef_ =  $\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_p \end{bmatrix}$ 
```

How to Create Model (Code)

**MATH**

$$\mathbf{w} = (X_b^T X_b)^{-1} X_b^T \mathbf{y}$$

**CODE**

```
In [1]: 1 reg = LinearRegression()  
        2 reg.fit(X, y)
```

```
Out[1]: LinearRegression()
```

```
In [2]: 1 reg.intercept_
```

```
Out[2]: 1.
```

```
In [3]: 1 reg.coef_
```

```
Out[3]: array([2., 3.])
```

Model

Assumption



**Real Face of the
Model**



**Cost Function and
Cost Landscape**



**How to Create
Model (Math)**



**How to Create
Model (Code)**



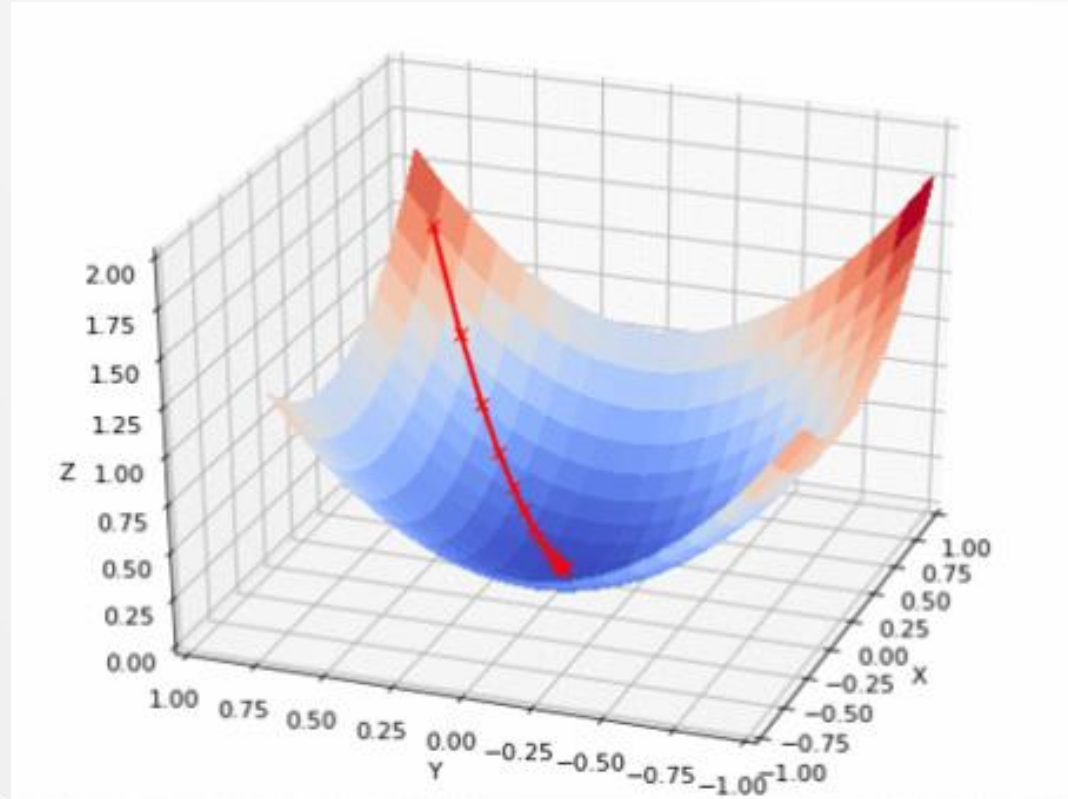
Further Reading



Further Reading

- Gradient Descent
- Moore–Penrose pseudo inverse
- Feature importance with p-value

Gradient Descent

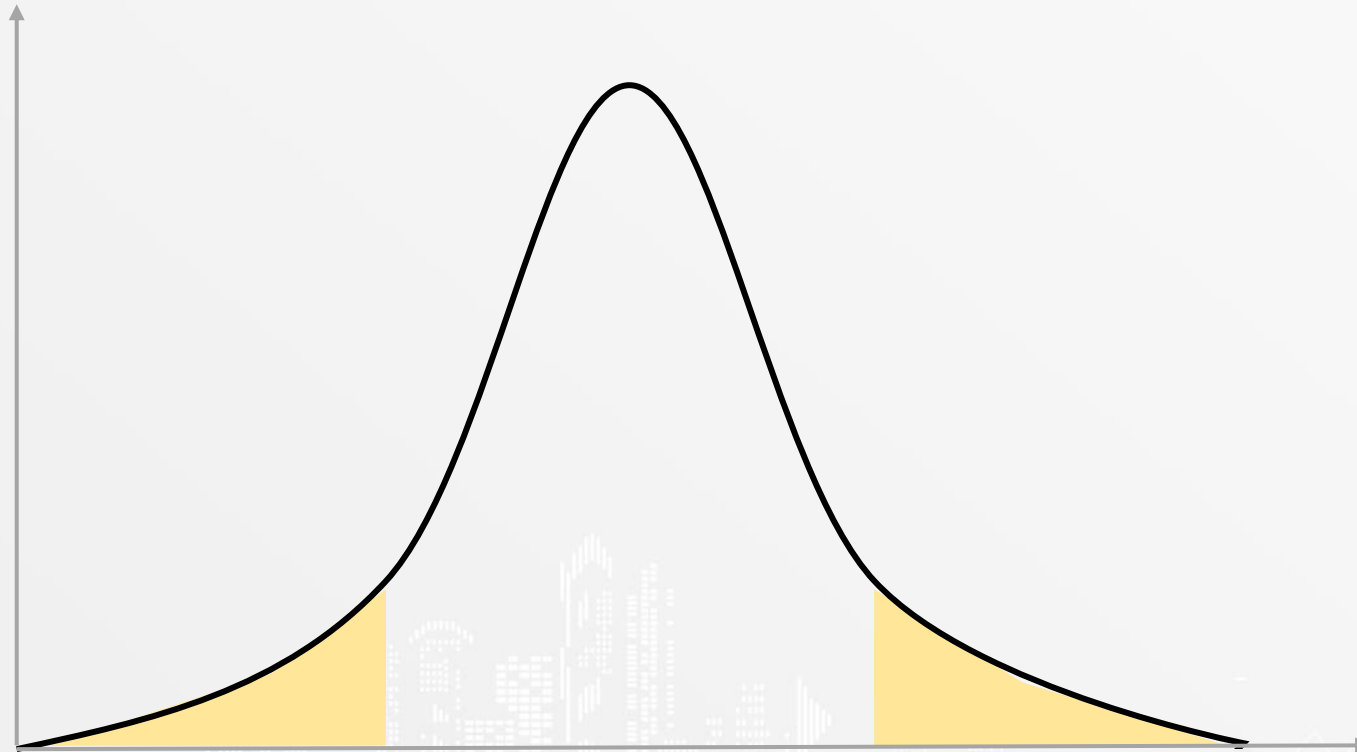


Moore-Penrose pseudo inverse

“ Normal Equation ”

$$\mathbf{w} = (X_b^T X_b)^{-1} X_b^T \mathbf{y}$$

Feature importance with p-value



Model

Assumption



**Real Face of the
Model**



**Cost Function and
Cost Landscape**



**How to Create
Model (Math)**



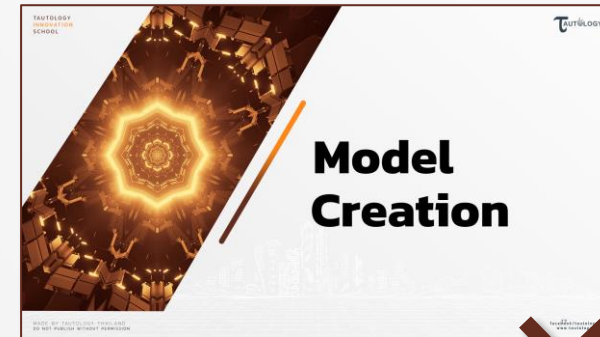
**How to Create
Model (Code)**



Further Reading



Linear Regression



Prediction

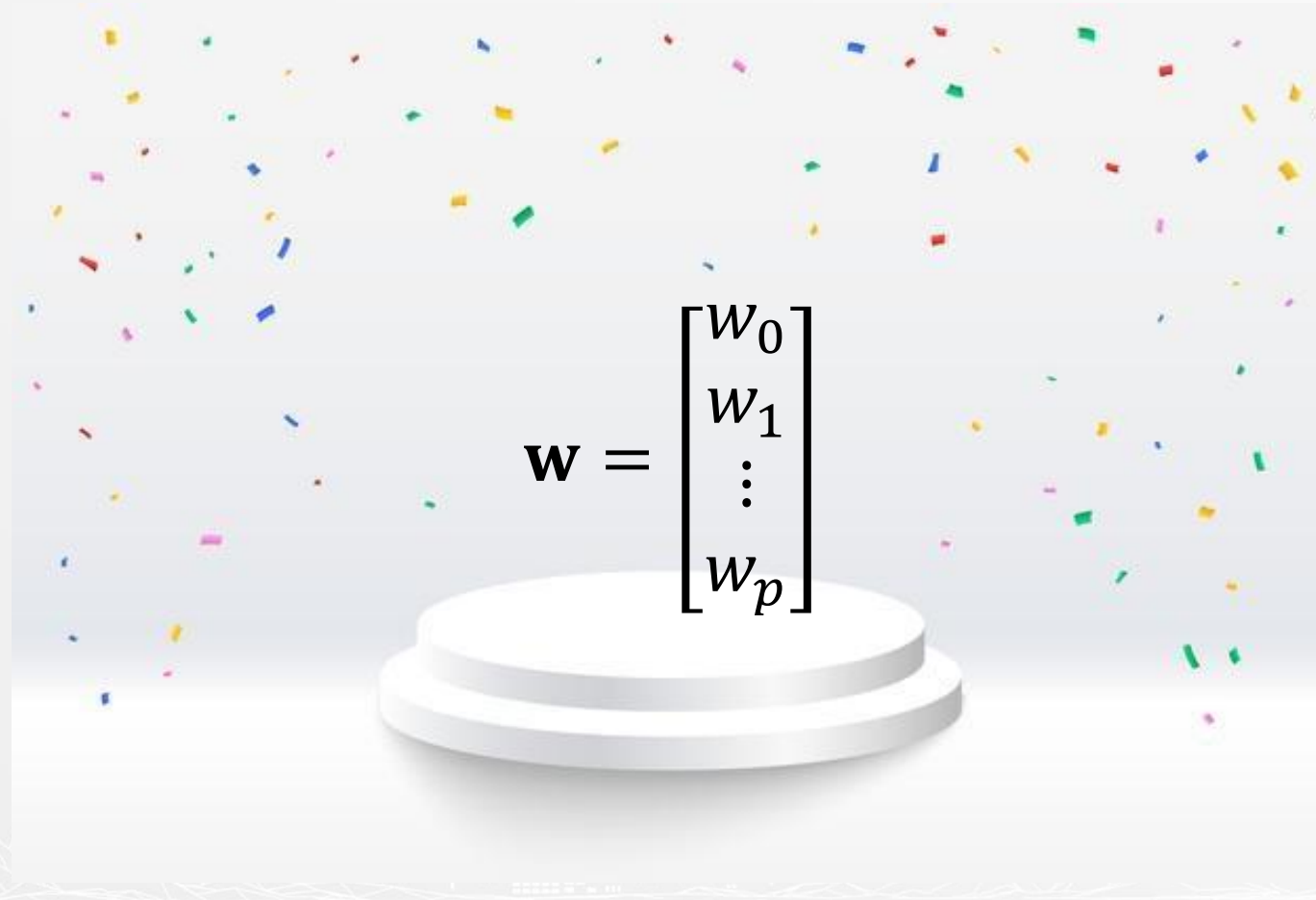
Prediction

Linear regression คือ สมการเชิงเส้นที่ใช้ตัวแปรต้นเพื่อพยากรณ์ตัวแปรตาม

$$\hat{y} = w_0 + w_1x_1 + w_2x_2 + \dots + w_px_p$$

- โดย
- ♦ \hat{y} คือ ค่าพยากรณ์ของตัวแปรตาม (predicted target)
 - ♦ $x_1, x_2, x_3, \dots, x_p$ คือ ตัวแปรต้น (feature)
 - ♦ $w_0, w_1, w_2, \dots, w_p$ คือ สัมประสิทธิ์ (coefficient)

Prediction



Prediction

1-Sample

Multi-Sample

Code

1-Sample

ตัวอย่างการคำนวณ \hat{y}

| x_1 | x_2 | | \hat{y} |
|-------|-------|---|-----------|
| 3 | 1 | → | ? |

1-Sample

สมมติว่า w ของปัญหานี้ที่เราหามาได้คือ

$$w = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

1-Sample

ซึ่งทำให้สามารถเขียนสมการ \hat{y} ได้ดังต่อไปนี้

$$\begin{aligned}\hat{y} &= 1 + 2x_1 + 3x_2 \\ &= 1 + 2(3) + 3(1) \\ &= 10\end{aligned}$$

1-Sample

ดังนั้น เราจะได้ \hat{y} ของข้อมูลชุดนี้คือ

| x_1 | x_2 |  | \hat{y} |
|-------|-------|---|-----------|
| 3 | 1 | | 10 |

Prediction

1-Sample



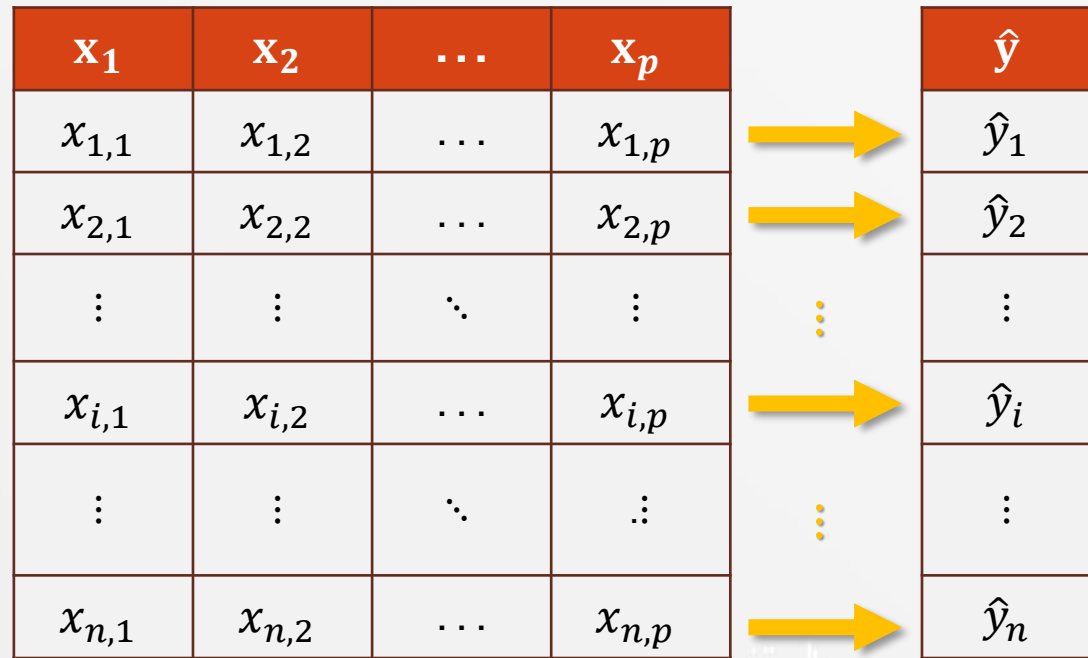
Multi-Sample



Code



Multi-Sample



Multi-Sample

$$\hat{y} = w_0 + w_1x_1 + w_2x_2 + \cdots + w_px_p$$



$$\hat{y}_1 = w_0 + w_1x_{1,1} + w_2x_{1,2} + \cdots + w_px_{1,p}$$

$$\hat{y}_2 = w_0 + w_1x_{2,1} + w_2x_{2,2} + \cdots + w_px_{2,p}$$

⋮

$$\hat{y}_i = w_0 + w_1x_{i,1} + w_2x_{i,2} + \cdots + w_px_{i,p}$$

⋮

$$\hat{y}_n = w_0 + w_1x_{n,1} + w_2x_{n,2} + \cdots + w_px_{n,p}$$

Multi-Sample

เพื่อให้สอดคล้องกับ format ของ data เราสามารถเขียนให้อยู่ในรูปของ matrix ได้ดังนี้

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix} = \begin{bmatrix} w_0 + w_1x_{1,1} + w_2x_{1,2} + \cdots + w_px_{1,p} \\ w_0 + w_1x_{2,1} + w_2x_{2,2} + \cdots + w_px_{2,p} \\ \vdots \\ w_0 + w_1x_{n,1} + w_2x_{n,2} + \cdots + w_px_{n,p} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \cdots & x_{1,p} \\ 1 & x_{2,1} & x_{2,2} & \cdots & x_{2,p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & x_{n,2} & \cdots & x_{n,p} \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_p \end{bmatrix}$$

Multi-Sample

$$\hat{\mathbf{y}} = X_b \mathbf{w}$$

โดยที่ $\hat{\mathbf{y}} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix}$, $X_b = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \cdots & x_{1,p} \\ 1 & x_{2,1} & x_{2,2} & \cdots & x_{2,p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n,1} & x_{n,2} & \cdots & x_{n,p} \end{bmatrix}$, $\mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_p \end{bmatrix}$

Multi-Sample

ตัวอย่างการคำนวณ \hat{y}

| x_1 | x_2 |
|-------|-------|
| 1 | 1 |
| 2 | 0 |
| 3 | 1 |
| 3 | 0 |



| \hat{y} |
|-----------|
| ? |
| ? |
| ? |
| ? |

Multi-Sample

- สมมติว่า w ของปัญหานี้ที่เราหามาได้คือ

$$w = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

- และจากข้อมูลใน dataset เราสามารถเขียน X_b ได้ดังต่อไปนี้

$$X_b = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 0 \\ 1 & 3 & 1 \\ 1 & 3 & 0 \end{bmatrix}$$

Multi-Sample

เราคำนวณค่า \hat{y} ได้จาก $\hat{y} = X_b w$

$$\begin{aligned}\hat{y} &= \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 0 \\ 1 & 3 & 1 \\ 1 & 3 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \\ &= \begin{bmatrix} (1 \times 1) + (2 \times 1) + (3 \times 1) \\ (1 \times 1) + (2 \times 2) + (3 \times 0) \\ (1 \times 1) + (2 \times 3) + (3 \times 1) \\ (1 \times 1) + (2 \times 3) + (3 \times 0) \end{bmatrix} \\ &= \begin{bmatrix} 6 \\ 5 \\ 10 \\ 7 \end{bmatrix}\end{aligned}$$

Multi-Sample

ดังนั้น เราจะได้ \hat{y} สำหรับข้อมูลชุดนี้คือ

| x_1 | x_2 |
|-------|-------|
| 1 | 1 |
| 2 | 0 |
| 3 | 1 |
| 3 | 0 |



| \hat{y} |
|-----------|
| 6 |
| 5 |
| 10 |
| 7 |

Prediction

1-Sample



Multi-Sample

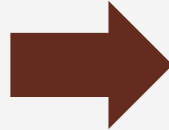


Code



Code

ตัวอย่าง code สำหรับหา \hat{y}

| x_1 | x_2 | | \hat{y} |
|-------|-------|---|-----------|
| 1 | 1 |  | ? |
| 2 | 0 | | ? |
| 3 | 1 | | ? |
| 3 | 0 | | ? |

ตารางแสดง toy dataset

Code

- Code สำหรับหา \hat{y} จากข้อมูลของเรา โดยที่ $X = \begin{bmatrix} 1 & 1 \\ 2 & 0 \\ 3 & 1 \\ 3 & 0 \end{bmatrix}$

```
In [ 4 ]: ▶ 1 reg.predict(X)
```

```
Out[ 4 ]: array([ 6.,  5., 10.,  7.])
```


Code

ดังนั้น เราจะได้ \hat{y} สำหรับข้อมูลชุดนี้คือ

| x_1 | x_2 |
|-------|-------|
| 1 | 1 |
| 2 | 0 |
| 3 | 1 |
| 3 | 0 |



| \hat{y} |
|-----------|
| 6 |
| 5 |
| 10 |
| 7 |

Code



Code for this section



Open File
Model Creation.ipynb

Code



MATH

$$\hat{y} = X_b w$$



CODE

```
In [ 4 ]: 1 reg.predict(X)
Out[ 4 ]: array([ 6.,  5., 10.,  7.] )
```

Prediction

1-Sample



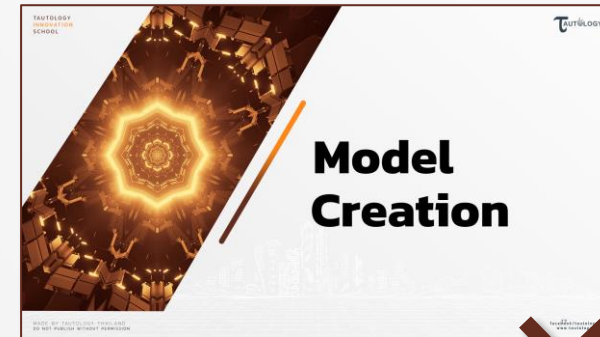
Multi-Sample



Code



Linear Regression



Workshop

Workshop

AI in Marketing

AI in Investment

AI in Insurance

AI in Marketing

- Abstract
- Why this project important?
- Who this project for?
- Ads Dataset
- What we learn from this project?

Abstract

สร้าง model เพื่อพยากรณ์รายได้ที่ได้รับจากการยิง ads โดย feature ที่นำมาใช้ คือ ข้อมูลต่าง ๆ ที่เก็บได้จากการยิง ads เช่น

- จำนวนเงินที่ใช้ในการยิง ads
- จำนวนครั้งที่มีคนคลิก ads



Why this project important?



- สามารถวางแผนรายได้ของบริษัท จาก performance ของ ads
- สามารถต่อยอดกับการทำ marketing บน platform อื่น ๆ

Who this project is for?

- ผู้ขายสินค้าออนไลน์
- ผู้ดูแลเพจ
- Marketing consultant
- นักวิเคราะห์ข้อมูล



Ads Dataset

chrisBow First commit Latest commit a7ffe73 on Oct 11, 2018 History

1 contributor

41 lines (41 sloc) 1.48 KB

Raw Blame

Search this file...

| | spend | clicks | impressions | display | transactions | revenue | ctr | con_rate |
|----|-------|--------|-------------|---------|--------------|---------|------|----------|
| 2 | 22.61 | 165 | 8672 | 0 | 2 | 58.88 | 1.9 | 1.21 |
| 3 | 37.28 | 228 | 11875 | 0 | 2 | 44.92 | 1.92 | 0.88 |
| 4 | 55.57 | 291 | 14631 | 0 | 3 | 141.56 | 1.99 | 1.03 |
| 5 | 45.42 | 247 | 11709 | 0 | 2 | 209.76 | 2.11 | 0.81 |
| 6 | 50.22 | 290 | 14768 | 0 | 3 | 197.68 | 1.96 | 1.03 |
| 7 | 33.05 | 172 | 8698 | 0 | 2 | 204.36 | 1.98 | 1.16 |
| 8 | 12.88 | 68 | 2924 | 0 | 1 | 117.32 | 2.33 | 1.47 |
| 9 | 23.87 | 112 | 5919 | 0 | 1 | 72.04 | 1.89 | 0.89 |
| 10 | 58.38 | 306 | 14789 | 0 | 3 | 290.4 | 2.07 | 0.98 |
| 11 | 48.16 | 300 | 14818 | 0 | 3 | 245.52 | 2.02 | 1 |
| 12 | 63.18 | 355 | 17488 | 0 | 4 | 264.6 | 2.03 | 1.13 |
| 13 | 81.28 | 533 | 26411 | 0 | 5 | 469.4 | 2.02 | 0.94 |
| 14 | 28.33 | 177 | 9011 | 0 | 2 | 66.72 | 1.96 | 1.13 |

Ads Dataset

Feature

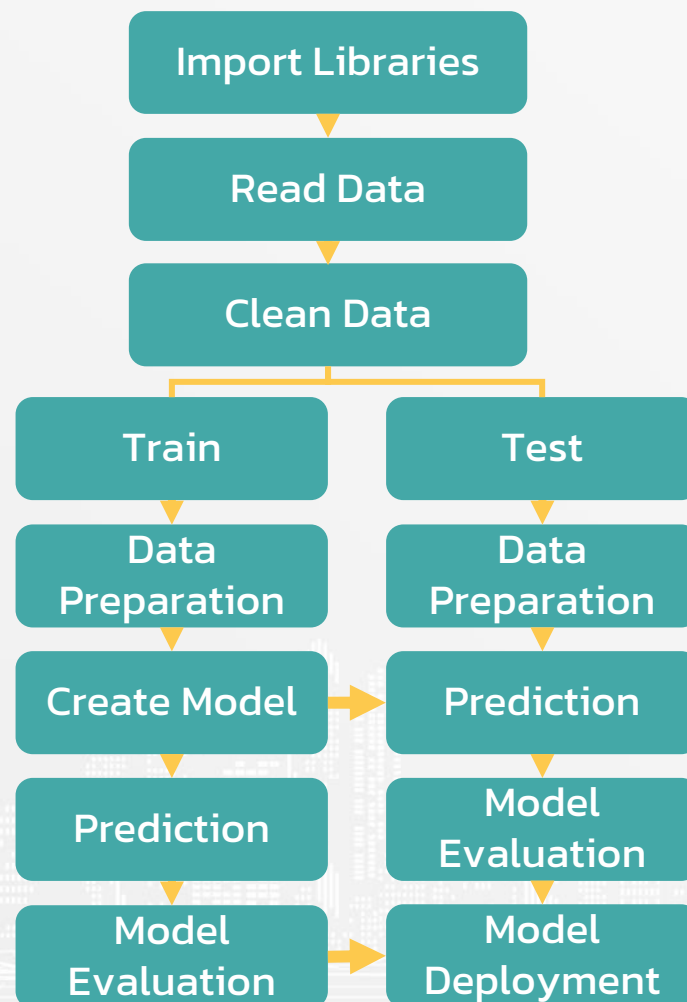
- spend : จำนวนเงินที่ใช้ในการยิง ads
- clicks : จำนวนครั้งที่มีการคลิก ads
- impressions : จำนวนครั้งที่ ads ปรากฏบน website
- display : มี web address ปรากฏบน ads หรือไม่
- transactions : จำนวนครั้งที่มีการทำธุรกรรม
- ctr : จำนวนครั้งที่คลิกโฆษณาต่อจำนวนครั้งที่โฆษณาปรากฏ (clickthrough rate)
- con_rate : จำนวนสินค้าที่ถูกซื้อต่อจำนวนคนที่เข้ามาดูสินค้า (conversion rate)



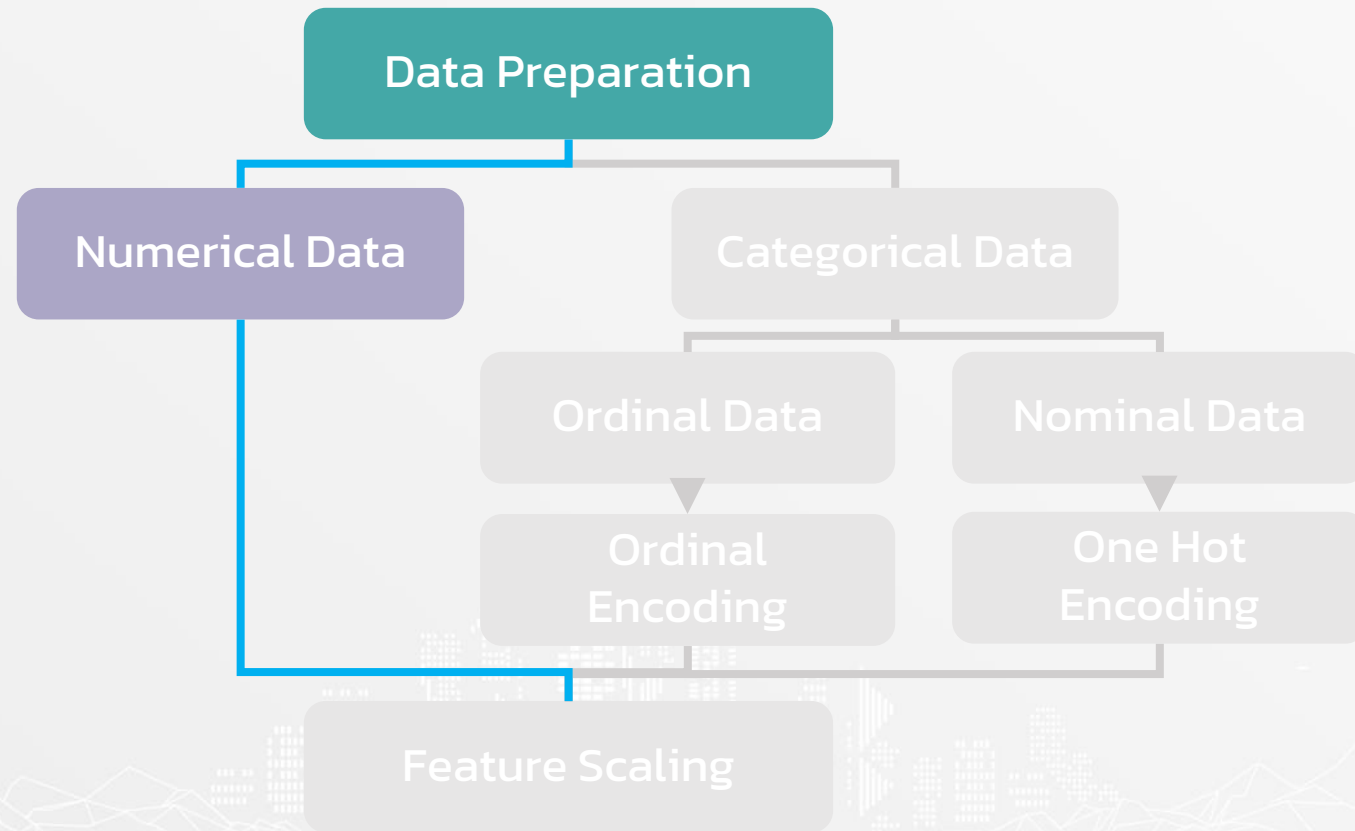
Target

- revenue : รายได้ที่ได้รับจากการยิง ads

What we learn from this project?



Data Preparation





01. MARKETING



ads_dataset.csv



marketing_mc.ipynb



marketing_md.ipynb



marketing_model.pickle

Workshop

AI in Marketing

AI in Investment

AI in Insurance

AI in Investment

- Abstract
- Why this project important?
- Who this project for?
- Set50 Dataset
- What we learn from this project?

Abstract

สร้าง model เพื่อพยากรณ์ราคาสูงสุดรายวันของ SET50 โดย feature ที่นำมาใช้ คือ technical indicator ต่าง ๆ เช่น

- standard deviation
- RSI



Why this project important?



- สามารถสร้างระบบการซื้อขายที่ปราศจากอารมณ์ของมนุษย์
- สามารถเป็นพื้นฐานสำหรับสร้างระบบการซื้อขายที่ robust มากขึ้น
- สามารถต่อยอดกับการเก็งกำไรบนสินทรัพย์ชนิดอื่น ๆ

Who is this project for?

- นักลงทุน
- ผู้ดูแลกองทุน
- นักวิจัยเชิงปริมาณ
- นักวิเคราะห์ข้อมูล



SET50 Dataset

ข้อมูลของ SET50 สามารถเก็บได้ที่เว็บไซต์ <https://www.investing.com>

| SET 50 Historical Data | | | | | | |
|------------------------|--------|---------------|--------|-------------------------|---------|----------|
| Time Frame: | | | | | | |
| Daily | | Download Data | | 09/06/2011 - 10/07/2021 | | |
| Date | Price | Open | High | Low | Vol. | Change % |
| Sep 06, 2011 | 734.49 | 724.53 | 734.49 | 722.61 | 836.64K | 0.66% |
| Sep 07, 2011 | 744.18 | 736.88 | 744.25 | 735.60 | 915.18K | 1.32% |
| Sep 08, 2011 | 747.06 | 740.70 | 747.06 | 738.38 | 844.77K | 0.39% |
| Sep 09, 2011 | 739.73 | 747.53 | 748.28 | 738.83 | 1.09M | -0.98% |
| Sep 12, 2011 | 723.30 | 728.89 | 730.86 | 720.25 | 925.92K | -2.22% |
| Sep 13, 2011 | 716.87 | 726.64 | 728.53 | 713.70 | 685.74K | -0.89% |
| Sep 14, 2011 | 711.34 | 720.18 | 720.69 | 701.79 | 845.61K | -0.77% |
| Sep 15, 2011 | 721.34 | 714.74 | 721.62 | 708.28 | 1.19M | 1.41% |
| Sep 16, 2011 | 718.64 | 726.70 | 728.32 | 716.88 | 663.14K | -0.37% |
| Sep 19, 2011 | 706.87 | 712.53 | 714.67 | 706.87 | 621.48K | -1.64% |
| Sep 20, 2011 | 714.70 | 706.12 | 714.70 | 700.93 | 643.67K | 1.11% |
| Sep 21, 2011 | 717.08 | 711.76 | 718.85 | 710.24 | 650.54K | 0.33% |

SET50 Dataset

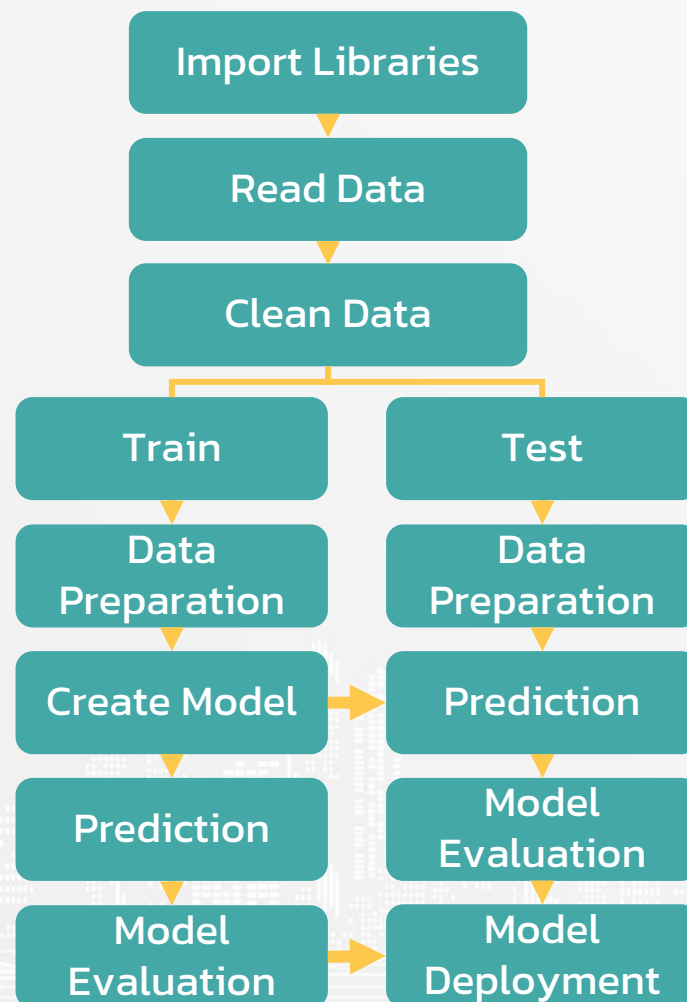
Feature

- Open_0 : ราคาเปิดของวันนั้น ๆ
- STD5_Open_0 : ส่วนเบี่ยงเบนมาตรฐานของราคาเปิด 5 วันล่าสุด
- RSI14_Open_0 : ค่าที่บอกการแกว่งของราคาเปิด 14 วันล่าสุด
- ATR14_0 : ค่าที่บอกความผันผวนของราคา 14 วันล่าสุด

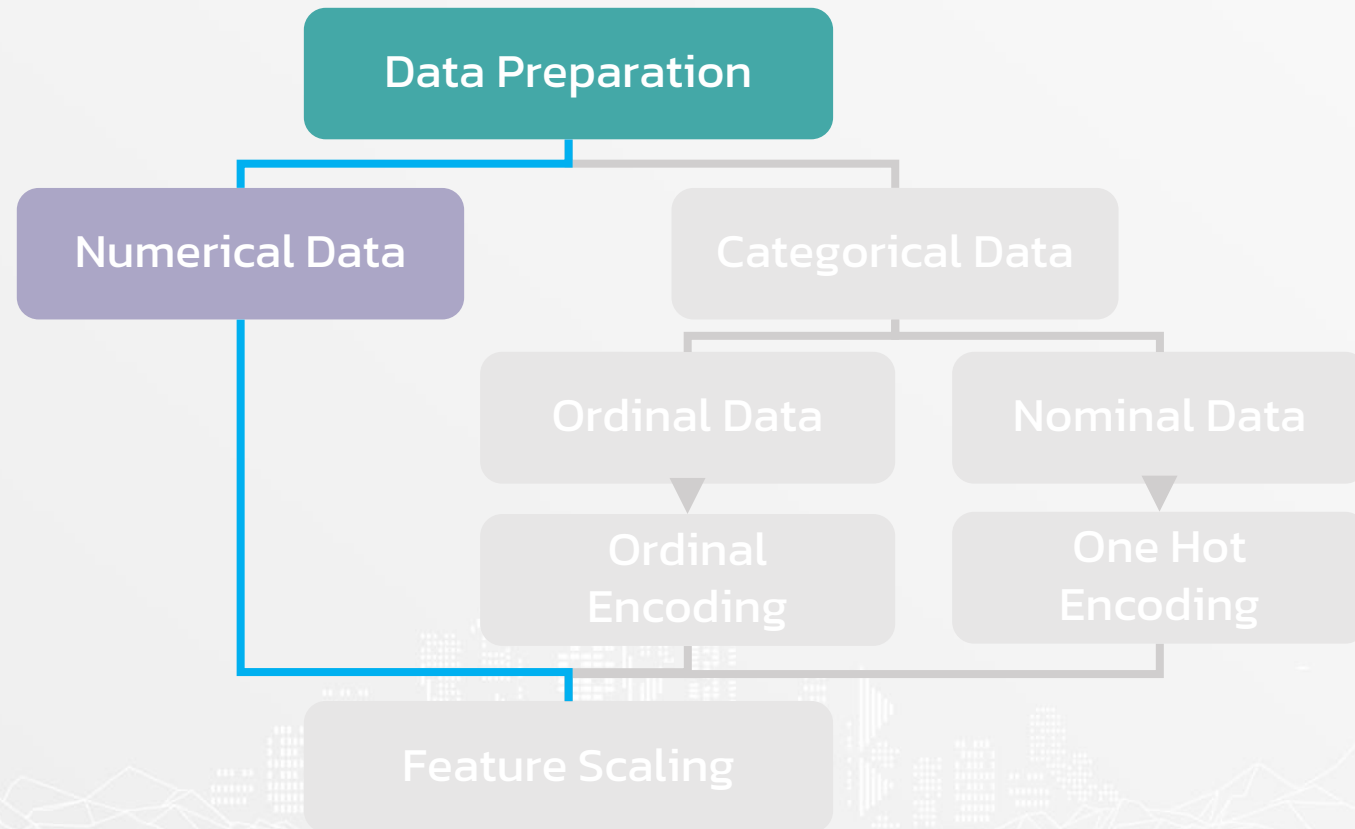
Target

- High_0 : ราคาสูงสุดของวันนั้น ๆ

What we learn from this project?



Data Preparation





02. INVESTMENT/**set50**



set50_dataset.csv



set50_mc.ipynb







set50_md.ipynb







set50_model.pickle



02. INVESTMENT/**btcsud**

-  **btcsud_dataset.csv**
-  **btcsud_mc.ipynb**
-  **btcsud_md.ipynb**
-  **btcsud_model.pickle**

02. INVESTMENT/**eurusd**

-  **eurusd_dataset.csv**
-  **eurusd_mc.ipynb**
-  **eurusd_md.ipynb**
-  **eurusd_model.pickle**

02. INVESTMENT/**xauusd**

-  **xauusd_dataset.csv**
-  **xauusd_mc.ipynb**
-  **xauusd_md.ipynb**
-  **xauusd_model.pickle**

****ข้อมูลสามารถเก็บได้ที่เว็บไซต์ <https://www.investing.com>**

Workshop

AI in Marketing

AI in Investment

AI in Insurance

AI in Insurance

- Abstract
- Why this project important?
- Who this project for?
- Insurance Dataset
- What we learn from this project?

Abstract

สร้าง model เพื่อประเมินค่าประกันภัยส่วนบุคคล โดย feature ที่นำมาใช้ คือ ประวัติทั่วไป และประวัติด้านสุขภาพ ของผู้ทำประกัน เช่น

- เพศ
- อายุ
- BMI



Why this project important?



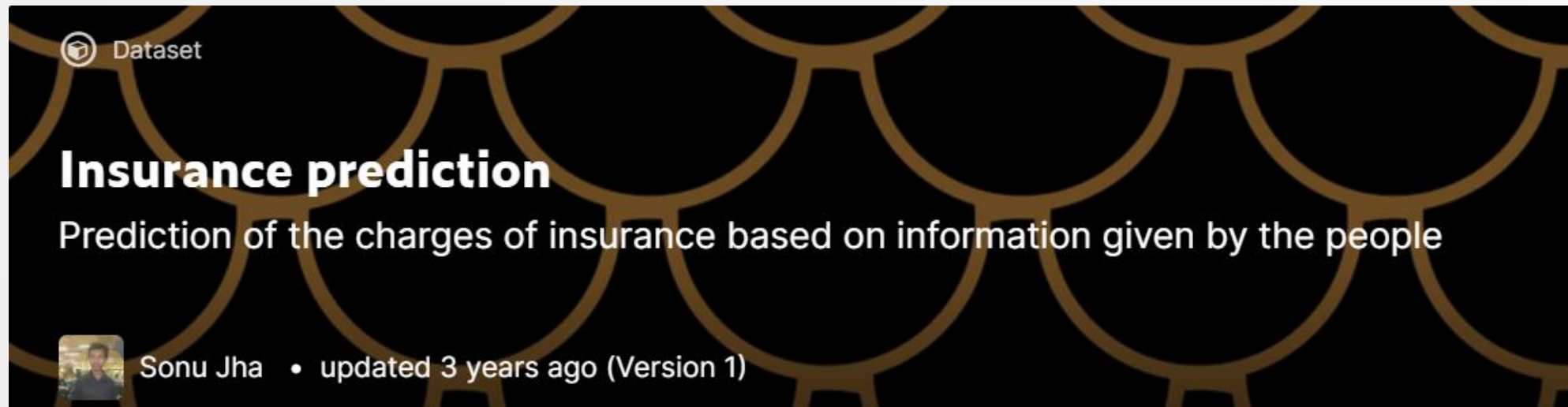
- สามารถวางแผนตั้งราคาเบี้ยประกันภัยที่เหมาะสมที่สุด
- สามารถเพิ่มขีดความสามารถในการแข่งขันกับบริษัทคู่แข่ง

Who this project is for?

- นักคณิตศาสตร์ประกันภัย
- ผู้วางแผนทำประกันภัย
- นักวิเคราะห์ข้อมูล



Insurance Dataset



<https://www.kaggle.com/sonujha090/insurance-prediction>

Insurance Dataset

Feature

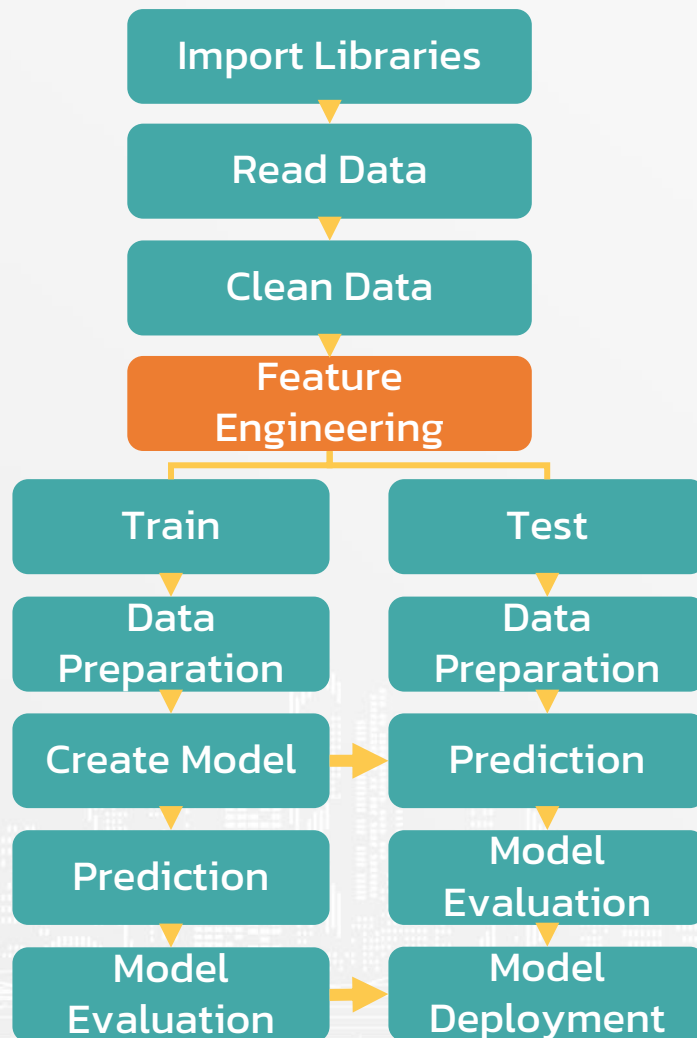
- age : อายุ
- sex : เพศ
- BMI : ค่า BMI
- children : จำนวนบุตร
- region : พื้นที่ที่พักอาศัย

Target

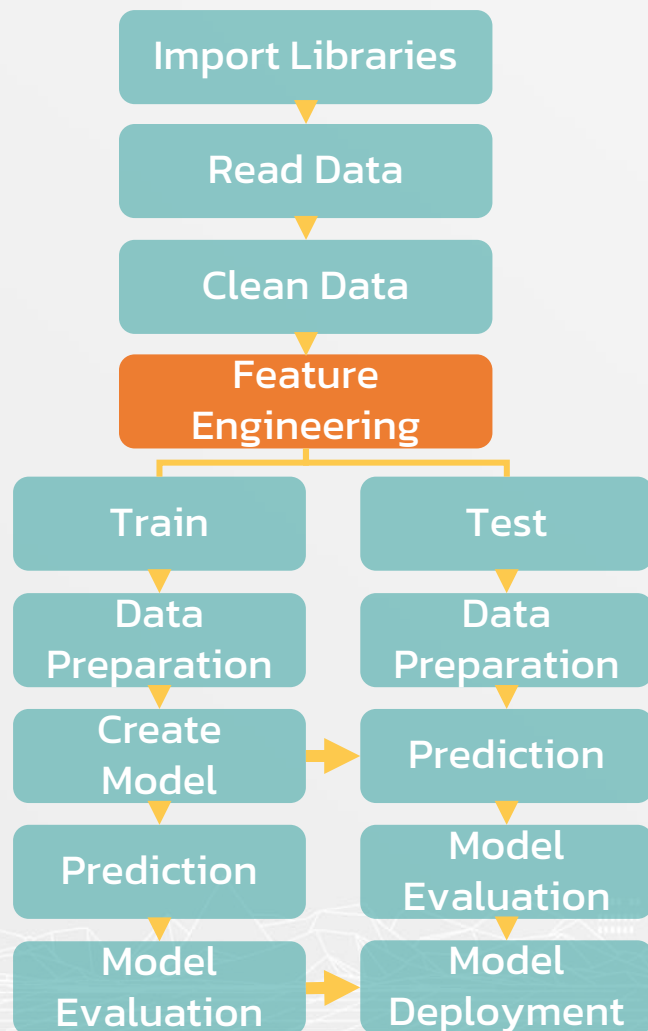
- charges : ค่าประกันที่จ่ายให้กับบริษัทประกันภัย



What we learn from this project?



What we learn from this project?



Feature Engineering

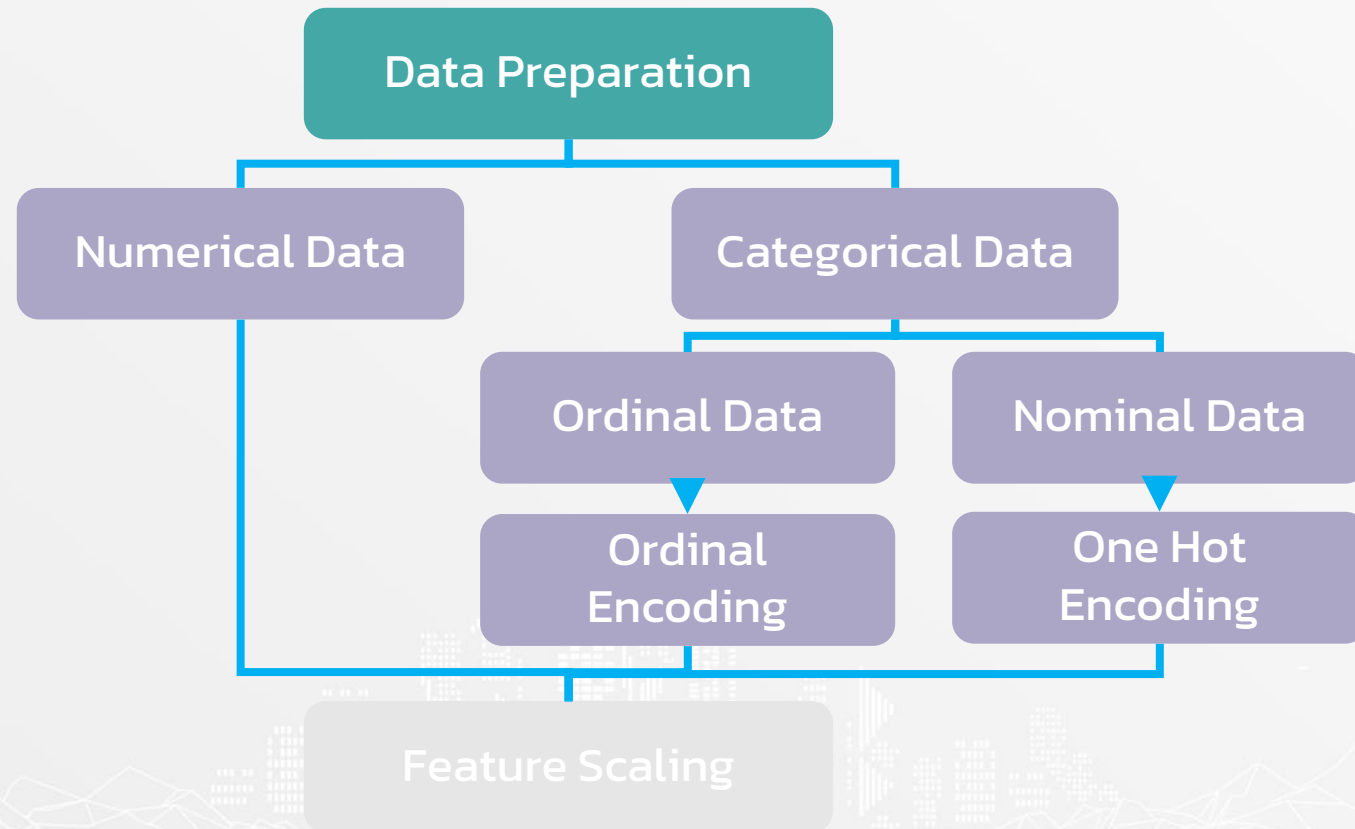
สร้าง feature ใหม่ เพื่อเก็บข้อมูลระดับภาวะน้ำหนัก
ตัวโดยใช้ค่า BMI ในการแบ่ง

| | age | sex | bmi | children | region | charges | bmiclass |
|---|-----|-----|--------|----------|--------|-------------|----------|
| 0 | 19 | 0 | 27.900 | 0 | 3 | 16884.92400 | 3 |
| 1 | 62 | 0 | 26.290 | 0 | 2 | 27808.72510 | 3 |
| 2 | 27 | 1 | 42.130 | 0 | 2 | 39611.75770 | 4 |
| 3 | 30 | 1 | 35.300 | 0 | 3 | 36837.46700 | 4 |
| 4 | 34 | 0 | 31.920 | 1 | 0 | 37701.87680 | 4 |

Code

```
1 data['bmiclass'] = (  
2     (data['bmi'] < 18.5) * 1  
3     + ((data['bmi'] >= 18.5) & (data['bmi'] < 23)) * 2  
4     + ((data['bmi'] >= 23) & (data['bmi'] < 30)) * 3  
5     + (data['bmi'] >= 30) * 4  
6 )
```

Data Preparation





03. INSURANCE



Insurance_dataset.csv



Insurance_mc.ipynb



Insurance_md.ipynb



Insurance_model.pickle

Workshop

AI in Marketing

AI in Investment

AI in Insurance

Linear Regression

