人工智能之自動化光學檢測實務

Yi-Yung Chen

適用影片單元02~單元03

回歸 (Regression)

- 簡單線性回歸
- 多元線性回歸
- 多項式回歸
- 評估回歸模型的表現

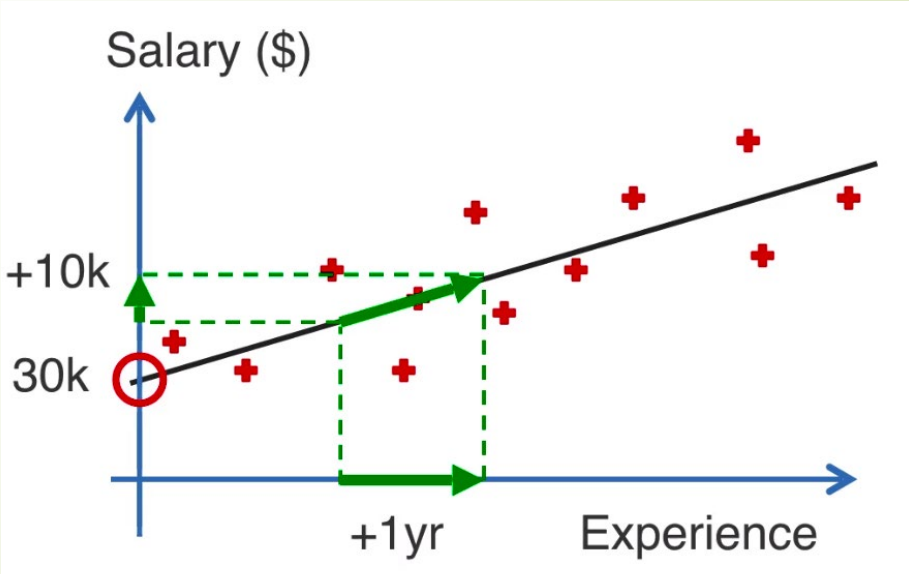
多元線性回歸 vs. 多項式線性回歸

- 簡單線性回歸 (Simple Linear)
  - $y = b_0 + b_1 \times x_1$
- 多元線性回歸 (Multiple Linear)
  - $y = b_0 + b_1 \times x_1 + b_2 \times x_2 + b_3 \times x_3 + \cdots + b_n \times x_n$
- 多項式回歸 (Polynomial Linear)
  - $y = b_0 + b_1 \times x_1 + b_2 \times x_1^2 + b_3 \times x_1^3 + \cdots + b_n \times x_1^n$

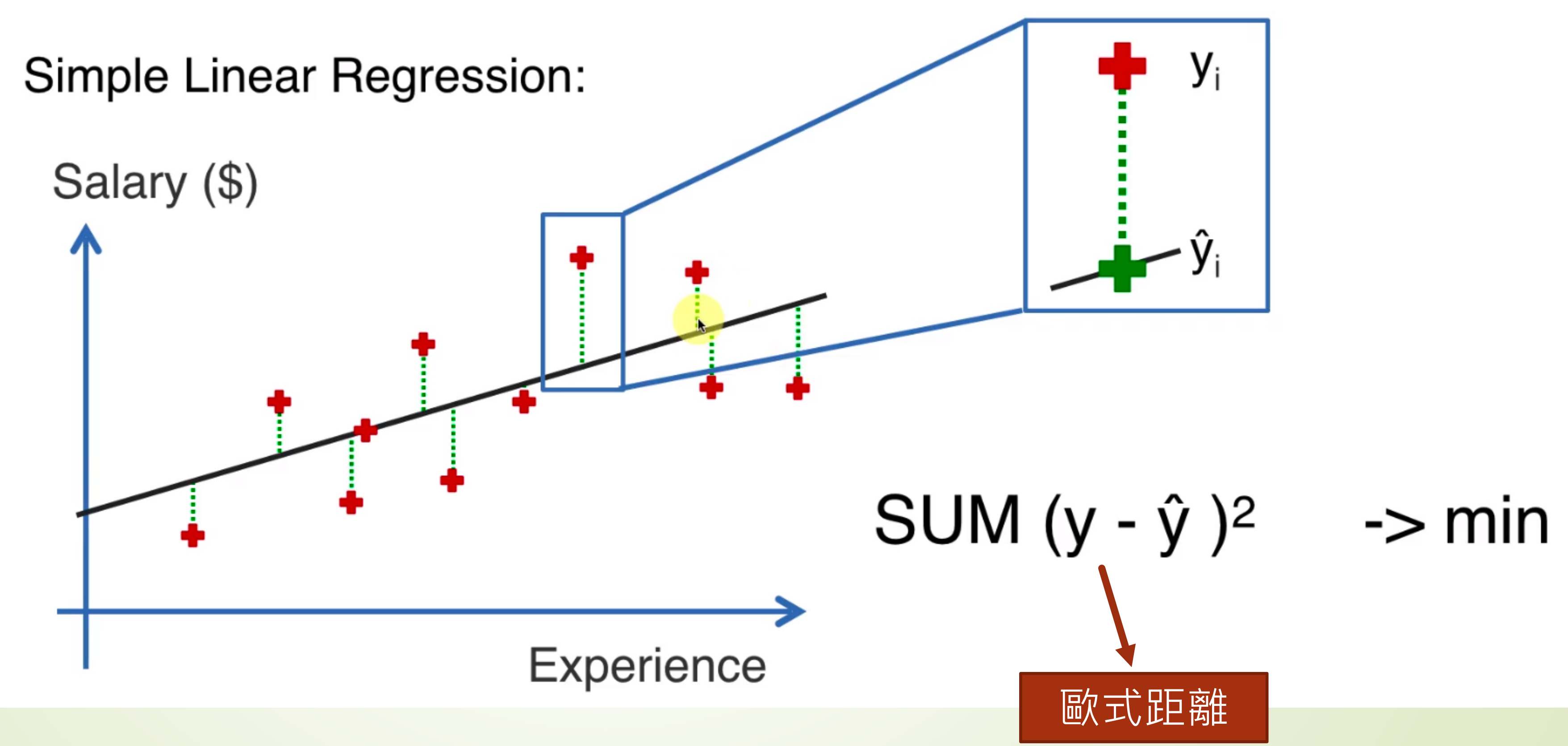
簡單線性回歸

簡單線性回歸

- $y = w_1 \times x + w_2$
  - $Salary = w_1 \times Experience + w_2$
  - y = 9450x + 25792, R^2 = 0.957 (from Excel)



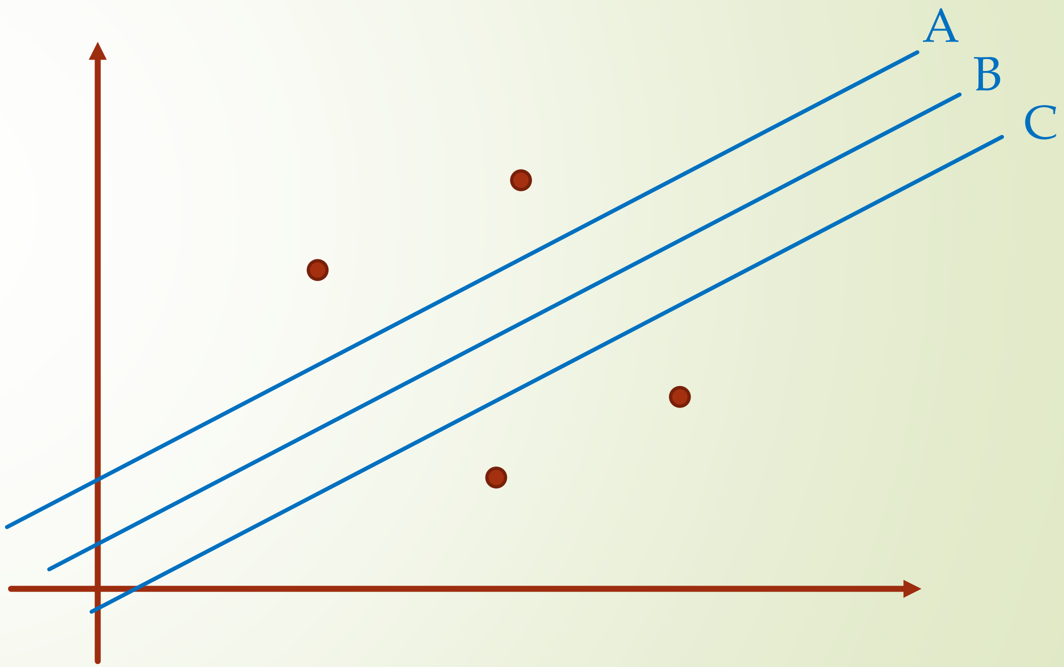
最小平方和


Simple Linear Regression:
Salary ($)
Experience
yi
ŷi
SUM (y - ŷ)² -> min
歐式距離

Error Function

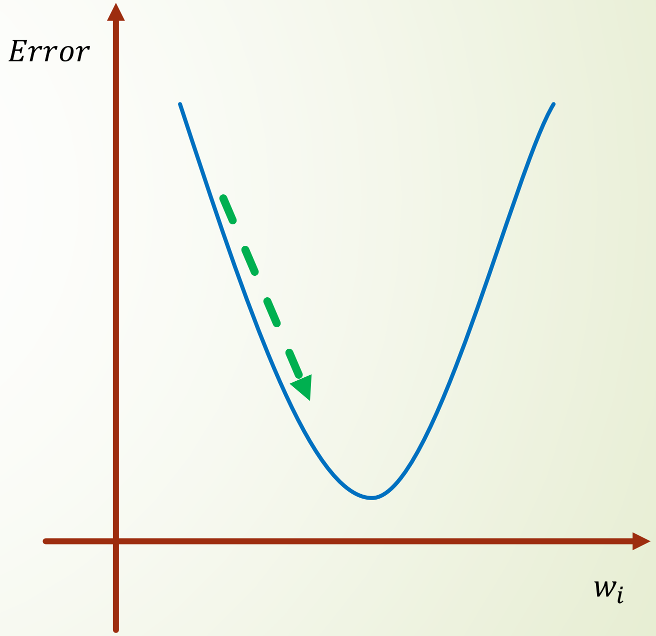
- Mean Absolute Error
  - $Error = \frac{1}{m}\sum_{i=1}^{m}|y - \hat{y}|$
- Mean Squared Error
  - $Error = \frac{1}{2m}\sum_{i=1}^{m}(y - \hat{y})^2$
- Question:
  - Which one will offer higher absolute error
  - Which one will offer higher squared error

$$1 + 3 = 2 + 2,$$
$$1^2 + 3^2 > 2^2 + 2^2$$



Gradient Descent

- Gradient of Error Function
  - $w_i \rightarrow w_i - \alpha \frac{\partial}{\partial w_i} Error$
- Mean Squared Error
  - $\frac{\partial}{\partial w_1} Error = -(y - \hat{y}) \times x$
  - $\frac{\partial}{\partial w_2} Error = -(y - \hat{y})$
- Mean Absolute Error
  - $\frac{\partial}{\partial w_1} Error = \pm x$
  - $\frac{\partial}{\partial w_2} Error = \pm 1$



多元線性回歸

多元線性回歸

- $y = b_0 + b_1 \times x_1 + b_2 \times x_2 + \cdots + b_n \times x_n$
  - 可視為兩個向量的內積
- 多元線性回歸的條件
  - Linearity (線性)
  - Homoscedasticity (等方差性)
  - Multivariate normality (多元常態分佈)
  - Independence of errors (誤差獨立)
  - Lack of multicollinearity (無多重共線性)

虛擬變量 (Dummy Variables)

Profit	R&D Spend	Administration	Marketing Spend	State	New York	California
192261.8	165349.2	136897.8	471784.1	New York	1	0
191792.1	162597.7	151377.6	443898.5	California	0	1
191050.4	153441.5	101145.6	407934.5	California	0	1
182902	144372.4	118671.9	383199.6	New York	1	0
166187.9	142107.3	91391.77	366168.4	California	0	1
$y = b_0$	$b_1 \cdot x_1$	$b_2 \cdot x_2$	$b_3 \cdot x_3$		$b_4 \cdot D_1$	$D_2 = 1 - D_1$

$$y = b_0 + b_1 \times x_1 + b_2 \times x_2 + b_3 \times x_3 + b_4 \times x_4$$

Building a Model

- All-in
  - Preparing for Backward Elimination
- Backward Elimination (反向淘汰)
- Forward Selection (順向選擇)
- Bidirectional Elimination (雙向淘汰)
- Score Comparison (訊息量比較)

= Stepwise Regression (逐步回歸)

Backward Elimination (反向淘汰)

- Step 1: Select a significance level to stay in the model
  - E.g. SL = 0.05
- Step 2: Fit the full model with all possible predictors
- Step 3: Consider the predictor with the highest P-value
  - If P > SL, to go Step 4, otherwise go to **FIN (your model is ready)**
- Step 4: Remove the predictor
- Step 5: Fit model without this variable

Forward Selection (順向選擇)

- Step 1: Select a significance level to enter the model
  - E.g. SL = 0.05
- Step 2: Fit all simple regression models $y \sim x_n$
  - Select the one with the lowest P-value
- Step 3: Keep this variable and fit all possible models with one extra predictor added to the one(s) you already have
- Step 4: Consider the predictor with the lowest P-value
  - If P < SL, to go Step 3, otherwise go to **FIN (your model is ready)**

Bidirectional Elimination (雙向淘汰)

- Step 1: Select a significance level to enter and to stay in the model
  - E.g. $SL_{enter} = 0.05$, $SL_{stay} = 0.05$
- Step 2: Perform the next step of Forward Selection
  - New variables must have: $P < SL_{enter}$ to enter
- Step 3: Perform ALL steps of Backward Elimination
  - Old variables must have $P < SL_{stay}$ to stay
- Step 4: No new variables can enter, and no old variables can exit
  - **FIN (your model is ready)**

Score Comparison (信息量比較)

- Step 1: Select a criterion of goodness of fit
  - Akaike criterion (AIC)
- Step 2: Construct All Possible Regression Models
  - 2^N-1 total combinations
  - E.g. 10 columns means 1,023 models
- Step 3: Select the one with the best criterion
  - **FIN (your model is ready)**

評估回歸模型的表現

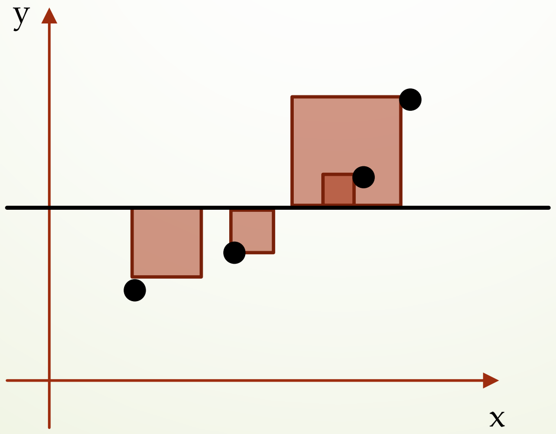
決定係數 R²

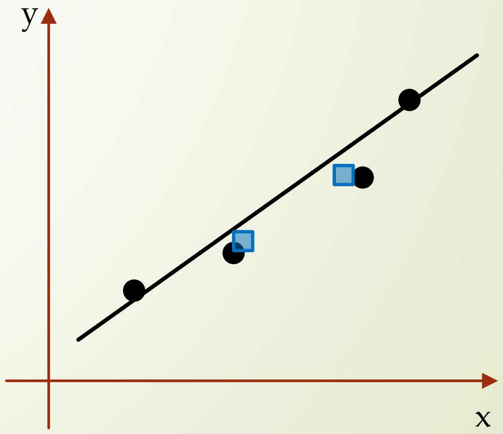
- 決定係數 (coefficient of determination)
  - **因變量**的變異中，可由**自變量**解釋部分所占的比例
- $R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$

$$SS_{tot} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

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

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$





決定係數的陷阱

- 自變量越多越好嗎？

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

$$y = w_0 + w_1 \times x_1$$

$$y = w_0 + w_1 \times x_1 + w_2 \times x_2$$

$$y = w_0 + w_1 \times x_1 + w_2 \times x_2 + w_3 \times x_3$$

$$y = w_0 + w_1 \times x_1 + w_2 \times x_2 + w_3 \times x_3 \cdots$$

R^2 will never decrease

Adjusted R^2

- 如何懲罰過多無用的自變量？

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

$$Adj\ R^2 = 1 - (1 - R^2)\frac{n-1}{n-p-1}$$

  - p: number of regressors
  - n: sample size

1. p 變大
2. 懲罰係數變大
3. Adj R^2變小

感謝聆聽