

# ■ 線性回歸模型（Linear Regression）起步教材

—— 進入機器學習最重要的第一個模型

編撰：屏東大學 周國華老師 2025/11/18 參考：ChatGPT

---

## ① 什麼是線性回歸？

線性回歸（Linear Regression）是機器學習中最基礎、最常用的預測模型之一。它的目標是：

利用已知資料（特徵），找出一條最佳的「直線」或「平面」，用來預測數值。

適用於「連續數值」的預測問題，例如：

- 房價預測
  - 成績預測
  - 銷售額預測
  - 體重、收入、生產量預測
  - 氣溫、需求量、時間成本預測
- 

## ② 模型的核心概念

線性回歸假設：

因變數（要預測的數值）與自變數（特徵）之間是「線性關係」的。

最簡單的一次線性方程式：

$$y = \beta_0 + \beta_1 x$$

- $y$ ：要預測的結果（例如 `final_score`）
- $x$ ：某個特徵（例如 `study_hours`）
- $\beta_0$ ：截距（intercept）
- $\beta_1$ ：斜率（slope）

當特徵不只一個時，就是「多元線性回歸」：

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$$

---

### ③ 線性回歸在做什麼？

線性回歸的任務只有一個：

👉 找到一組最好的係數（ $\beta$ ），讓預測值與實際值差距最小。

換句話說：

- 如果「學習時間」愈高，成績就愈高 →  $\beta$  會是正的
- 如果「缺席數」愈高，成績就愈低 →  $\beta$  會是負的
- 若某欄位與成績沒有關係 →  $\beta$  會接近 0

這些  $\beta$ （權重）構成了「模型」。

---

### ④ Training：如何學會 $\beta$ ？

電腦會使用演算法（如 最小平方法 OLS）

自動找出最佳的  $\beta$ ，使得：

誤差（error）最小化

常見的誤差衡量指標有：

- MSE（Mean Squared Error）
  - RMSE（Root Mean Squared Error）
  - MAE（Mean Absolute Error）
  - $R^2$ （決定係數，用來判斷模型準確度）
- 

### ⑤ 什麼時候使用線性回歸？

線性回歸適合：

- ✓ 資料之間呈現「大致線性」關係
- ✓ 特徵不太複雜
- ✓ 想知道「每個變數的重要性」

例如：

特徵	解釋
study_hours	學習時間增加 1 小時，成績大約 +2.5 分
sleep_hours	每天睡眠多 1 小時，成績約 +1.3 分

➔ 線性回歸能清楚告訴你「每個變數的效果」（可解釋性高）。

---

## 6 Regularization 正規化 (Ridge / Lasso / Elastic Net)

為避免模型過度擬合 (overfitting)，線性回歸常加入 Regularization：

模型	特性
Ridge (L2)	讓模型更穩定，避免係數過大
Lasso (L1)	會讓不重要的係數 $\rightarrow 0$ (可做特徵選擇)
Elastic Net (L1 + L2)	同時穩定模型，又能做特徵選擇 (最實用)

在您的教材案例中，採用 Elastic Net 是最恰當的選擇。

---

## 7 Linear Regression 的優點 🍊

優點	教學說明
容易理解	方程式簡單、透明，學生容易上手
可解釋性高	可知道哪個特徵影響最大、方向如何
運算快	適合大量資料
基準模型 (Baseline Model)	常作為比較的基準，測試其他模型好壞

---

## 8 Linear Regression 的侷限 ⚠️

侷限	說明
只能捕捉「線性」關係 若資料非線性 → 效果差	
對異常值敏感	Outliers 影響係數很大
特徵共線性會影響	Ridge / Lasso 可補強
難以擬合複雜模式	複雜問題常需用 Random Forest、XGBoost 等模型

---

## 9 教材用總結（一句話版）

線性回歸就是利用已知特徵，找一條最佳直線來預測未來數值的模型。  
它簡單、透明、可解釋，是每一個資料科學與機器學習課程的起點。

---