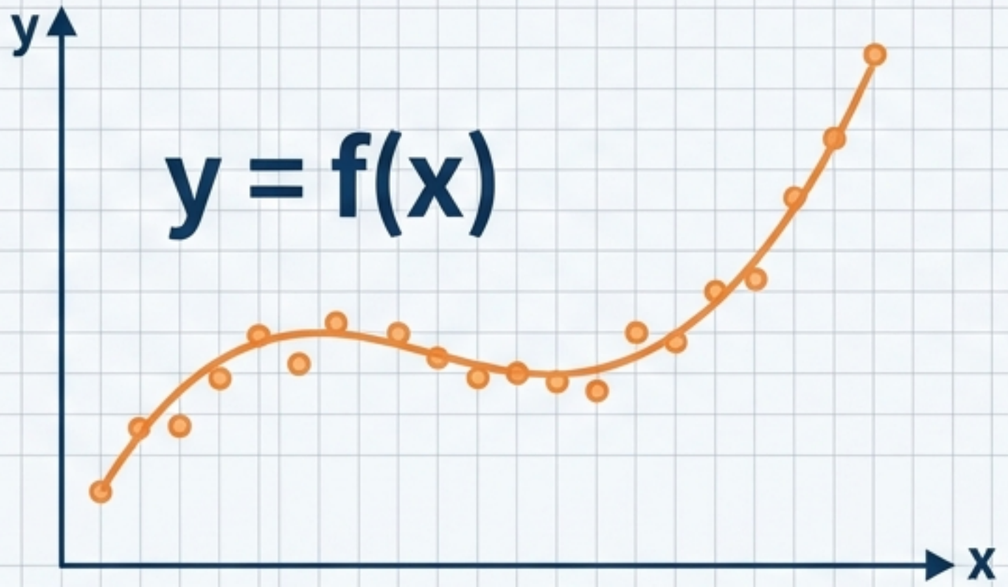
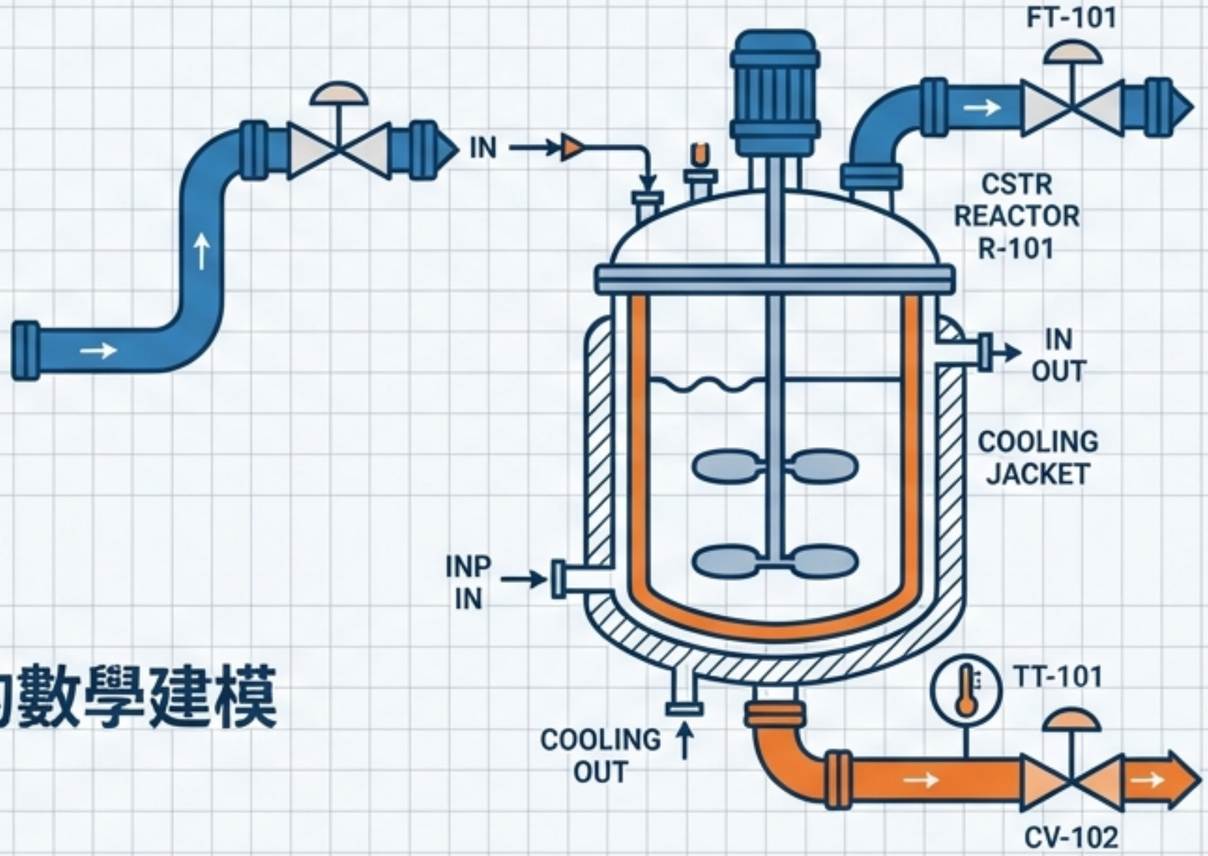
Course: AI在化工上之應用 | Unit 11

多項式回歸 (Polynomial Regression)



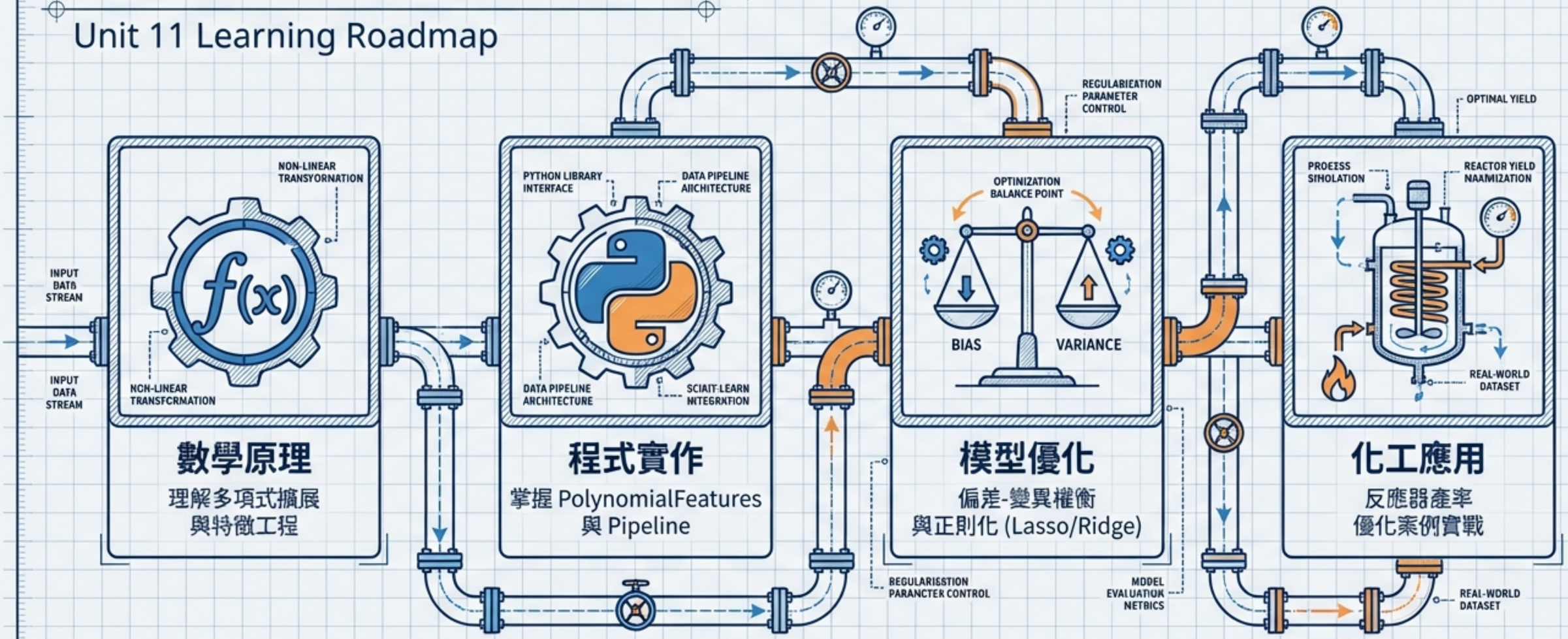


從線性到非線性：捕捉複雜化工製程的數學建模

Noto Sans TC	Roboto Mono
授課教師：莊曜禎 助理教授	
單位：逢甲大學 化工系 智慧程序系統工程實驗室 (PSE Lab)	
更新日期：2026-01-28	

學習導航：從原理到實踐

Unit 11 Learning Roadmap



數學原理

理解多項式擴展與特徵工程

程式實作

掌握 PolynomialFeatures 與 Pipeline

模型優化

偏差-變異權衡與正則化 (Lasso/Ridge)

化工應用

反應器產率優化案例實戰

PROBLEM STATEMENT

核心問題：如何用數據模型描述真實世界的非線性現象？

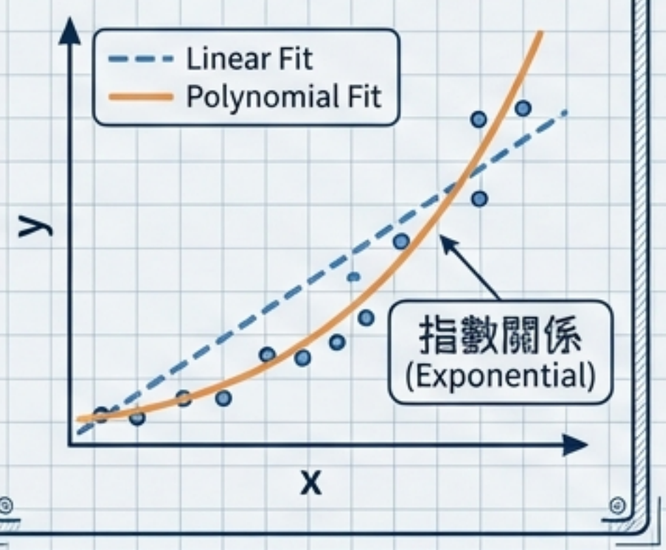
LEARNING OBJECTIVE

APPROVED BY: ENGINEERING DE... DATE:

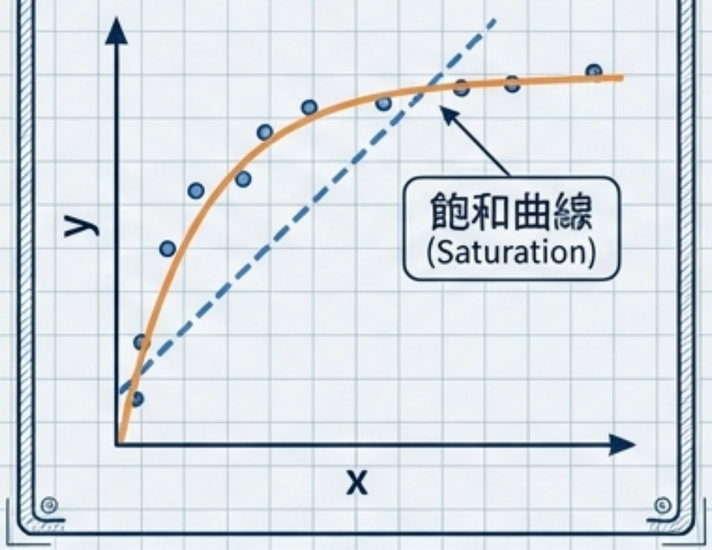
為何需要多項式回歸？

真實世界的化工數據很少是直線 (Real-world ChemE data is rarely a straight line)

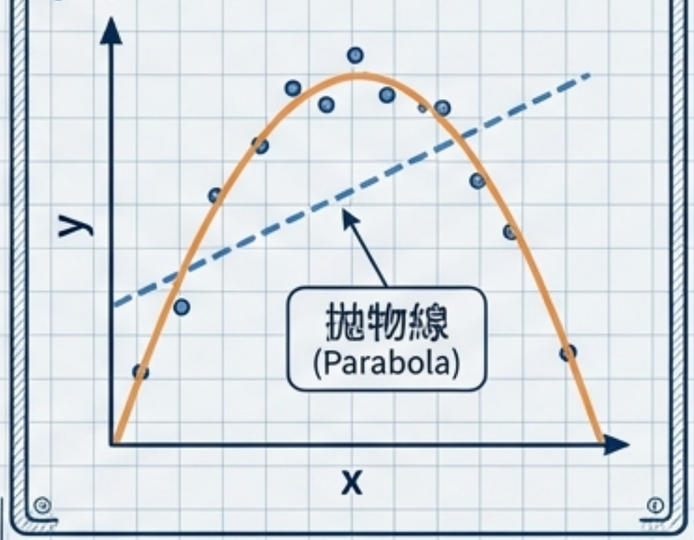
反應動力學
(Reaction Kinetics)



吸附等溫線
(Adsorption Isotherms)



製程產率
(Process Yield)



線性回歸 (Linear Regression) 無法捕捉這些趨勢，導致嚴8重的 **欠擬合** (Underfitting)。

數學原理：線性模型的延伸

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \varepsilon$$

線性參數 (Linear Parameters)
係數是線性的，可用 OLS 求解

非線性特徵 (Non-linear Features)
原始數據的轉換 (Transformation)

核心概念 (Key Concept)

Linear in Parameters, Non-linear in Features

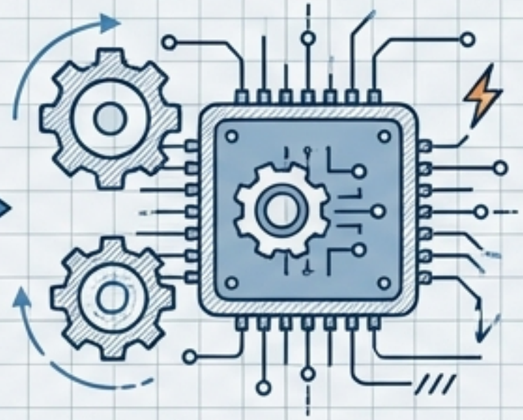
雖然看起來是曲線，但對係數 β 而言，這仍是一個線性方程式。

特徵工程：維度的擴展

Original Features
(n=2)

[a, b]

`PolynomialFeatures(degree=2)`



Expanded Features
(Roboto Mono)

[1,
a,
b,
a^2,
ab,
b^2
]

特徵數量公式： $\text{Count} = C\binom{n+d}{d},d)$

警告：維度爆炸 (Dimension Explosion)

隨著次數 (d) 增加，特徵數量呈指數成長。例如 n=5, d=3 → 56 個特徵。
風險：計算成本增加、過擬合 (Overfitting)。

建立流水線 (Building the Pipeline)

code.py

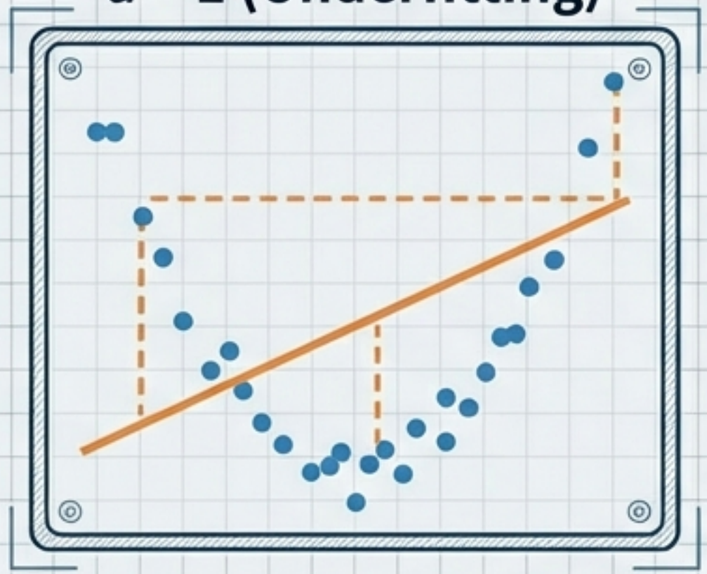
```python
from sklearn.pipeline import Pipeline
from sklearn.preprocessing import StandardScaler, PolynomialFeatures
from sklearn.linear_model import LinearRegression

# 建立 Pipeline：標準化 -> 特徵擴展 -> 線性回歸
pipe = Pipeline([
    ('scaler', StandardScaler()),              # Step 1: 縮放 (Scaling)
    ('poly', PolynomialFeatures(degree=2)),    # Step 2: 擴展 (Expansion)
    ('model', LinearRegression())              # Step 3: 回歸 (Modeling)
])
```

為什麼要縮放？

若 x=10，則 x^3=1000。
避免高次項數值主導模型。

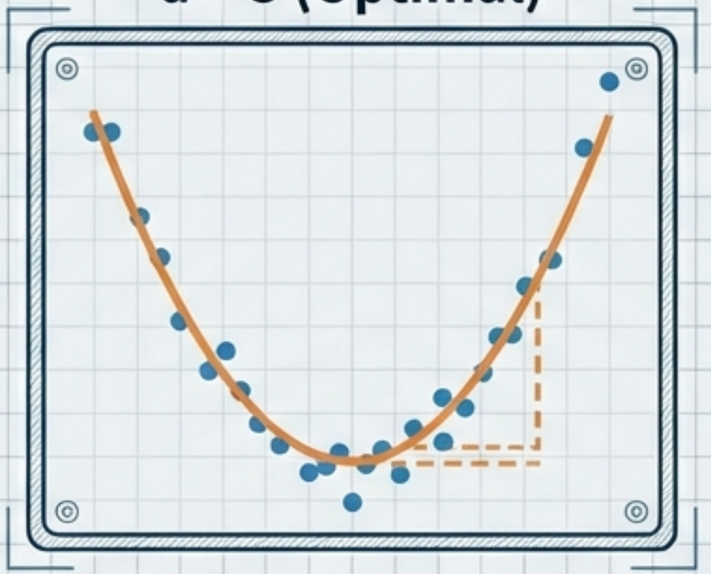
選擇最佳次數：偏差與變異的權衡

$d = 1$ (Underfitting)



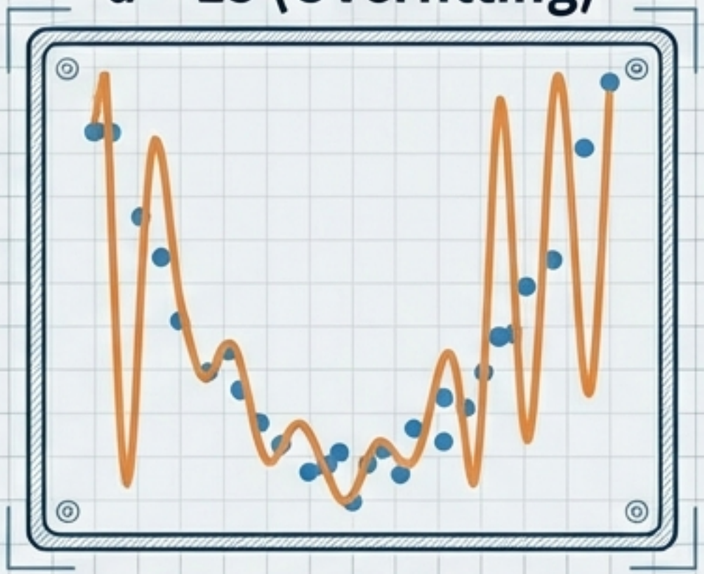
高偏差 (High Bias)

$d = 3$ (Optimal)



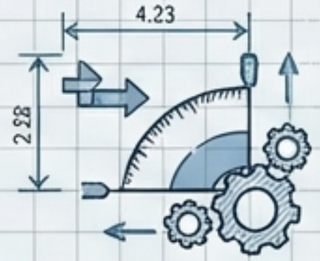
最佳平衡 (Sweet Spot)

$d = 15$ (Overfitting)



高變異 (High Variance)

Low Degree	Model too simple	Misses patterns
High Degree	Model too complex	Memorizes noise



化工案例：反應器優化 (Case Study: Reactor Optimization)

任務摘要 (MISSION BRIEF)

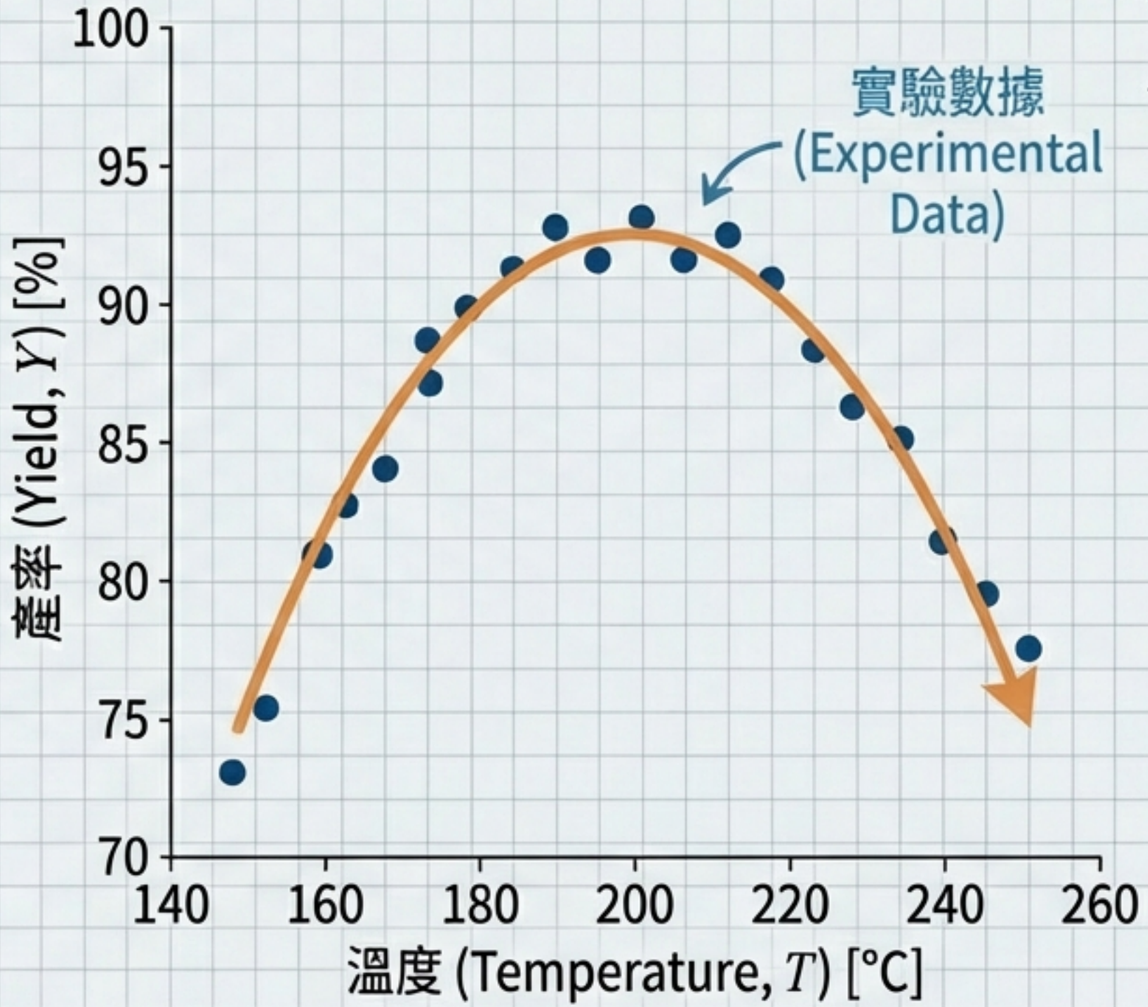
目標 (Goal): 尋找最佳操作溫度
以最大化產率 (Maximize Yield)

系統 (System): 催化反應器
(Catalytic Reactor)

輸入 (Input T): 150 - 250°C

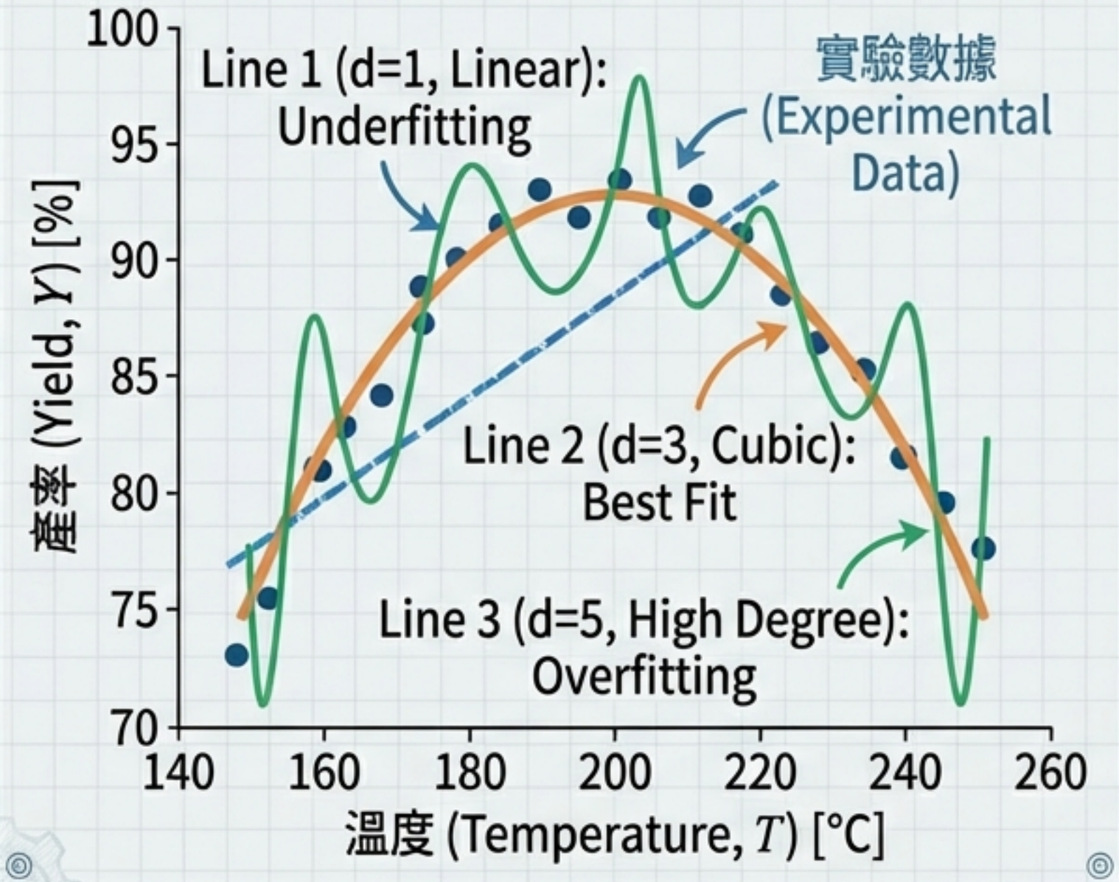
輸出 (Output Y): 77 - 93%

「我們需要找到這條曲線的最高點。」



模型結果比較 (Model Comparison Results)

曲線擬合比較



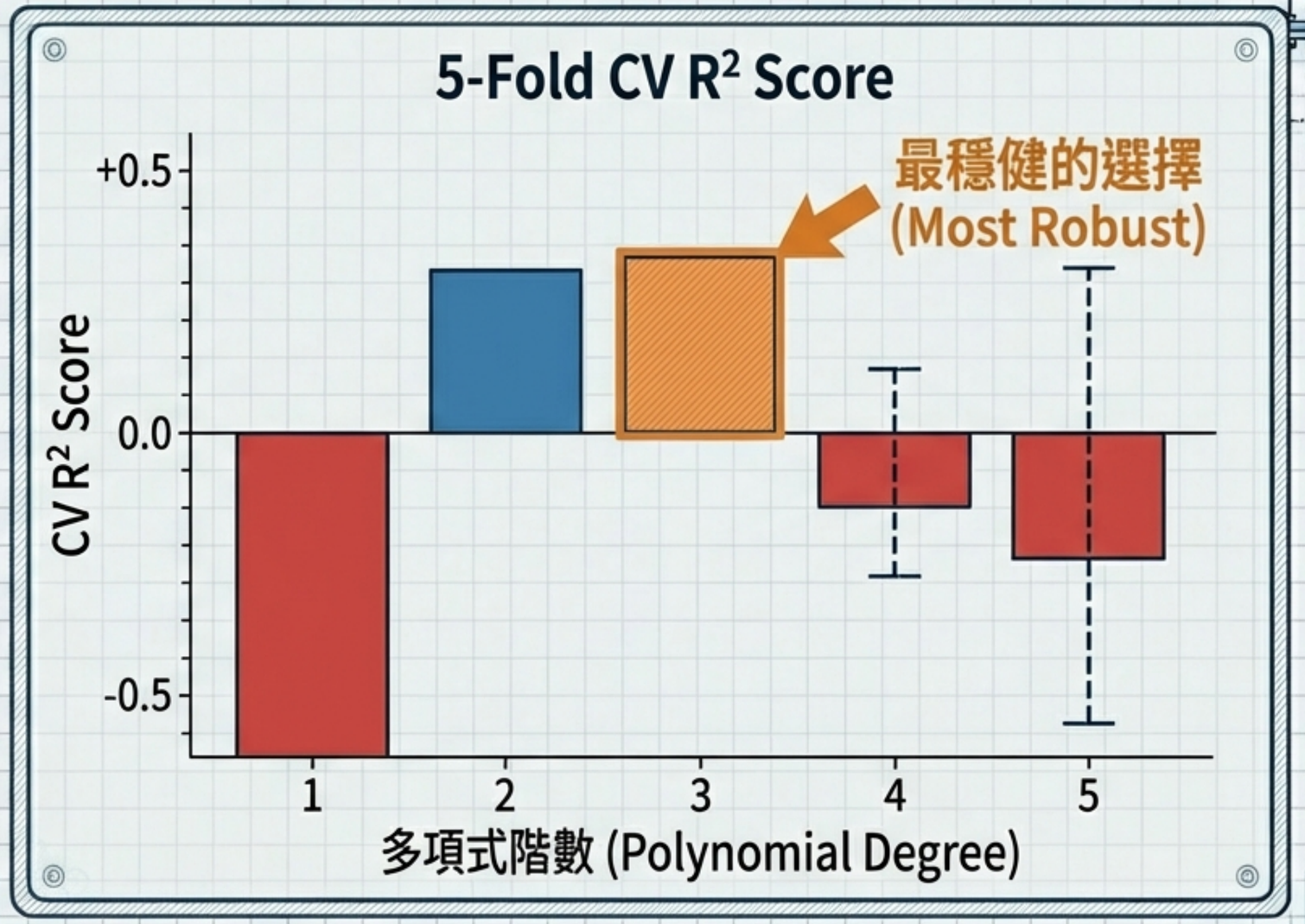
模型評估表

Degree	Train R^2	Test R^2	Status
1	0.37	-2.83	嚴重欠擬合 (Underfit)
3	**0.94**	**0.63**	**最佳 (Best Fit)**
5	0.95	-0.21	過擬合 (Overfit)

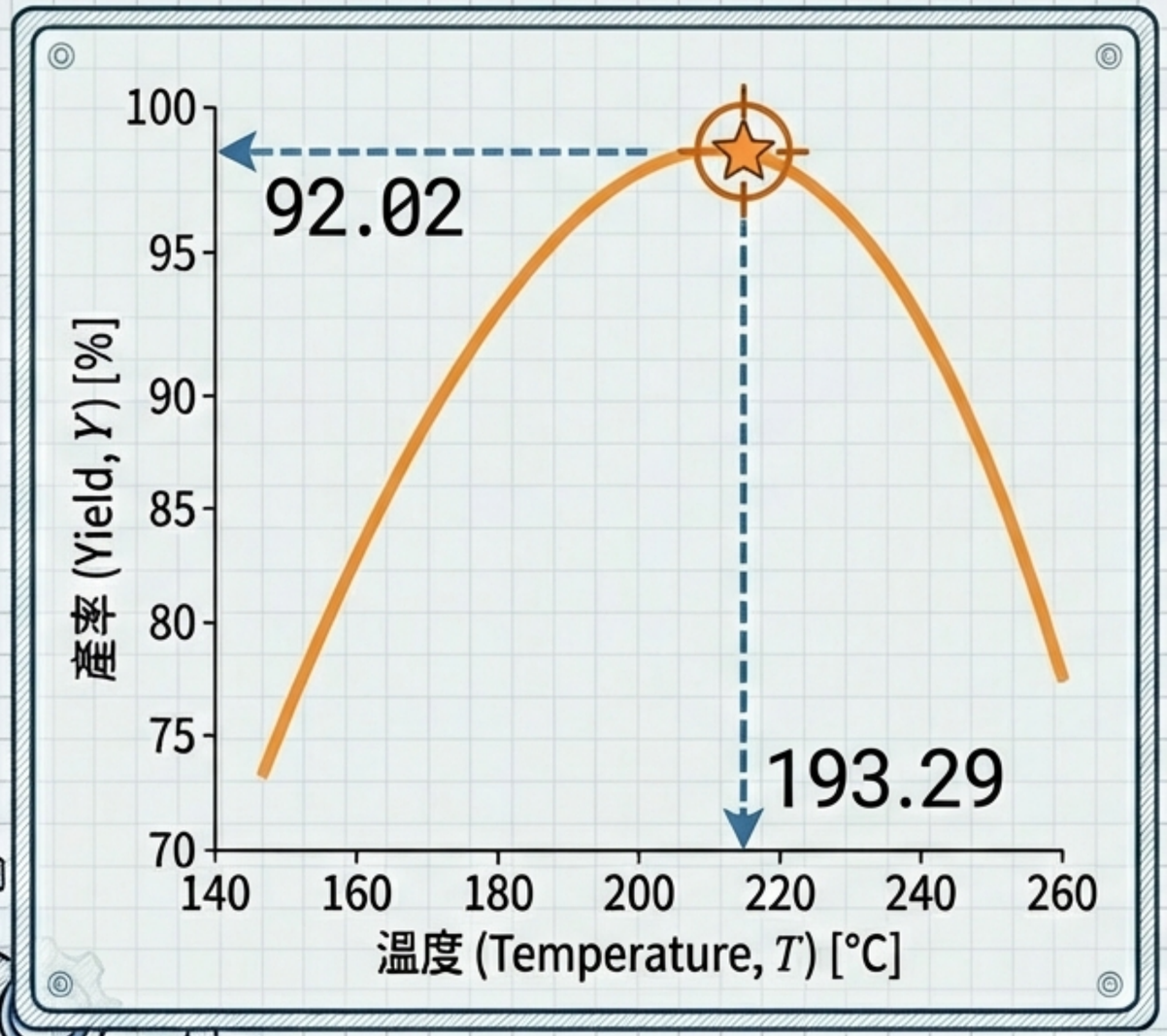
Insight: Degree 1 完全失敗。Degree 5 在訓練集表現好，但在測試集崩潰 (Test $R^2 < 0$)。

驗證策略：交叉驗證 (Validation Strategy: Cross-Validation)

高次多項式在小數據集上極不穩定，交叉驗證 (Cross-Validation) 幫助我們避開過擬合陷阱。



尋找最佳操作條件 (Finding the Optimum)



OPTIMIZATION RESULT

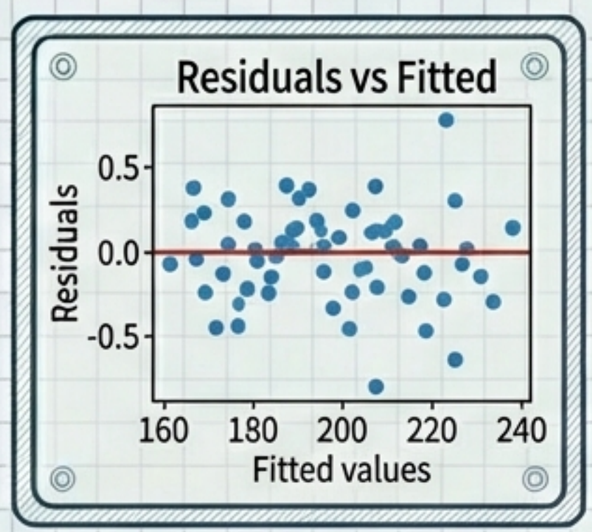
Optimum Temperature: **193.29 °C**

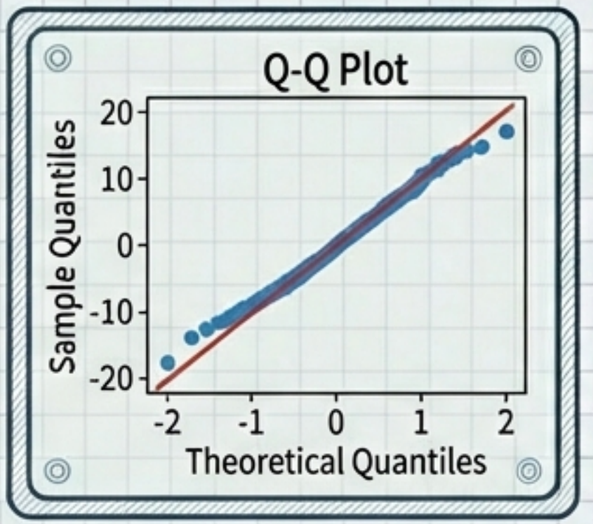
Max Yield: **92.02 %**

$$Y = 91.78 - 2.17T - 4.73T^2 + 0.57T^3$$

該模型成功定位了比實驗數據點更精確的最佳溫度。

模型診斷 (Model Diagnostics)





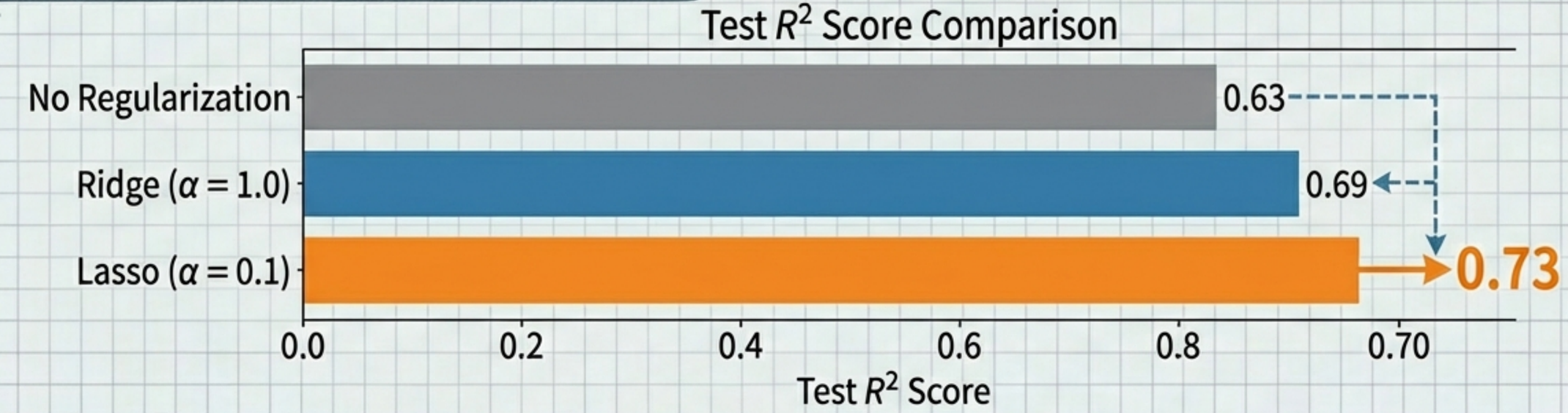
Assumption Check

- ☑ **均值為 0 (Mean ≈ 0):**
  無系統性偏差
- ☑ **常態性 (Normality):**
  Q-Q 圖顯示殘差符合常態分佈
- ☑ **同質變異性 (Homoscedasticity):**
  誤差一致

模型假設成立，預測可信
(Reliable Predictions)

進階改進：正則化 (Regularization)

使用 Lasso (L1) 或 Ridge (L2) 懲罰過大的係數，防止過擬合的係數，防止過擬合。



Lasso 不僅防止了過擬合，還進一步提升了模型的泛化能力 (+15%)。

實務建議與檢查清單 (Best Practices)

CHECKLIST

- ☑ **標準化 (Standardize)：**
  務必使用 StandardScaler 處理特徵
- ☑ **使用 Pipeline：**
  封裝預處理與模型，防止資料洩漏
- ☑ **交叉驗證 (Cross-Validation)：**
  使用 CV 選擇最佳次數 d
- ☑ **正則化 (Regularization)：**
  高次項建議搭配 Lasso/Ridge
- ☒ **禁止外推 (No Extrapolation)：**
  多項式在訓練範圍外會劇烈震盪

您的數位工具箱已升級

非線性建模

捕捉真實物理現象

PolynomialFeatures

靈活的特徵擴展

Pipeline & Regularization

穩健的工程實踐

下一步 (Next Step)：開啟 Unit11_Polynomial_Regression.ipynb

現在就開始您的實作練習！