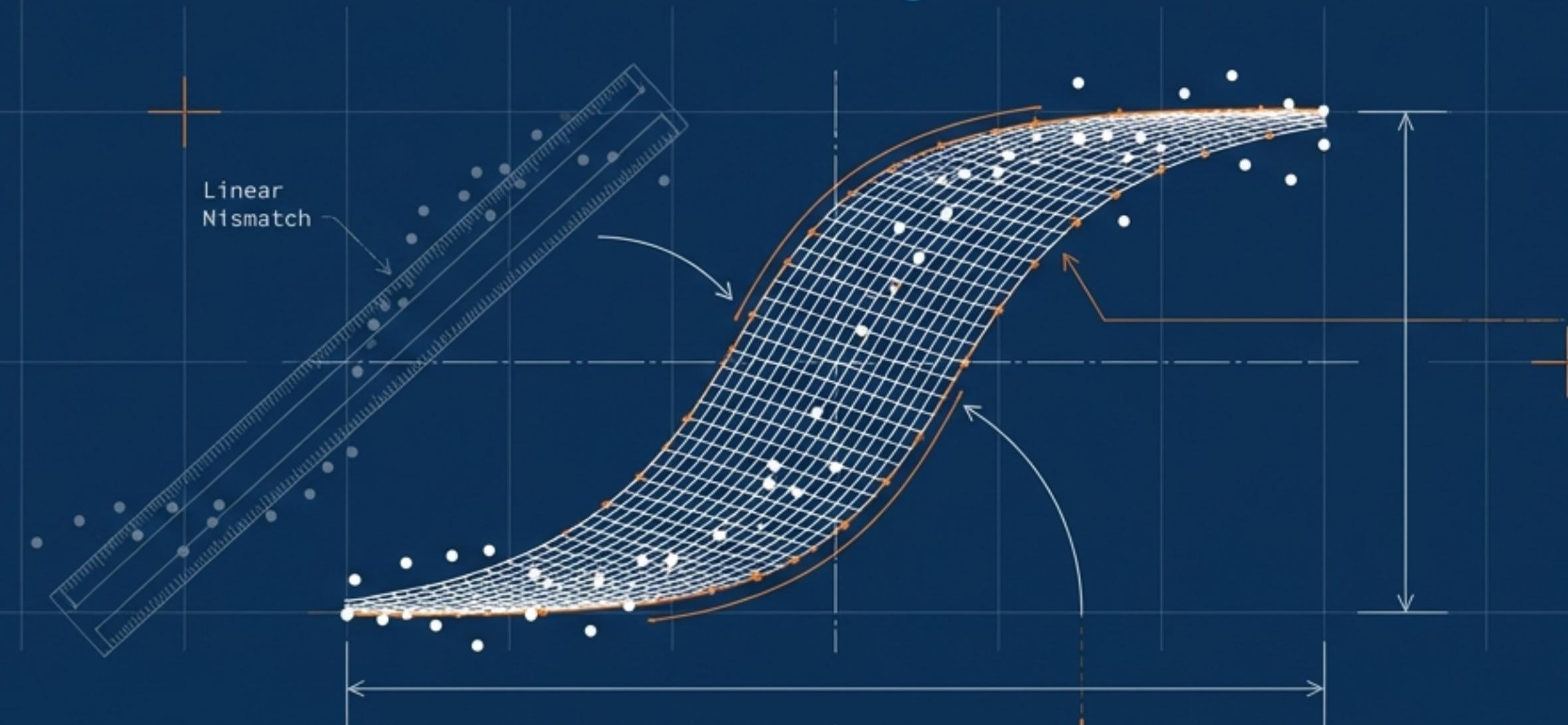


Unit 11 非線性模型回歸總覽

Non-Linear Models Regression Overview



突破線性限制：捕捉化工製程中的複雜模式

為什麼需要非線性模型？物理世界的真實樣貌

化工現象本質上是非線性的

1. 化學反應動力學 (Reaction Kinetics)

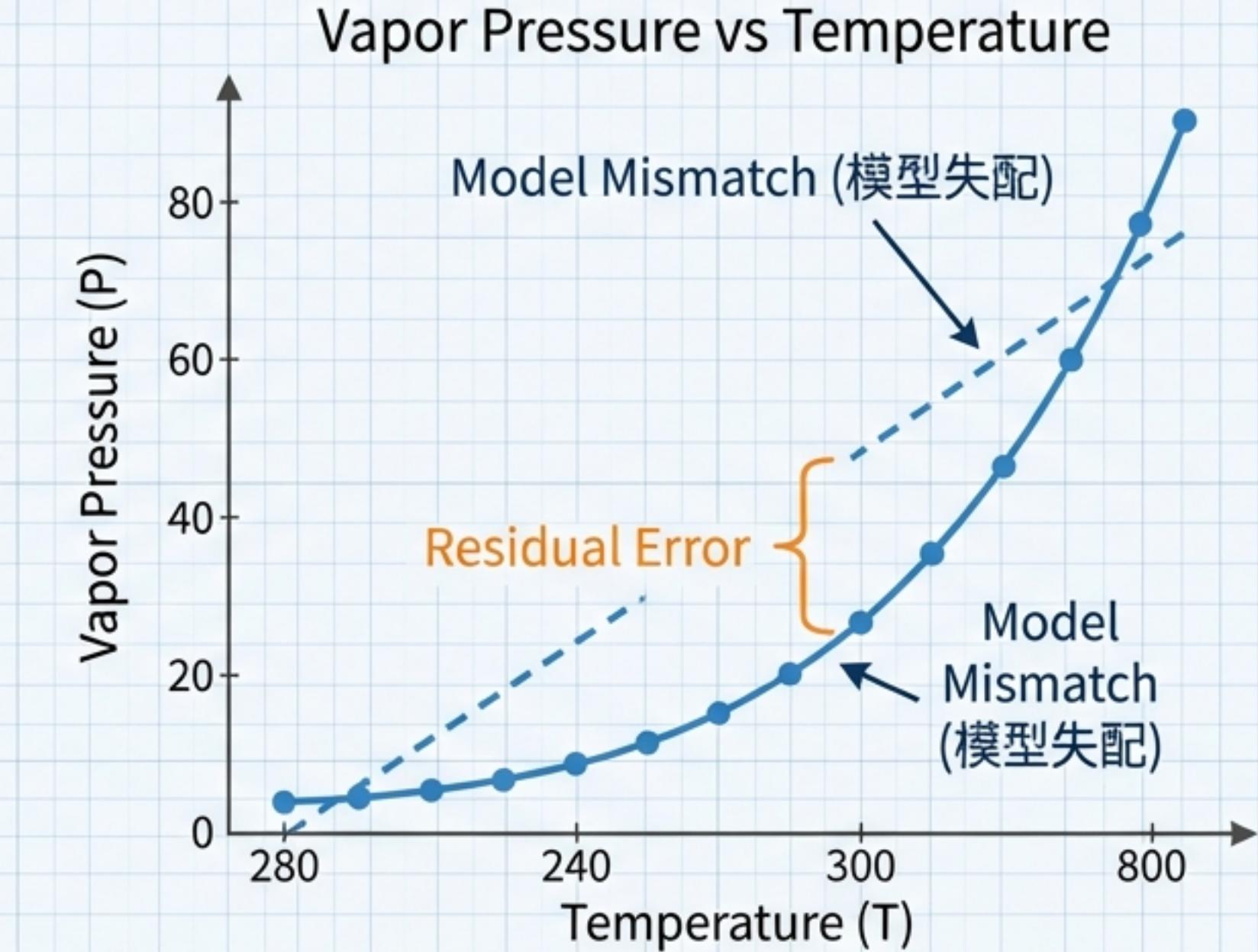
$$k = A e^{-\frac{E_a}{RT}} \text{ (Arrhenius)}$$

2. 相平衡 (Phase Equilibrium)

$$\log_{10} P = A - \frac{B}{C + T} \text{ (Antoine)}$$

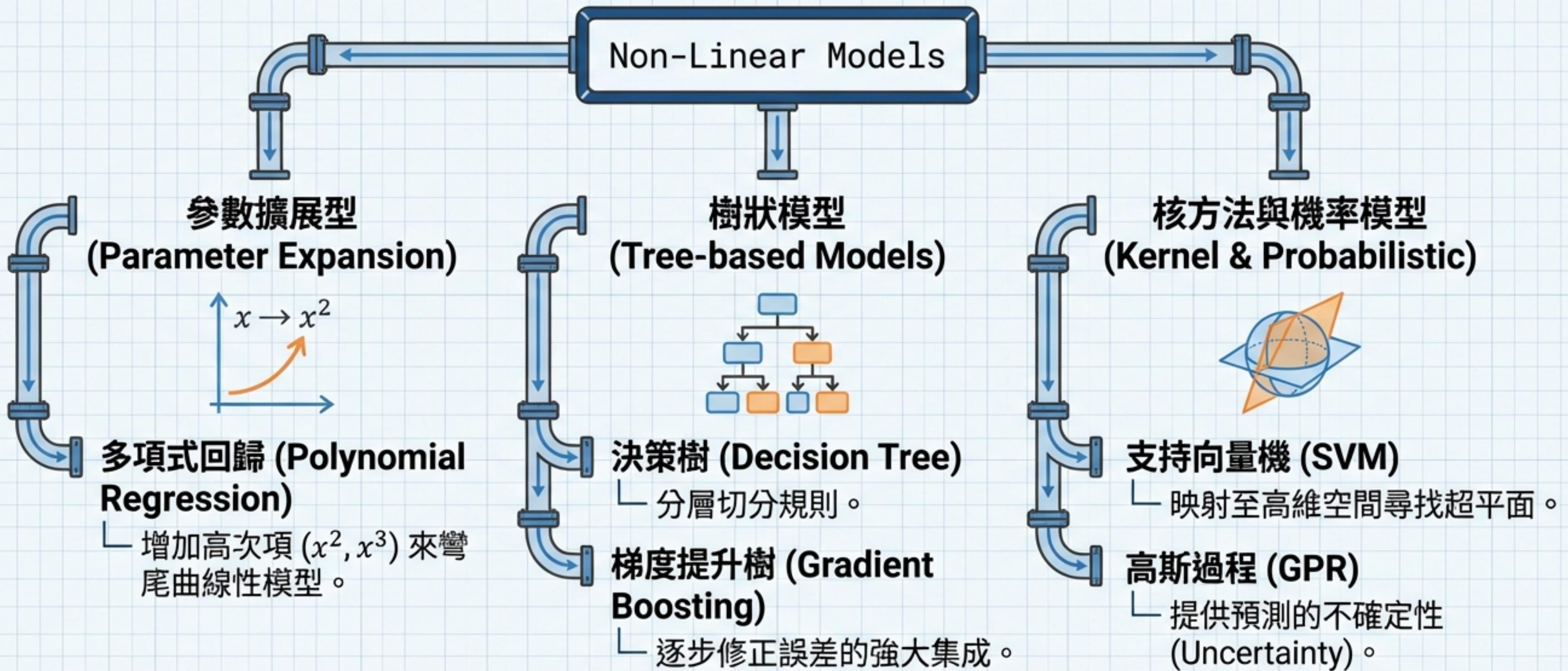
3. 吸附等溫線 (Adsorption)

$$q = \frac{q_{\max} K C}{1 + K C} \text{ (Langmuir)}$$

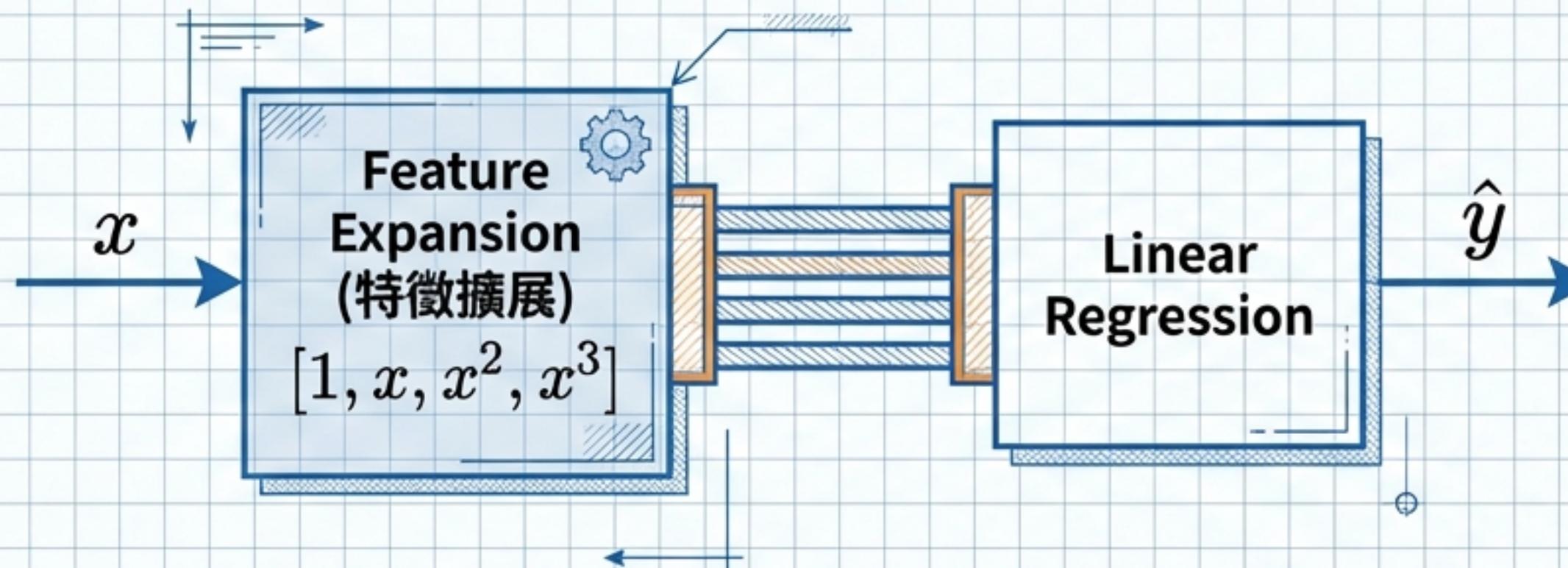


線性模型無法捕捉曲線關係、飽和效應與相變臨界點。

非線性模型工具箱：演算法分類



多項式回歸 (Polynomial Regression)：彎曲的線性模型



$$\hat{y} = w_0 + w_1x + w_2x^2 + w_3x^3$$

Technical Specs

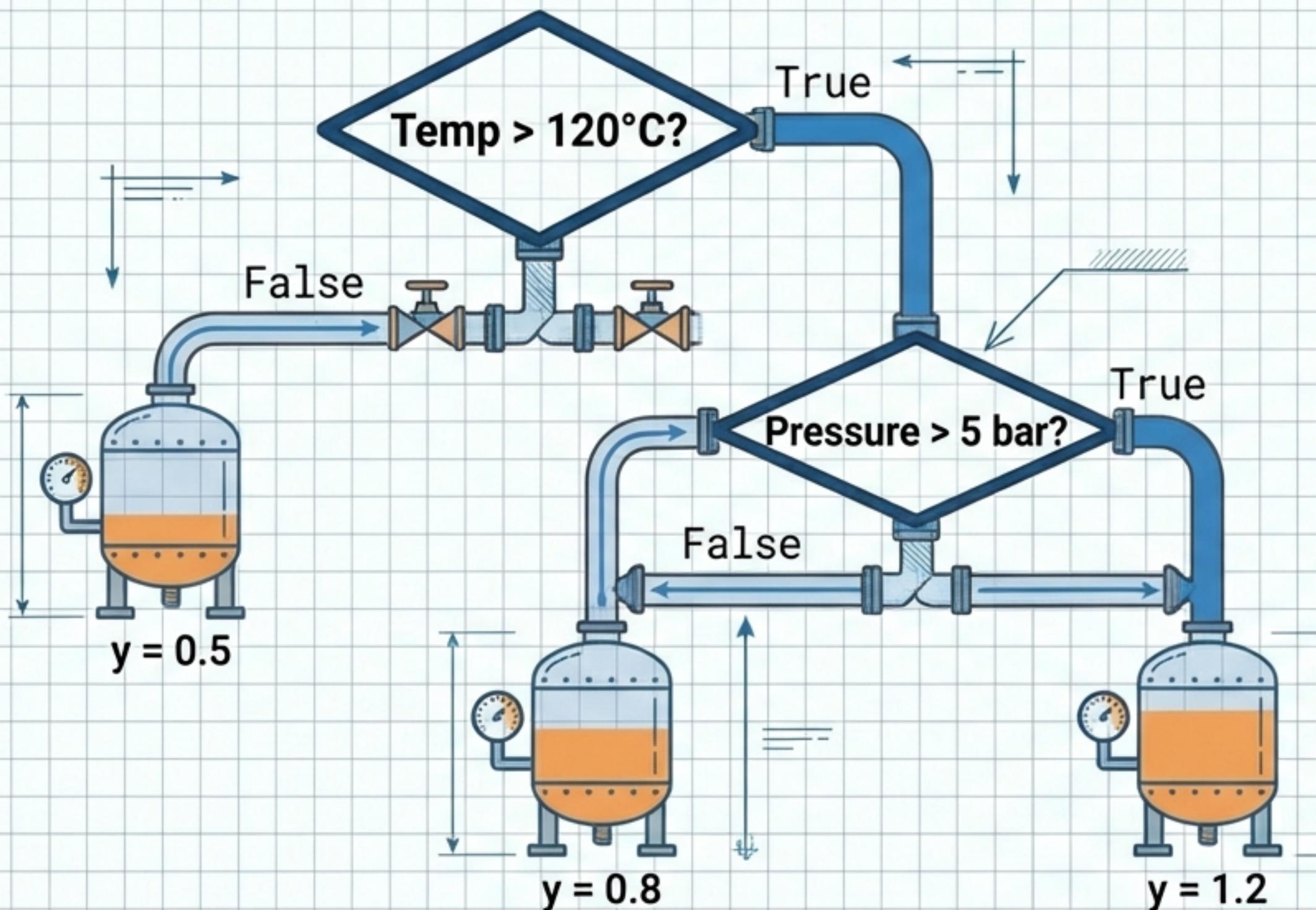
[+] 優點 (Pros):

- 實作簡單 (Simple implementation)
- 可解釋性高 (Interpretable)
- 適合平滑曲線 (Fits smooth curves)

[-] 缺點 (Cons)

- 高次項導致特徵爆炸 (Feature explosion)
- ! • 外推能力極差 (Poor Extrapolation)

決策樹 (Decision Tree)：分層決策的流程圖



**核心機制 (Mechanism) : **

- Concept: 透過一系列 If-Then 規則切割特徵空間。
- Optimization: 最小化子節點的 MSE (均方誤差)。

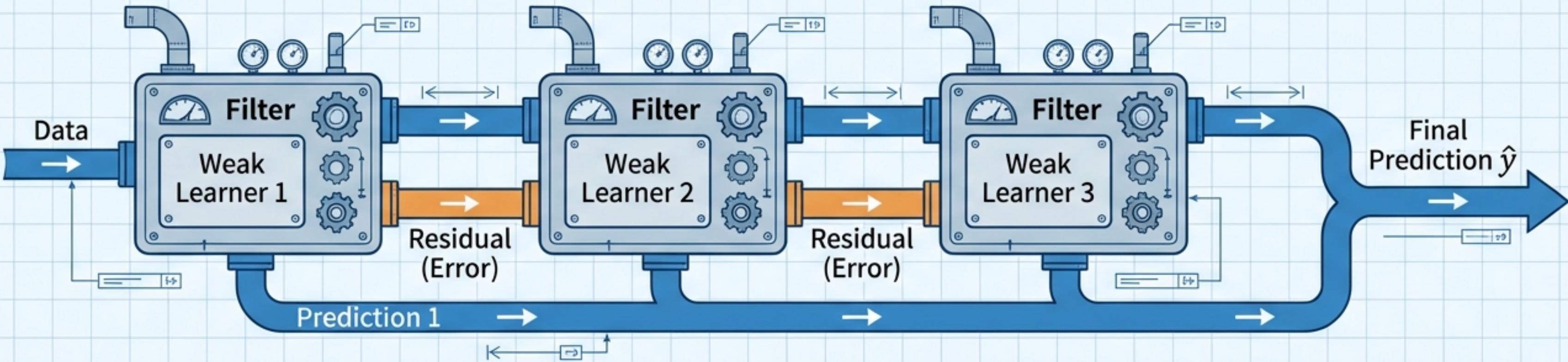
[+] 優點 (Pros):

- 無需特徵縮放 (No feature scaling)
- 自動處理交互作用 (Automatic interaction handling)
- 高度可視化 (White-box)

[−] 缺點 (Cons):

- 容易過擬合 (Overfitting) - Need Pruning
- 預測呈階梯狀 (Step-function), 不夠平滑 (Not smooth)

梯度提升樹 (Gradient Boosting Trees)：逐步修正的精確度之王



核心機制 (Mechanism):

Concept: 序列式集成學習 (Sequential Ensemble)

Process: 模型 1 預測目標 → 模型 2 預測「模型 1 的誤差」

$$\hat{y} = f_0 + \eta \sum f_t(\mathbf{x})$$

Technical Specs

**工業標準: XGBoost / LightGBM

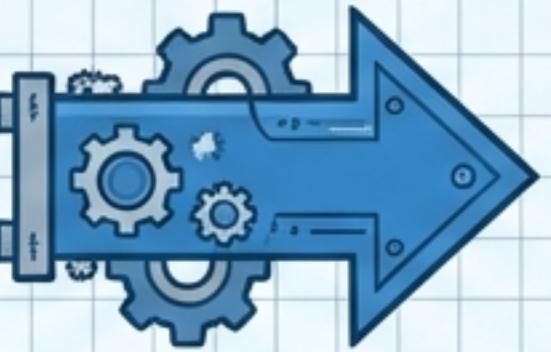
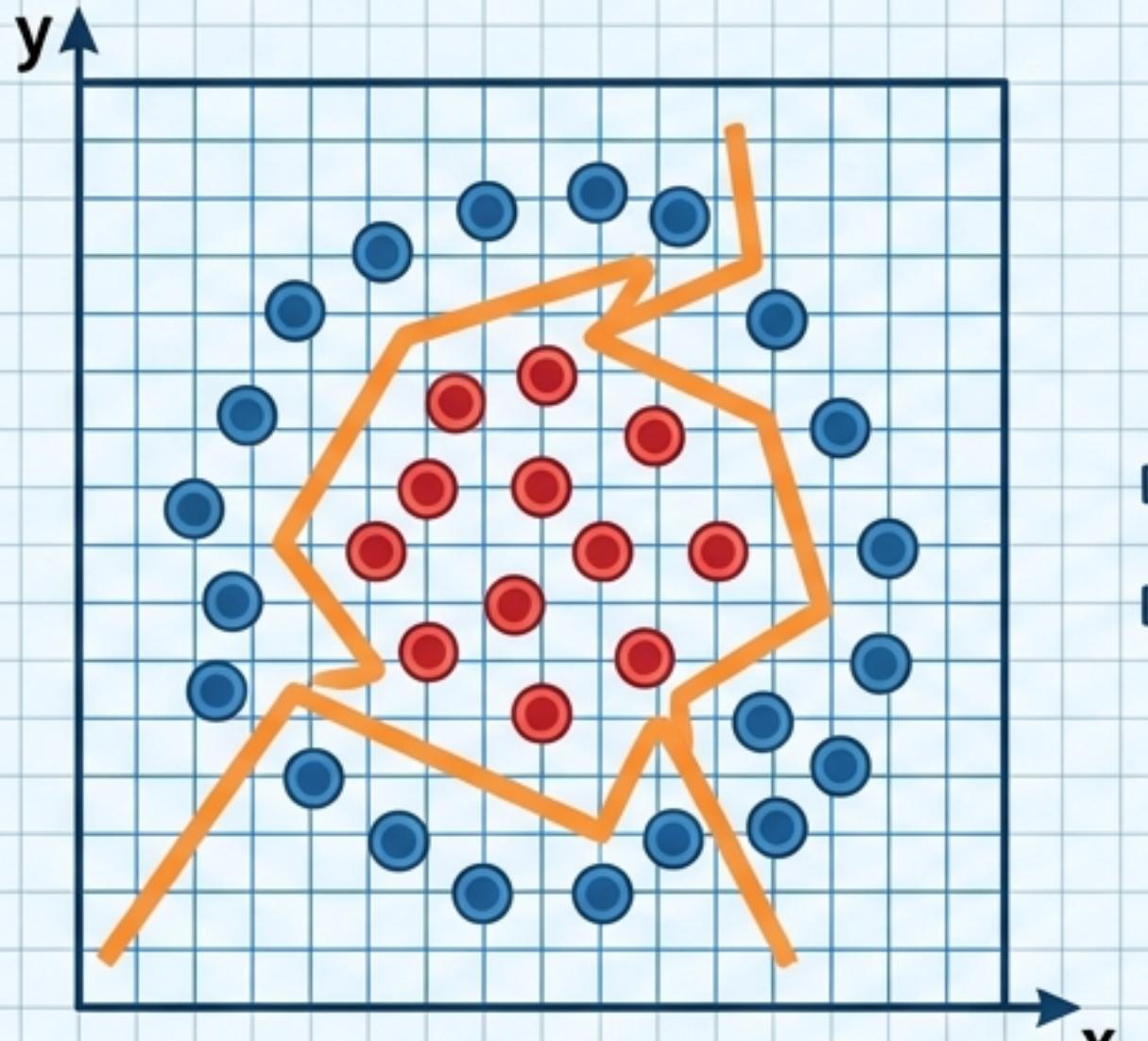
[+] 優點:

- 極高的預測準確度 (Kaggle Winner)
- 對異常值穩健

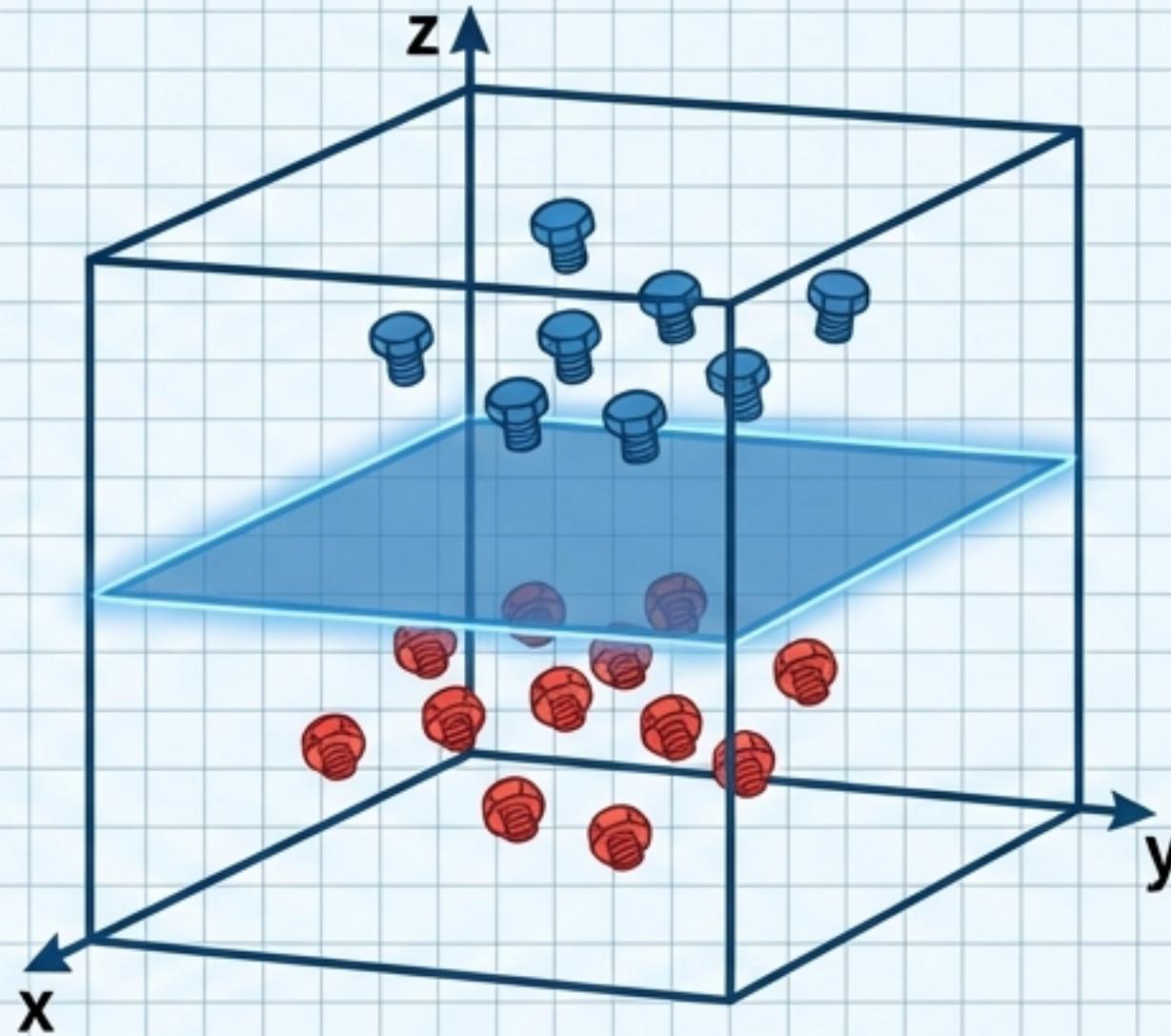
[−] 缺點:

- 黑盒模型 (Black Box)
- 訓練慢，參數多 !

支持向量機 (SVM)：高維空間的幾何切割



Mapping $\phi(x)$



Feature Space (Hyperplane Separation)

核心概念 (Core Concept):
Kernel Trick (核技巧)



重要規範 (Specs):

- Critical: 必須進行特徵縮放 (Scaling Required)。
- Best Use: 中小型資料集，高維特徵 (如光譜資料)。

高斯過程回歸 (GPR) : 帶有不確定性的預測



核心機制

基於貝氏推論 (Bayesian Inference) 的機率模型。

Output:

預測值 μ + 信賴區間 σ 。

策略應用 (Strategic Use):

- 主動學習 (Active Learning) : 利用不確定性指導下一次實驗。
- 限制 (Constraint) : 計算成本 $O(n^3)$, 僅適合小資料 (<1000 筆)。

模型選擇指南：如何挑選正確的工具？



模型 (Model)	準確度 (Accuracy)	可解釋性 (Interpretability)	資料量 (Data Size)	不確定性 (Uncertainty)
Polynomial	Low (低)	High (高)	Any (任意)	No (無)
Decision Tree	Med (中)	High (高)	Med (中)	No (無)
Gradient Boosting	High (高)	Low (低)	Large (大)	No (無)
SVM	High (高)	Low (低)	Small/Med (中/小)	No (無)
GPR	High (高)	Low (低)	Small (小)	Yes (有)



資料前處理：模型的安全聯鎖 (Preprocessing)

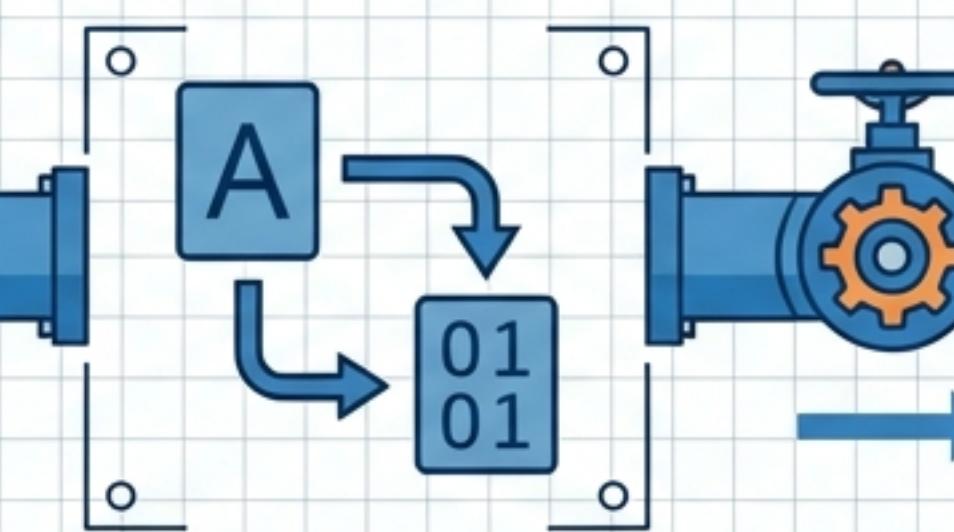
特徵縮放 (Feature Scaling)



Mandatory for:
SVM, GPR, Poly.

對數值範圍敏感 ($z = \frac{x-\mu}{\sigma}$)

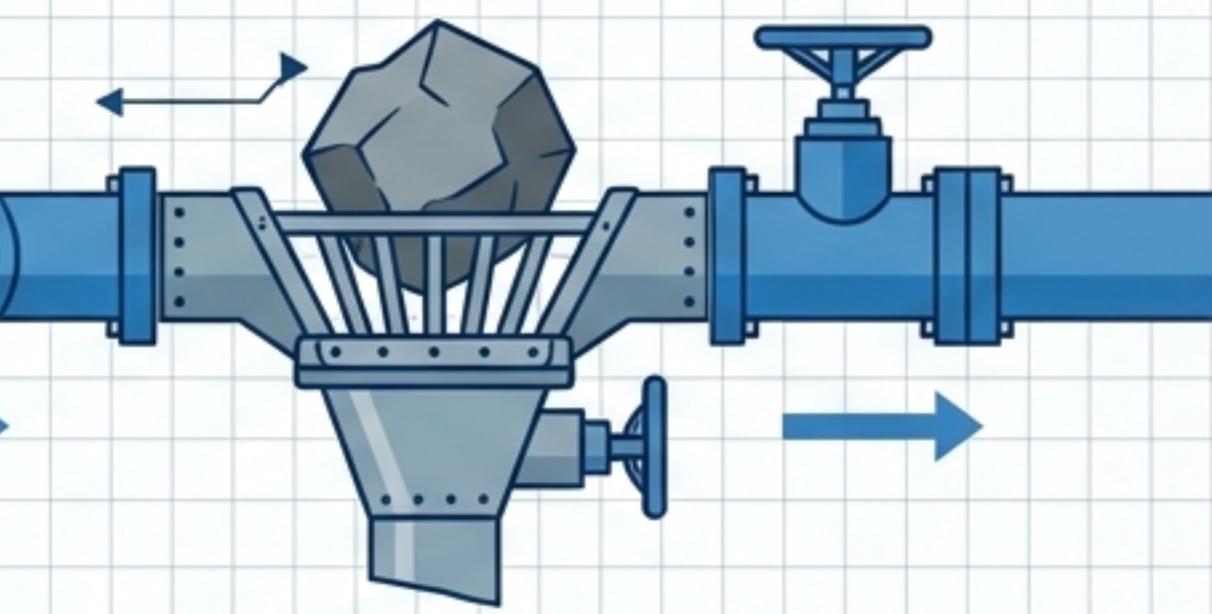
編碼 (Encoding)



One-Hot Encoding /
Target Encoding

類別轉數值

異常值 (Outliers)



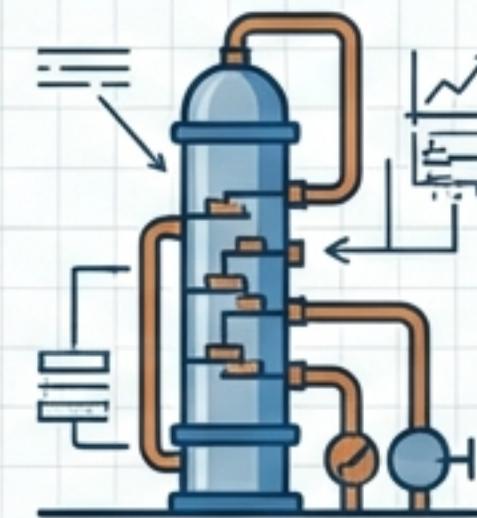
Robust Scaling /
Clipping

防止模型偏差



化工領域實際應用案例 (Real-World Applications)

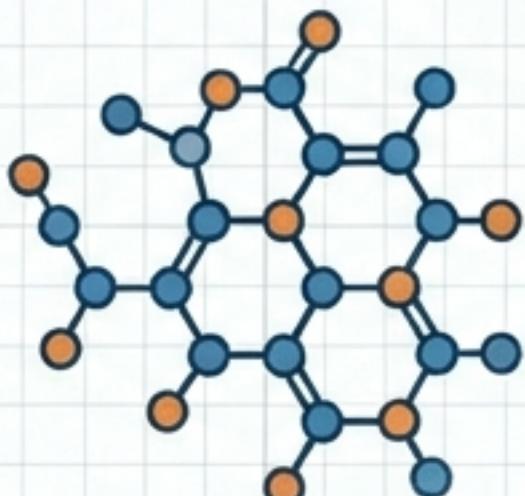
蒸餾塔控制 (Distillation Control)



Model: XGBoost

Impact: 預測 15 分鐘後的產品純度，波動降低 40%。

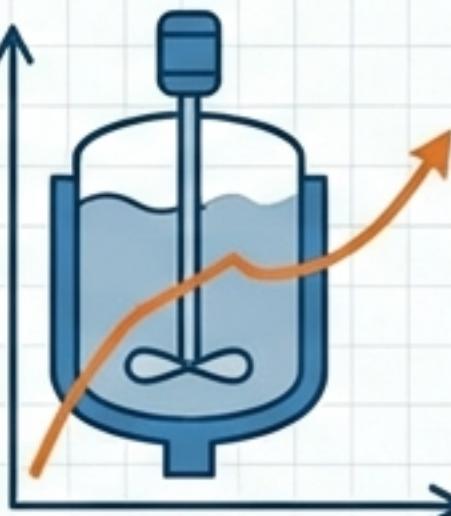
藥物溶解度預測 (Solubility QSAR)



Model: SVM (RBF)

Impact: 篩選 200+ 個分子描述符，虛擬篩選減少 80% 候選分子。

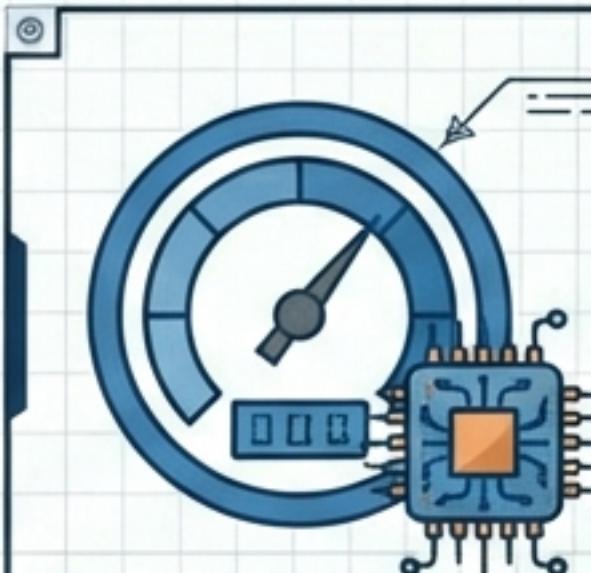
反應器產率優化 (Reactor Optimization)



Model: GPR + Bayesian Opt

Impact: 在 10 維參數空間尋找最佳溫度曲線，實驗減少 60%。

更多應用場景 (Extended Application Scenarios)



軟感測器 (Soft Sensors)

預測難以實時測量的變數 (如聚合物分子量)。

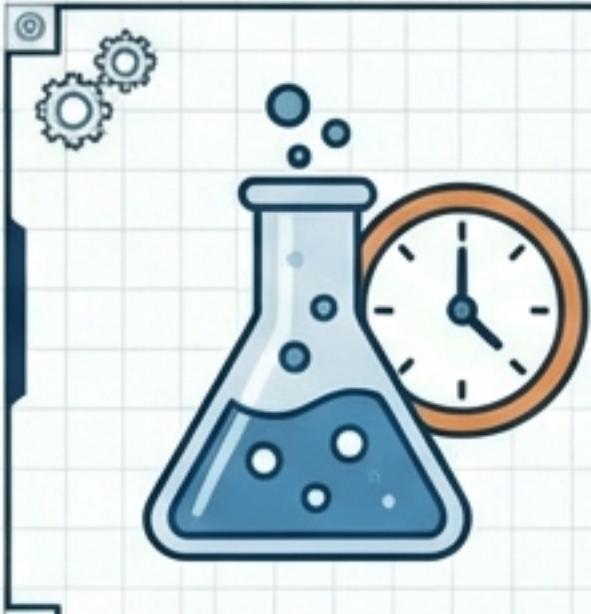
Roboto Mono



故障診斷 (Fault Detection)

識別泵浦或壓縮機的異常振動訊號。

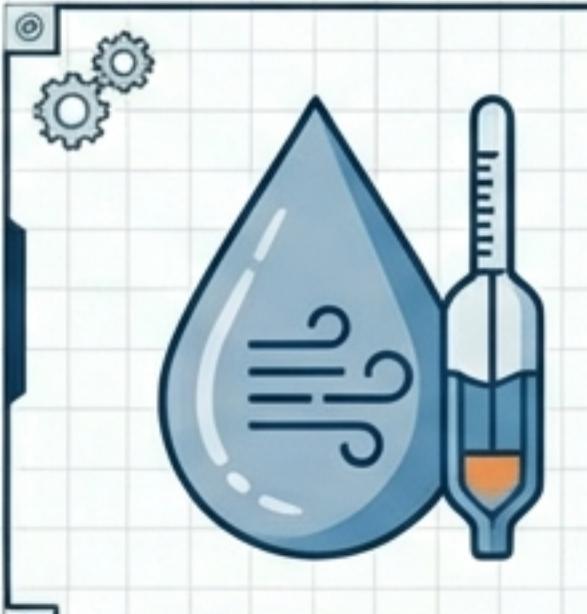
Roboto Mono



反應動力學 (Kinetics)

擬合複雜的多步驟反應速率常數。

Roboto Mono



物性預測 (Property Prediction)

預測混合物黏度、密度與熱值。

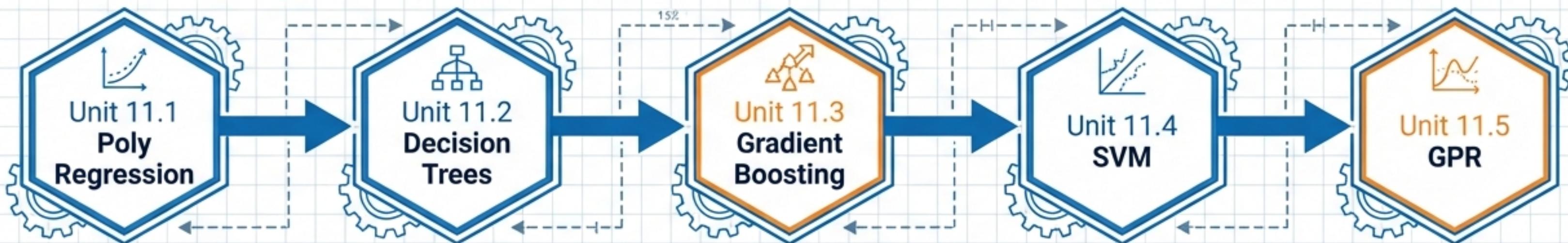
Roboto Mono

總結與學習路徑 (Summary & Learning Path)

Key Takeaways

- ✓ 沒有最好的模型，只有最適合的模型 (No Free Lunch)。
- ✓ 工業界大數據首選：**XGBoost (準確度)**。
- ✓ 實驗設計小數據首選：**GPR (不確定性)**。

The Roadmap



→ Next Step: 開啟 `Unit11_Polynomial_Regression.ipynb` 進行實作

課程資訊與授權 (Course Info & Credits)

Course: AI在化工上之應用

Unit: Unit 11 Non-Linear Models Regression

Instructor: 莊曜禎 助理教授 (Prof. Yao-Jen Zhuang)

Institution: 逢甲大學 化工系 智慧程序系統工程實驗室
(FCU ChemE, PSE Lab)

License: CC BY-NC-SA 4.0

Date: 2026-01-28



Roboto Mono