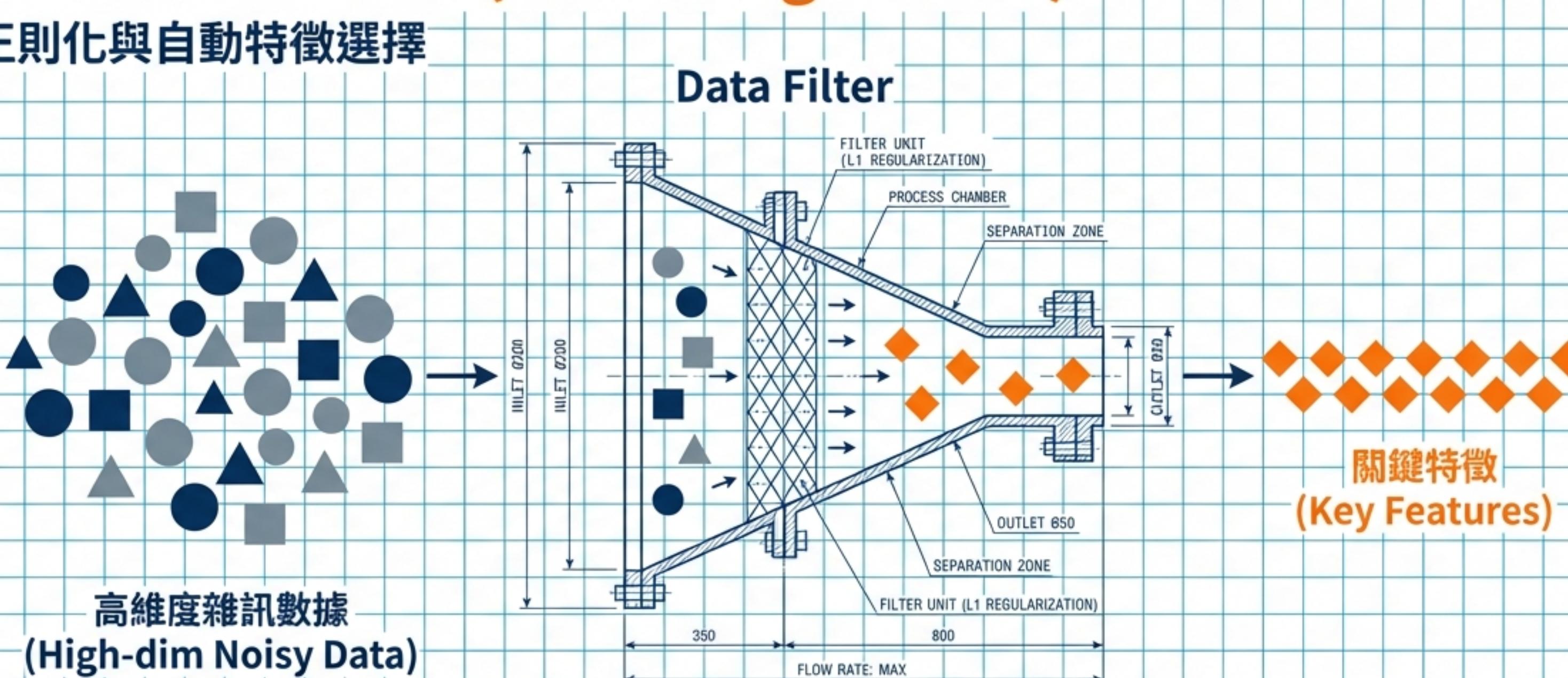


AI 在化工上的應用：從原理到實踐

Unit 10 Lasso 回歸 (Lasso Regression)

L1 正則化與自動特徵選擇



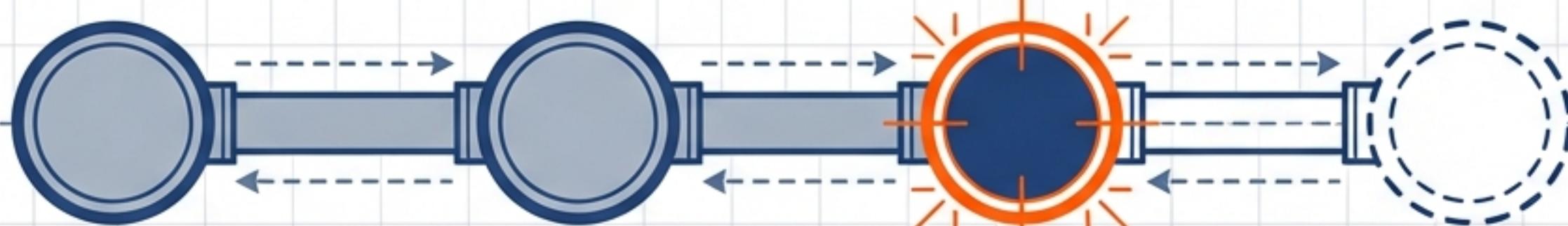
課程地圖與學習路徑 (Learning Roadmap)

Phase 1:
基礎
Foundation

Phase 2:
非監督式學習
Unsupervised

Phase 3:
監督式學習
Supervised

Phase 4:
進階
Advanced

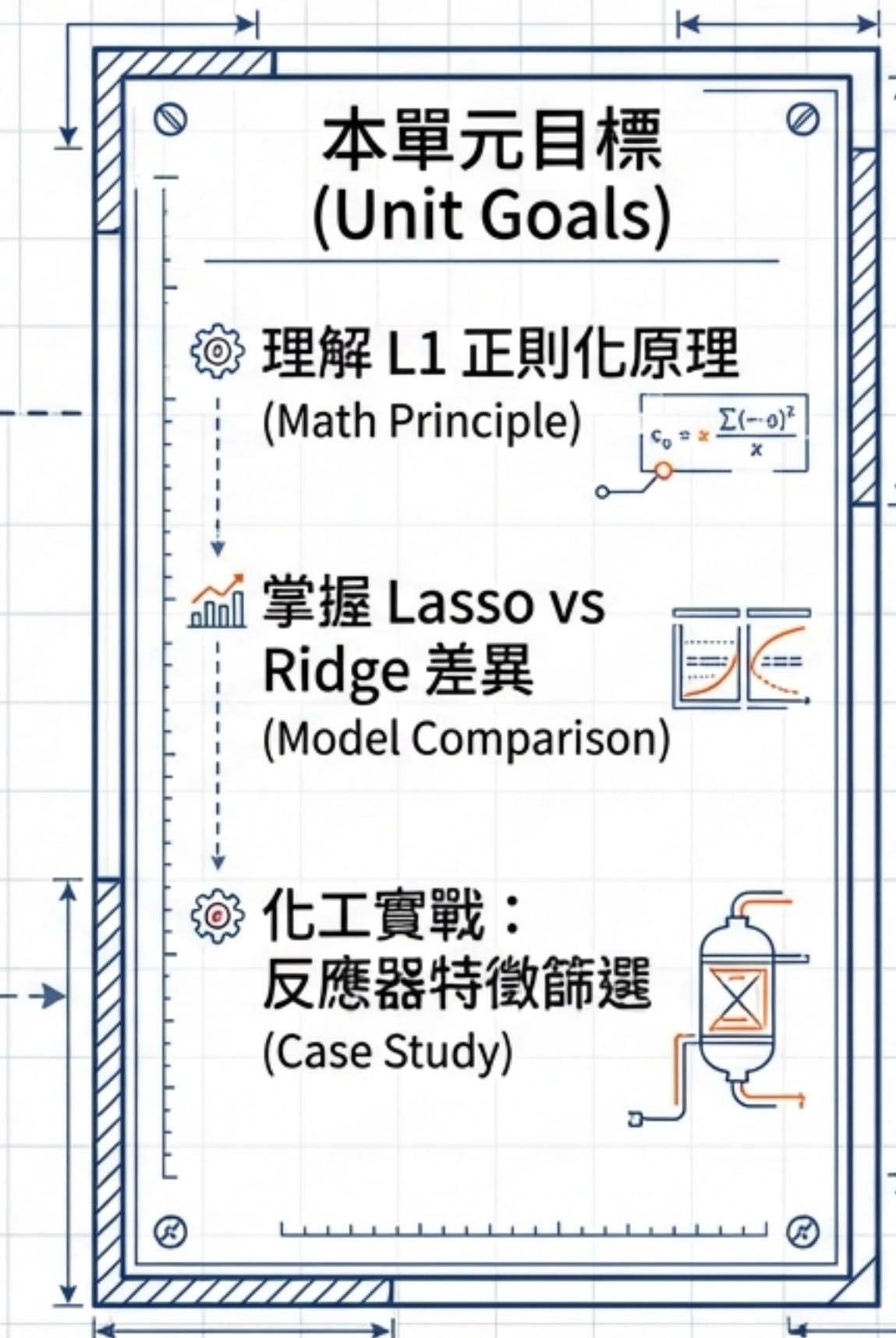


專注單元
(Focus Unit)

Unit 09:
OLS Linear Regression

Unit 10:
Lasso Regression

Unit 11: Elastic Net



核心概念：什麼是 Lasso 回歸？(What is Lasso?)

定義 (Definition)

Least Absolute Shrinkage and Selection Operator

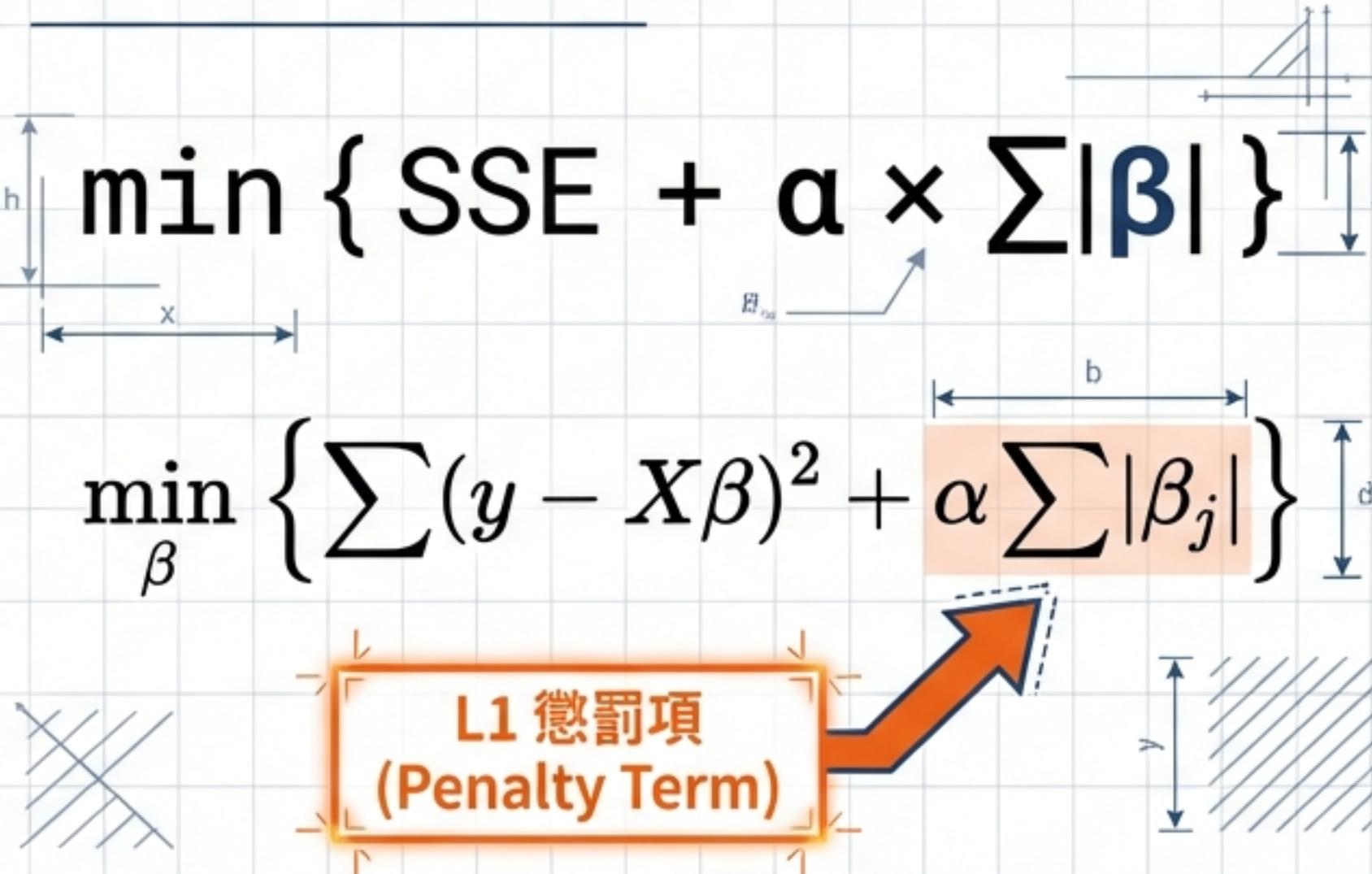
- 在普通最小平方法 (OLS) 中加入 L1 正則化項。
- 核心功能：不僅防止過擬合 (Overfitting)，更能將不重要特徵的係數壓縮為 0。



數學 Formula

$$\min \{ \text{SSE} + \alpha \times \sum |\beta| \}$$

$$\min_{\beta} \left\{ \sum (y - X\beta)^2 + \alpha \sum |\beta_j| \right\}$$

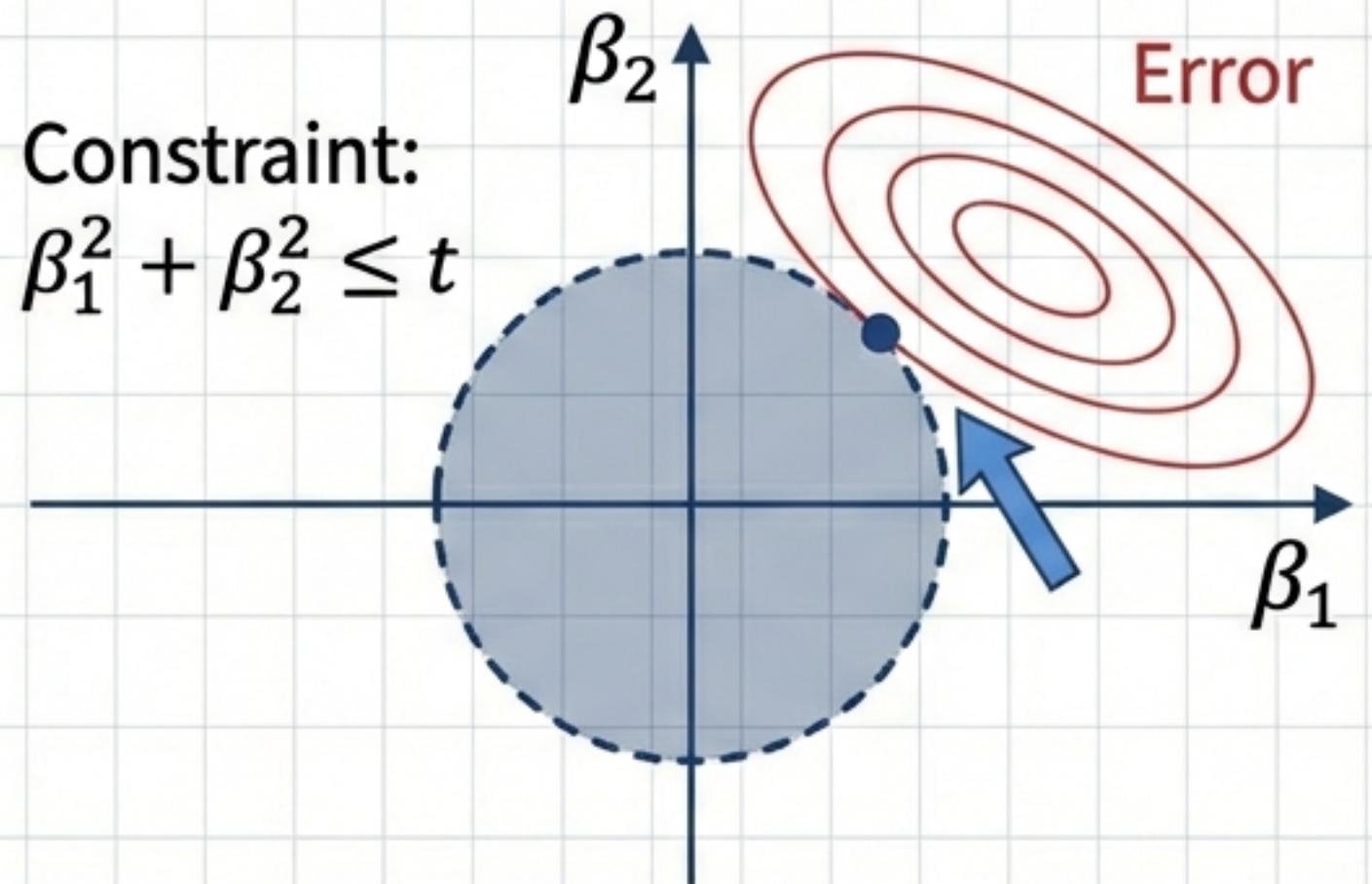


L1 懲罰項
(Penalty Term)

關鍵洞察：Lasso 能產生稀疏解 (Sparse Solution) — 模型會自動剔除冗餘變數，只保留關鍵特徵。

幾何直覺：為什麼 L1 能產生稀疏解？(Geometric Intuition)

Ridge Regression (L2)

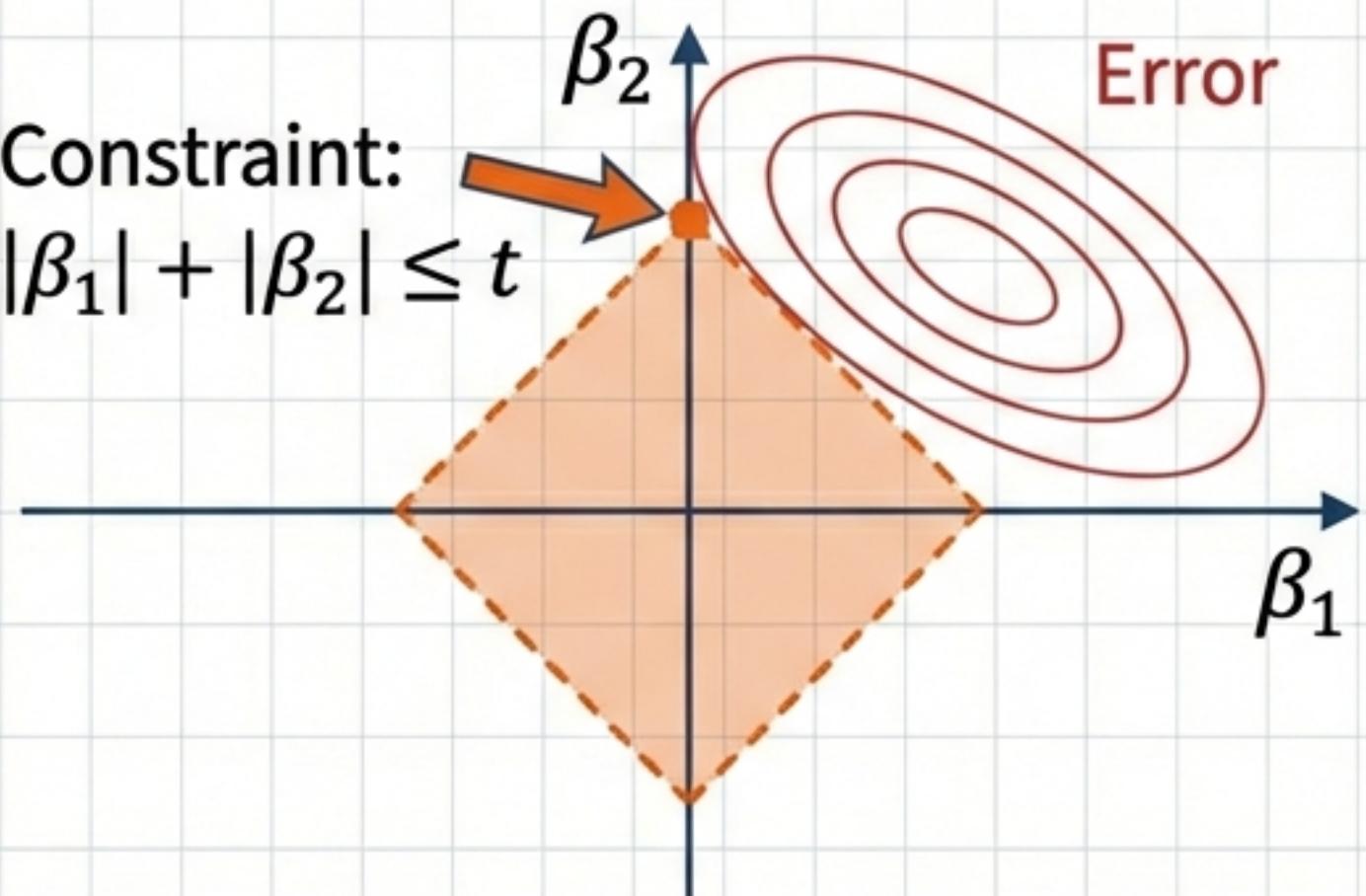


Constraint:

$$\beta_1^2 + \beta_2^2 \leq t$$

觸碰點不在軸上 → 係數變小，但不為 0

Lasso Regression (L1)



Constraint:

$$|\beta_1| + |\beta_2| \leq t$$

觸碰點在尖角 (Corner) → β_1 變為 0

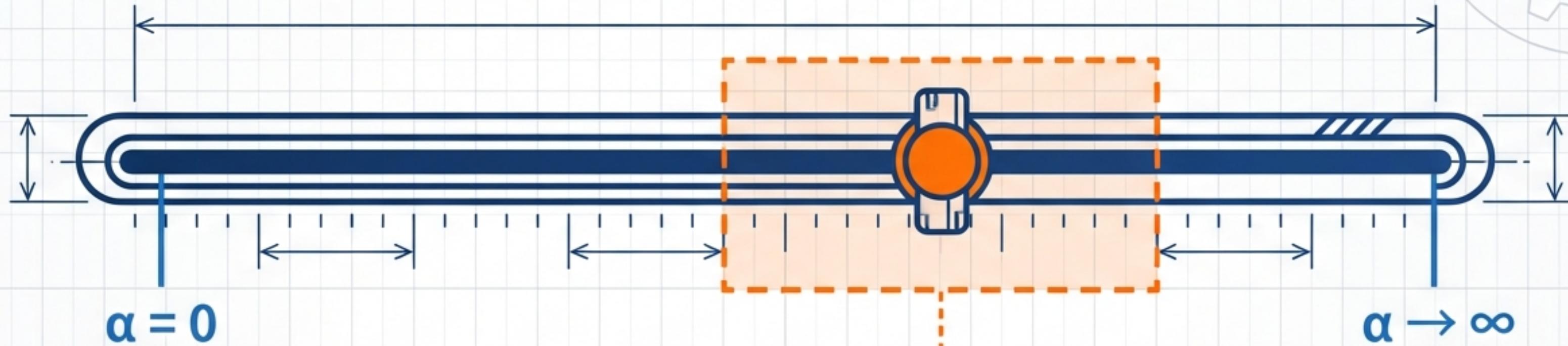
菱形的尖角是 Lasso 具備「特徵選擇」能力的幾何根源。

模型大對決：OLS vs Ridge vs Lasso



關鍵洞察：Lasso 能產生稀疏解 (Sparse Solution) — 模型會自動剔除冗餘變數，只保留關鍵特徵。

控制閥：正則化強度 α (Regularization Strength)



OLS 模型

無特徵選擇，模型最複雜

Optimal α

甜蜜點 (Sweet Spot)
平衡預測準確度與模型簡潔性

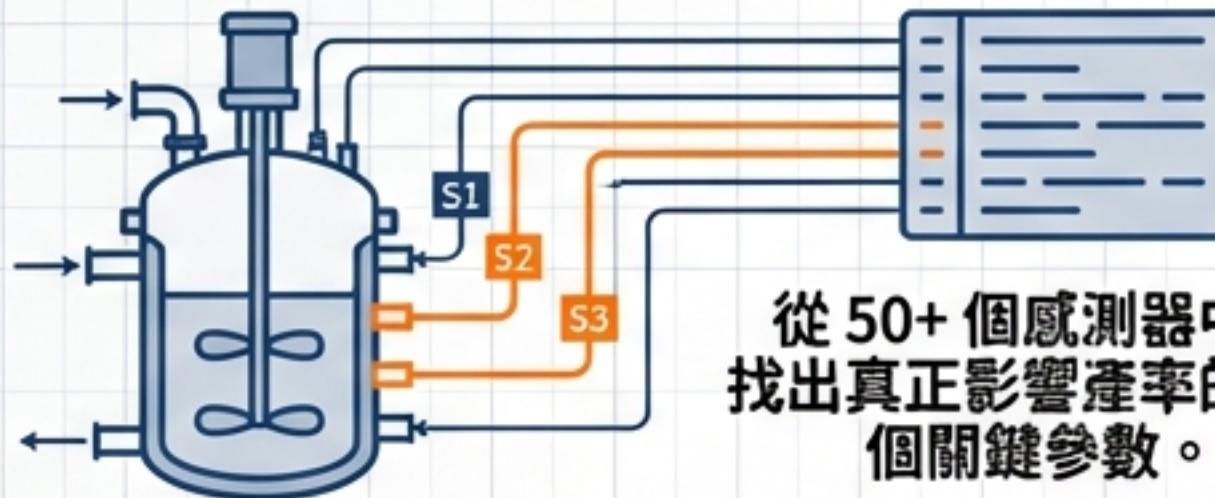
Null Model

所有係數為 0，發生欠擬合
(Underfitting)

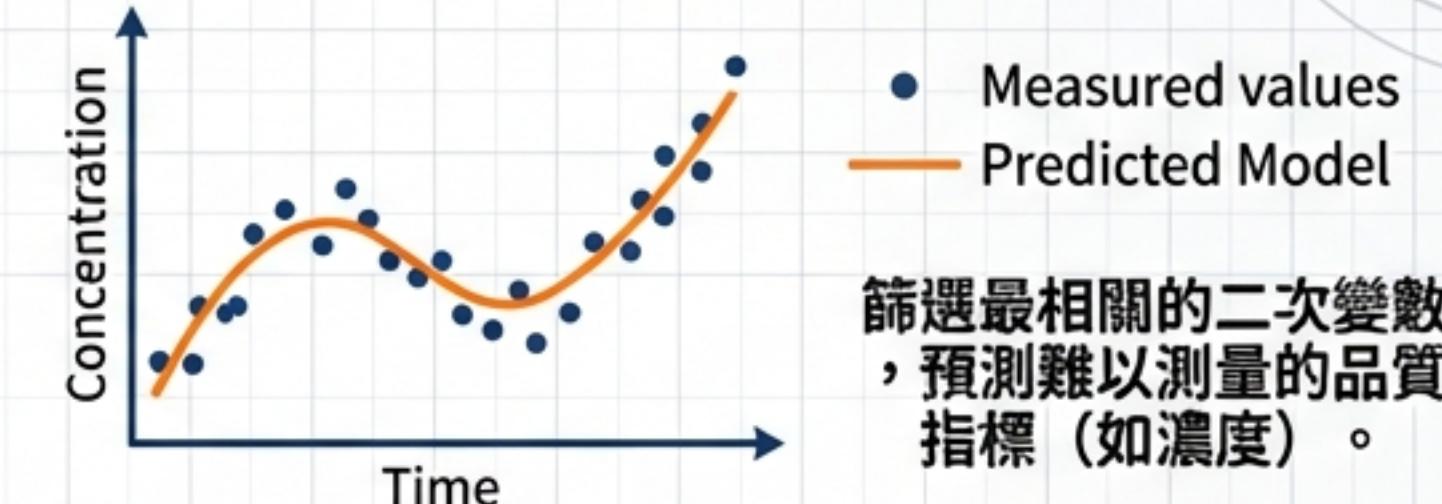
策略建議：不要憑感覺猜測 α 。請使用 **LassoCV** 進行交叉驗證
(Cross-Validation) 自動掃描最佳參數。

為什麼化工工程師需要 Lasso ? (Applications in ChemE)

反應器優化 (Reactor Optimization)



軟感測器開發 (Soft Sensors)



可解釋性 (Interpretability)



$$Y = \beta_1 * \text{Pressure} + \beta_2 * \text{Temperature}$$

主要因素
(Main Driver)

向操作員解釋『為什麼』製程發生偏移
(例如：是壓力導致的，而非溫度)。

降低成本 (Cost Reduction)



停止維護那些沒有預測價值的昂貴儀表。

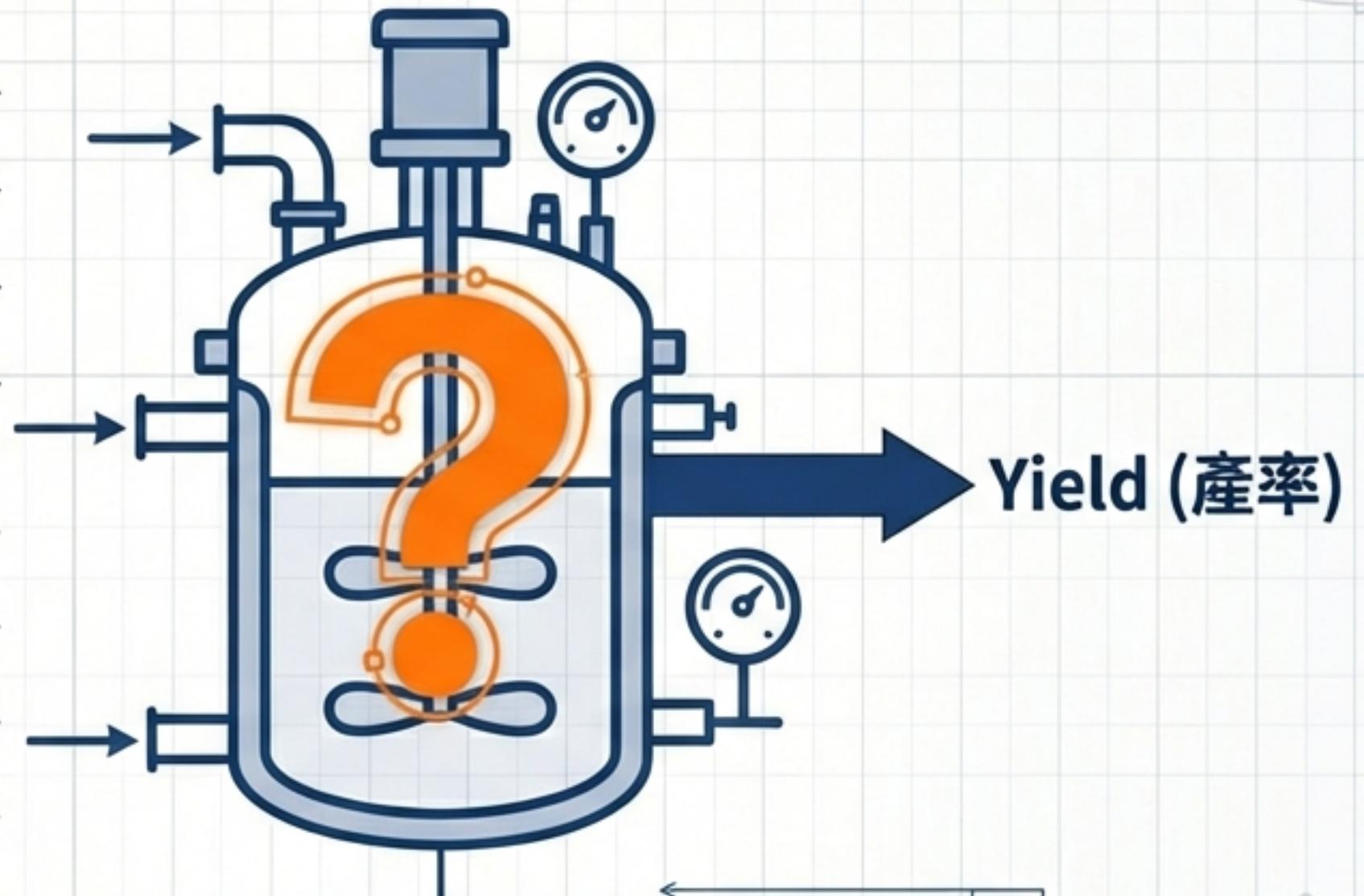
關鍵洞察：Lasso 透過自動特徵篩選，幫助化工工程師從複雜數據中找出關鍵驅動因子，實現更高效、更經濟的製程控制。

實戰案例：化學催化反應產率預測

目標：預測產率 (Yield)，並從 8 個操作變數中自動找出真正影響因子。

True Signals
- 真實訊號

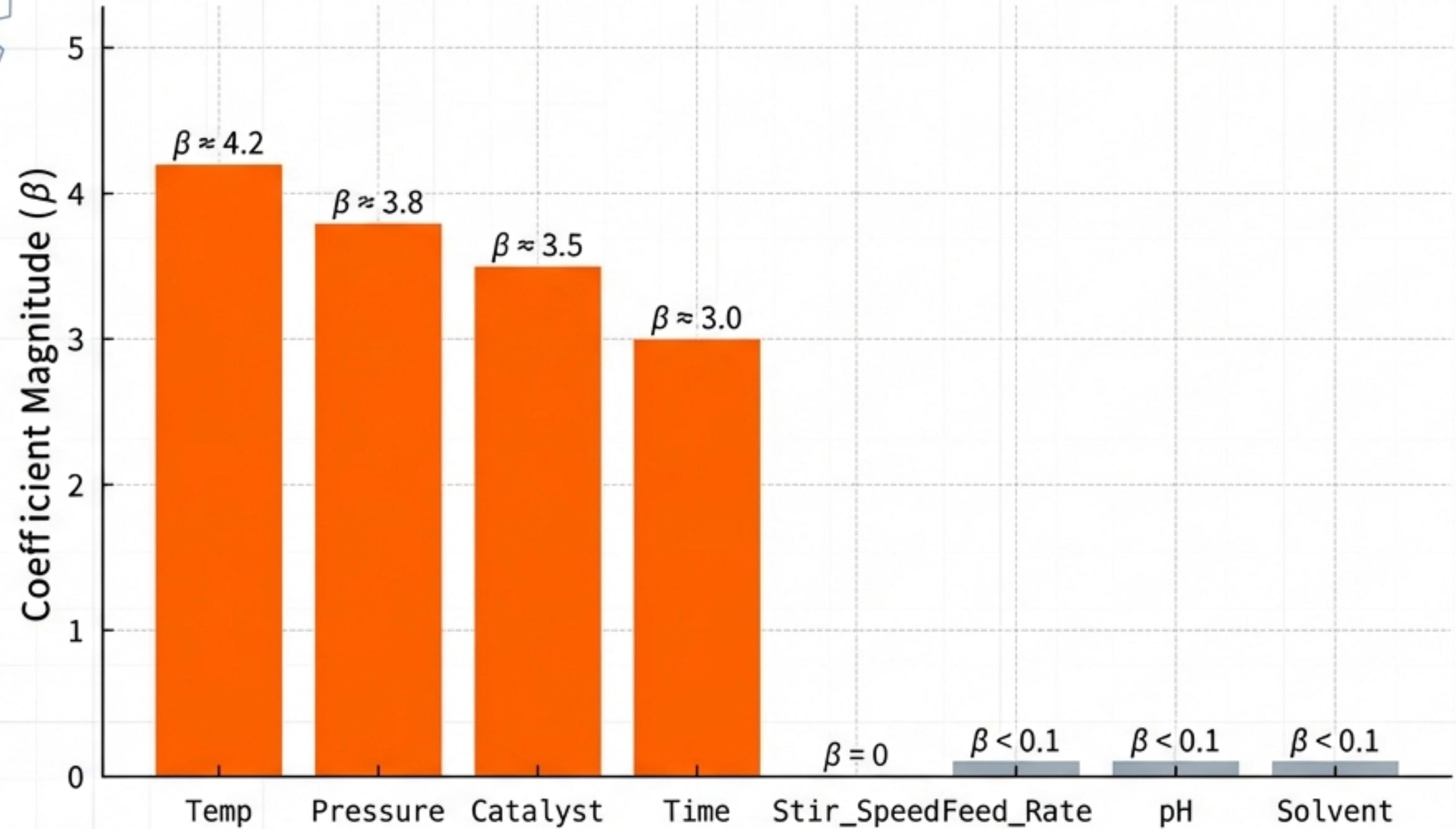
- Temperature (T)
- Pressure (P)
- Catalyst (C)
- Time (t)
- Stirring Speed
- Feed Rate
- pH
- Solvent Ratio



Noise
- 雜訊干擾

挑戰：AI 能否在不依賴人類經驗的情況下
，自動發現下面 4 個變數是無效的？

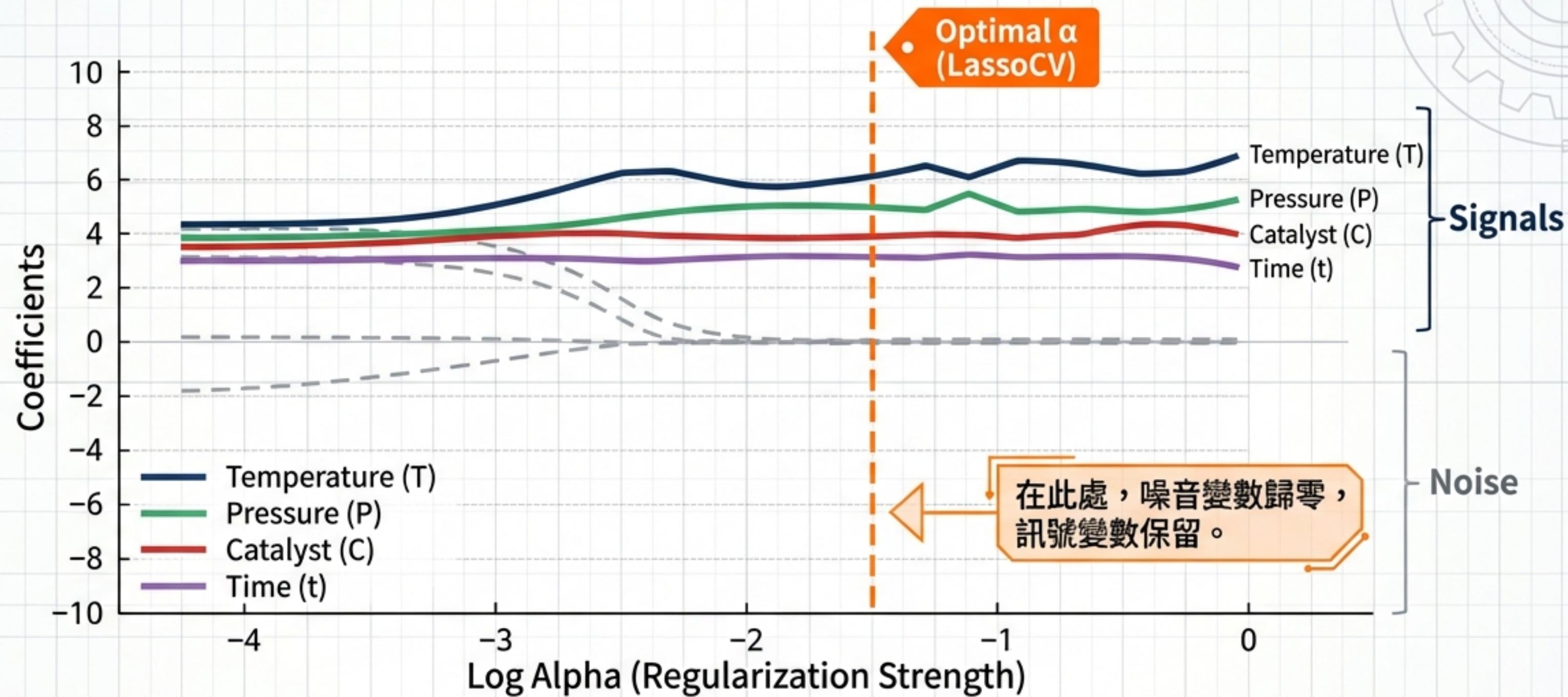
實驗結果：Lasso 揭示真相 (The Truth Revealed)



分析結果

1. 特徵篩選成功：Lasso 將攪拌速度 (Stirring Speed) 的係數壓縮為 0。
2. 雜訊抑制：其餘無關變數的權重極低，幾乎被忽略。
3. 洞察：模型自動學會了哪些物理量是無關的，無需人工干預。

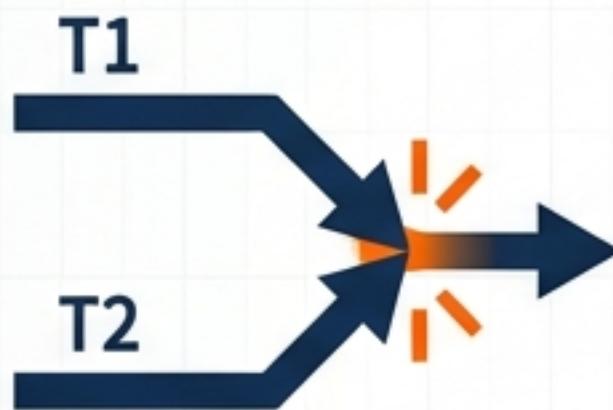
可視化正則化路徑 (Regularization Path)



此圖證明了模型的穩健性 (Robustness)：
重要特徵在各種 α 強度下持續保持顯著。

現實世界的挑戰 (Real World Challenges)

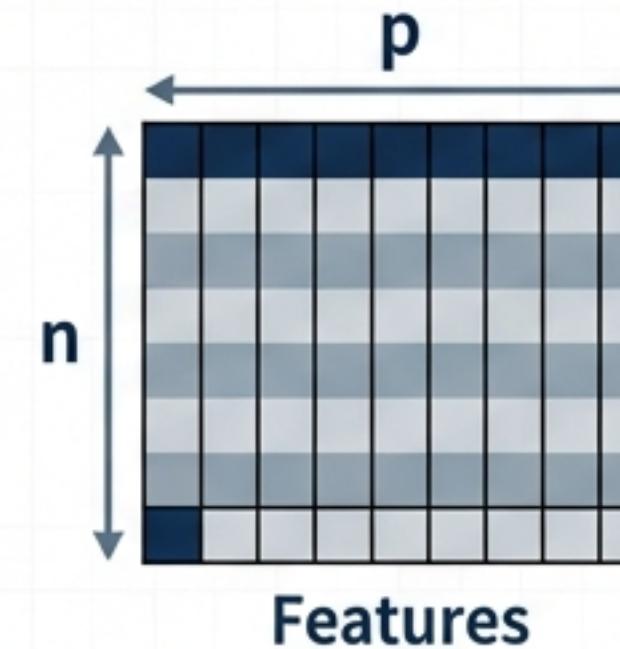
多重共線性 (Multicollinearity)



問題：如果特徵 A 與特徵 B 高度相關（例如兩個溫度計 T_1, T_2 ），Lasso 會隨機選一個並丟棄另一個。

結果不穩定。

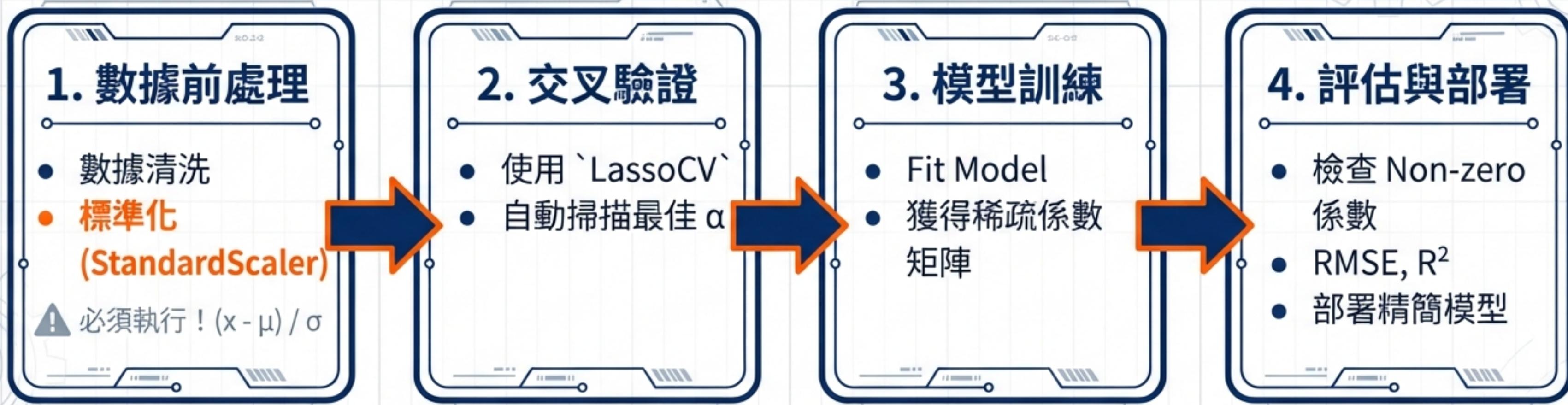
數據稀缺 (Data Scarcity)



問題：當特徵數 (p) 多於樣本數 (n) 時，Lasso 最多只能選出 n 個特徵。

解決方案：若遇到上述問題，
請改用 Elastic Net (結合 L1 + L2)。

標準工作流程 (Standard Operating Procedure)



總結：何時該使用 Lasso ? (Checklist)

推薦使用 (Yes)

- 懷疑數據中包含大量無關特徵 (Noise > Signal)
- 需要具備「可解釋性」的稀疏模型
- 作為複雜建模前的「特徵篩選器」
- 高維度數據 ($p \gg n$)

避免/注意 (Caution)

- 特徵間存在高度相關性 (改用 Elastic Net)
- 所有特徵依物理原理皆為重要 (改用 Ridge)
- 忘記做數據標準化 (Standardization)

下一步與學習資源 (Next Steps)

Coming Up: Unit 11 Elastic Net

結合 L1 與 L2 的混合策略，解決共線性問題。

課後練習 (Homework)

- 開啟 Unit10_Lasso_Regression.ipynb
- 任務：使用 Distillation Column 數據集，比較 Lasso 與 Ridge 的特徵選擇差異。

參考資源 (Resources)

- Scikit-learn Documentation: [LassoCV](#)
- Textbook: *The Elements of Statistical Learning*

「簡單即是美：Lasso 幫我們在
雜訊中找到訊號。」
(Simplicity is beauty: Lasso helps us
find the signal in the noise.)

AI 在化工上的應用 (AI in Chemical Engineering)
逢甲大學 智慧程序系統工程實驗室 | 莊曜禎 助理教授

