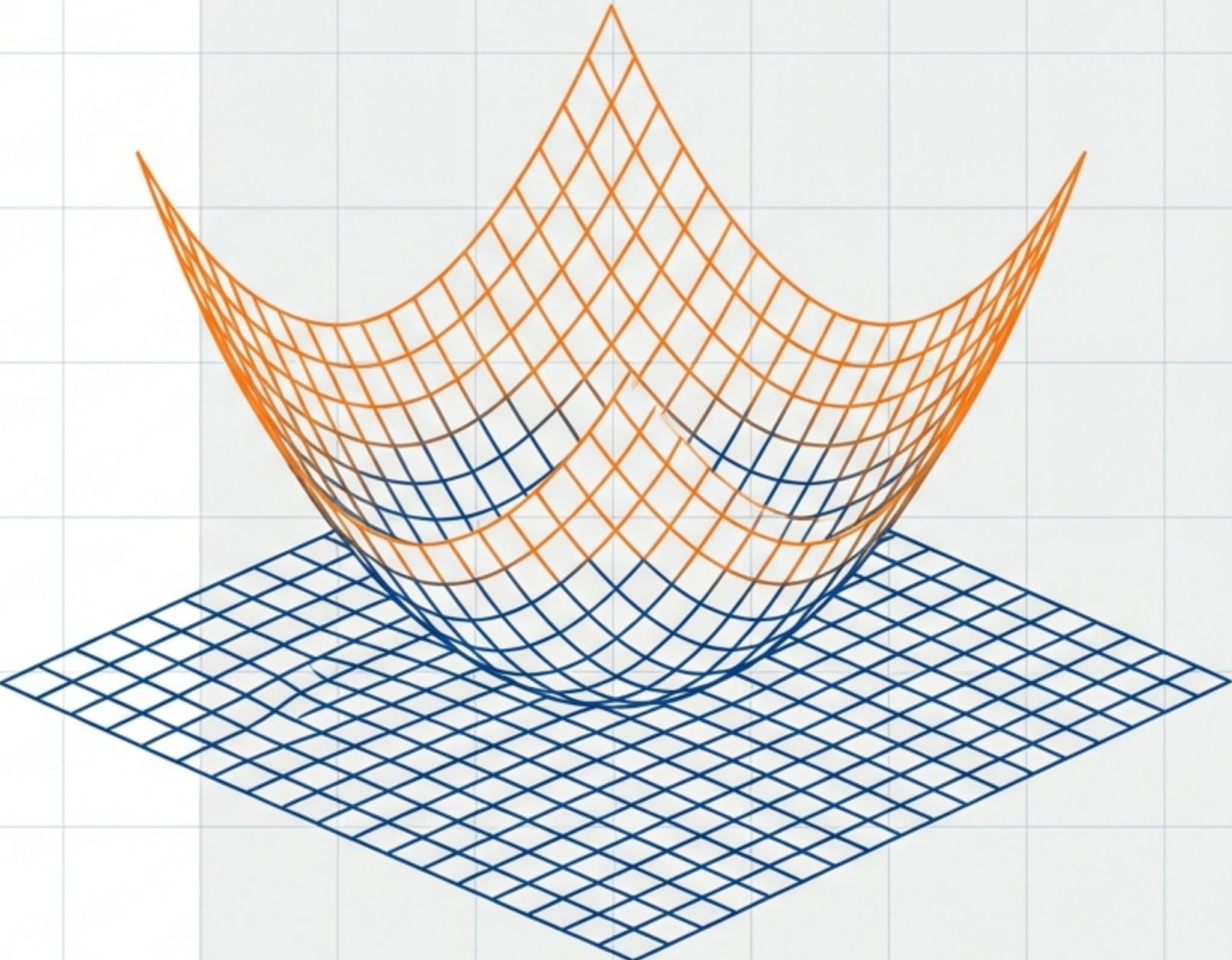


AI 在化工上的應用 : Unit 06

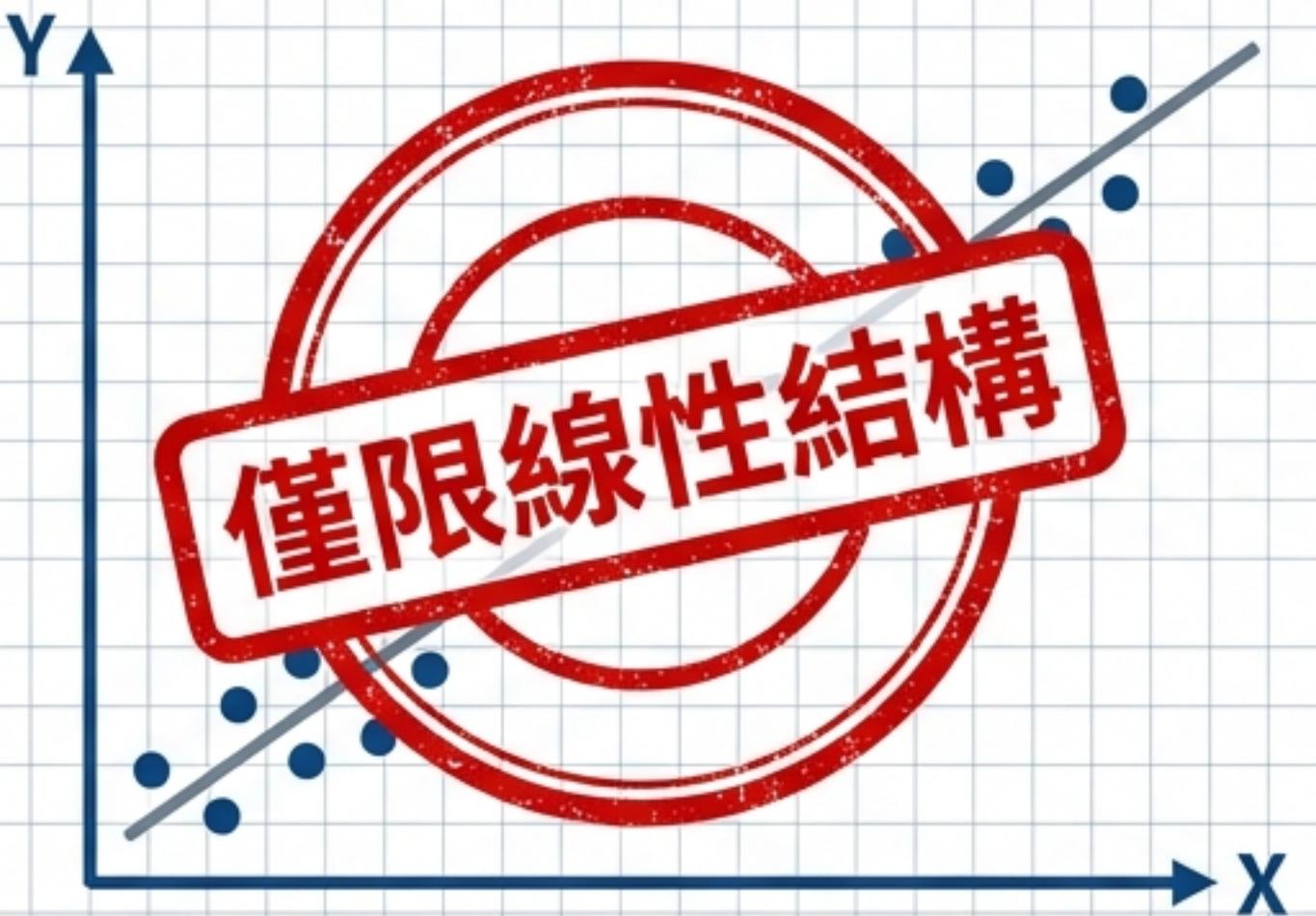
核主成分分析 (Kernel PCA)

駕馭非線性製程數據



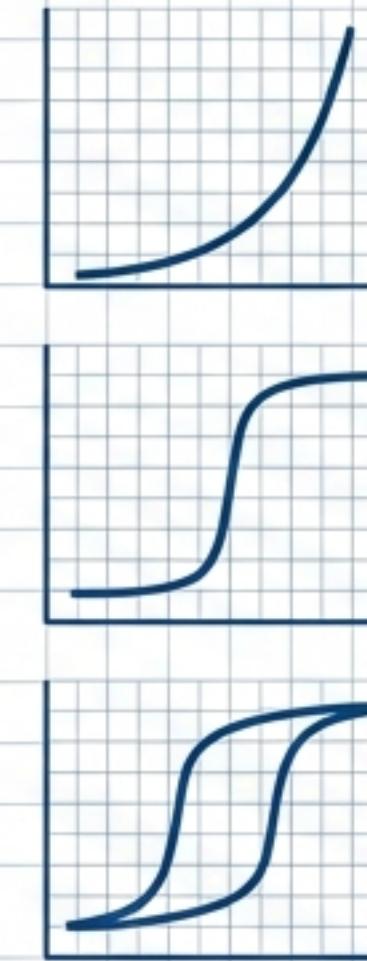
為什麼需要 Kernel PCA？化工世界的「非線性」本質

傳統 PCA 的侷限



- 僅能捕捉數據的線性相關性
- 面對彎曲、螺旋結構時解釋力不足

化工製程的真實世界

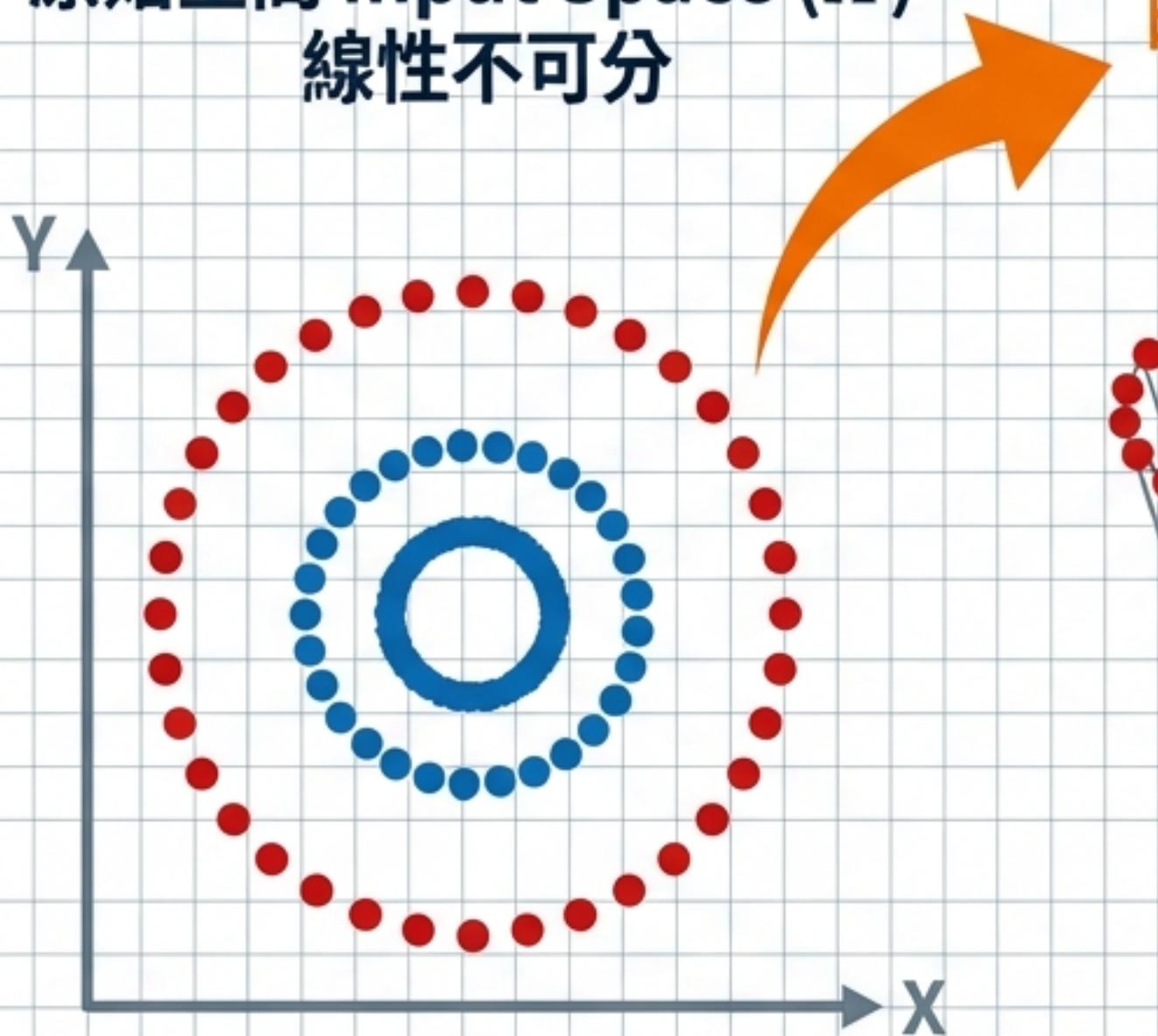


- 高度非線性：反應速率 $k = A e^{-E/RT}$
- 複雜交互作用：相變、吸附、動態響應

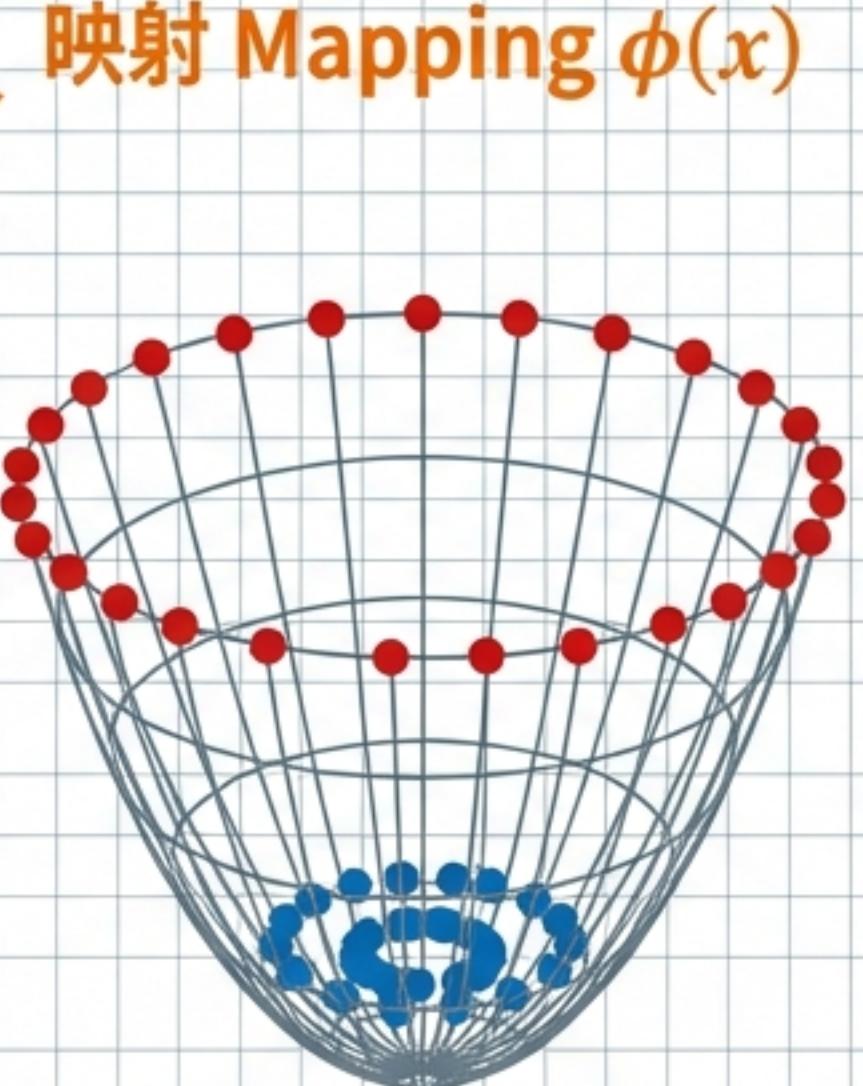
結論：我們需要一把能處理「彎曲空間」的尺

核心概念：高維空間的線性化

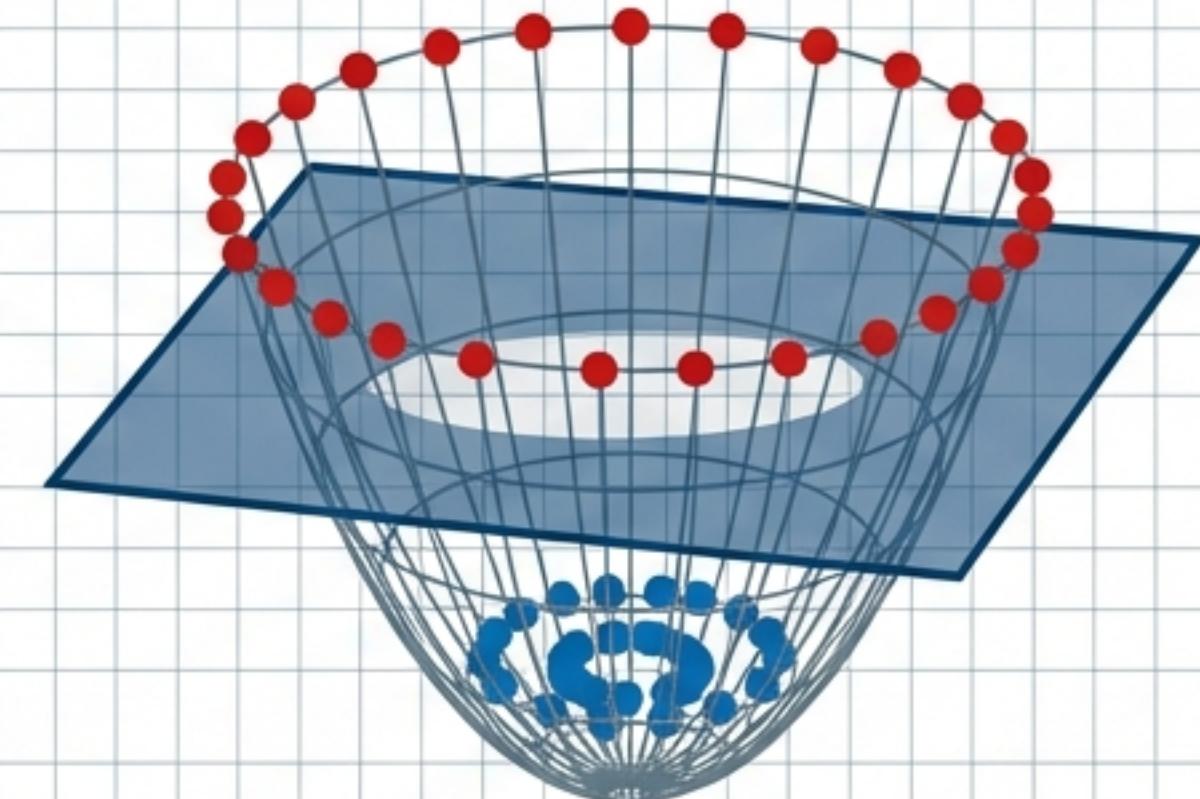
原始空間 Input Space (X)
線性不可分



映射 Mapping $\phi(x)$

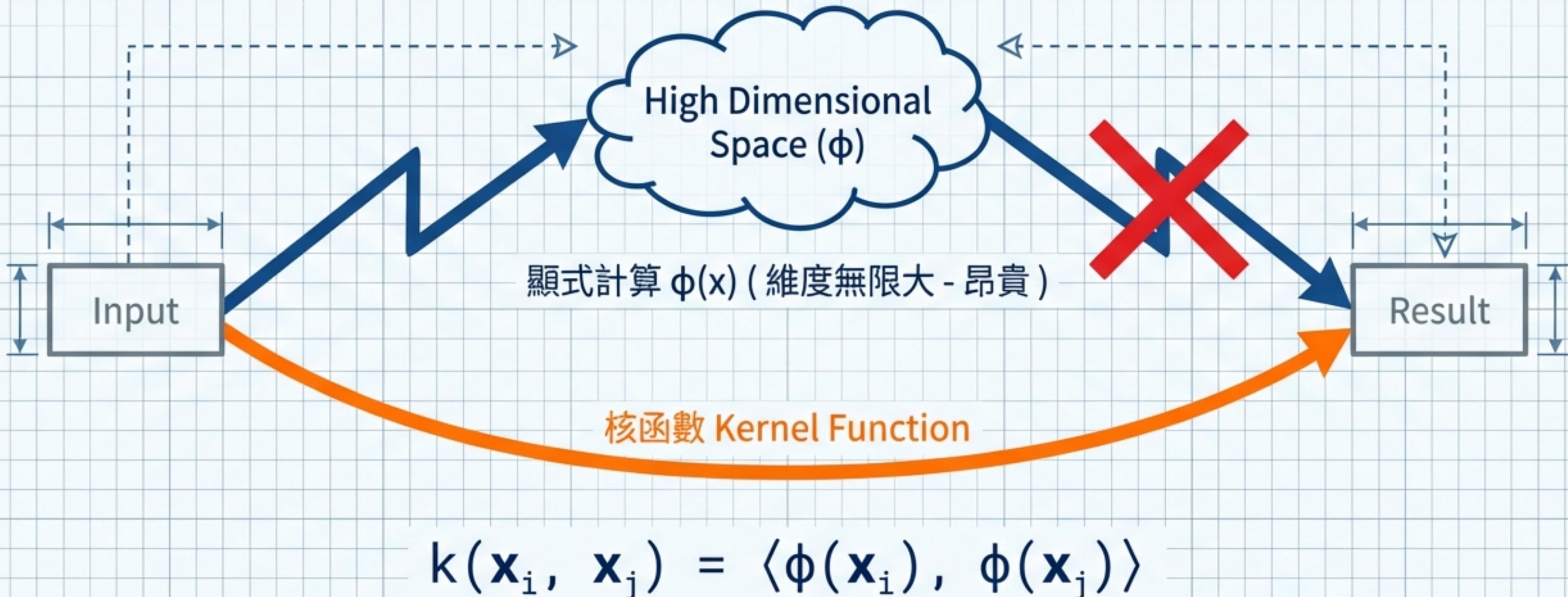


特徵空間 Feature Space (F)
線性可分



核心理念：透過非線性函數將數據映射至高維空間，使原本複雜的結構變得平坦可分。

解密 AI 術語：核技巧 (The Kernel Trick)



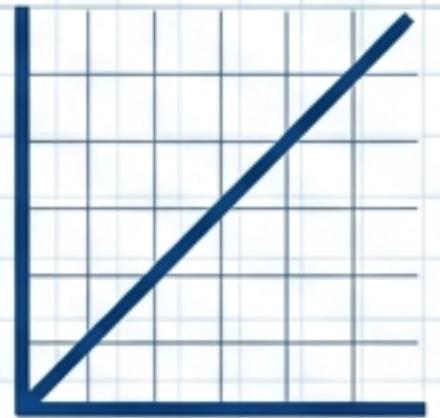
定義：我們不需要知道 $\phi(x)$ 是什麼，只需要知道如何計算特徵空間的「內積」。

優勢：將複雜的高維計算簡化為原始空間的函數運算 (隱式映射)。

選擇正確的工具：常見核函數 (Kernel Functions)

線性核 (Linear Kernel)

$$k(x, y) = x^T y$$

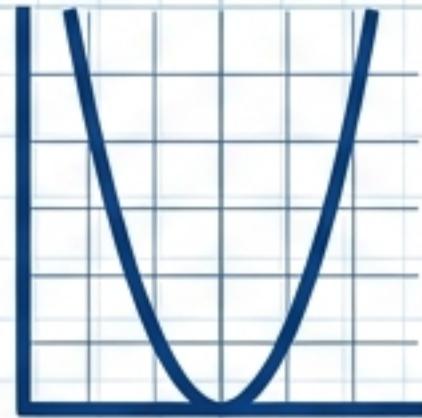


Noto Sans TC Regular

等同於傳統 PCA。適用於已知
線性關係的基準測試。

多項式核 (Polynomial Kernel)

$$k(x, y) = (\gamma x^T y + c_0)^d$$

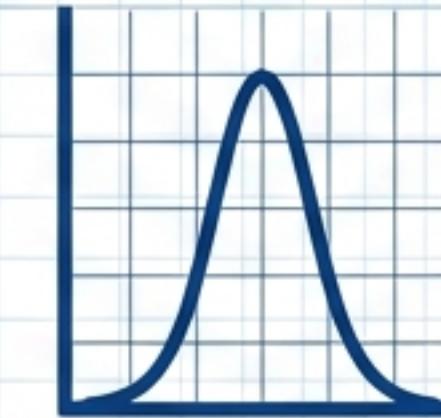


Noto Sans TC Regular

捕捉高階交互作用。適用於反
應動力學擬合 ($d = 2, 3$)。

徑向基函數核 (RBF / Gaussian) [👑 首選]

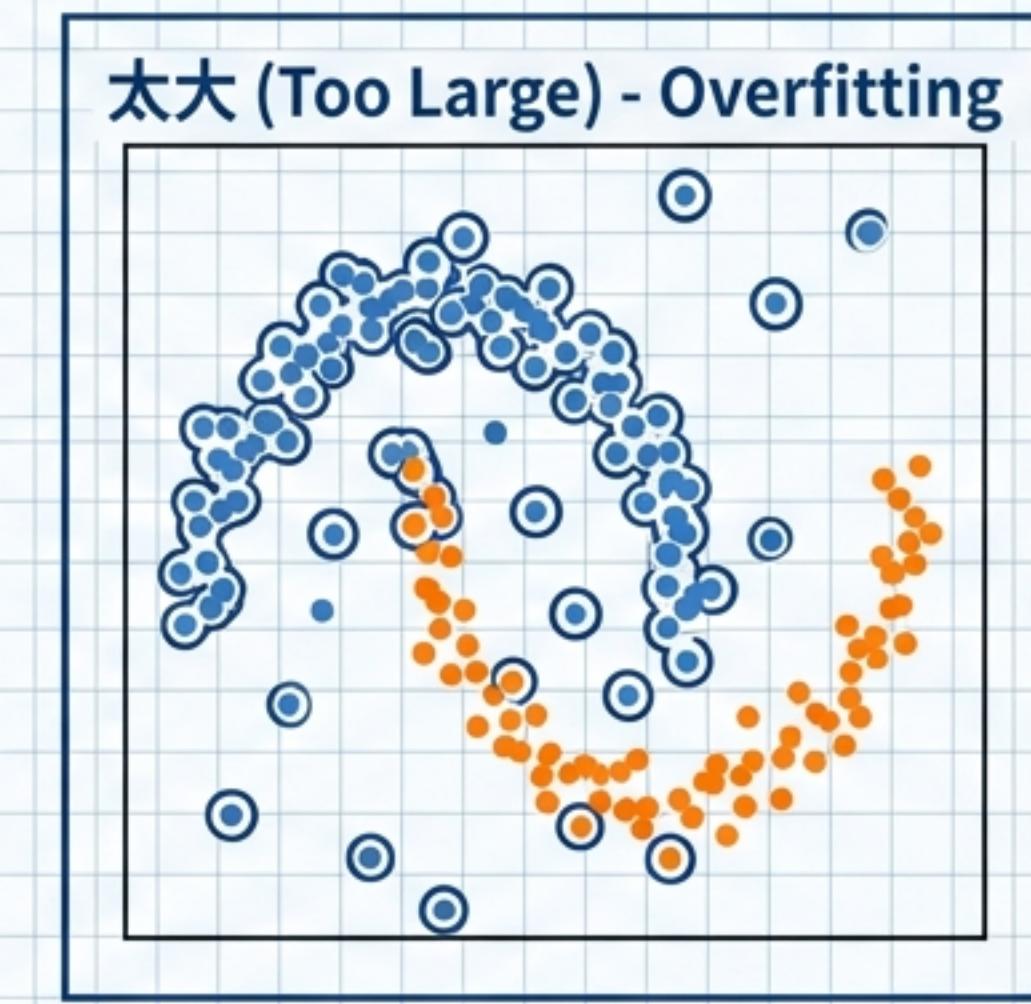
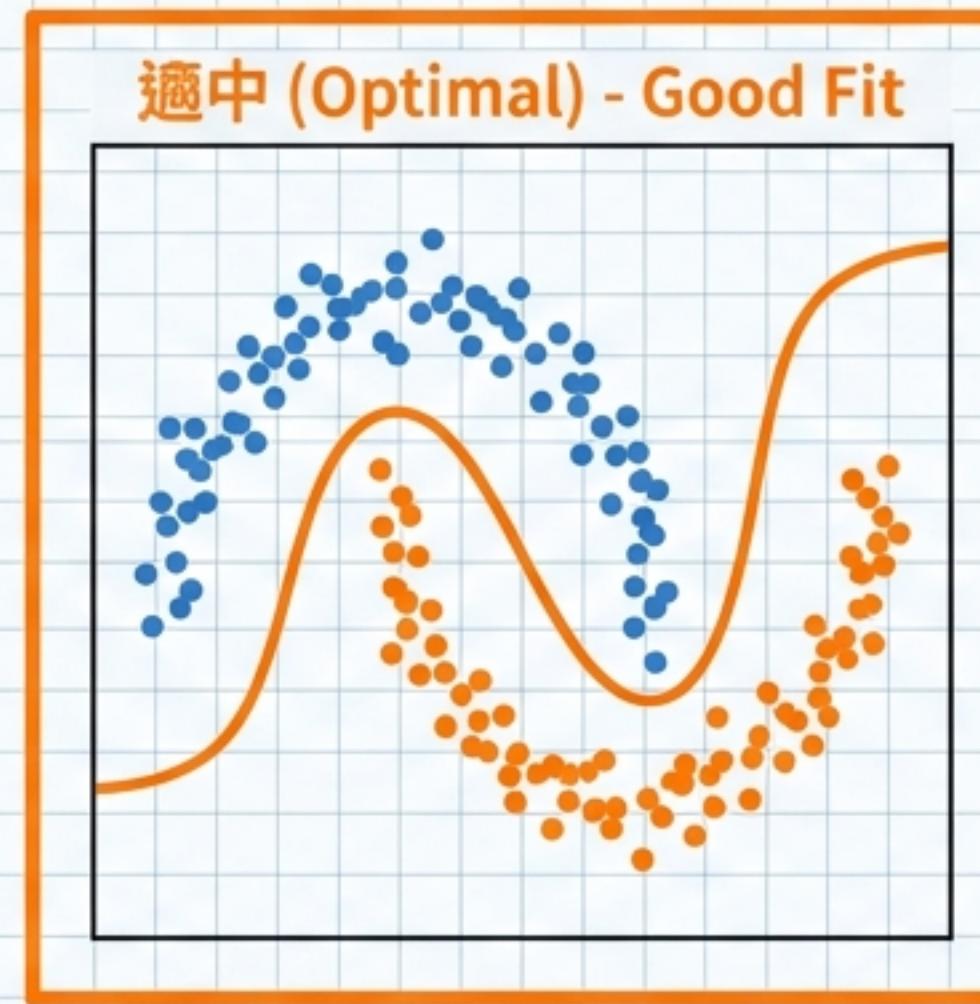
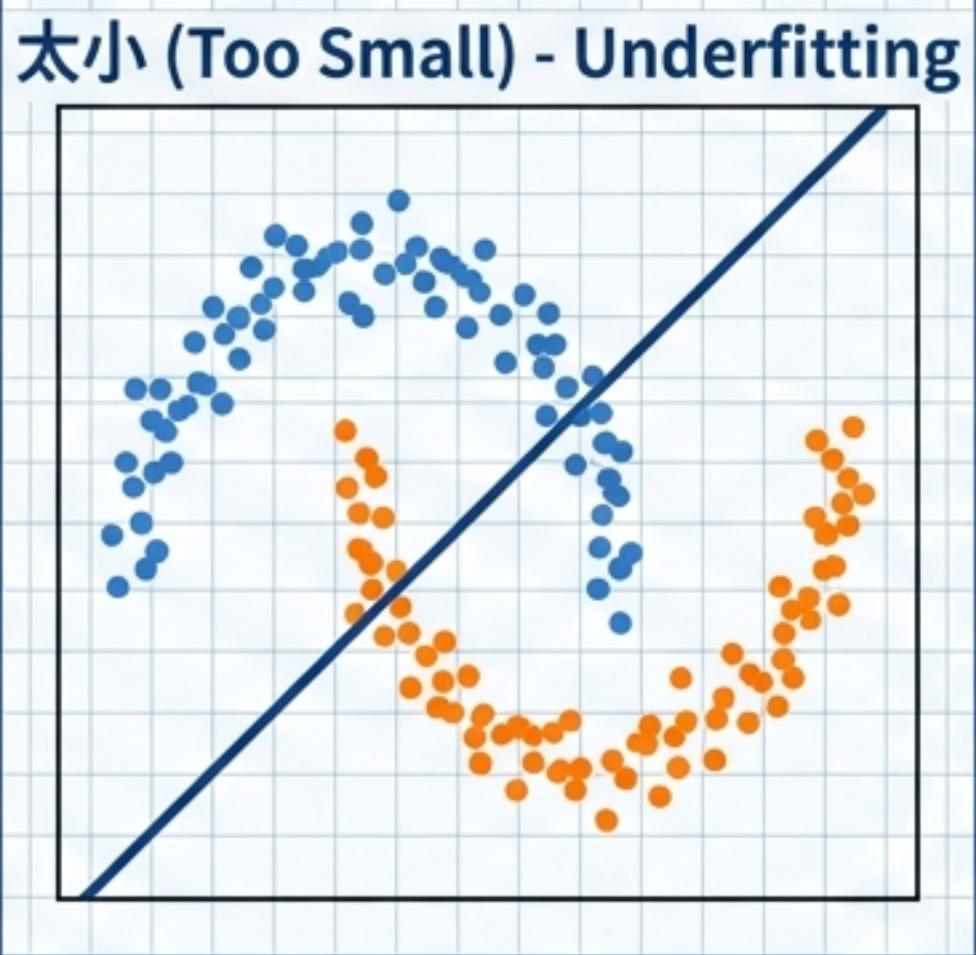
$$k(x, y) = \exp(-\gamma ||x - y||^2)$$



Noto Sans TC Regular

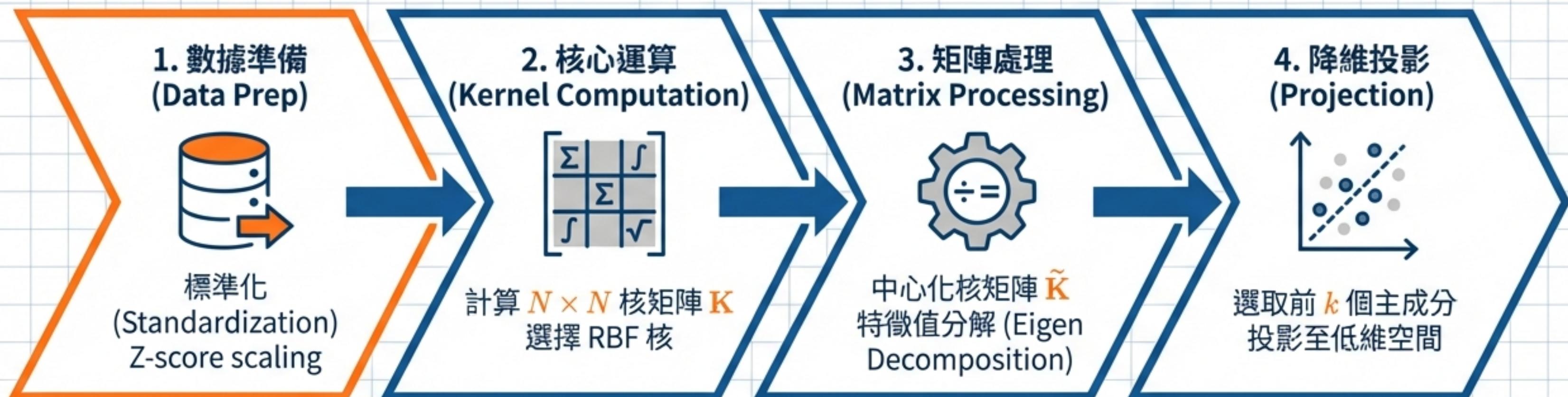
映射至無限維空間，局部性強。
化工非線性監控的最通用選擇。

關鍵參數調校：Gamma (γ) 的影響



γ 決定了單一數據樣本的影響範圍。建議使用 Grid Search 配合 Cross-Validation 尋找最佳值。

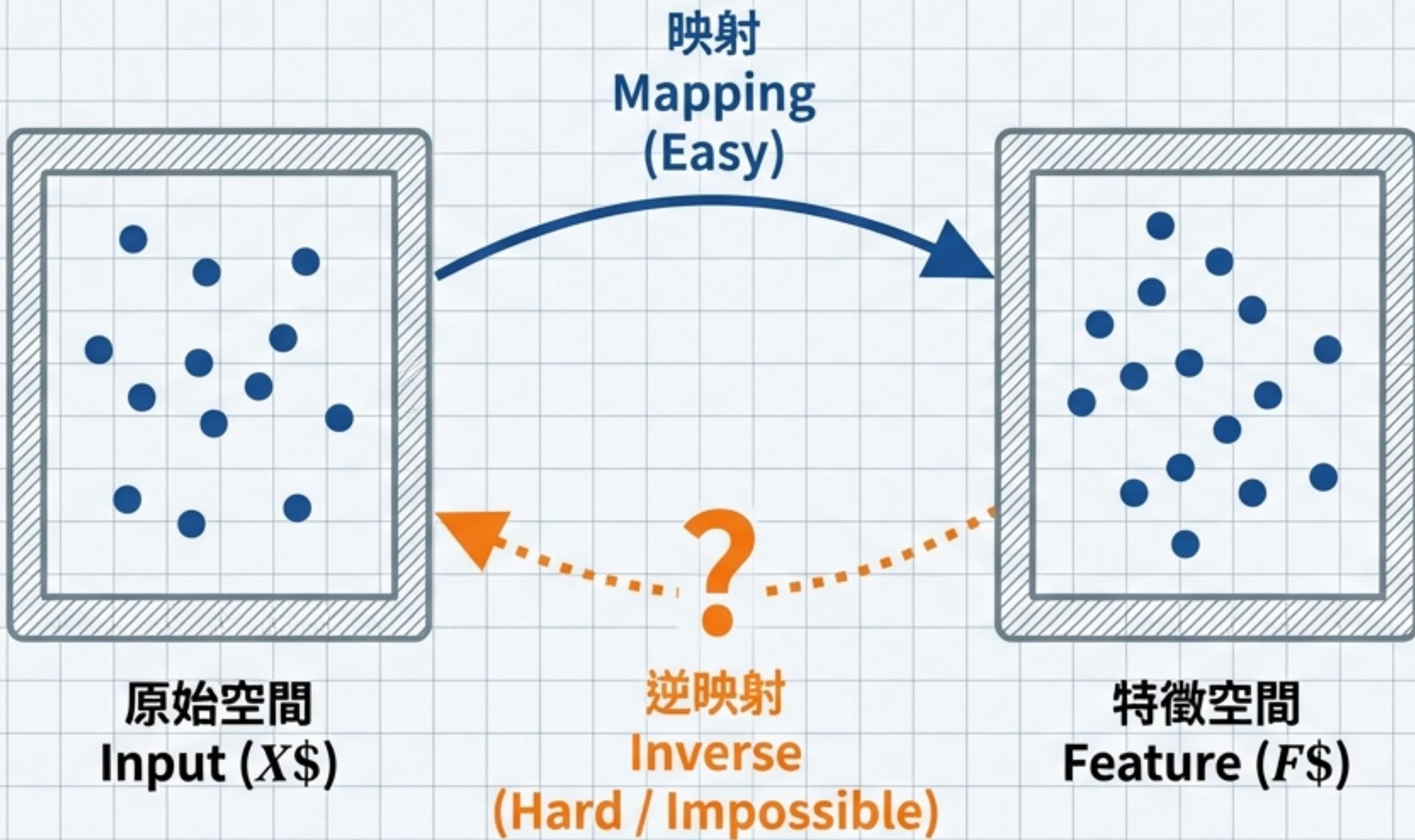
實作流程：從數據到非線性特徵



典範對決：傳統 PCA vs. Kernel PCA

特性 Feature	傳統 PCA (Linear)	Kernel PCA (Non-Linear)
映射方式	線性投影 (旋轉座標)	非線性映射 (彎曲空間)
適用數據	簡單混合、穩態操作	複雜反應、相變、動態製程
可解釋性	高 (Loadings 有物理意義)	低 (特徵隱晦)
計算成本	低 ($O(p^2)$)	高 ($O(n^2)$) – 隨樣本增長
超參數	無	核函數類型, Gamma
主要用途	概觀分析、貢獻度診斷	複雜分群、異常檢測

現實世界的挑戰：Pre-image 問題



Pre-image 問題：我們可以將數據映射到高維空間並降維，但很難「逆映射」回原始物理空間。

影響：Kernel PCA 擅長「檢測」與「分類」，但較不適合用於「數據重建」或直觀的變數解釋。

AI 在化工全生命週期的價值：KPCA 應用



製程監控 (Monitoring)

偵測反應失控前的非線性
徵兆。

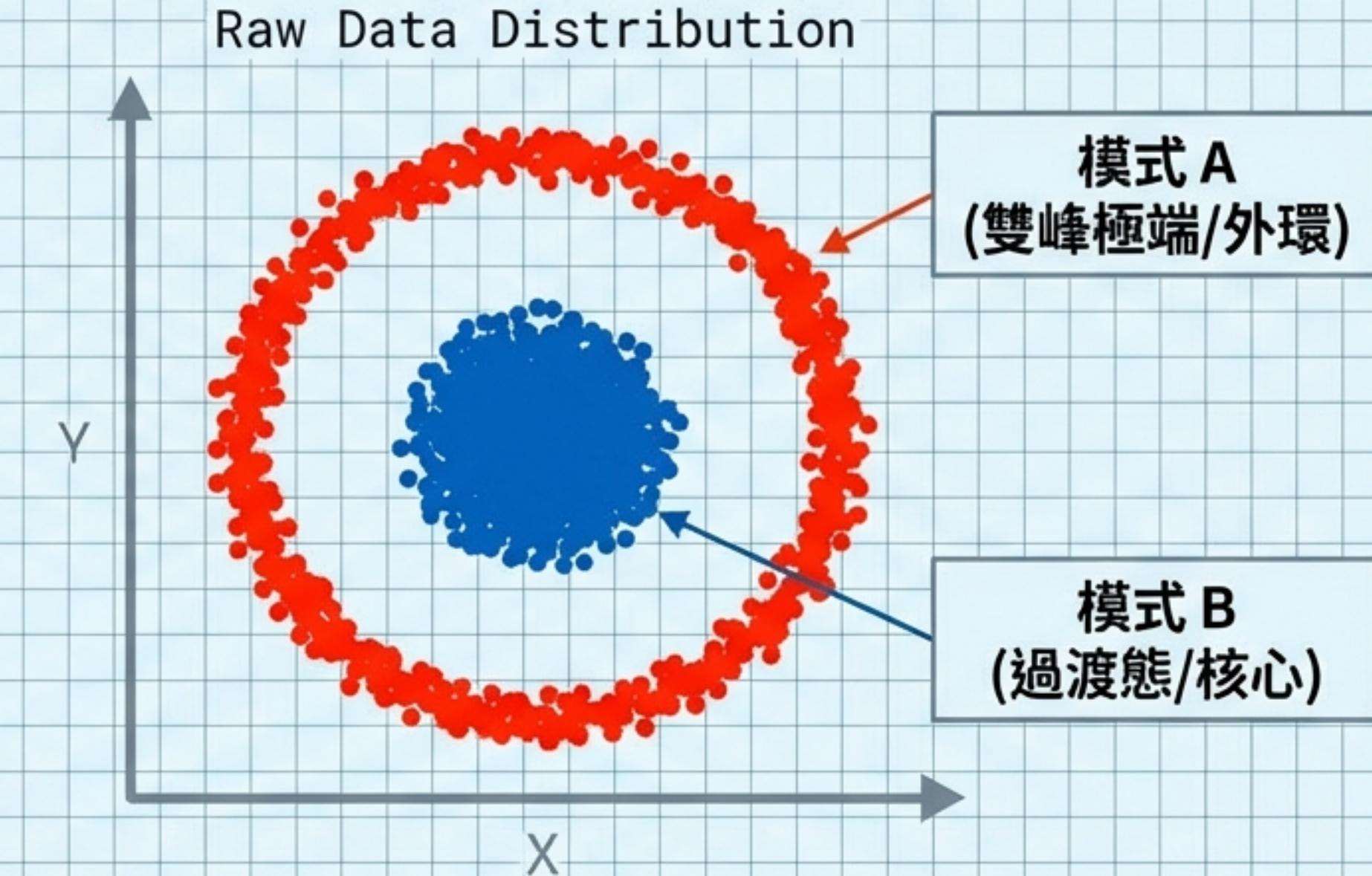
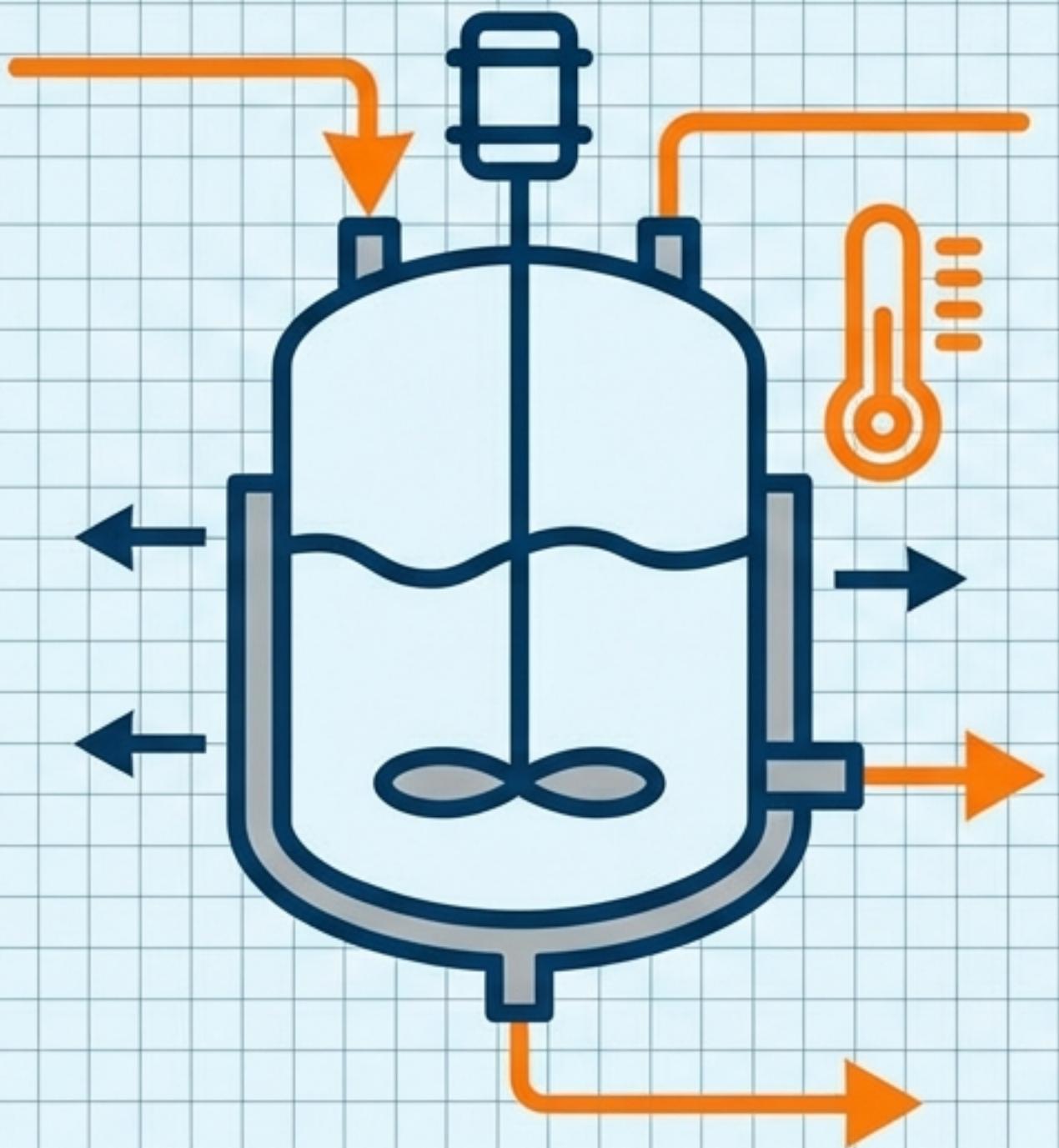
軟感測器 (Soft Sensors)

作為特徵提取器，處理多感
測器間的非線性共線性。

模式識別 (Pattern Recognition)

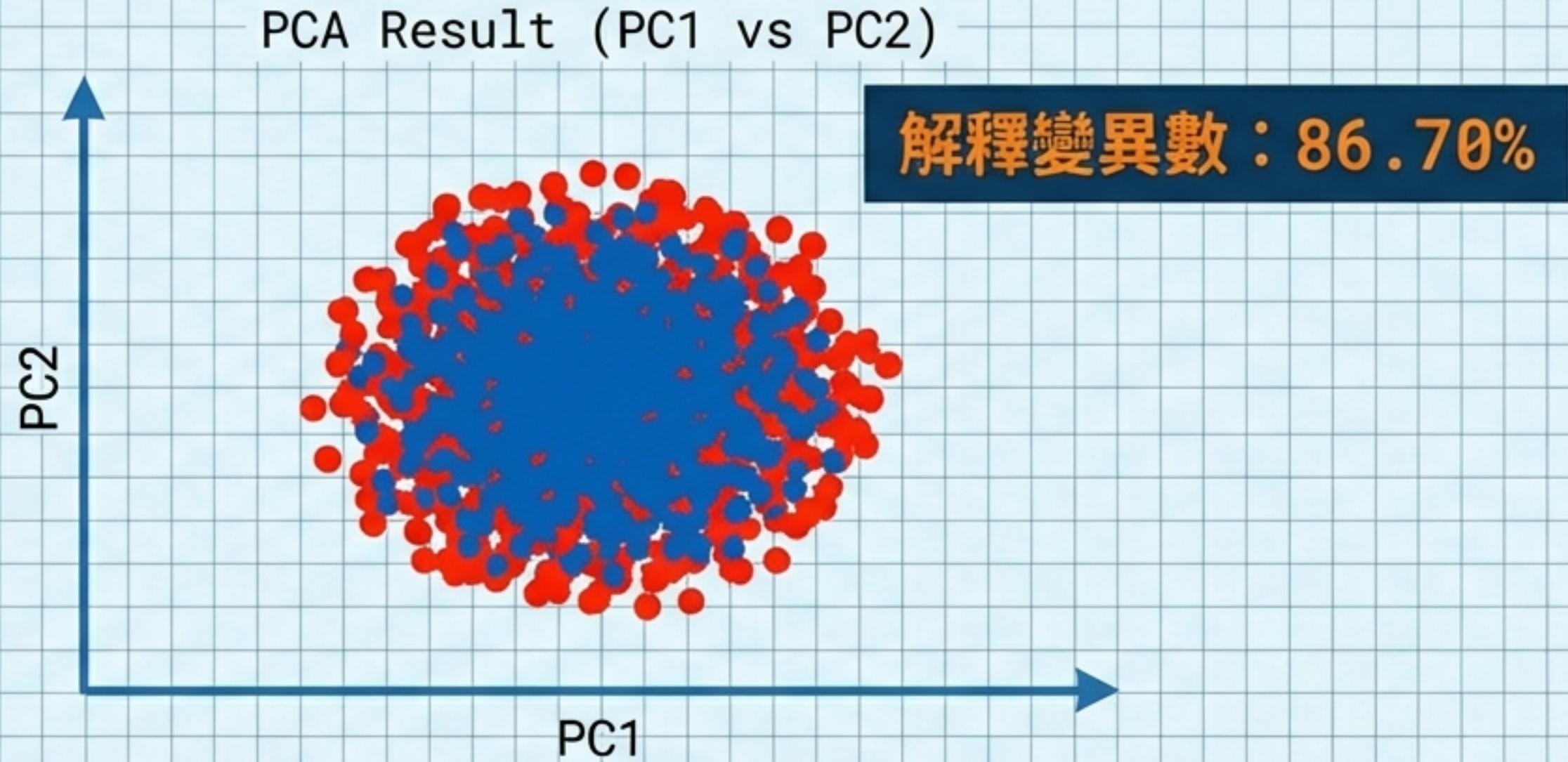
區分複雜的操作模式（如：
不同產品等級）。

實作範例：化工反應器的多模式操作



挑戰：典型的「同心圓」結構。在 X 軸與 Y 軸的投影上，兩者完全重疊，線性 PCA 無法區分。

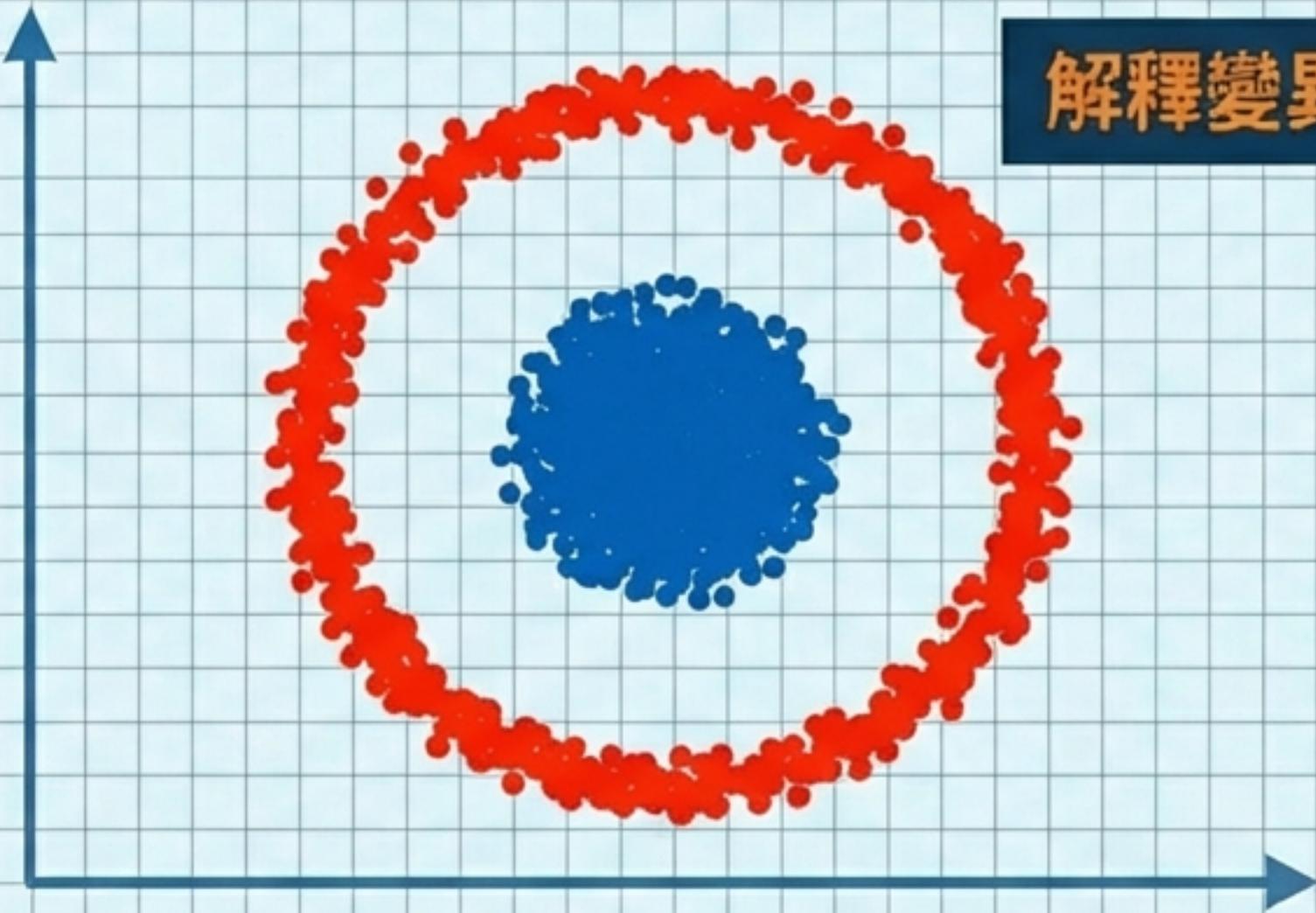
傳統 PCA 的困境：高變異數的陷阱



警示：高解釋變異數 ≠ 良好的分類效果。
PCA 僅捕捉了數據的擴散範圍，卻遺失了結構資訊。

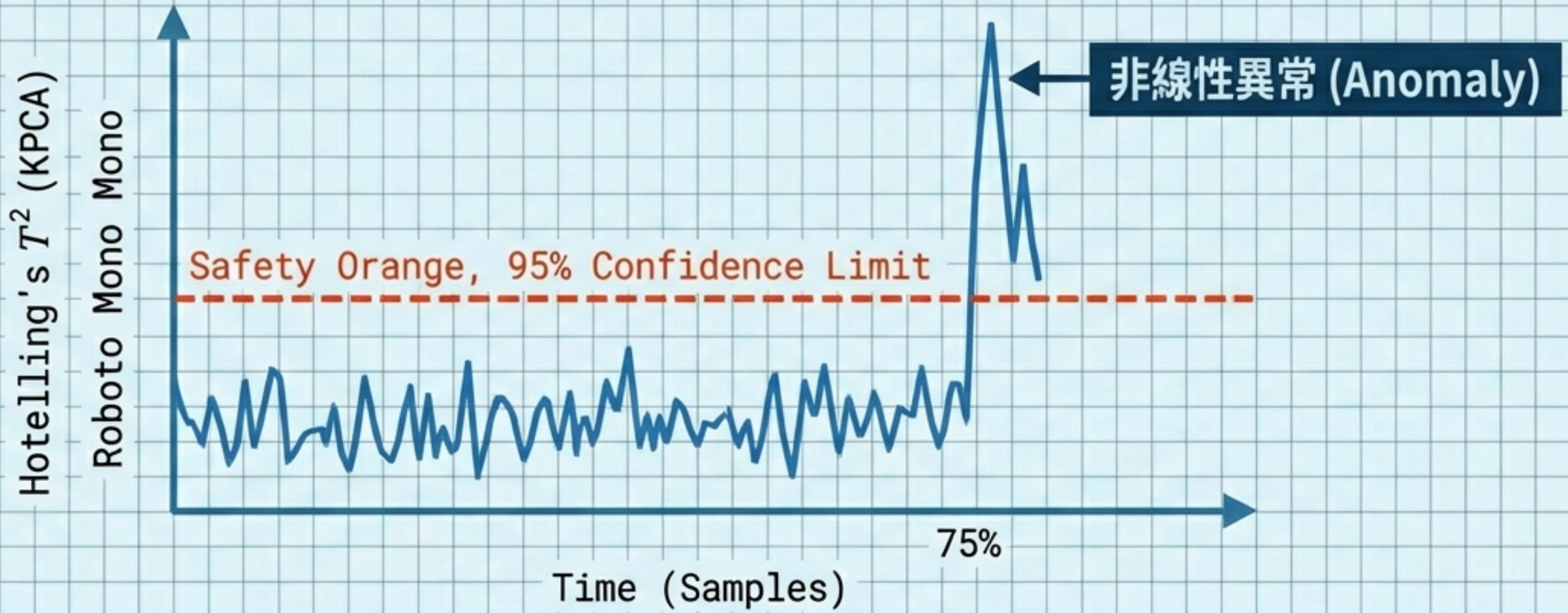
Kernel PCA 的威力：看見看不見的結構

Kernel PCA Result (RBF Kernel)



結論：利用 RBF 核，我們成功在特徵空間中分離了「核心」與「外環」。

進階應用：非線性製程監控



監控邏輯：將新數投影至 KPCA 空間，計算統計量。
相比線性 PCA，此方法能偵測出違反變數間非線性關係的微妙故障。

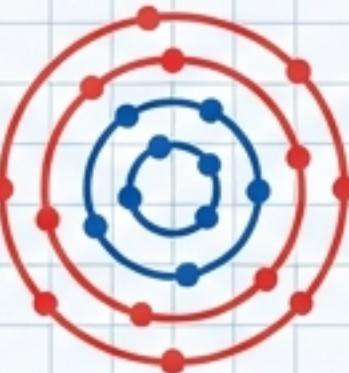
結語：從簡單開始，向複雜邁進

優先使用 PCA



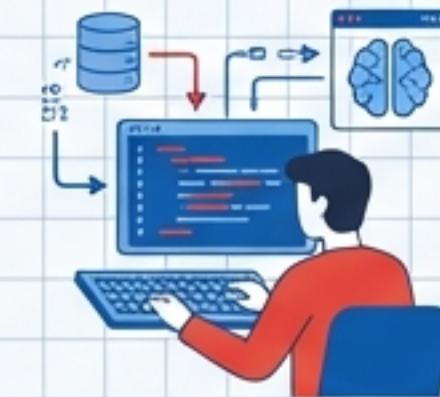
如果線性方法有效，不要使其複雜化。簡單即是美。

識別非線性



當 PCA 失效或面對複雜反應動力學時，引入 KPCA。

動手實踐 (Hands-On)



使用 Python scikit-learn。注意樣本數 ($N < 10,000$) 以免計算過載。

下一步：前往 Unit 07 - 深度學習基礎 (Deep Learning Foundations)