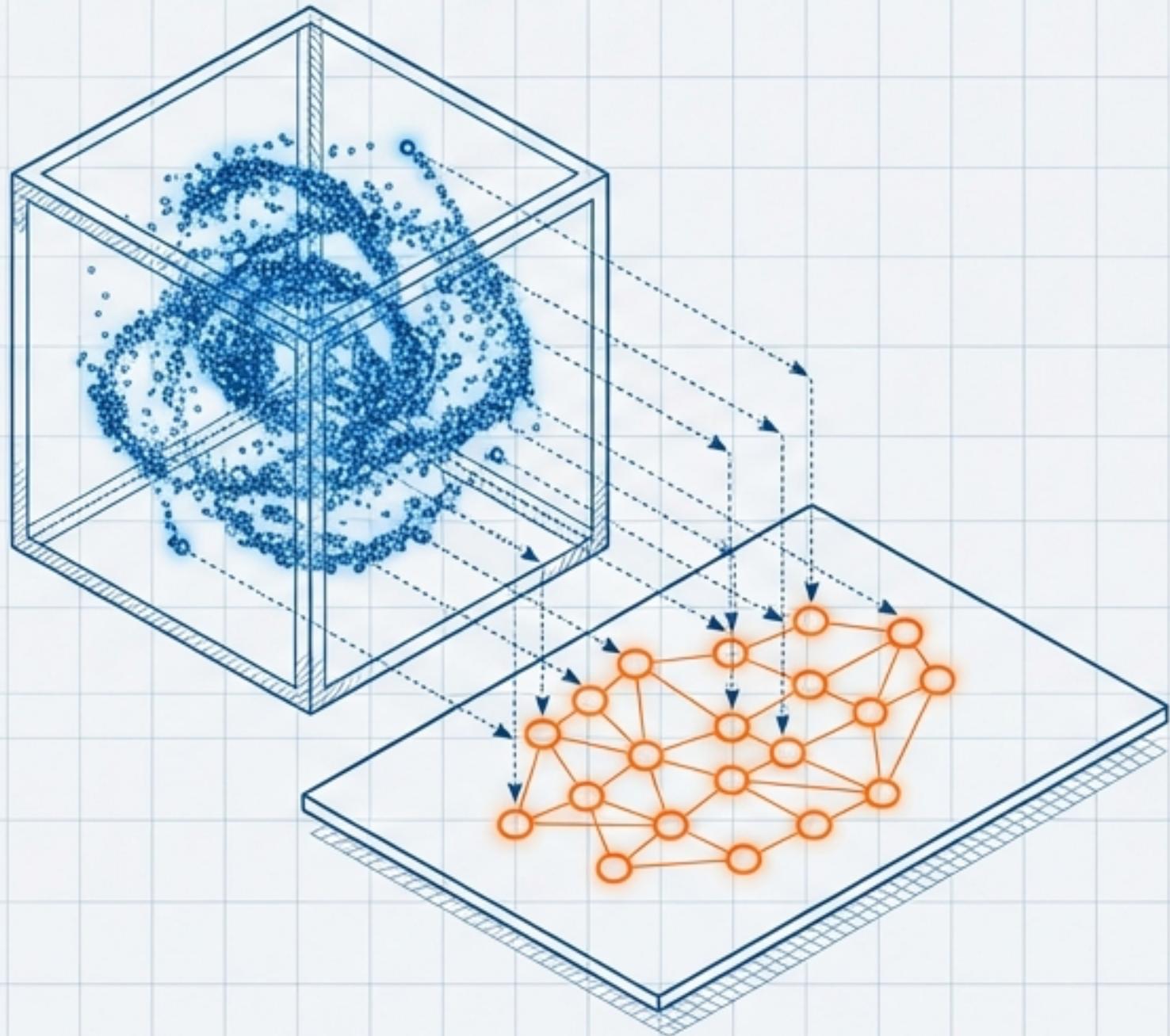


AI 在化工上的應用：降維技術

Unit 06 課程概論與導讀

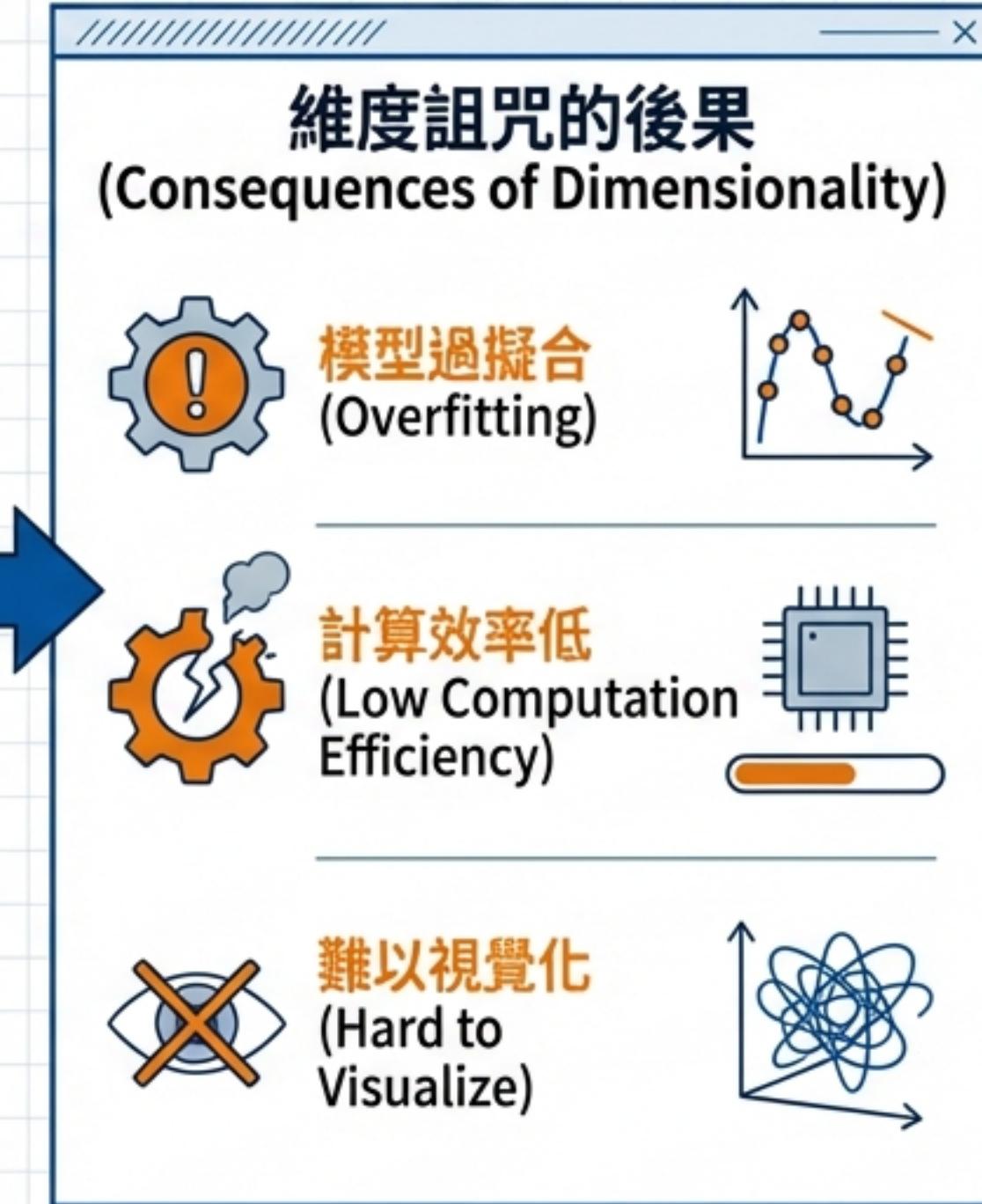
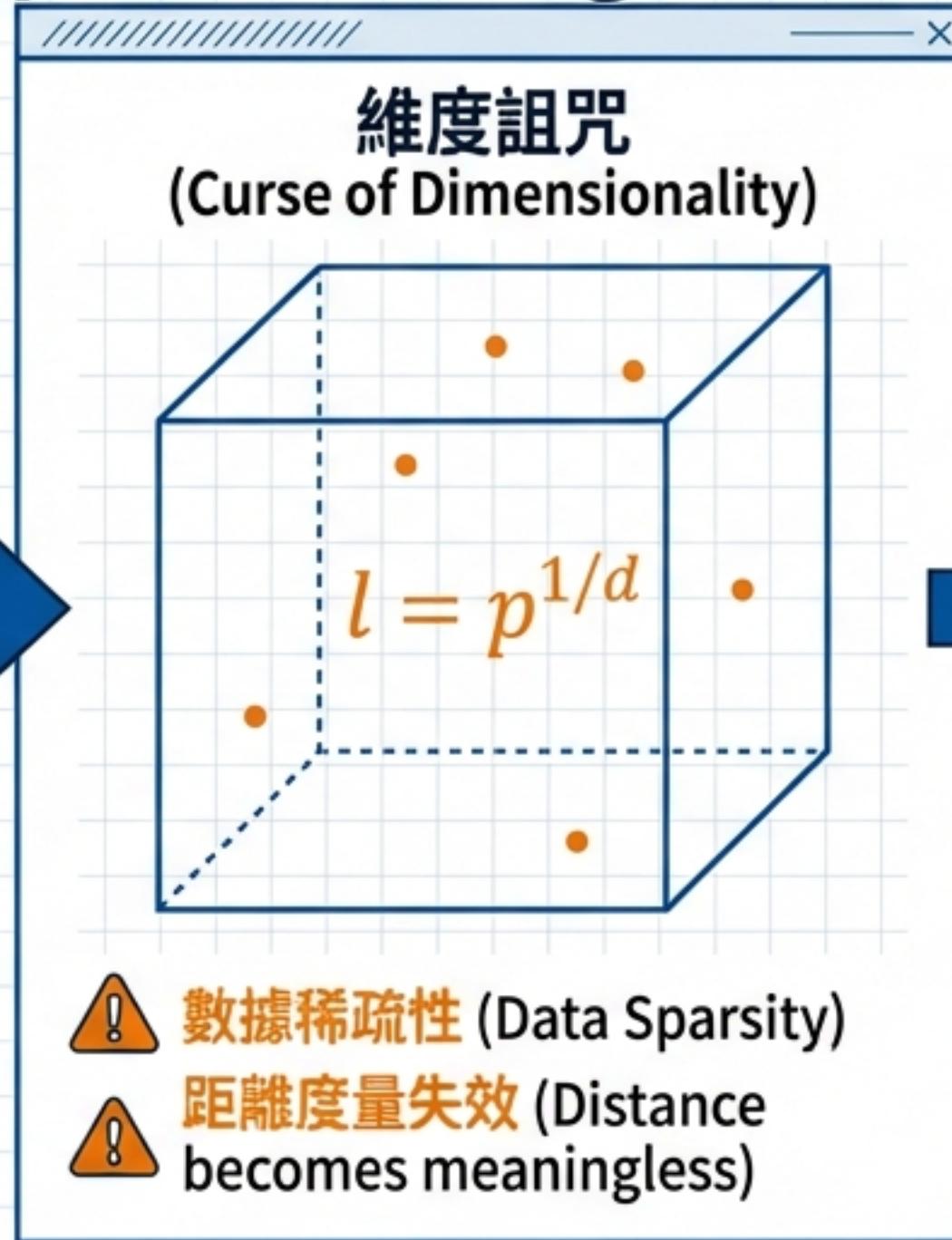
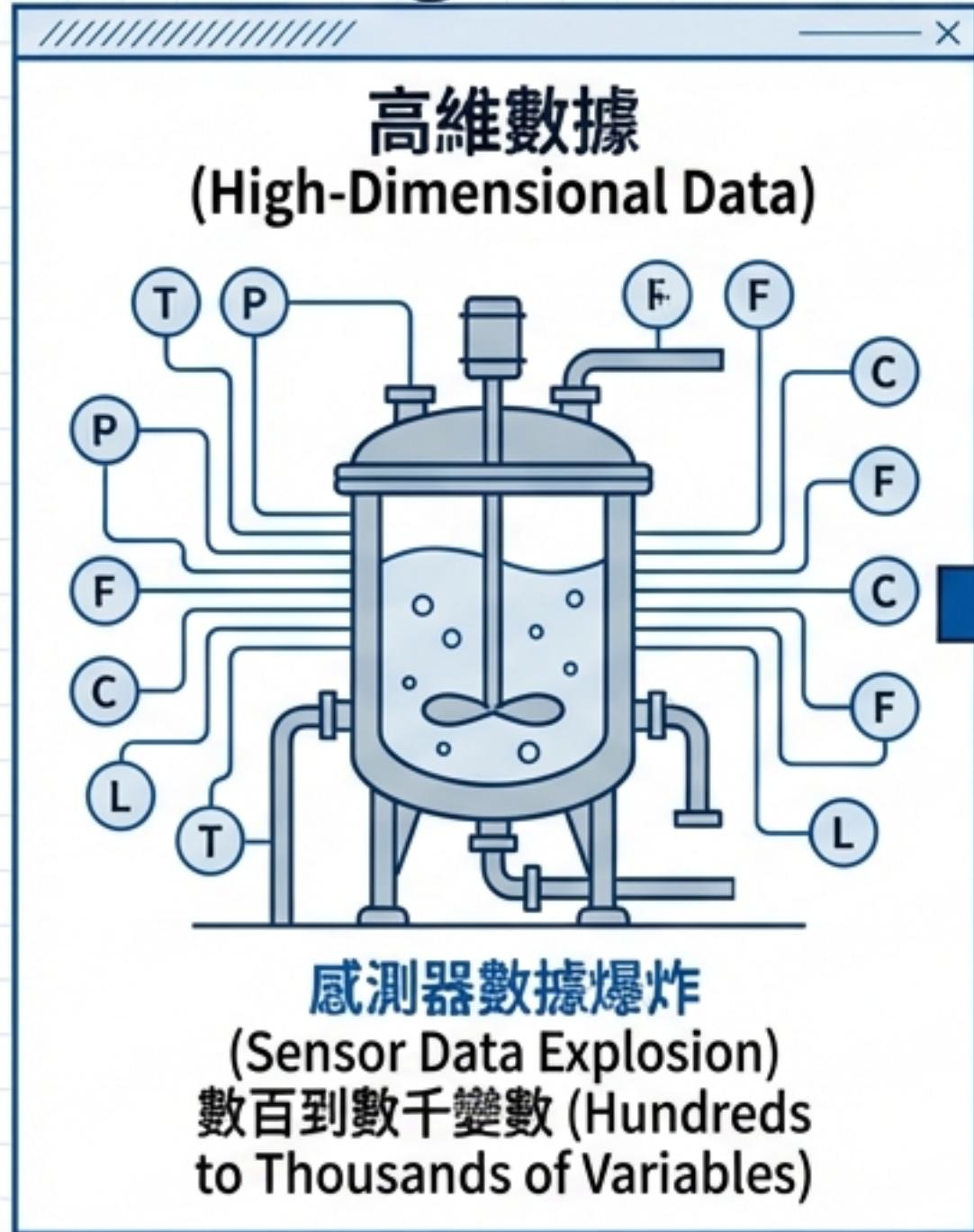
AI Applications in ChemE: Dimensionality Reduction



授課教師：莊曜禎 助理教授

學期：114學年度第2學期

維度詛咒：高維數據的挑戰與後果 (The Curse of Dimensionality: Challenges and Consequences of High-Dimensional Data)

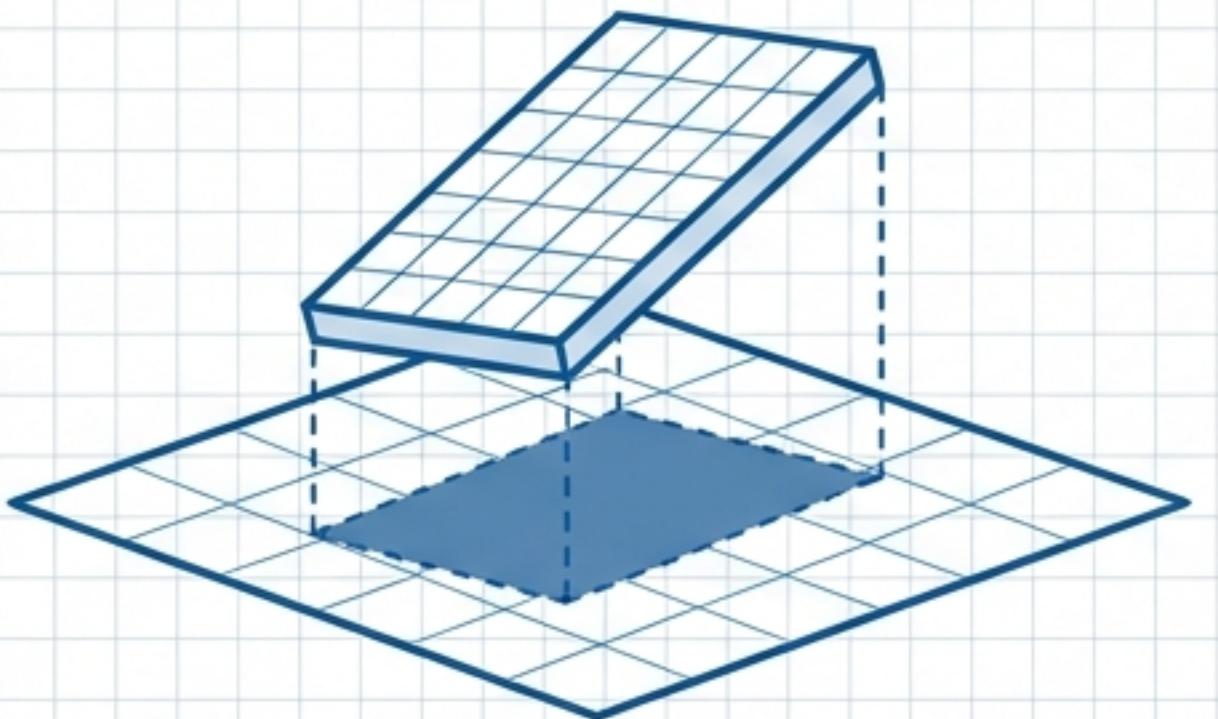


Key Insight: **More Data ≠ Better Information.** (更多數據不代表更多資訊)

核心概念：將高維資訊投影至低維空間

Mapping $f : \mathbb{R}^D \rightarrow \mathbb{R}^d, d \ll D$

線性降維 (Linear)

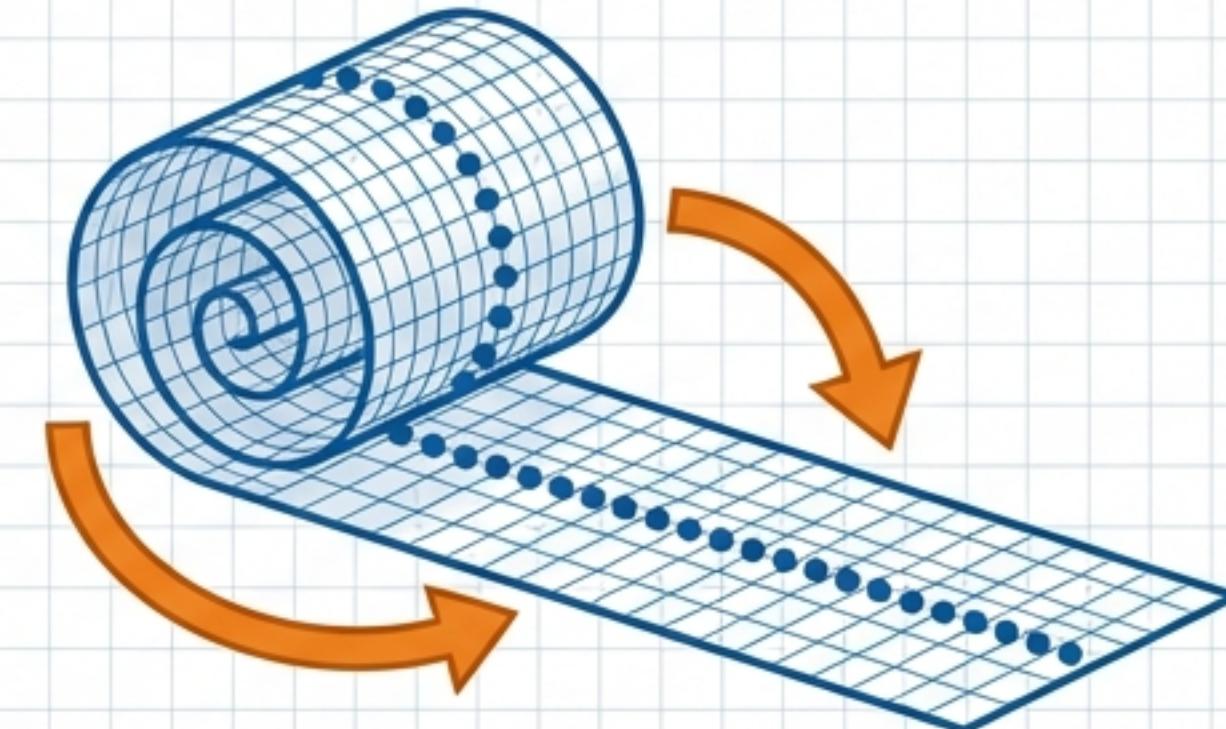


概念： $y = W^T x$

適用：簡單結構、全局模式

範例：PCA, LDA

非線性降維 (Non-Linear)



概念： $y = f(x)$

適用：複雜流形、非線性關係

範例：t-SNE, UMAP, Kernel PCA

ChemE Context: Transforming '50 Sensor Readings' into '1 Health Index'.

演算法地圖：化工領域的三大降維利器



PCA (The Workhorse)



類型：線性 (Linear)

特點：速度快、可解釋性

高

最佳應用：製程監控
(Process Monitoring),
MSPC

t-SNE (The Microscope)

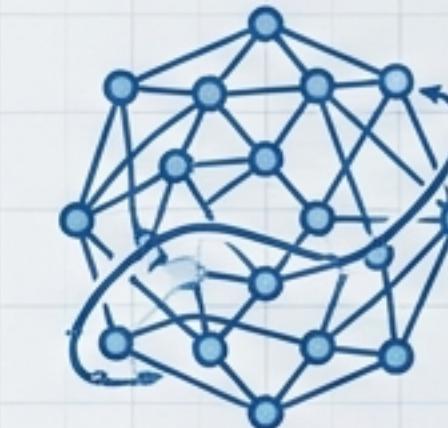


類型：非線性 (Non-Linear)

特點：視覺化群集極強

最佳應用：探索性分析
(EDA)，故障模式識別

UMAP (The Big Picture)

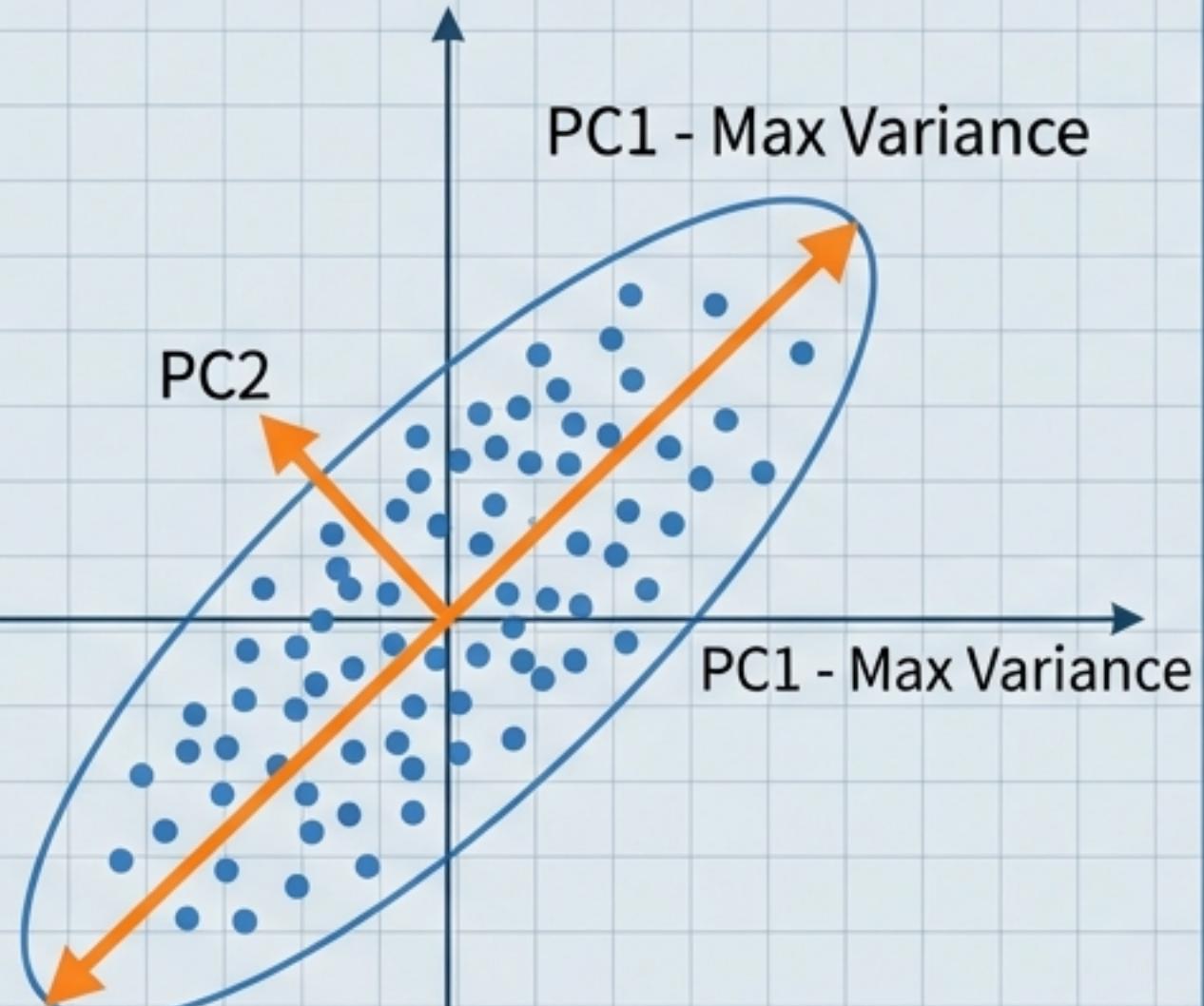


類型：非線性 (Non-Linear)

特點：平衡全局結構與速度

最佳應用：大規模數據，跨廠區比較

主成分分析 (PCA)：製程監控的黃金標準

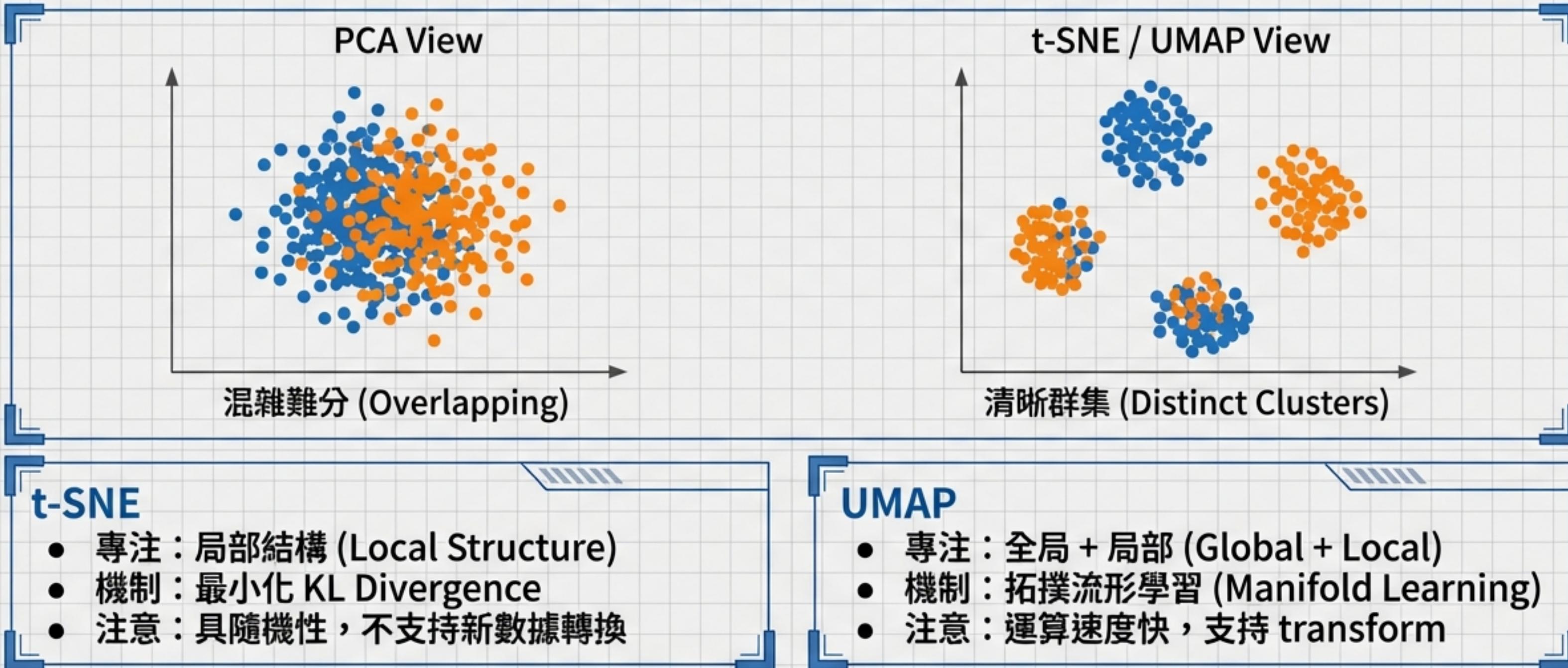


- 1. 數學核心 (Mathematical Core)
 - 目標：最大化變異數 Maximize Variance $\text{Var}(Xw)$
 - 特徵值 (λ)：代表解釋的變異程度
 - 投影： $y = W^T x$
- 2. 化工應用 (ChemE Application - MSPC)
 - Hotelling's T^2 ：測量正常操作範圍內的偏差
 - SPE (Squared Prediction Error)：測量模型外的異常偏差

保留變異數最大的方向 = 保留最多資訊

非線性方法：當線性假設失效時 (t-SNE & UMAP)

視覺化比較 (Visual Comparison)



應用案例：識別不同的操作模式 (如：開機、穩態、停機)

戰略決策：如何選擇合適的演算法？

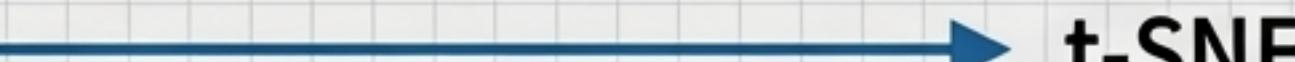
演算法 (Algorithm)	計算速度 (Speed)	可解釋性 (Interpretability)	全局結構 (Global)	新數據投影 (New Data)
PCA	快 (Fast)	高 (High)	佳 (Good)	Yes
Kernel PCA	慢 (Slow)	低 (Low)	佳 (Good)	Yes
t-SNE	慢 (Slow)	低 (Low)	差 (Poor)	No
UMAP	快 (Fast)	中 (Medium)	佳 (Good)	Yes

Decision Guide (決策指南)

需要解釋物理意義？



單純做視覺化/找群集？

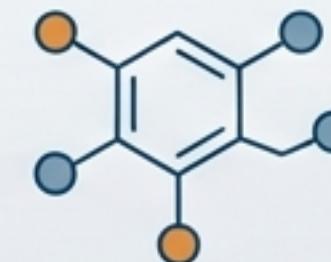


數據量巨大且需保留結構？



降維在化工全生命週期的價值

研發 (R&D)



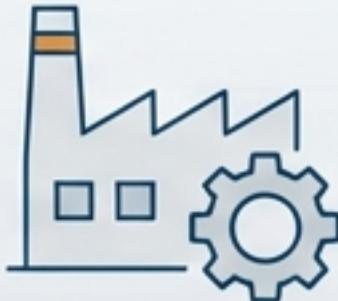
- 觸媒篩選
(Catalyst Screening)
- 高維性質視覺化

製程設計 (Design)



- 反應路徑分析
- 複雜動力學降維
(Kernel PCA)

工廠營運 (Operation)



- 故障偵測
(Fault Detection)
- 訊號增強與降噪
(PCA)

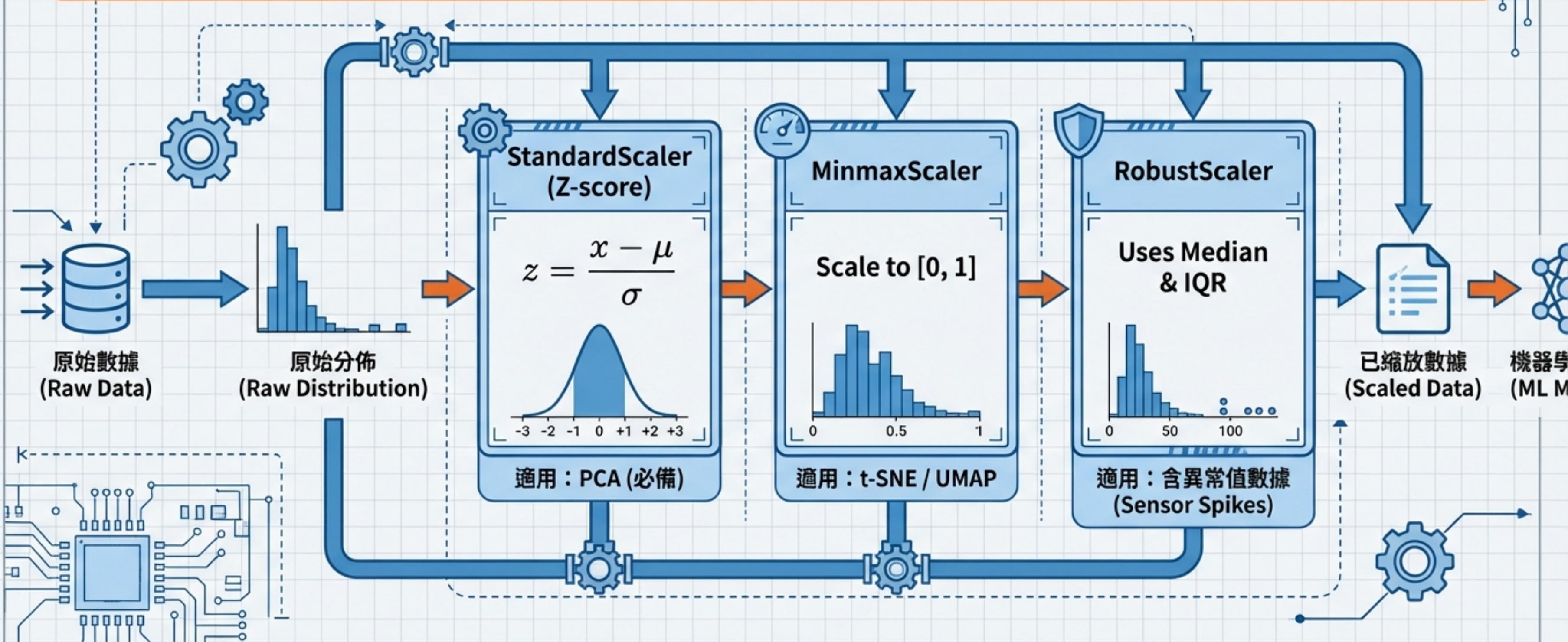
分析與優化 (Analysis)



- 批次軌跡分析
(Batch Trajectory)
- 黃金批次 vs 異常批次比較

工作流程 Part 1：數據前處理 (Preprocessing)

⚠ 原則：Garbage In, Garbage Out (垃圾進，垃圾出) - Scale Matters!

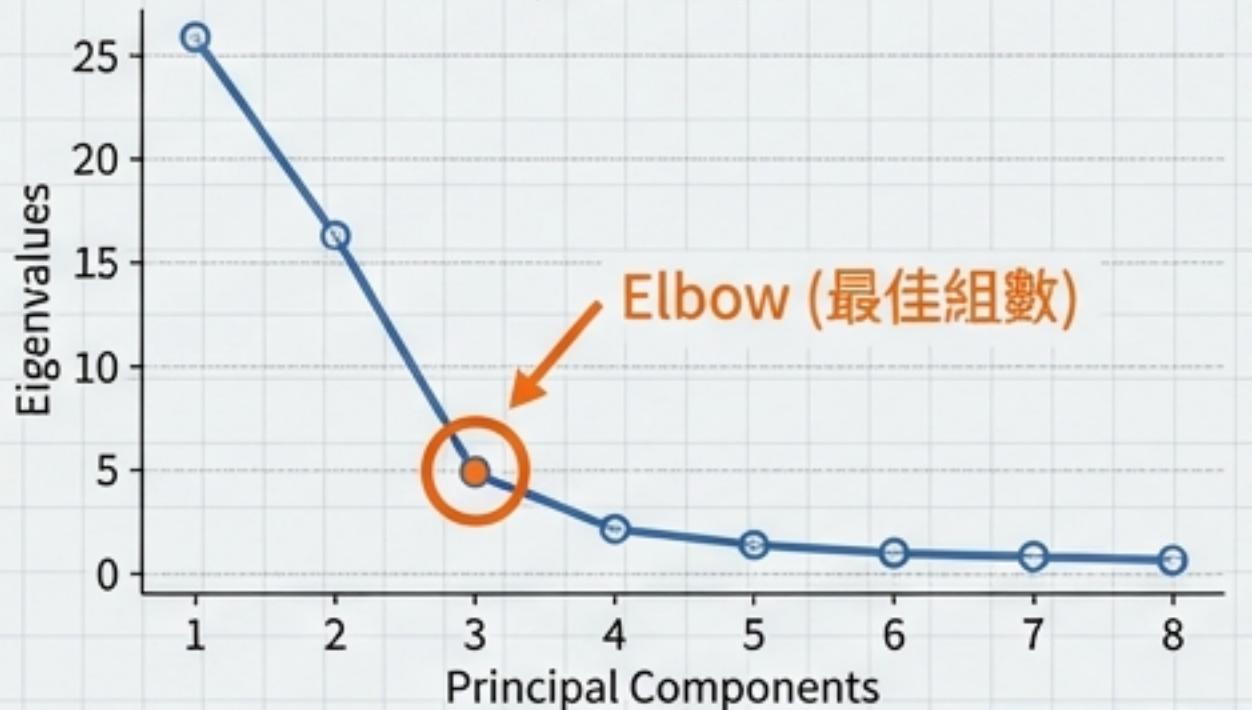


評估指標：如何知道降維效果好壞？



PCA 評估 (PCA Metrics)

Scree Plot
(Line chart)



解釋變異量 (Explained Variance Ratio) $> 85\%$



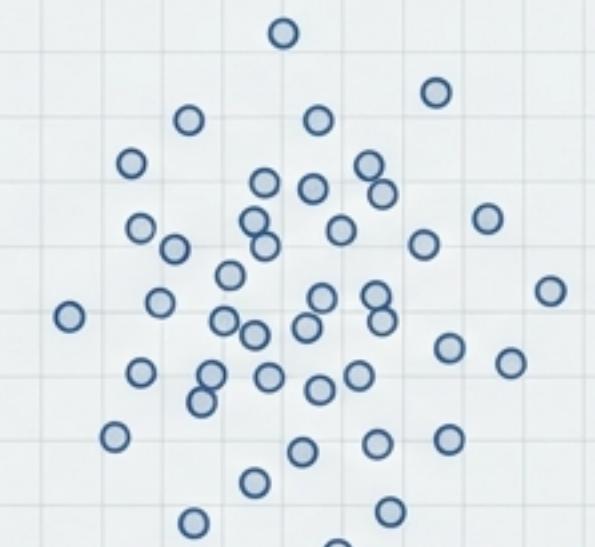
重建誤差 (Reconstruction Error): Low is better



分群評估 (Clustering Metrics)



Good Clustering
(Tight)



Bad Clustering
(Spread Out)



Silhouette Score : 衡量群集分離度



Trustworthiness : 衡量局部結構保持度

現實世界的挑戰 (Real World Challenges)

可解釋性 (Interpretability)



Problem: "Why is PC1 high?"

Solution: "Loading Plots
(負荷圖) - 追溯回物理變數
(如溫度、壓力)"

動態特性 (Dynamics)



Problem: "Chemical processes have time lags."

Solution: "Dynamic PCA
(Lagged variables)"

非穩態 (Non-Stationarity)



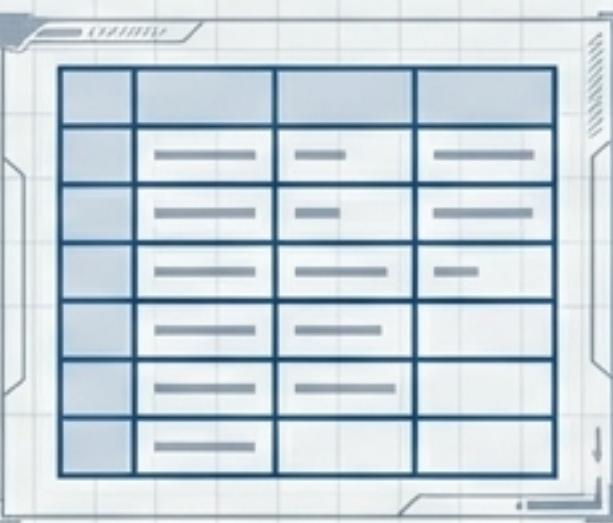
Problem: "Catalyst decay changes normal state."

Solution: "Adaptive models / Retraining"

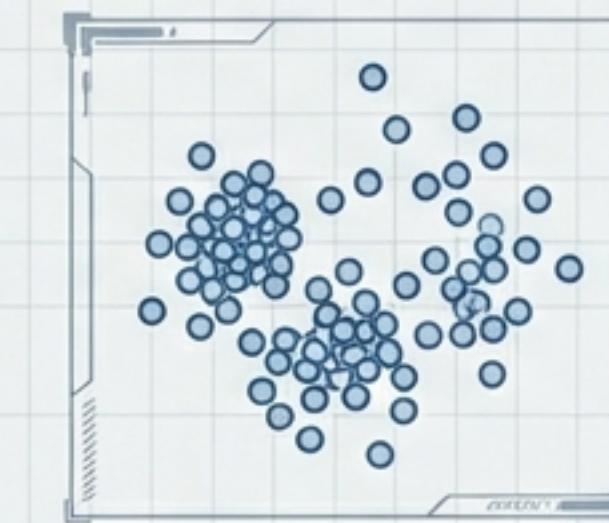
實作路徑：使用 Python (sklearn)

必備套件 Libraries: scikit-learn, umap-learn

```
1. Scale: X_scaled = scaler.fit_transform(X_raw)
2. Fit: model = PCA(n_components=3)
3. Transform: X_reduced = model.fit_transform(X_scaled)
```



Raw Data (50 cols)



Visual Output (2D)

總結：化繁為簡，洞見自現

核心要點 (Key Takeaways)



概念：降維是去除噪音、保留信號的過程 (Signal vs Noise)。



工具選擇：

- 監控與解釋 -> PCA
- 視覺化與探索 -> t-SNE / UMAP



效益：提升模型效能，實現早期異常檢測。

AI 不會取代化工程師，
但懂得使用 AI 降維工具的工程師將取代不懂的人。



學習地圖與下一步

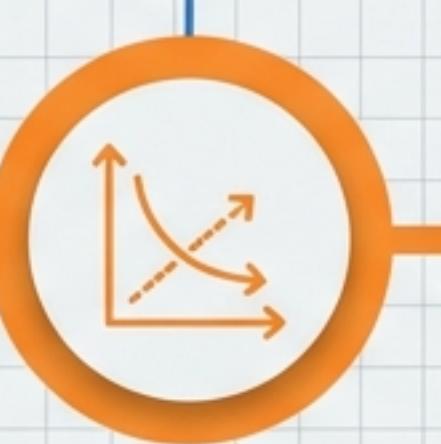


Unit 01-04

(基礎)

Roboto Mono

Noto Sans TC

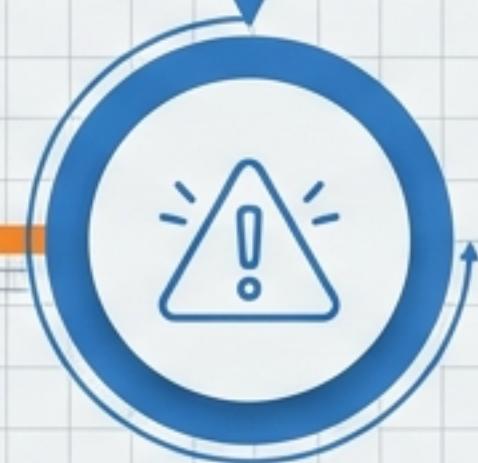


Unit 06
降維
(Unsupervised)

Noto Sans TC

Current Location

我們將使用本單元學到的 PCA
重建誤差 與 低維投影 技術，
在下個單元構建工廠的自動警
報系統。



Unit 07
異常檢測
(Anomaly Detection)

Roboto Mono

Next Stop



NotebookLM