

Unit 09 進階補充教材：製程安全異常偵測

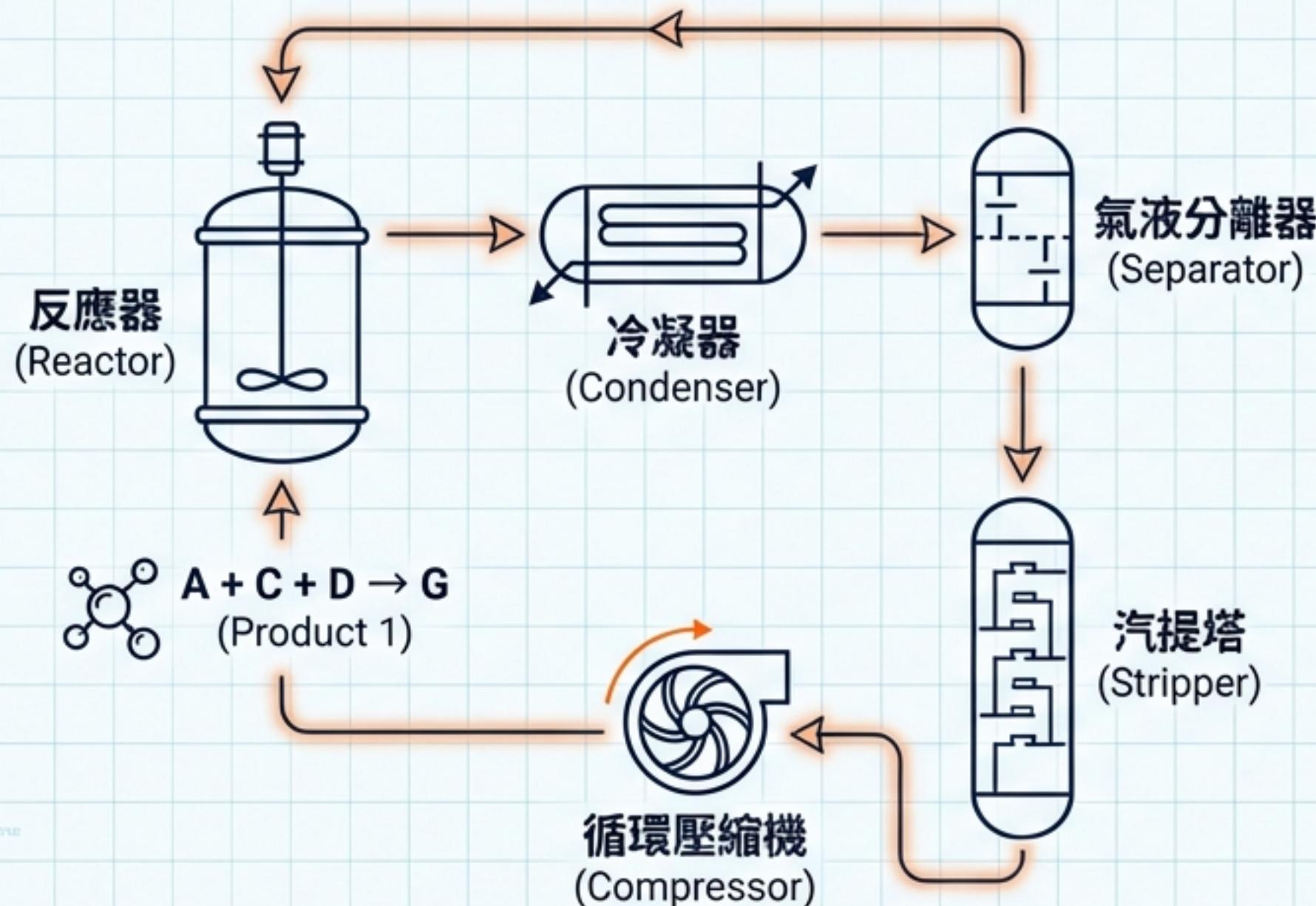
TEP 案例實戰、進階 PCA 與多種演算法深度評析

spec box

COURSE: 化工資料科學與機器學習實務 (CHE-AI-101)
INSTRUCTOR: 莊曜禎 助理教授 | 逢甲大學 化工系

戰場定義：Tennessee Eastman Process (TEP) 綜合案例

The Process



The Data Specs

變數 (Variables):

52 Total (41 Measured XMEAS,
11 Manipulated XMV)

故障 (Faults):

21 Types (IDV1-20)

採樣頻率 (Sampling):

3 minutes/sample

重點場景 (Focus Scenario):

IDV5 (Condenser Cooling Water
Inlet Temperature Step Change)

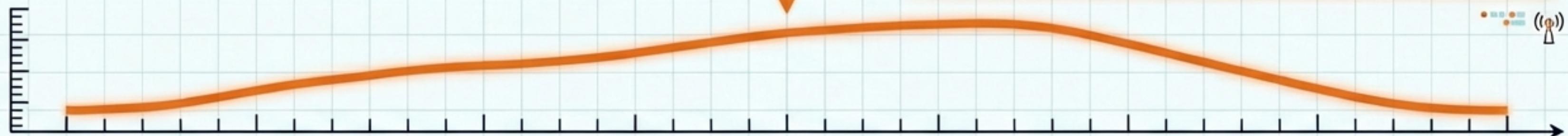


TEP 不僅是數據集，它是真實化工廠的模擬，包含多重迴路與物理限制。

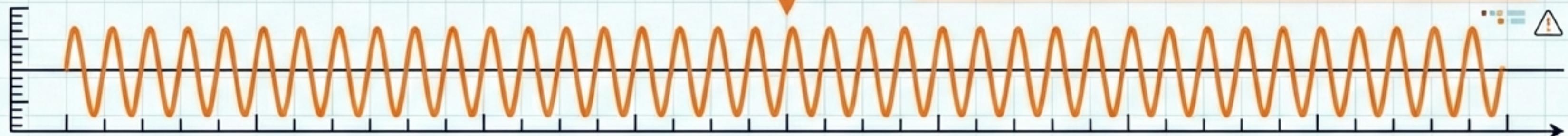


數據特徵：多重時間尺度與動態行為

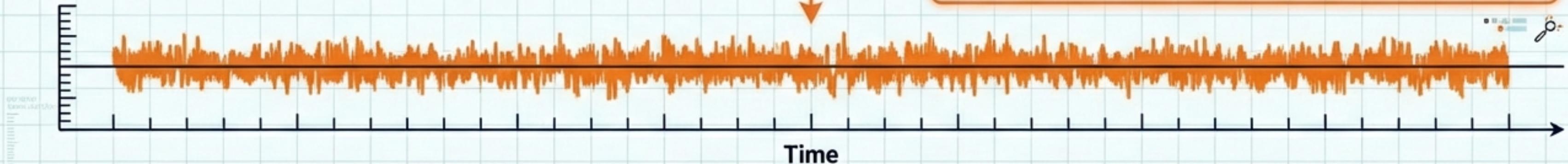
XMEAS_07 (Reactor Pressure)



XMEAS_08 (Reactor Level)



XMEAS_09 (Reactor Temp)

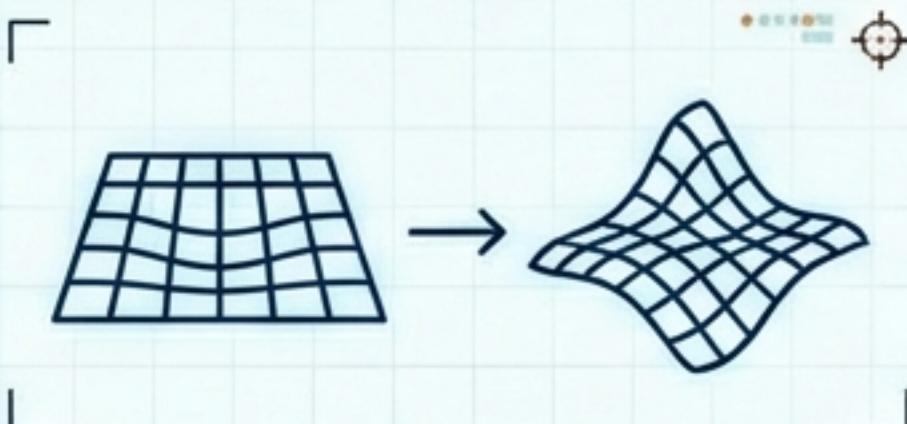


Challenge : 異常偵測模型必須能區分「正常的操作動態」與「真正的故障偏離」。



武器庫升級：進階 PCA 技術譜系

Kernel PCA (非線性)



透過核函數 $K(x, x')$ 映射至高維空間。

$$K(x, x') = \exp(-\gamma \|x - x'\|^2)$$

TEP Insight: 對於線性故障 (IDV5) 效果有限。

Sparse PCA (可解釋性)

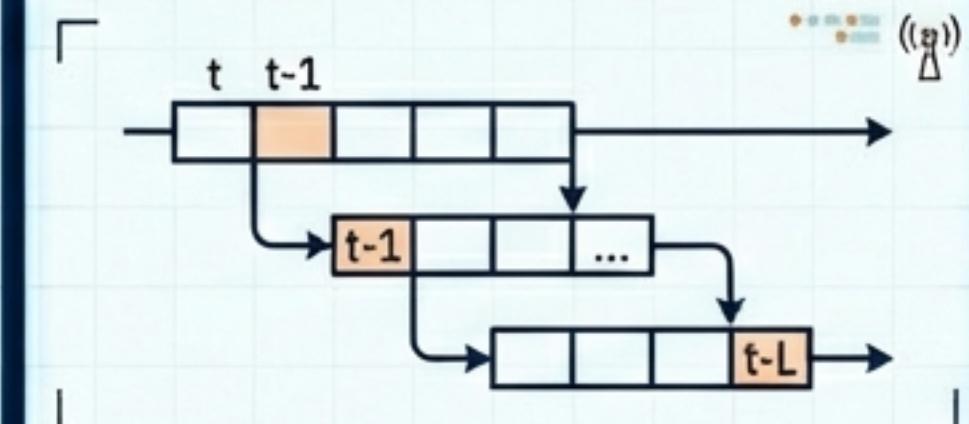
1	0	0	0	0	0	0
0	2	3	0	0	0	0
0	0	5	1	0	0	0
0	0	0	0	4	0	0
0	0	0	0	0	1	0

引入 L1 正則化，強制非關鍵變數係數歸零。

$$\min_{\mathbf{W}} \|\mathbf{X} - \mathbf{T}\mathbf{W}^T\|_F^2 + \alpha \|\mathbf{W}\|_1$$

TEP Insight: 將 52 變數縮減為 ~27 個關鍵貢獻者。

Dynamic PCA (動態性)



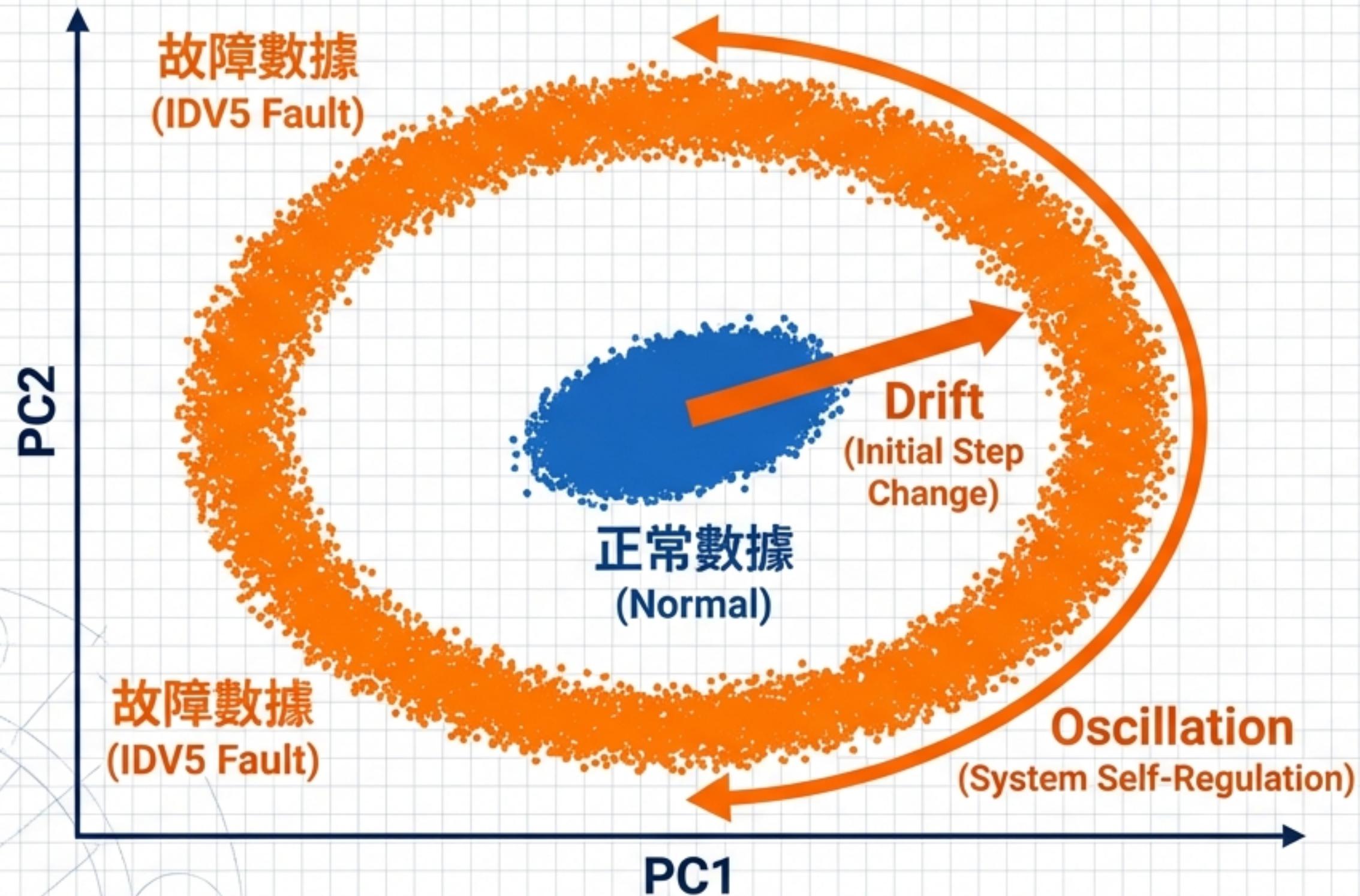
處理時間序列自相關 (Autocorrelation)。

$$\mathbf{X}(t) = [x(t), x(t-1), \dots, x(t-L)]$$

Use Case: 適用於具備時間滯後的批次製程。



降維後的真相：IDV5 故障的幾何結構



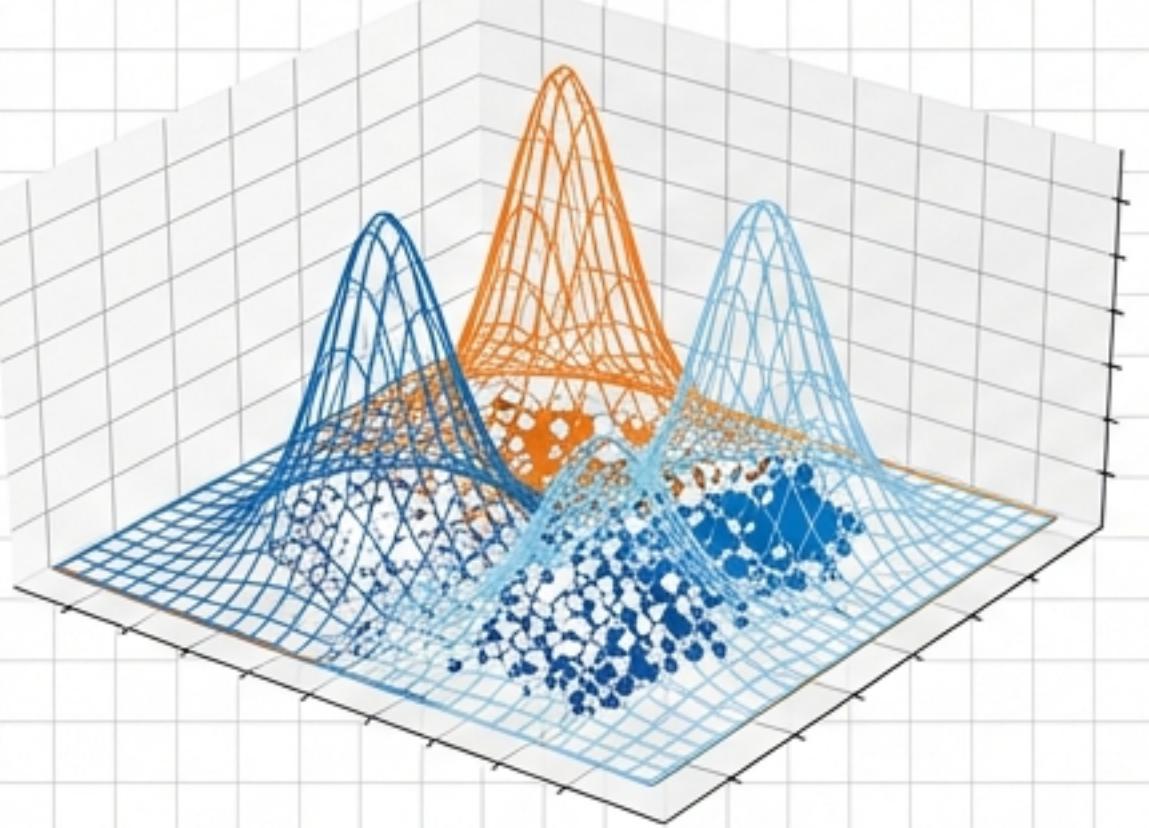
工程解讀

冷凝器效率下降 → 壓力上升 → 壓縮機負荷增加 → 形成環形軌跡。

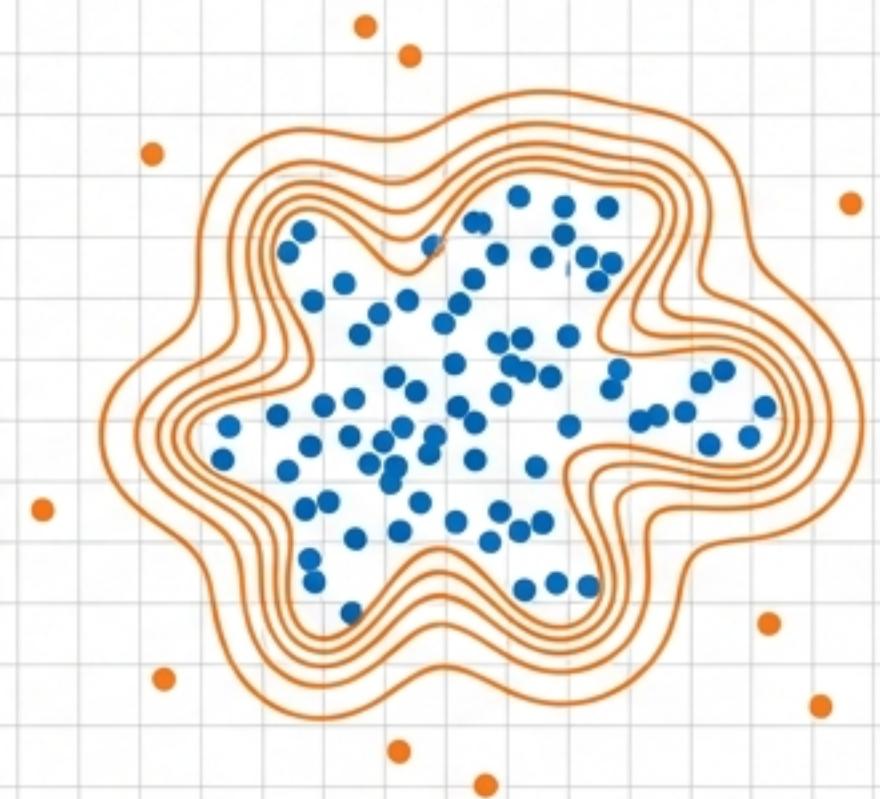
Mahalanobis Distance:
Normal $D^2 < 9.21$
vs Fault $D^2 > 50$

密度與邊界法：捕捉多模態與高維邊界

GMM (Gaussian Mixture Model)



One-Class SVM



邏輯：假設數據由多個高斯分佈混合而成。

Why it works: TEP 正常操作包含穩態、過渡態與振盪態 (Multimodal).

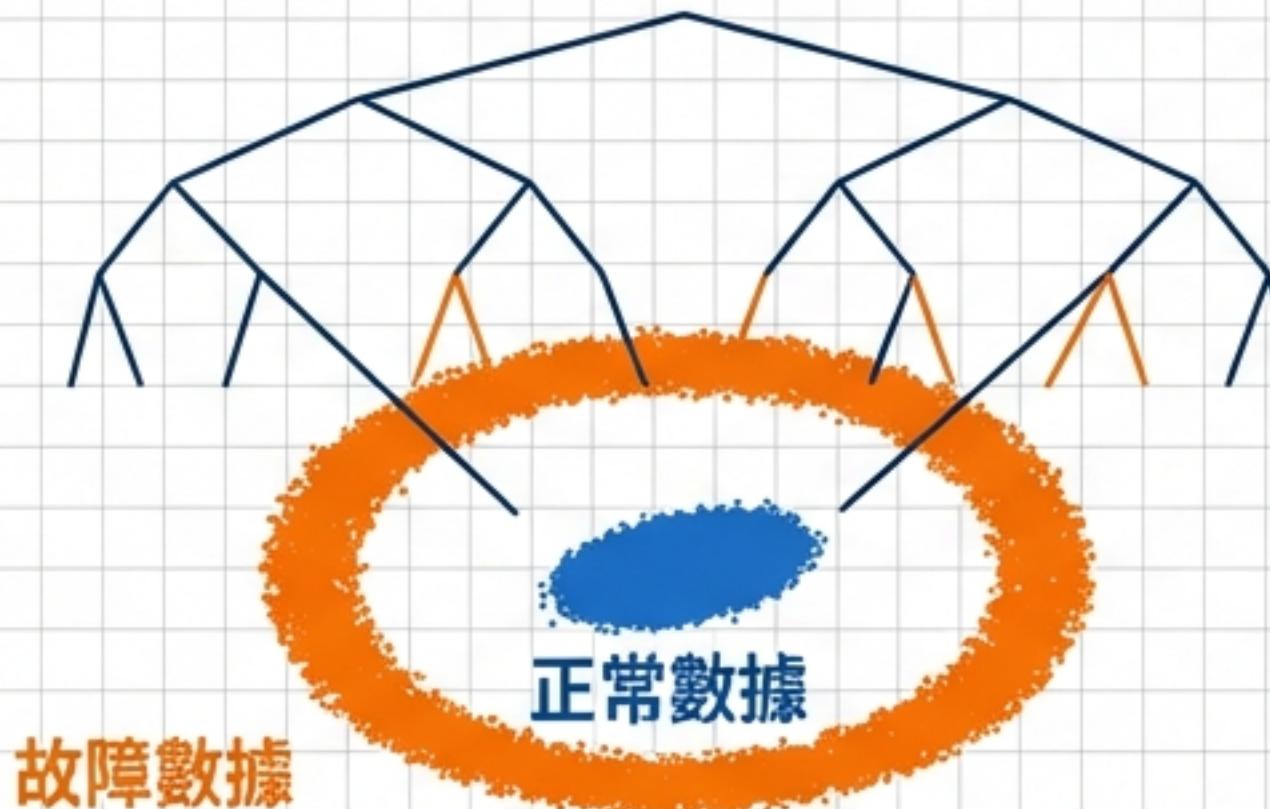
邏輯：尋找包含正常數據的最小超球體或超平面。

Key Param: $\nu = 0.04$

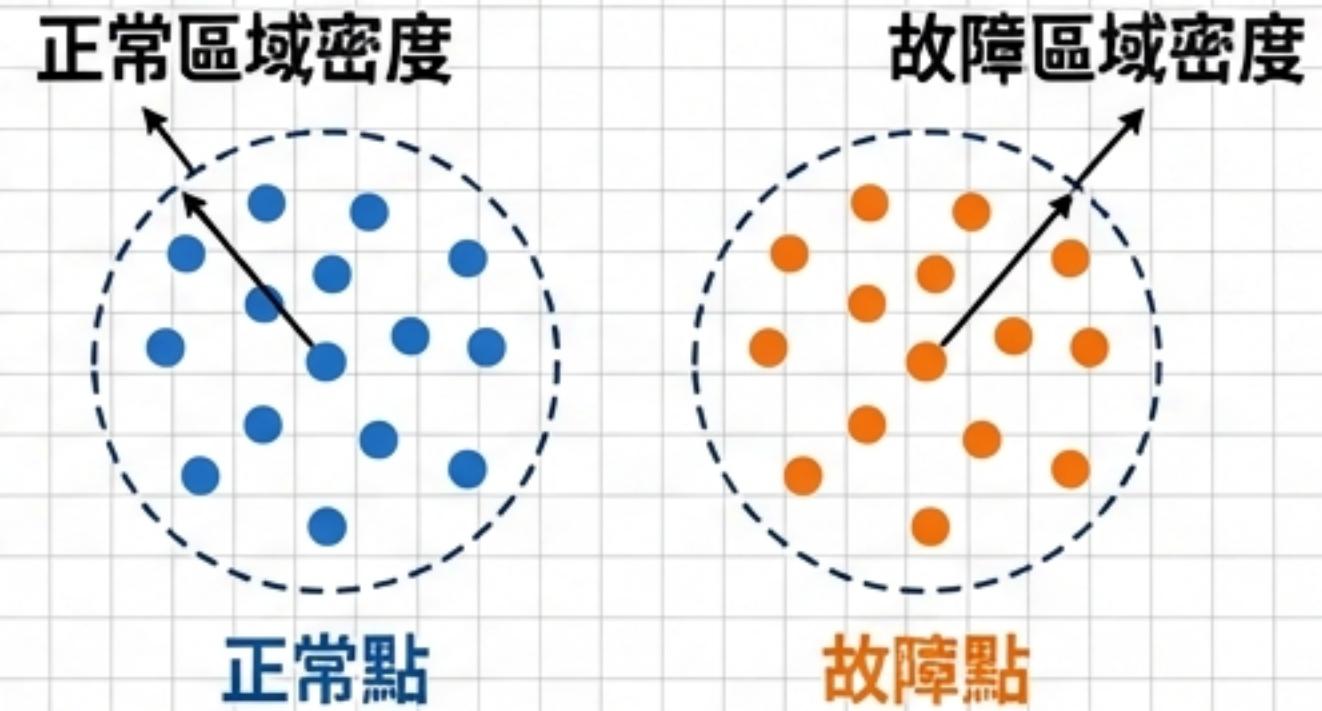
Why it works: 在高維空間建立穩健邊界，無需假設分佈形狀。

為什麼它們失敗了？(Why Others Failed?)

Isolation Forest (準確率 < 50%)



LOF (Local Outlier Factor)



Failure: 演算法假設異常是『孤立且稀疏』的。但 IDV5 故障形成密集叢集，導致路徑長度與正常點相似。

Failure: 故障區域密度 \approx 正常區域密度。LOF 比值接近 1 (False Negative)。



警示：化工故障往往是密集的結構性偏移，而非孤立的噪聲點。

終極評比：演算法競技場

準確率 (Accuracy) 排名

GMM (K=3) - 86.77% [Winner]

One-Class SVM - 68.75% [Robust]

LOF - 50.00% [Failed]

Isolation Forest - 49.90%

Metric	GMM	SVM
--------	-----	-----

Recall	100%	75%
--------	------	-----

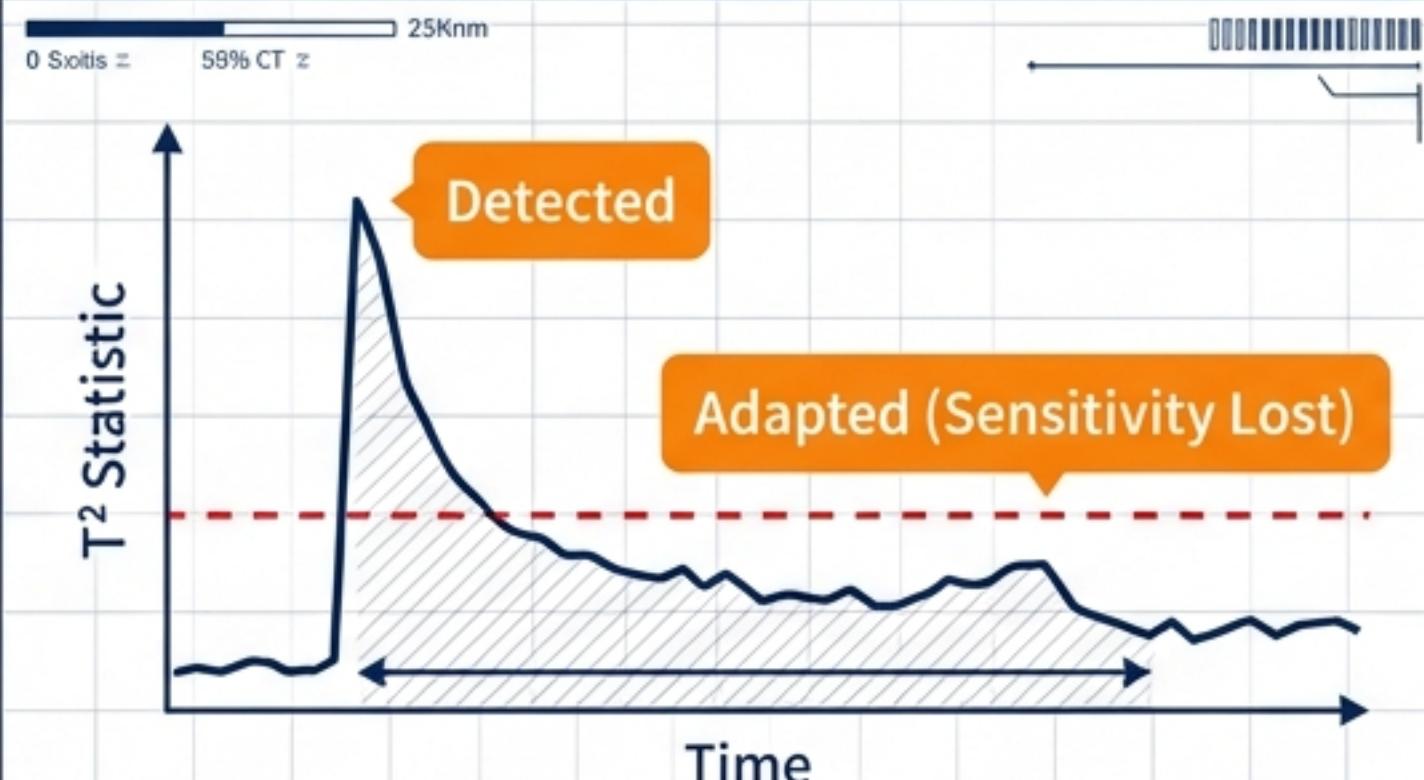
Speed	58ms	71ms
-------	------	------

GMM 因成功捕捉 TEP 的多模態特性而奪冠，但 SVM 提供了較佳的理論穩健性。

工程師的診斷工具：MSPC 統計製程管制

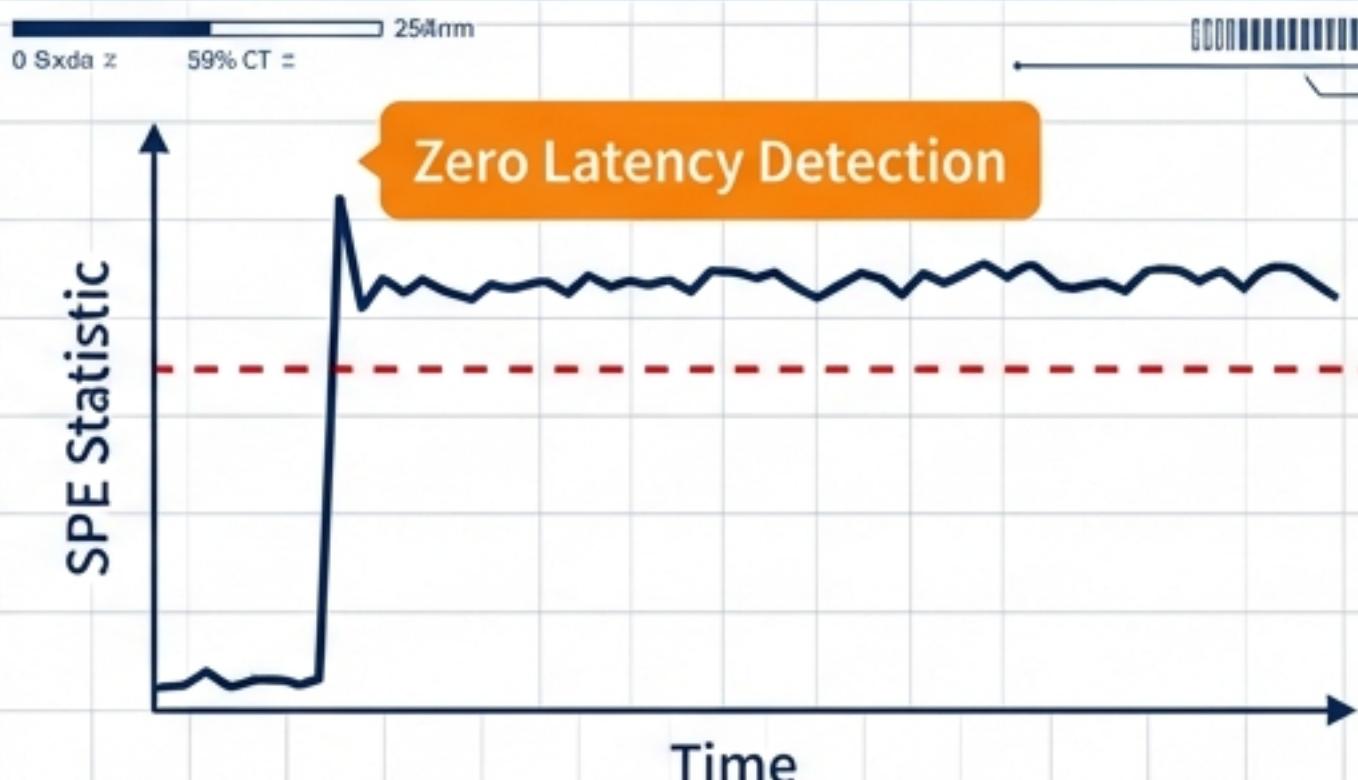
從純預測走向物理診斷

T² Statistic (Hotelling)



監控模型空間 (Model Space) 的均值偏移。

SPE Statistic (Q-statistic)



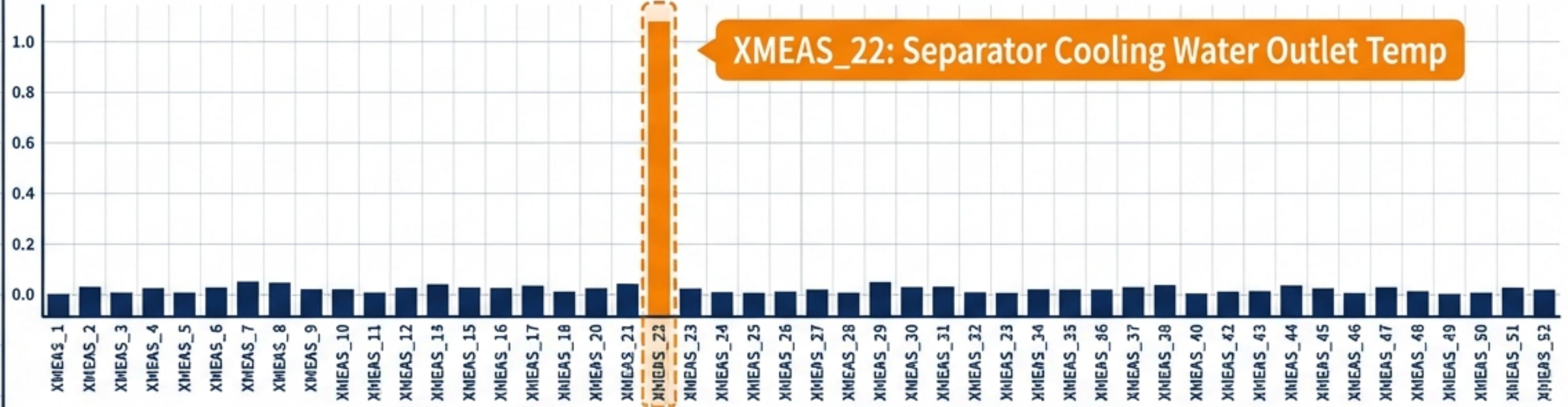
監控殘差空間 (Residual Space) 的結構破壞。



針對 IDV5 關聯結構破壞，SPE 表現出零延遲與持續報警的能力。

根因分析：誰是罪魁禍首？(Root Cause Analysis)

SPE Contribution Plot (Sample 160)



[Step 1:
SPE Alarm]

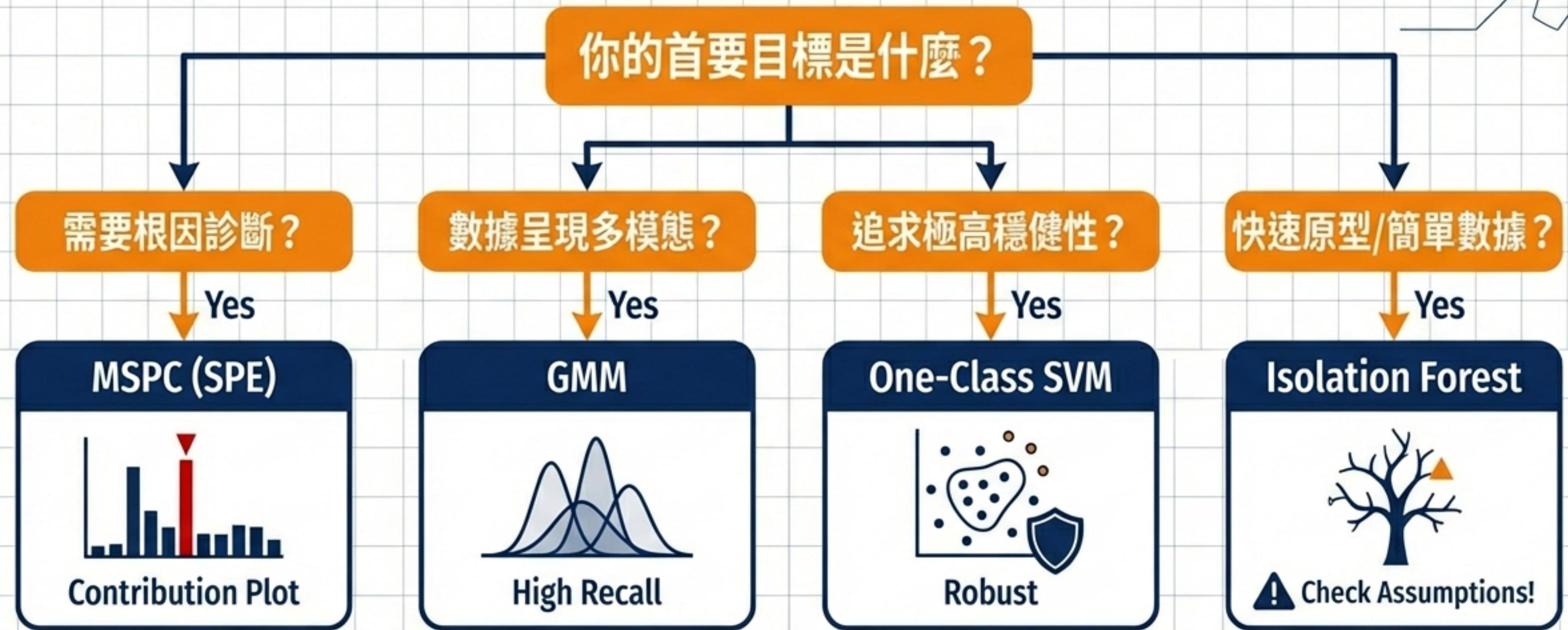
[Step 2:
Check Contribution]

[Step 3:
Identify XMEAS_22]

[Step 4:
Fix Cooling Valve]

工程價值：不只告訴操作員『出問題了』，還精確指出是『分離器冷卻水』異常 (貢獻度 > 48%)。

實戰指南：異常偵測方法選擇決策樹



沒有最好的模型，只有最適合場景的模型。

實務部署分層策略 (Deployment Strategy)



Level 1: 實時監控 (Real-time Monitoring)

⌚ Tools: GMM / SVM

⌚ Metric: Speed < 100ms

⚠ Action: 快速報警 (Is something wrong?)

Level 2: 根因診斷 (Root Cause Diagnosis)

🔍 Tools: MSPC (Contribution Plots)

🔴 Trigger: Only when Level 1 alarms

❓ Action: 定位變數 (What is wrong?)

Level 3: 模型迭代 (Model Maintenance)

⟳ Action: 監控概念漂移 (Concept Drift)
並利用新數據重訓模型

結論與核心洞察 (Key Takeaways)

Context Matters



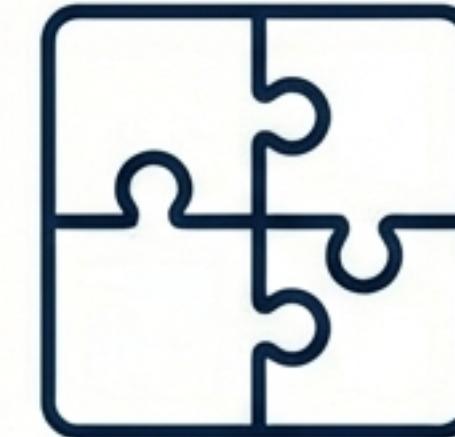
TEP 案例證明，盲目使用通用算法（如 IsoForest）在密集型化工故障上可能失效。

Physics > Prediction



準確率不是唯一指標。能解釋「哪個閥門壞了」(Contribution Plot) 比單純報警更有價值。

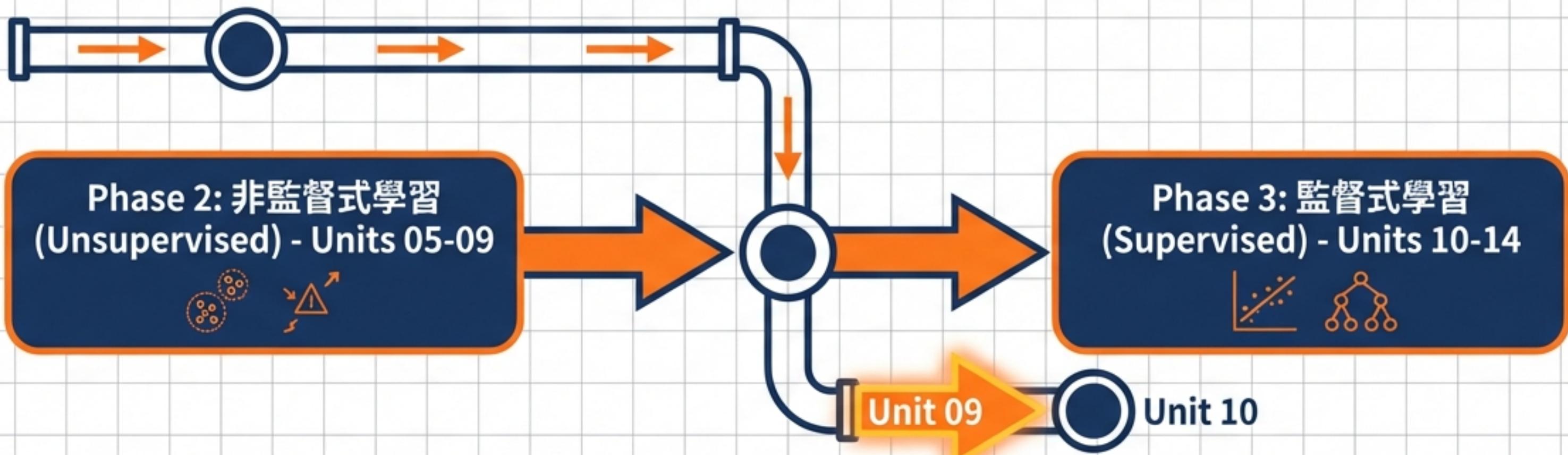
The Winning Combo



推薦 GMM (高靈敏度) + MSPC (高可解釋性) 的組合拳策略。

AI 不會取代化工程師，但懂得用 AI 診斷的工程師將無可取代。

延伸學習路徑 (Learning Path)



Action Items

- Code:** 前往 GitHub 下載 `Unit09_Advanced_Topics.ipynb` 進行 TEP 實作。
- Next:** 準備進入 Unit 10：監督式學習與軟感測器開發。