

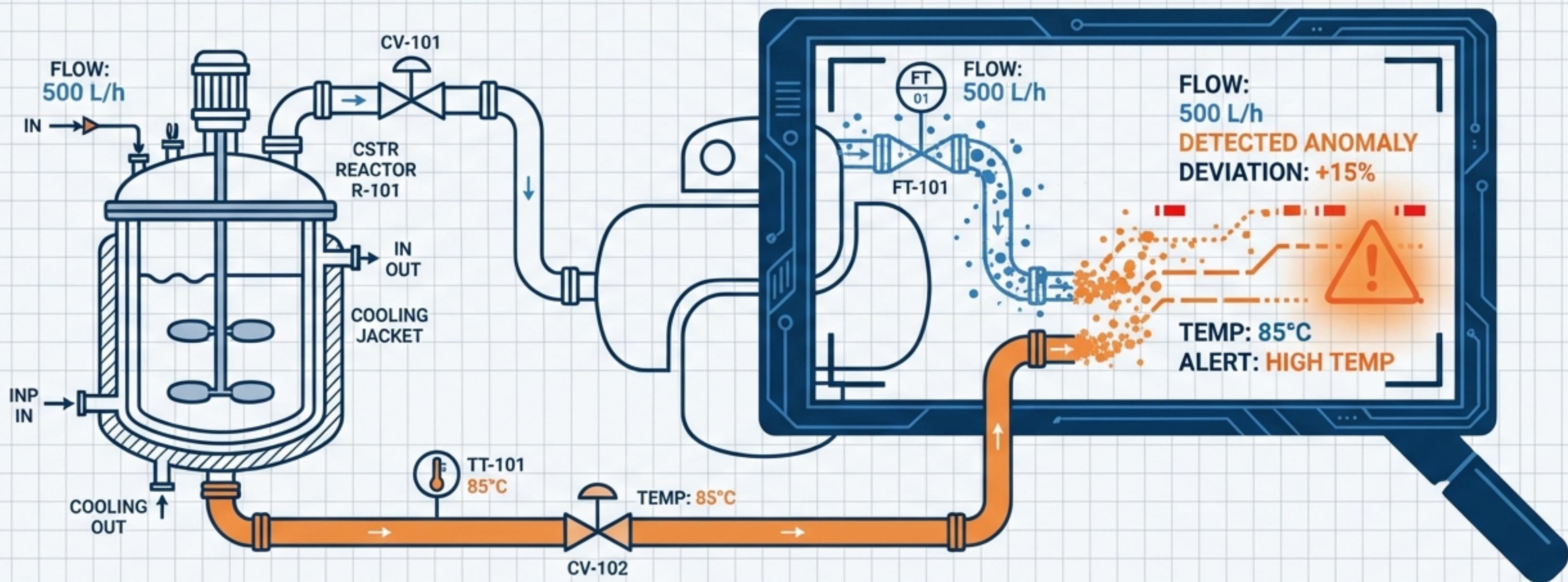
Unit 07: 異常檢測總覽

Anomaly Detection Overview

Roboto Mono

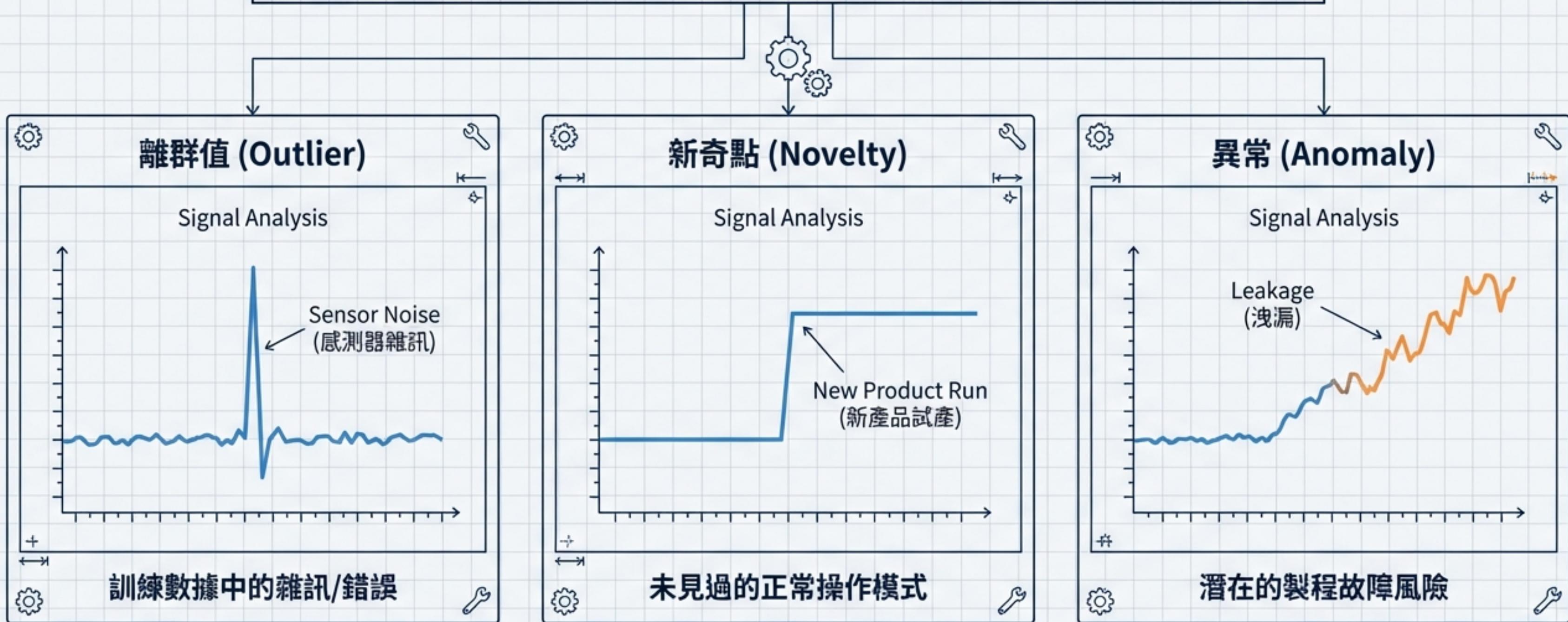
Course: CHE-AI-114

Instructor: 莊曜禎 助理教授

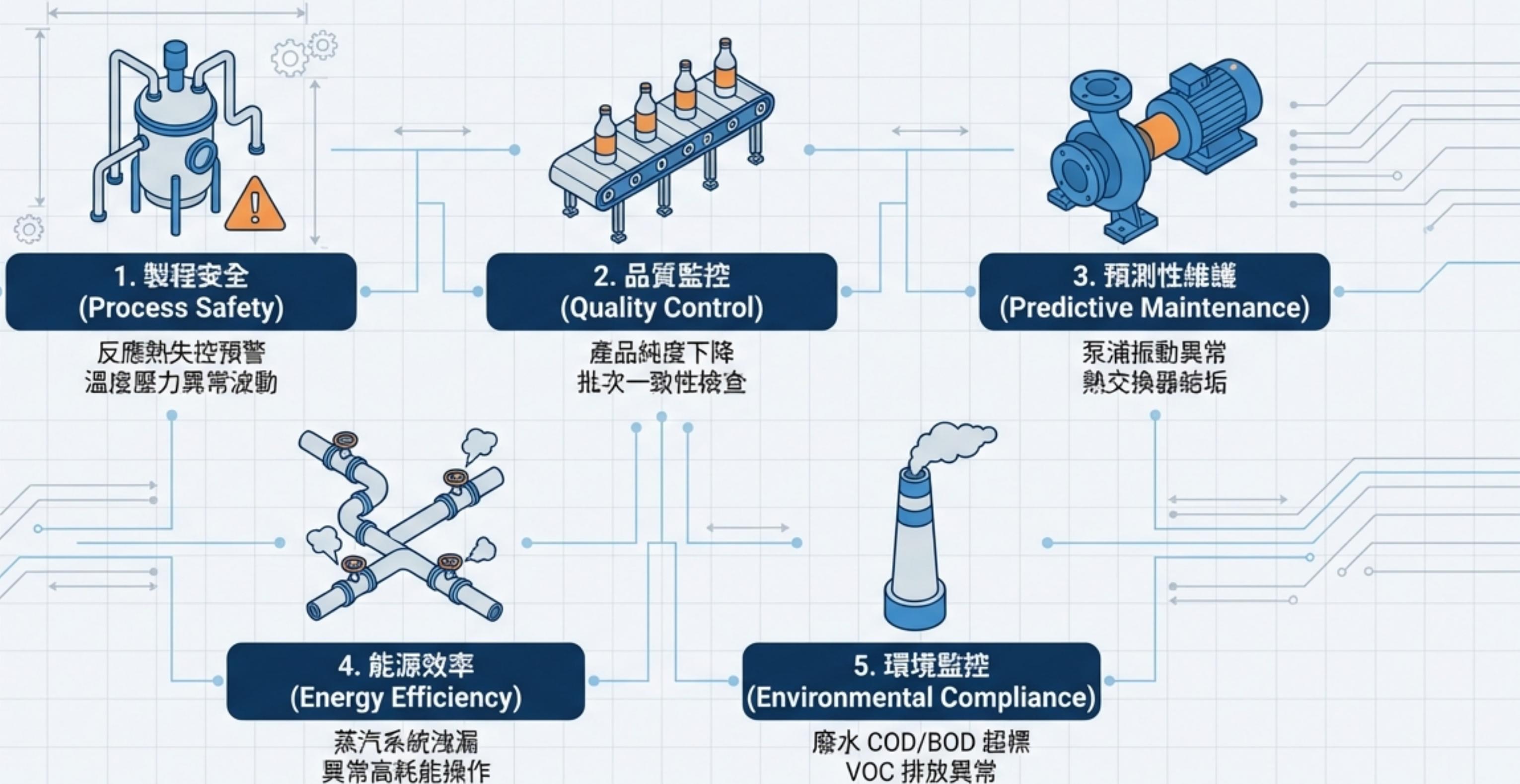


定義與核心概念 (Definition & Core Concepts)

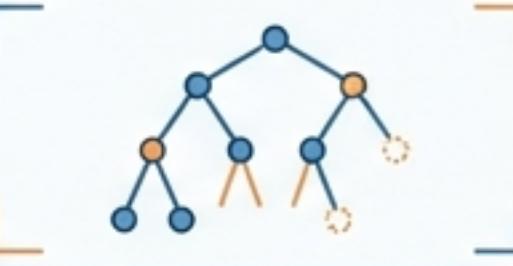
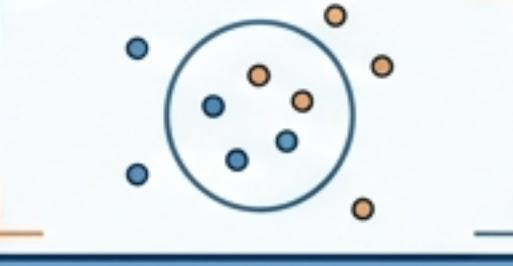
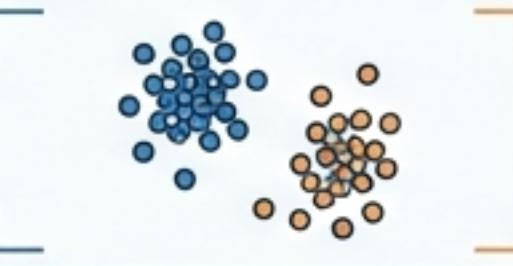
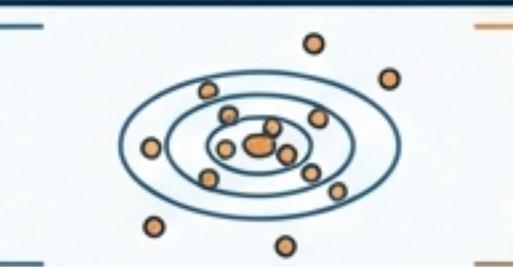
異常檢測 (Anomaly Detection)：識別與大多數數據顯著不同的樣本技術。



化工領域的五大應用戰場 (5 Battlefields in Chemical Engineering)

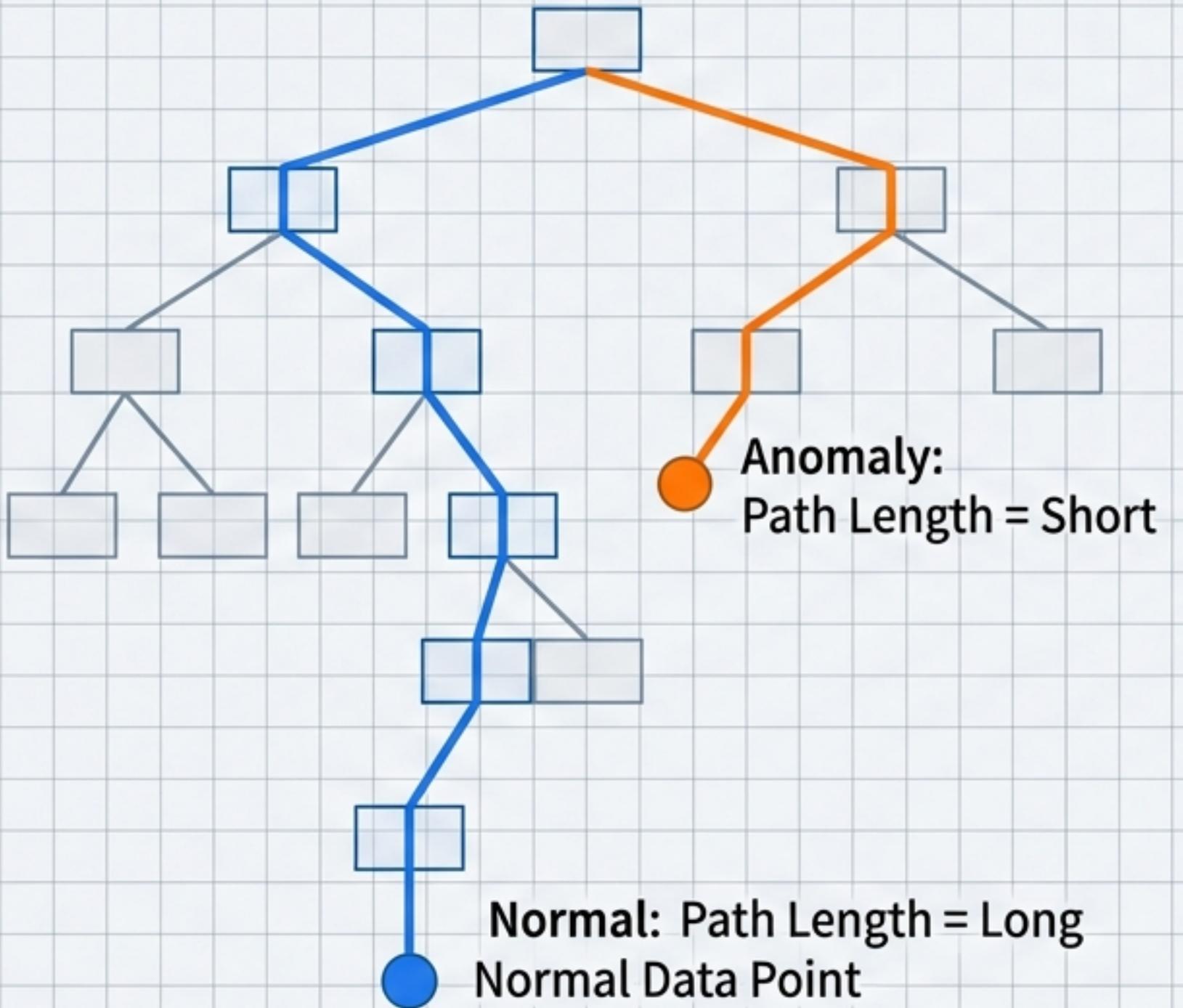


sklearn 異常檢測模組工具箱 (The Algorithm Toolbox)

演算法 (Algorithm)	核心圖示 (Icon)	類型 (Type)	最佳適用場景 (Best Use Case)
Isolation Forest (孤立森林)		基於樹 (Tree-based)	大規模 / 高維數據 (Large Scale / High Dim)
One-Class SVM		基於邊界 (Boundary-based)	小樣本 / 高精度需求 (Small Data / Precision)
Local Outlier Factor (LOF)		基於密度 (Density-based)	密度不均 / 多模式操作 (Uneven Density)
Elliptic Envelope		基於統計 (Statistical)	高斯分布數據 (Gaussian Distribution)

Isolation Forest (孤立森林)

異常點容易被孤立，正常點難以被孤立



特點 (Features):

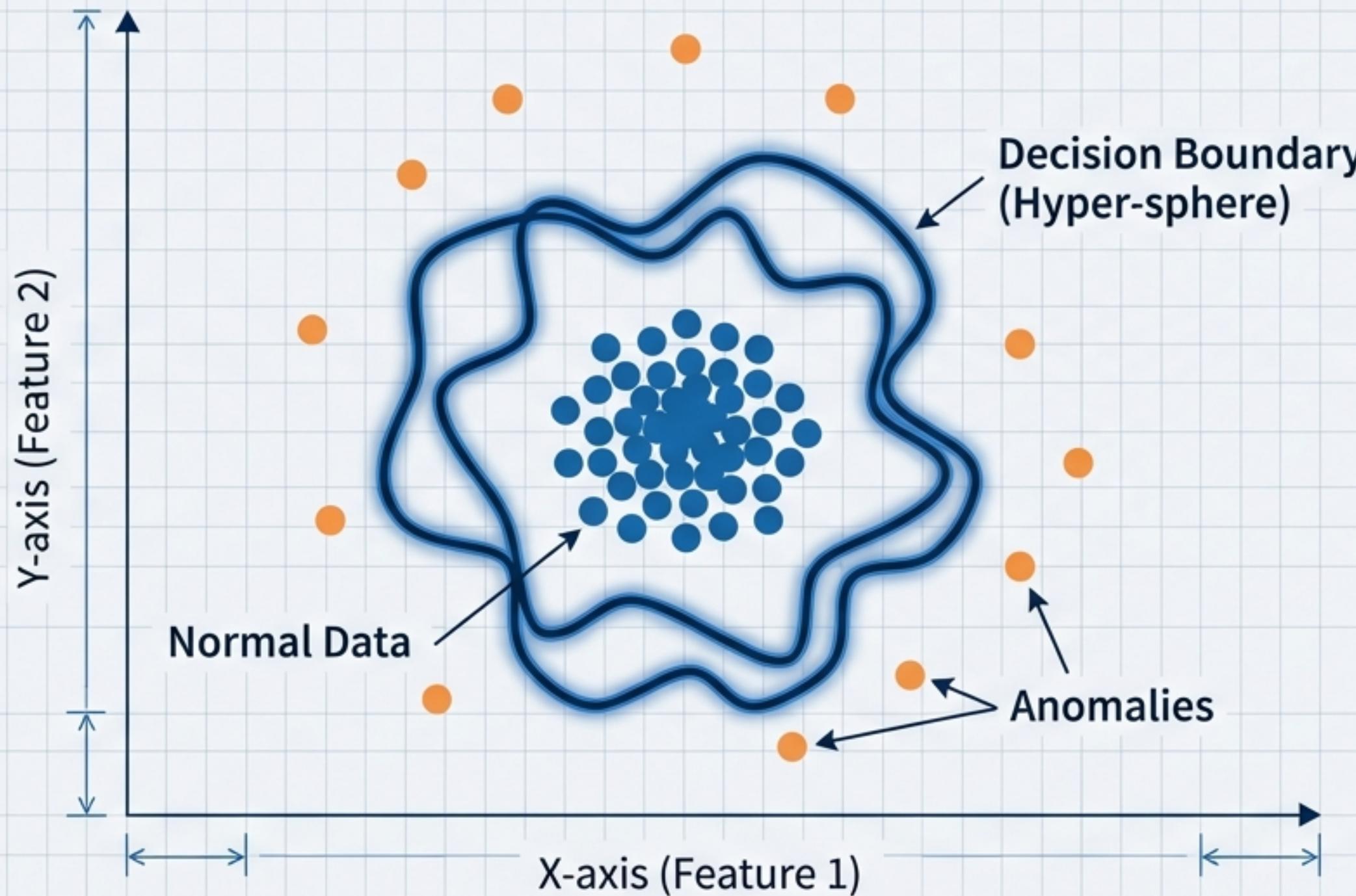
- 時間複雜度: $O(n \log n)$
- 適用: 高維度數據 (High Dimensions)
- 無需假設數據分布

化工應用 (ChemE Use Case):

- 蒸餾塔監控 (Distillation Column)
 - 監控 50+ 個感測器 (Temp, Pressure, Flow)
 - 成功檢測 12 起石化廠異常事件

One-Class SVM (一類支持向量機)

尋找包含正常數據的最小超球面



特點 (Features):

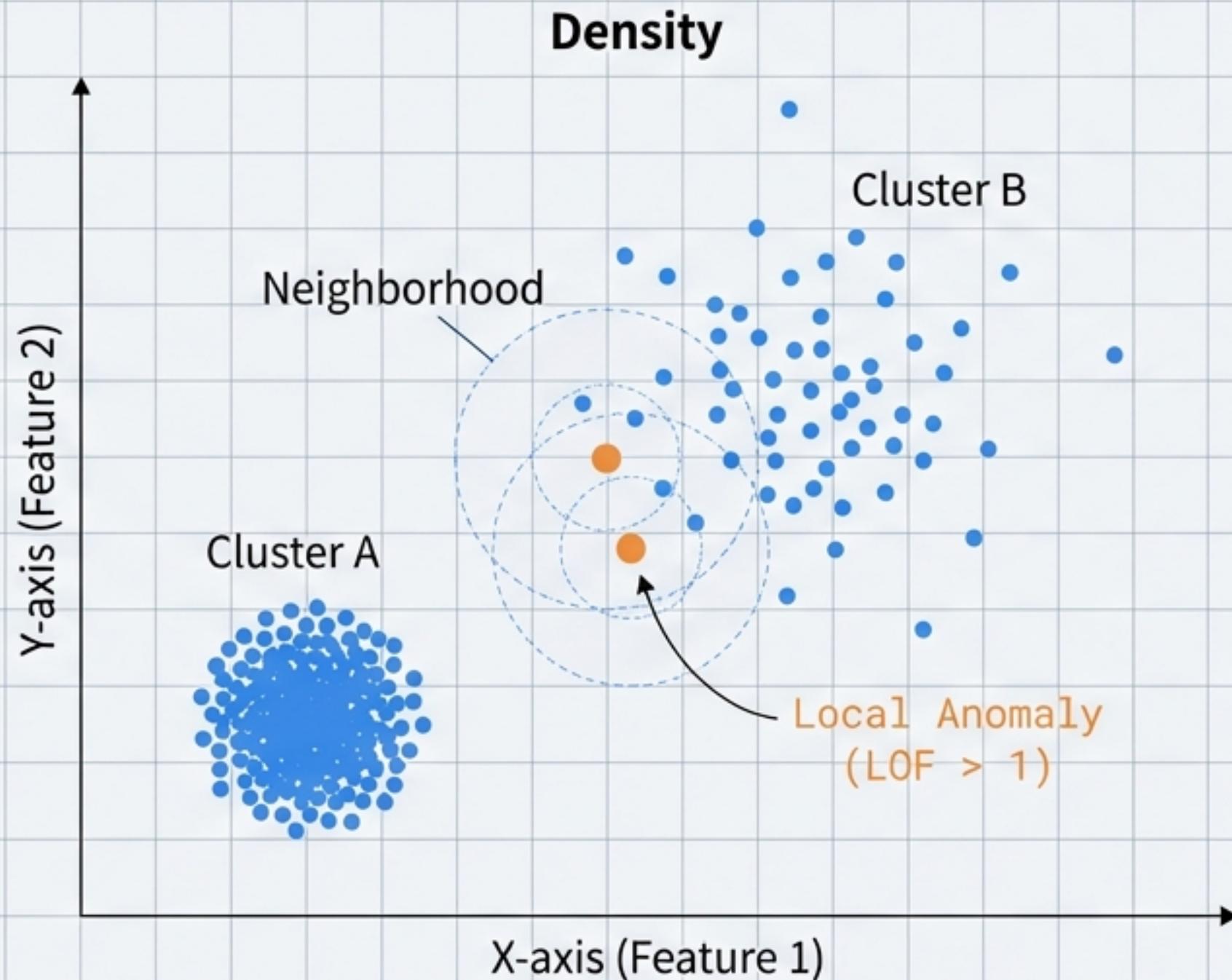
- 機制: Kernel Function 映射至高維空間
- 優勢: 邊界極其精確
- 適用: 小樣本數據 (Small Data)

化工應用 (ChemE Use Case):

- 高價值藥品批次監控 (Pharma Batch)
- 僅用 200 批次正常數據訓練
 - 精確檢測微量純度偏差

Local Outlier Factor (LOF - 局部離群因子)

比較每個點與其鄰居的密度比值



Noto Sans TC

特點 (Features):

- 機制: K-Nearest Neighbors 密度計算
- 優勢: 適應非均勻密度分布
- 缺點: 計算成本較高

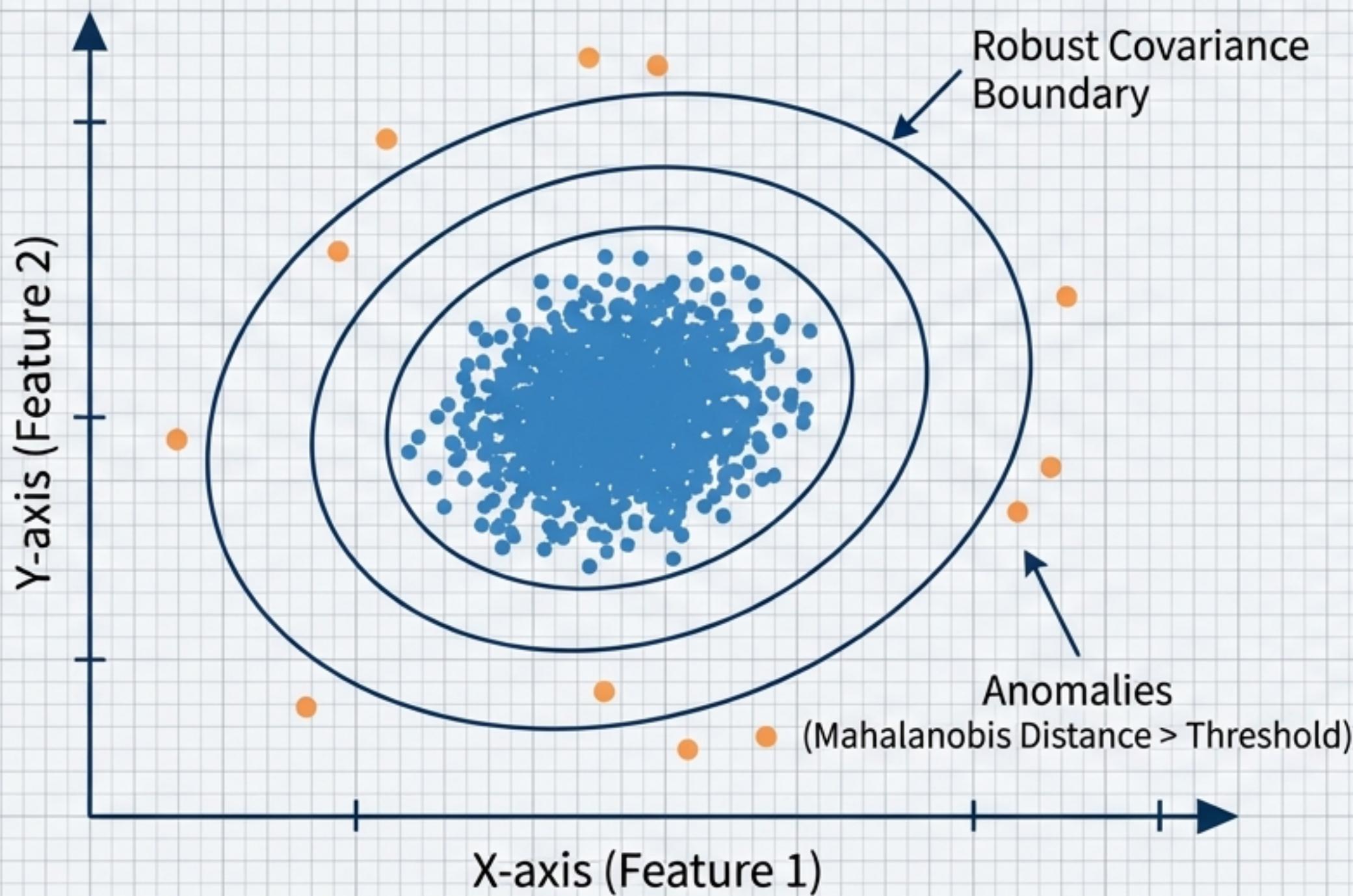
Roboto Mono #0F3460

化工應用 (ChemE Use Case):

- 反應器多模式操作 (Multi-mode Reactor)
- 識別特定操作階段下的局部異常

Elliptic Envelope (橢圓包絡)

基於高斯分布假設的統計邊界



特點 (Features):

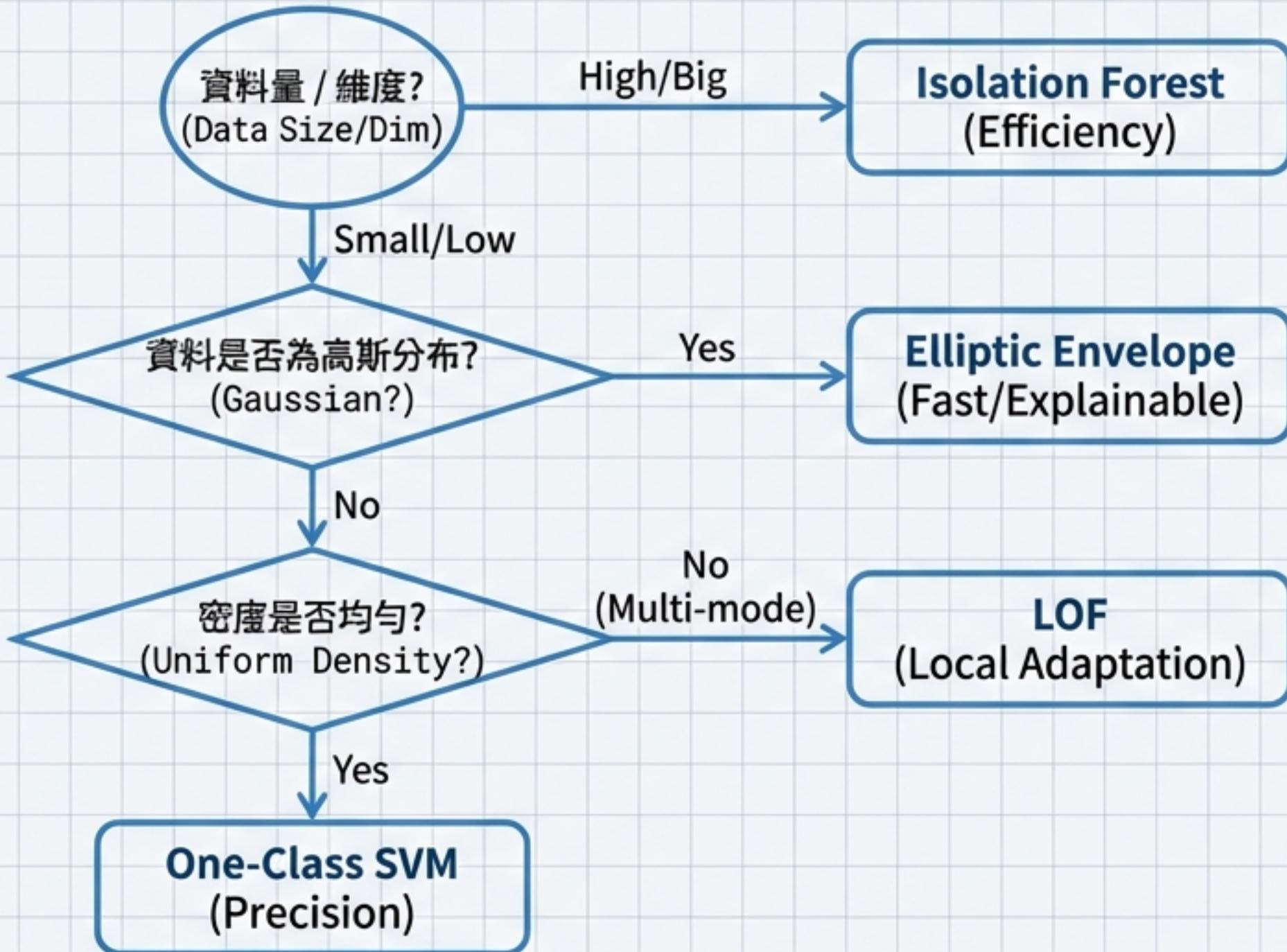
- 假設: 數據服從高斯分布 (Gaussian)
- 優勢: 運算快速，具備統計解釋性

化工應用 (ChemE Use Case):

穩態品質管制 (Steady-state QC)

- 監控食品加工的水分與糖度

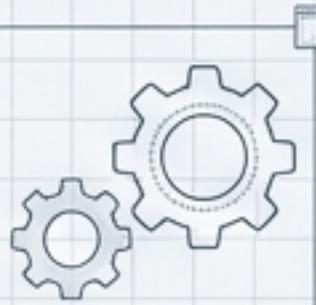
演算法選擇決策指南 (Selection Strategy)



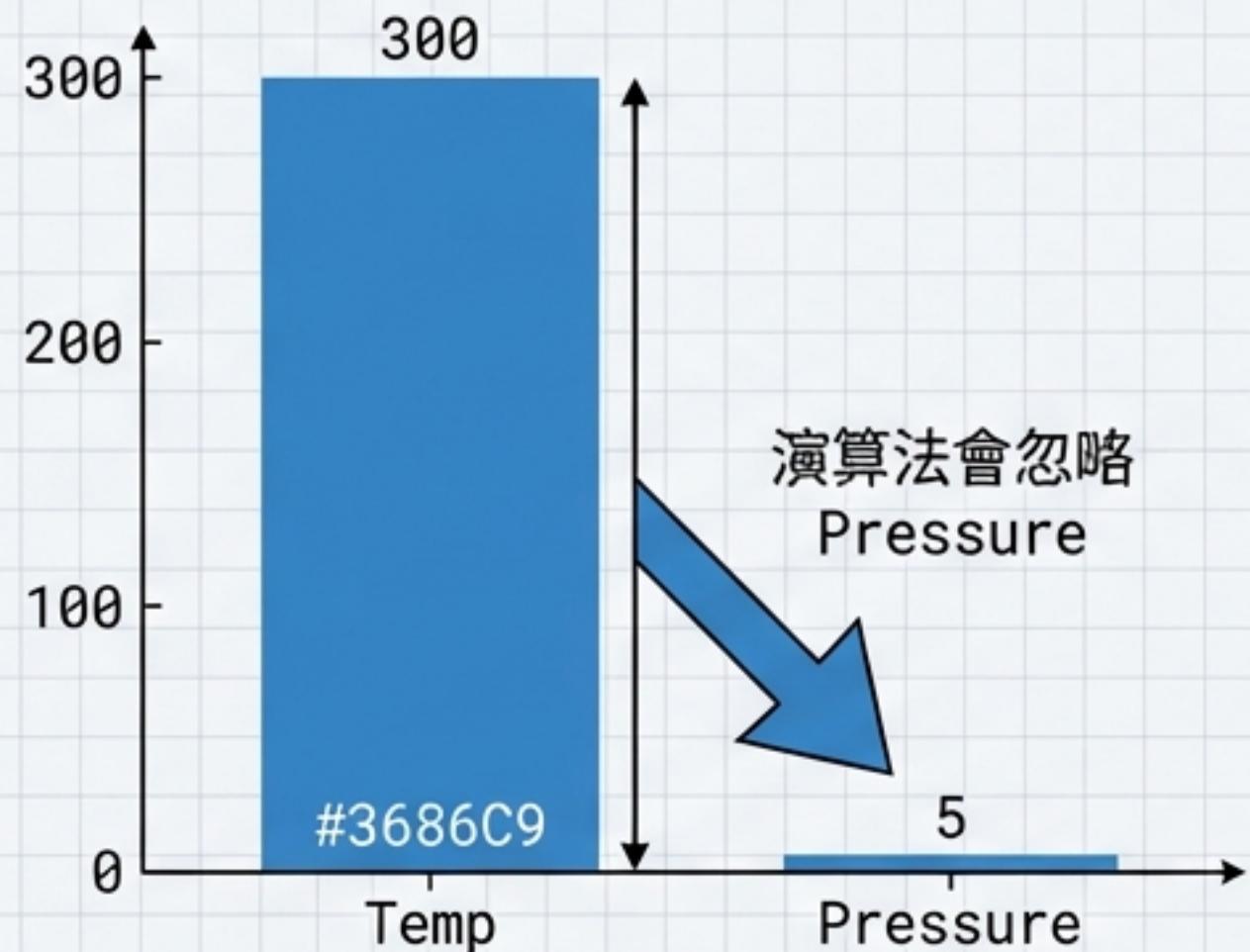
化工實務建議: 先從簡單的 Isolation Forest 或 Elliptic Envelope 開始嘗試，並建立基準線 (Baseline)。

資料前處理：標準化與正規化 (Data Preprocessing)

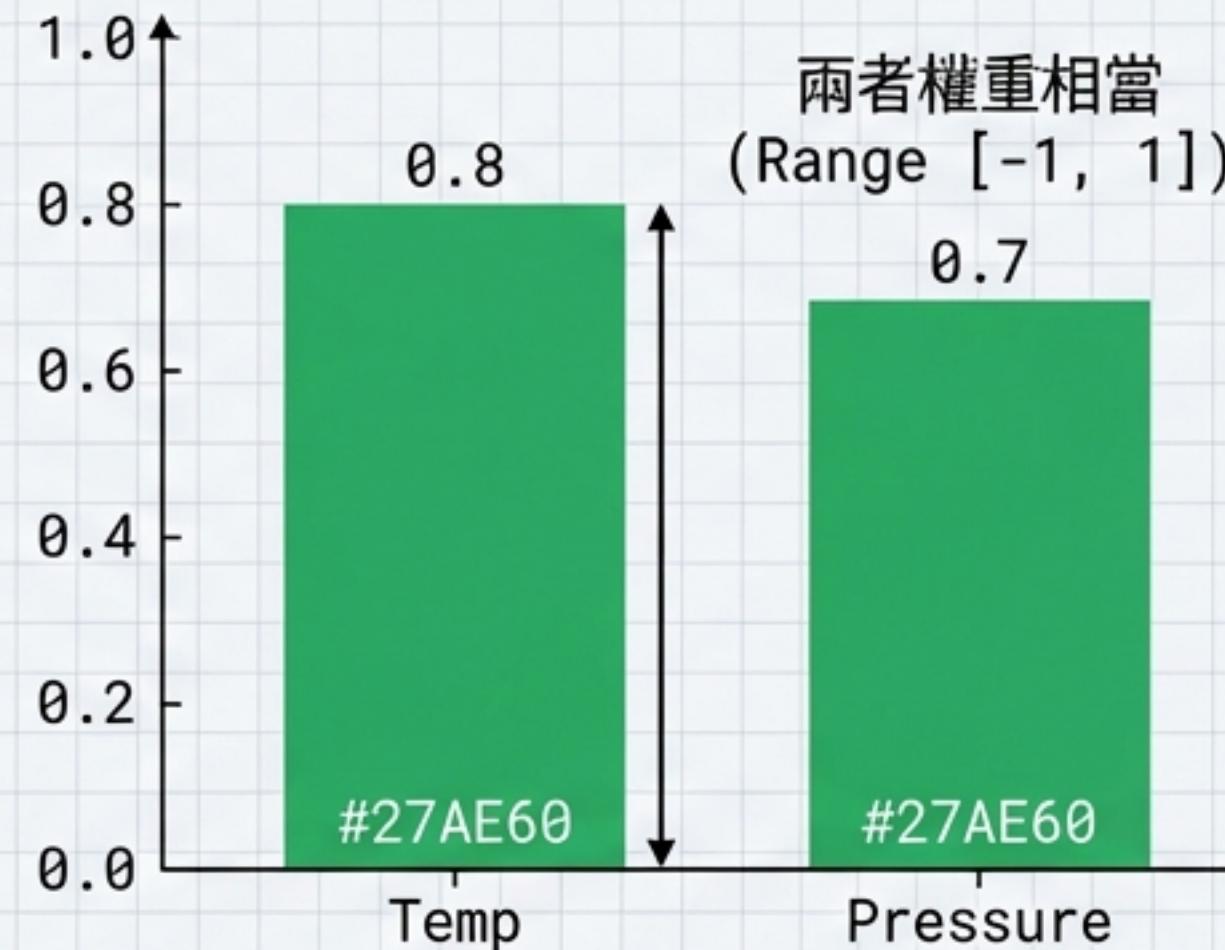
消除尺度差異 (Scale Invariance)



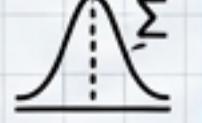
Before Scaling (原始數據)



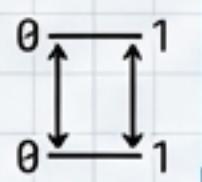
After Scaling (標準化後)



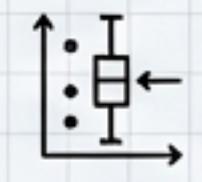
Standardization (Z-score)
適用於高斯分布 (Elliptic Envelope, OC-SVM)。



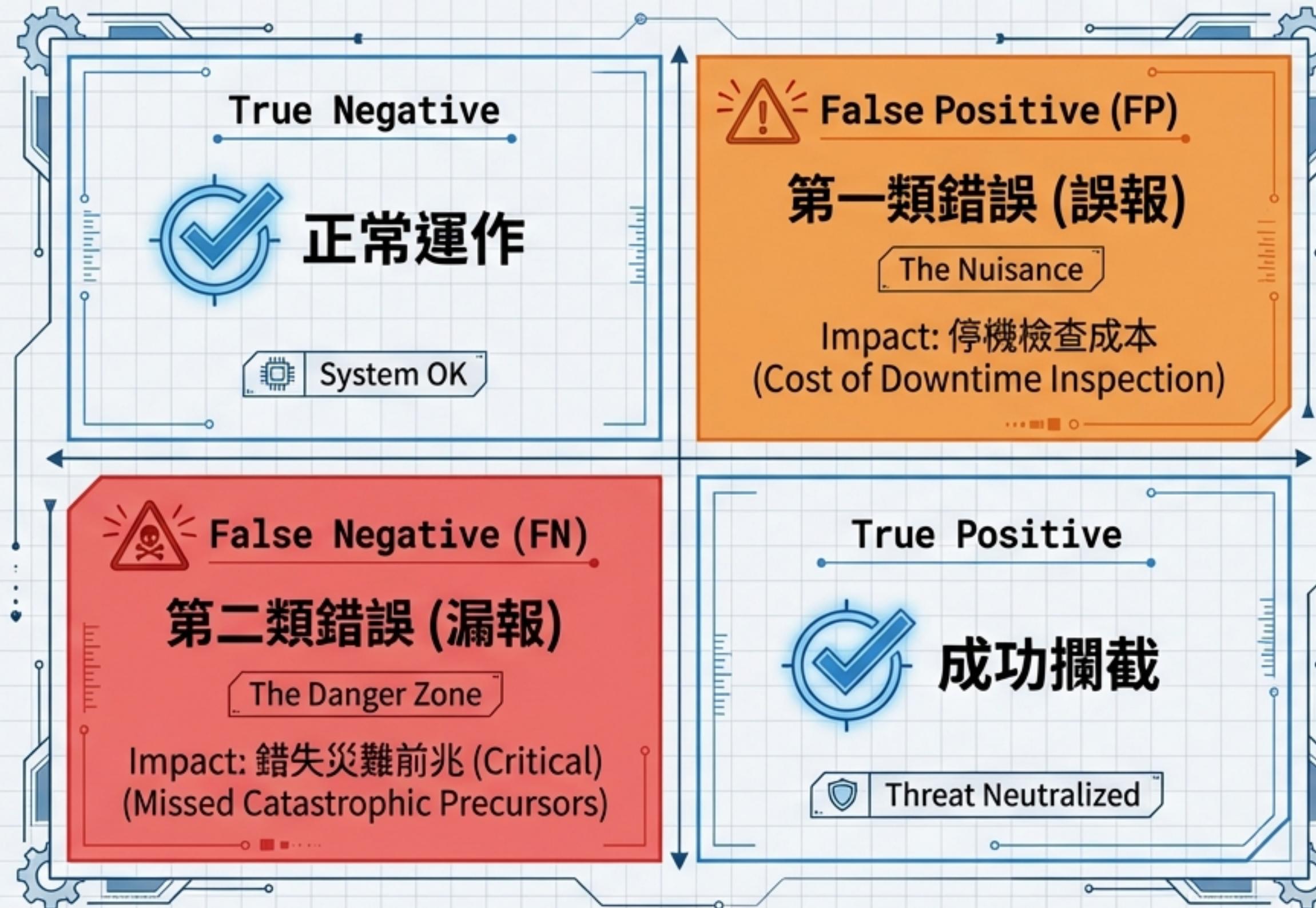
Normalization (Min-Max)
適用於有邊界數據。



Robust Scaling
使用 IQR，抵抗離群值干擾。



模型評估：有標籤場景 (Evaluation with Labels)



評估指標 (Evaluation Metrics)

- **Precision** : 告警的可信度?
(How trustworthy is the alert?)
- **Recall** : 抓到了多少故障?
(How many faults were caught?)
Safety Critical
- **F1-Score** : 綜合指標
(Harmonic mean of Precision and Recall)
- **ROC-AUC** : 整體效能
(Overall Performance, Area Under Receiver Operating Characteristic Curve)

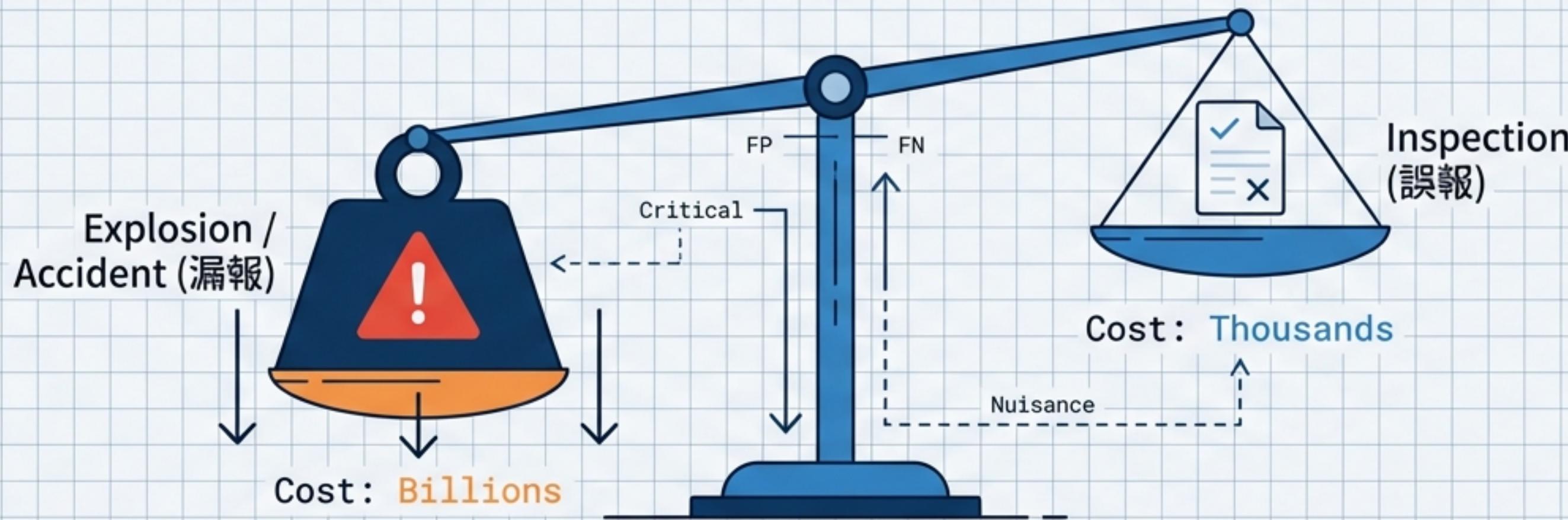
無標籤評估策略 (Evaluation without Labels)

Human-in-the-loop Validation



Industrial Blueprint v1

化工實務考量：成本不對稱性 (Cost Asymmetry)



策略：容忍較高誤報 (FP) 以換取極低漏報 (FN)。

告警疲勞 (Alarm Fatigue) 對策



時間窗口抑制
(1 alert / 10 mins)

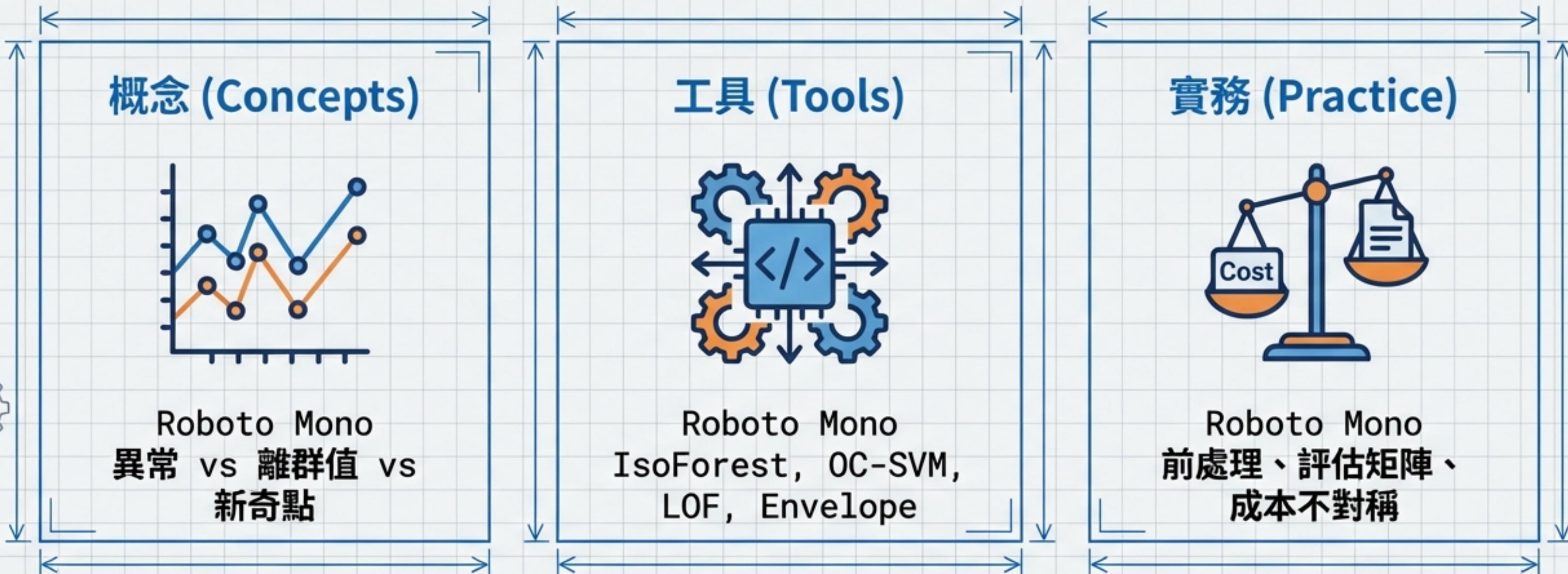


多變數確認
(Multi-variable Confirmation)

異常檢測實務檢核清單 (Implementation Checklist)



單元總結與下一步 (Summary & Next Steps)



下一步 (Next Steps) : Unit 08: 實作演算法 (Implementation)

Unit 09: 時間序列異常檢測 (Time Series Anomaly Detection)