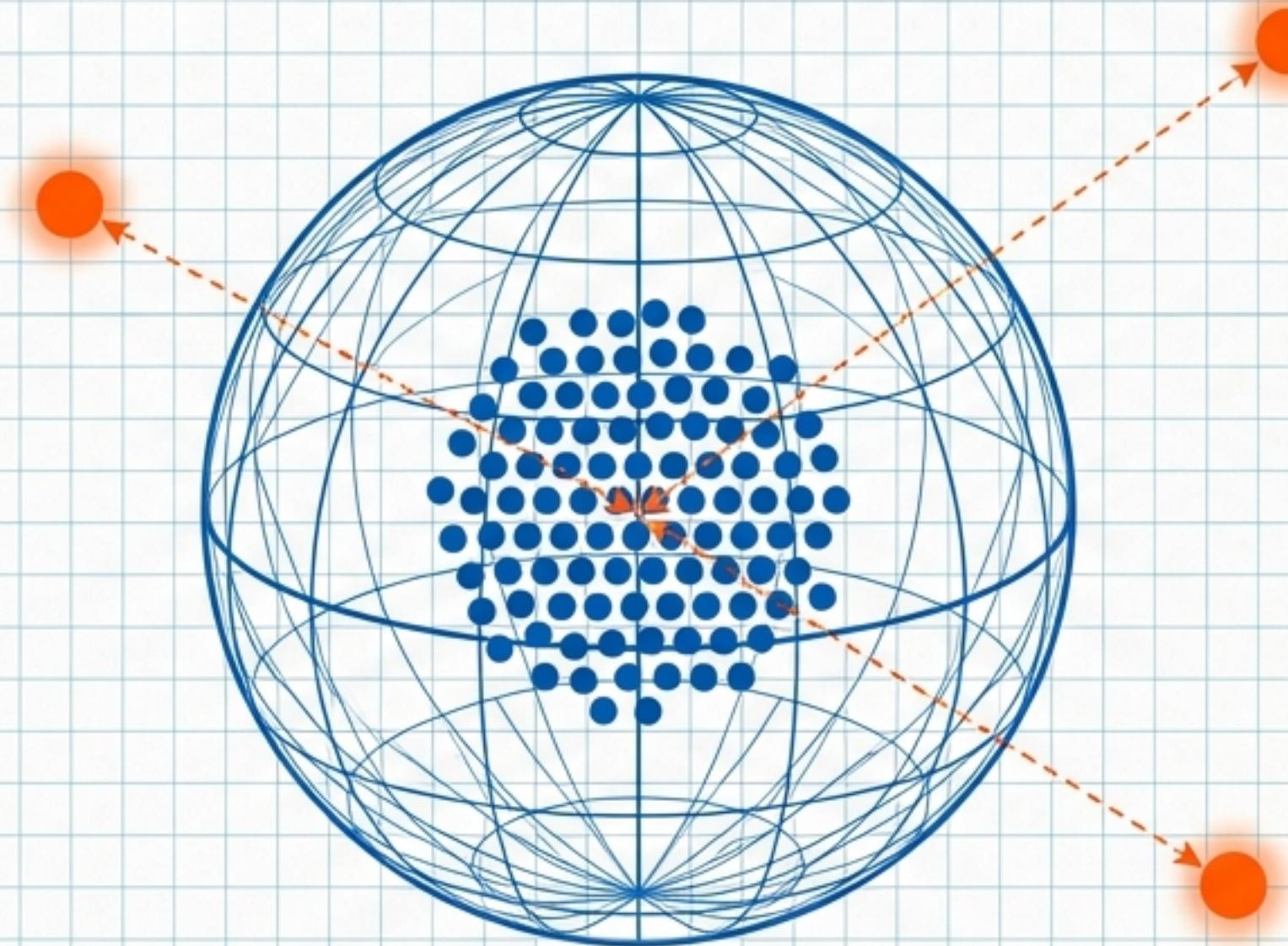


AI 在化工上的應用：從原理到實踐

Unit 07: 一類支持向量機 (One-Class SVM)

基於支持向量理論的精確異常檢測



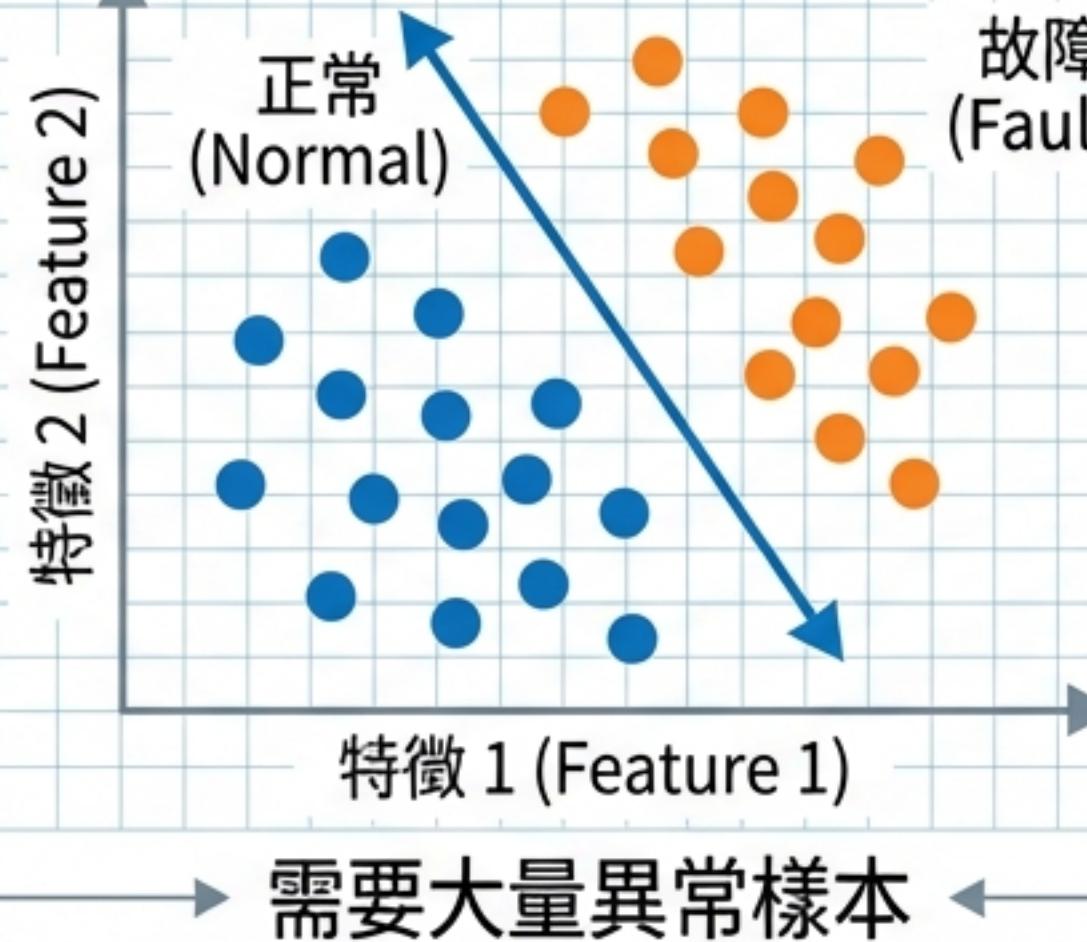
授課教師：莊曜禎 助理教授

學期：114學年度第2學期

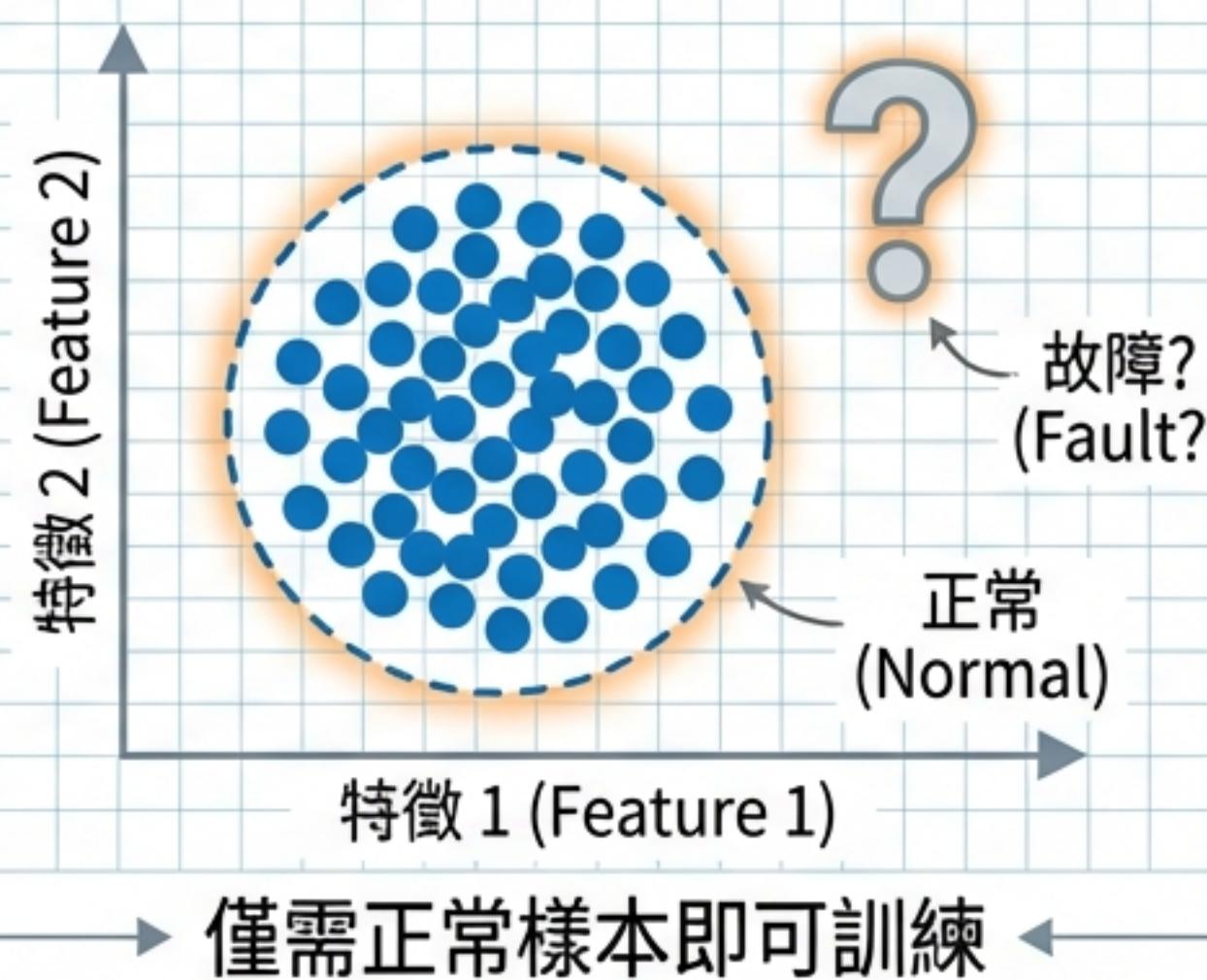
異常檢測的挑戰：數據不平衡

定義『正常』，是為了暴露『異常』。

二元分類 (Binary Classification)

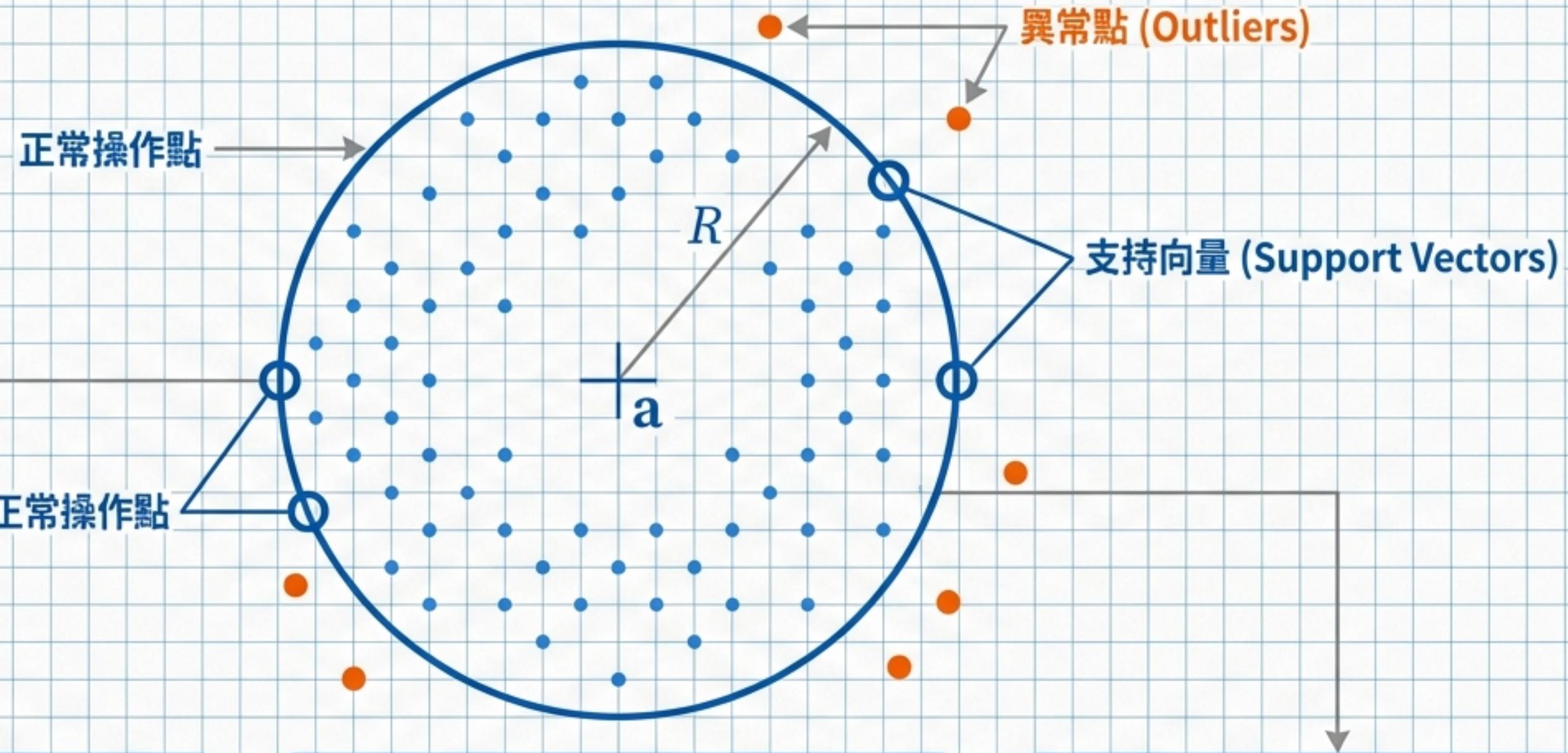


一類分類 (One-Class Classification)



- 化工現狀：正常數據充足 (Millions)，異常數據稀少 (Few)
- 核心理念：在特徵空間中找到一個最小的超球面，將大部分正常數據包含在內

核心原理：超球面決策邊界



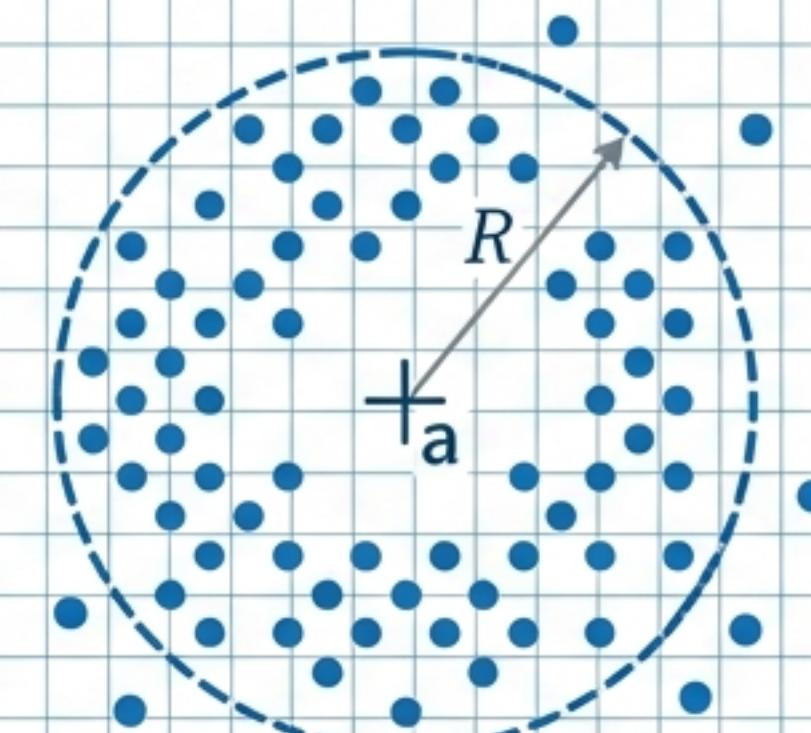
目標 (Goal)：最小化半徑 R^2 ，尋找包圍正常數據的最緊湊邊界。

支持向量 (Support Vectors)：位於邊界上的關鍵數據點，它們如同建築的柱子，獨自決定了邊界的形狀。

應用場景 (Application)：反應器操作空間、高價值產品品質監控。

數學引擎：兩種幾何觀點

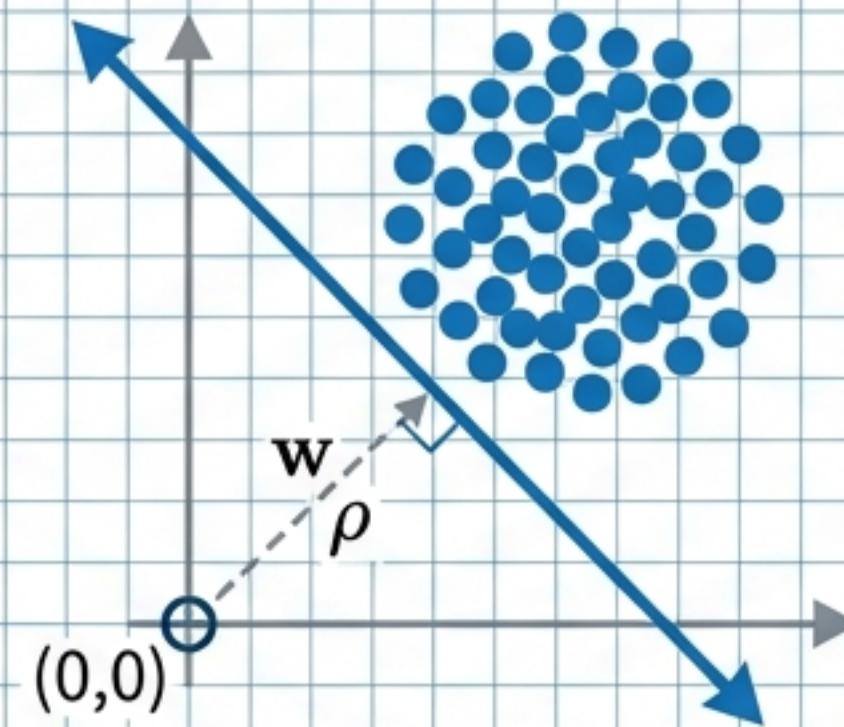
超球面方法 (Hypersphere Approach)



$$\min_{R,a} R^2 + C \sum \xi_i$$

尋找最小半徑

超平面方法 (Hyperplane Approach)



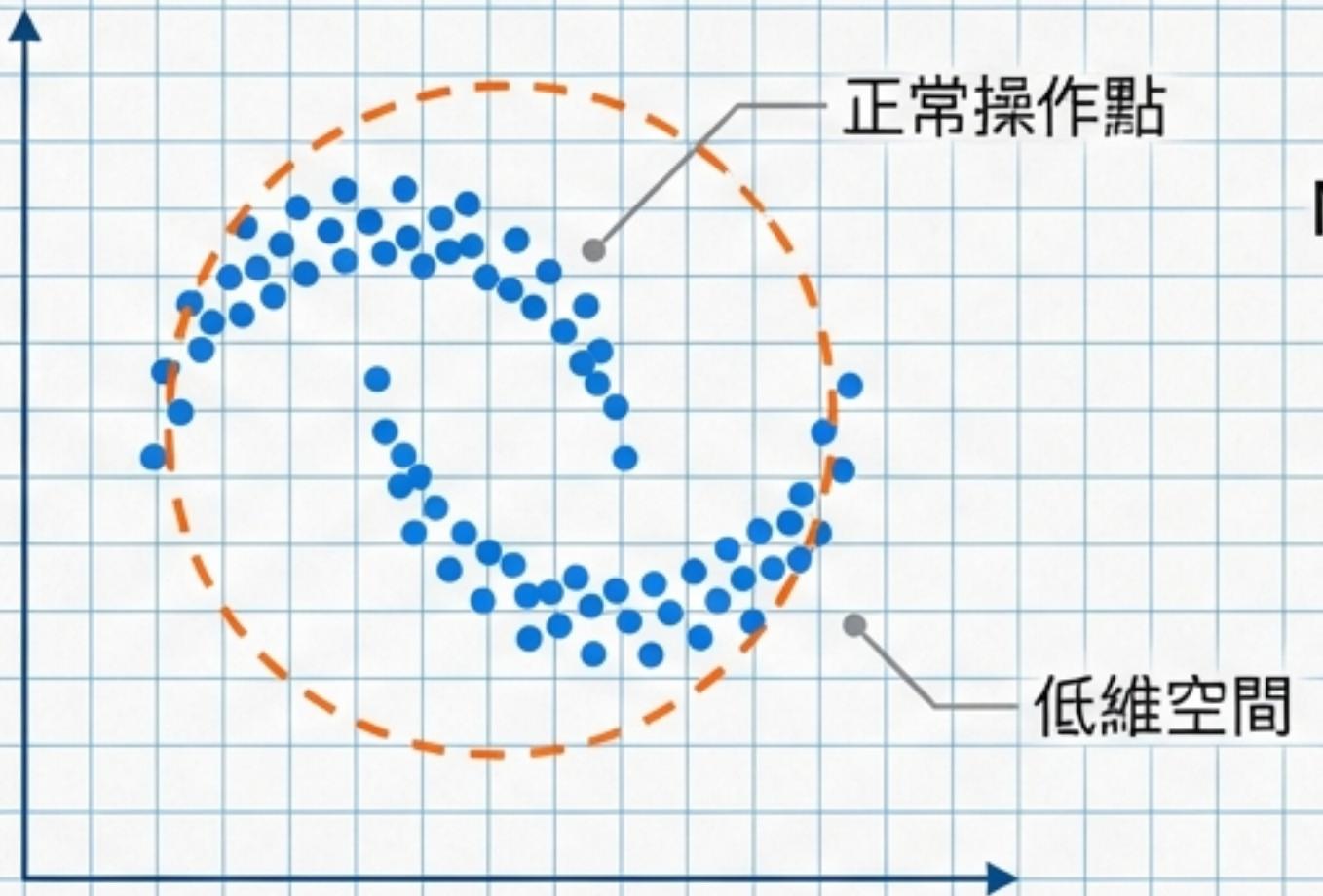
$$\min_{\mathbf{w}, \rho} \frac{1}{2} \|\mathbf{w}\|^2 - \rho$$

最大化與原點的距離

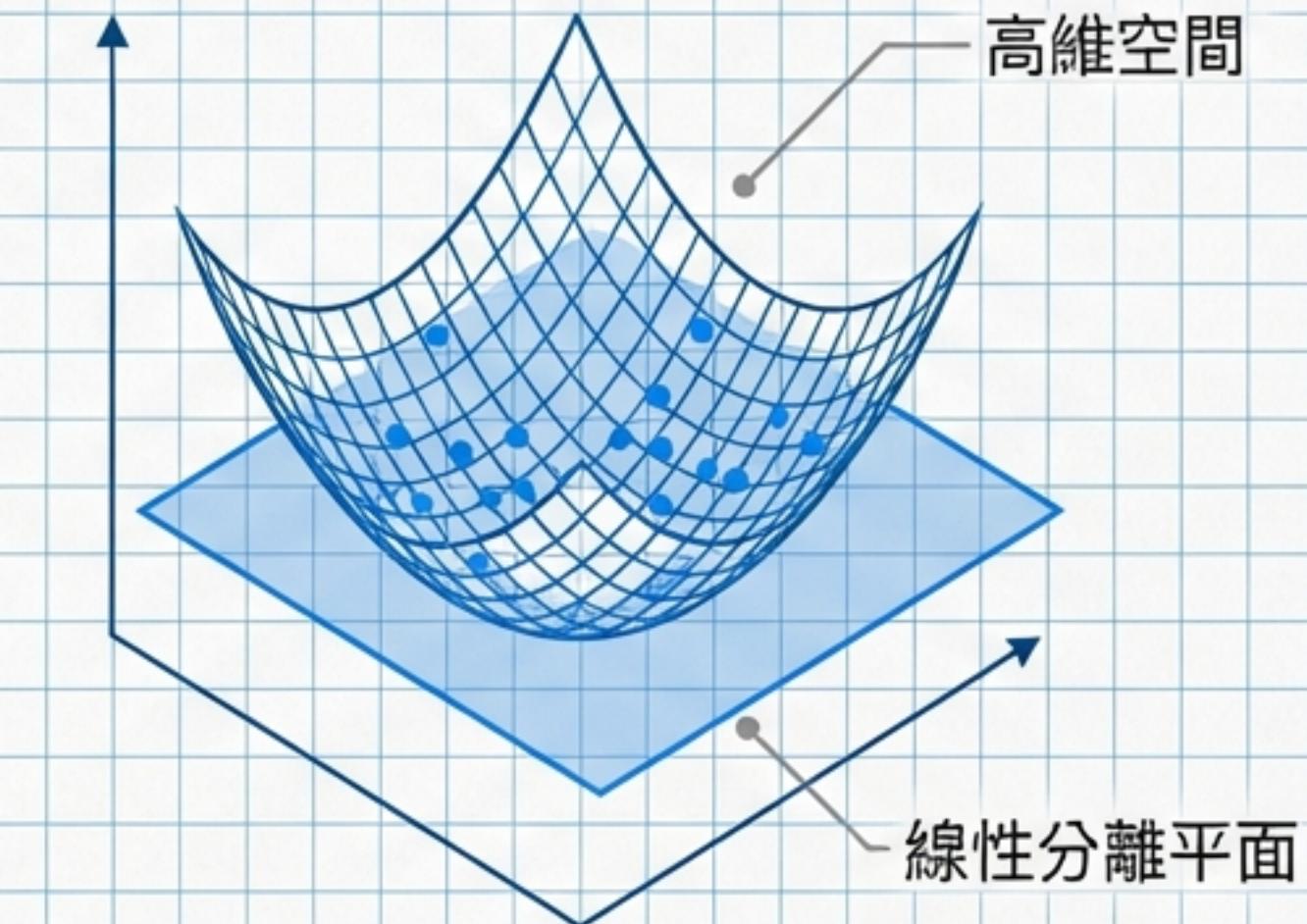
判定法則：決策函數 $f(\mathbf{x}) < 0$ 即判定為異常

核技巧 (Kernel Trick)：處理非線性化工製程

低維空間不可分 (In數不可分)



Mapping $\phi(x)$



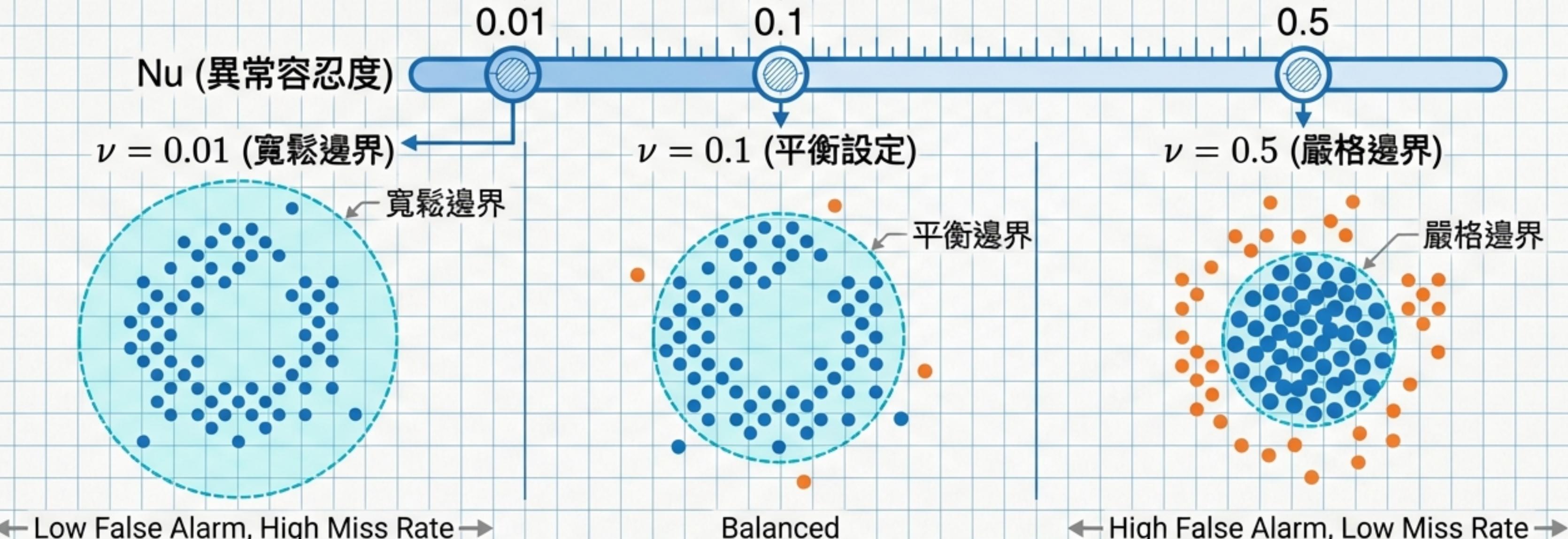
RBF 核函數 (Gaussian Kernel) : $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2)$

化工應用 (Chemical Engineering Application) : 適合處理非線性操作空間 (如溫度-壓力耦合關係)。

線性核 (Linear Kernel) : 僅適用於高維感測器數據降維。

參數調校 1： ν (Nu) - 異常容忍度

$\nu \in (0, 1]$ 控制支持向量的比例下界與訓練誤差的上界

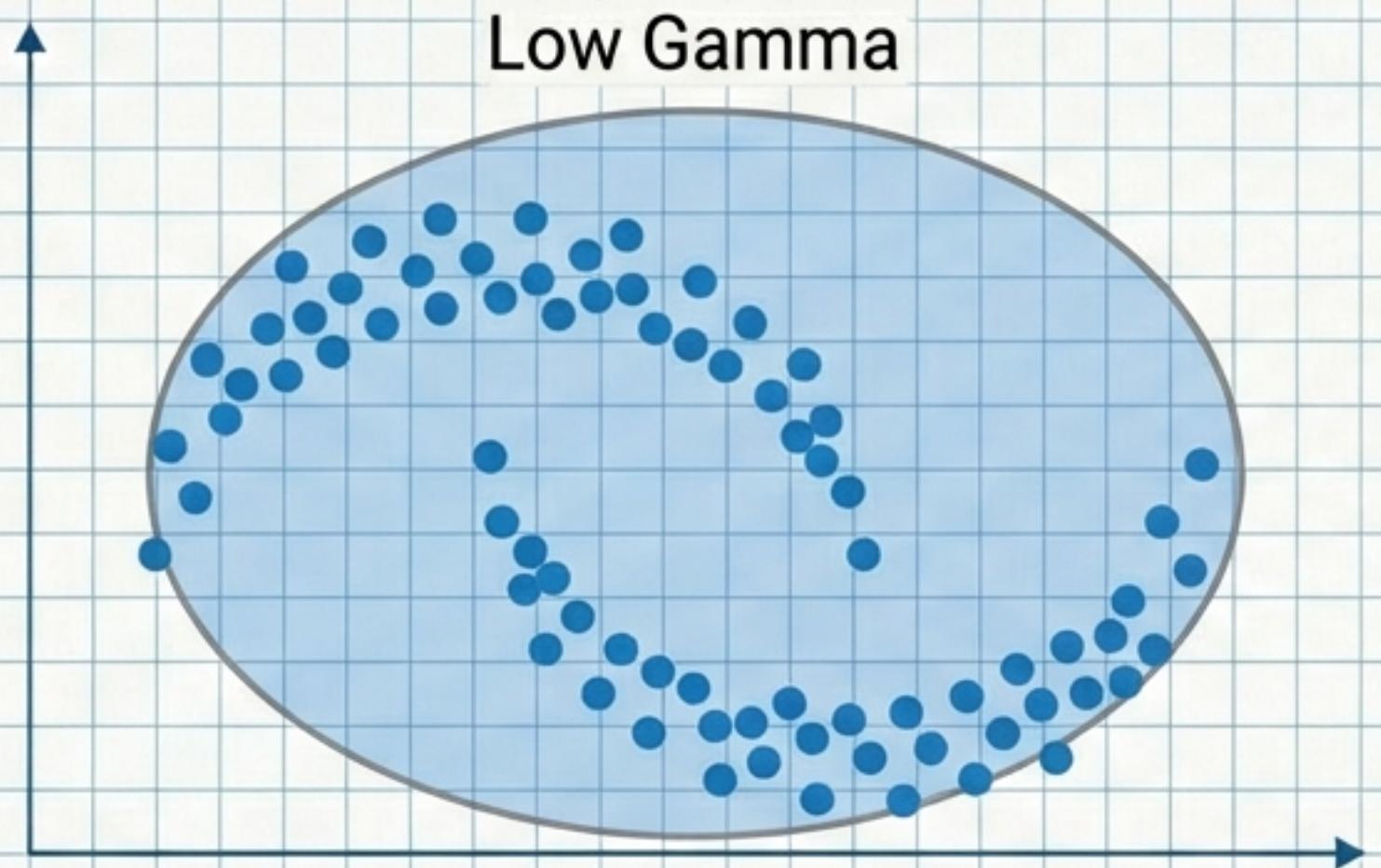


化工應用建議：

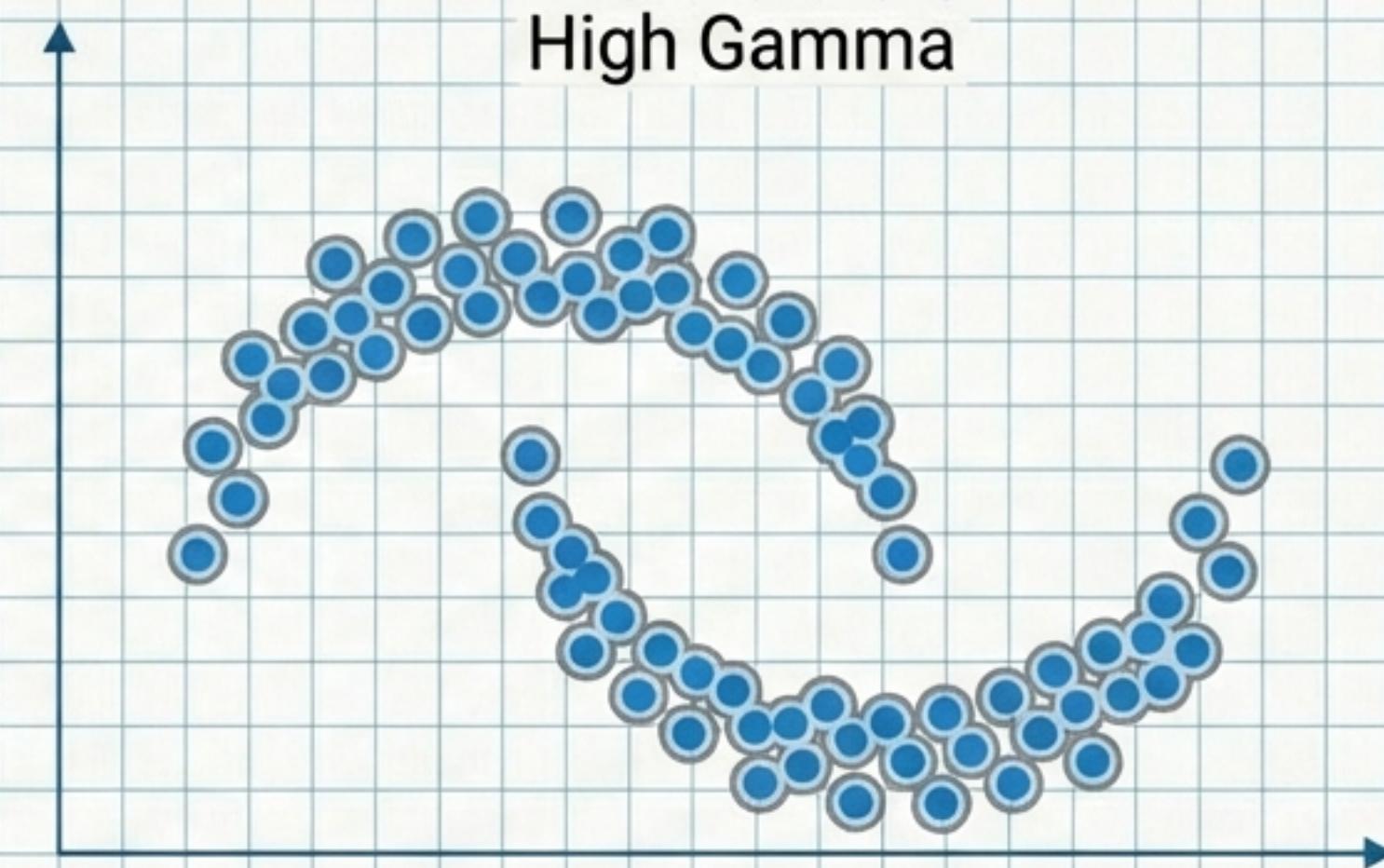
- 穩定製程： $\nu = 0.01 \sim 0.05$ (預期異常率 1-5%)
- 新製程/複雜製程： $\nu = 0.05 \sim 0.1$

參數調校 2 : γ (Gamma) - 邊界複雜度

控制單個訓練樣本的影響範圍



平滑 (Smooth) - 泛化能力強



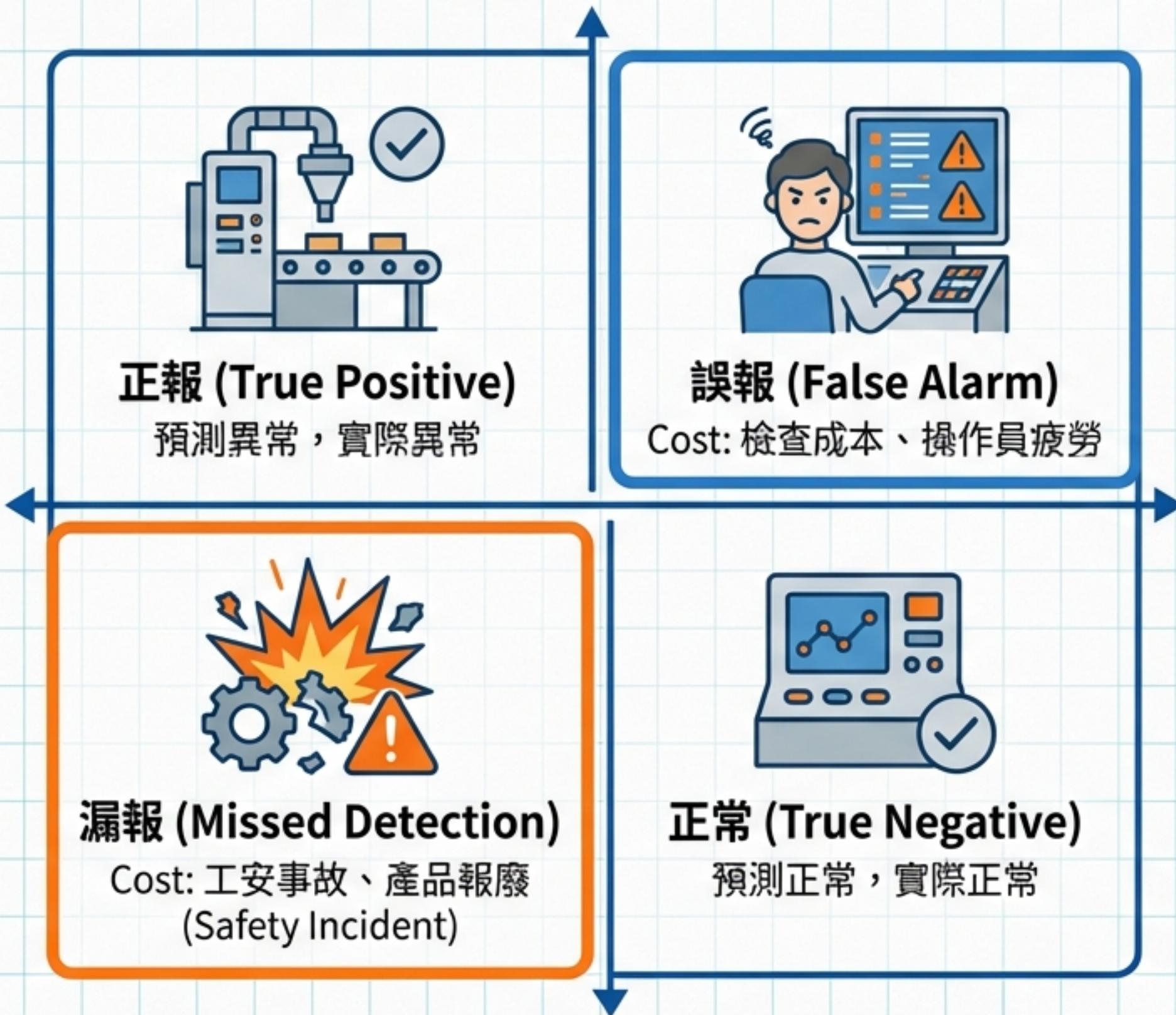
過擬合 (Overfitting) - 僅記憶數據 (Island Hopping)

最推薦設定性箱：

推薦設定：使用 `gamma='scale'`

公式： $\gamma = \frac{1}{n_{\text{features}} \times \text{Var}(X)}$ (自動根據特徵變異數調整)

模型評估：從混淆矩陣到化工風險



評估指標與策略

$$\text{Precision} = \frac{\text{TP}}{(\text{TP}+\text{FP})} \quad (\text{告警可信度})$$

$$\text{Recall} = \frac{\text{TP}}{(\text{TP}+\text{FN})} \quad (\text{異常捕捉率})$$

應用策略

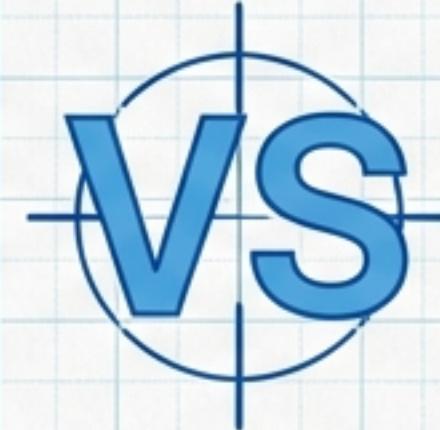
- **高風險製程 (High Risk)**：優先提高 Recall (寧可誤報，不可漏報)
- **自動化產線 (Automation)**：優先提高 Precision (避免頻繁停機)

算法選擇：One-Class SVM vs. Isolation Forest



One-Class SVM

- Strength: 精確邊界 (Precise Boundary) ↗, 小樣本 ($< 10k$)
- Weakness: 計算慢 ($O(n^3)$)
↘, 對維度敏感
- Use Case: 批次反應品質監控 (Batch Quality)



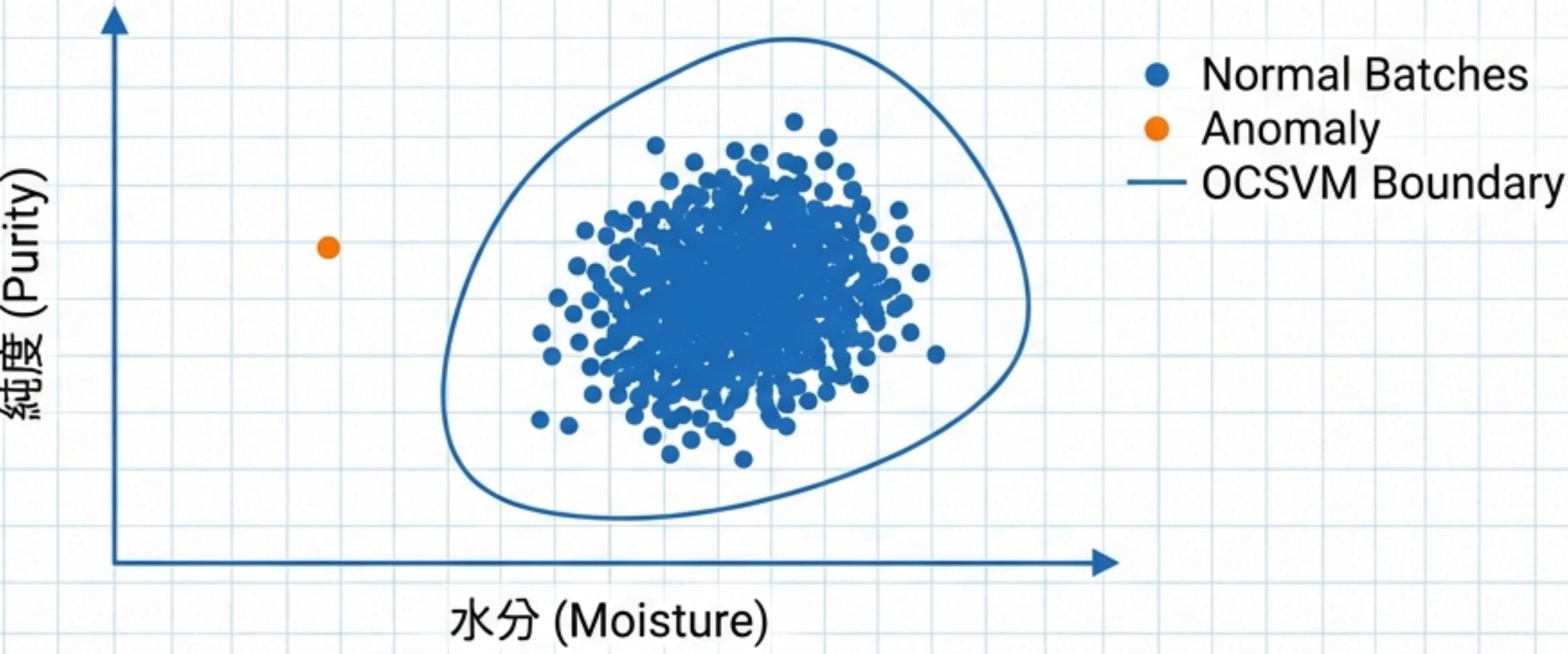
Isolation Forest

- Strength: 速度快 ($O(n \log n)$)
↗ 處理大數據/高維
- Weakness: 邊界較粗糙
↘ (Approximate boundary)
- Use Case: 全廠感測器異常篩查 (Plant-wide Screening)



應用案例 1：高價值產品品質監控

醫藥中間體生產 (僅有 300 批正常數據)

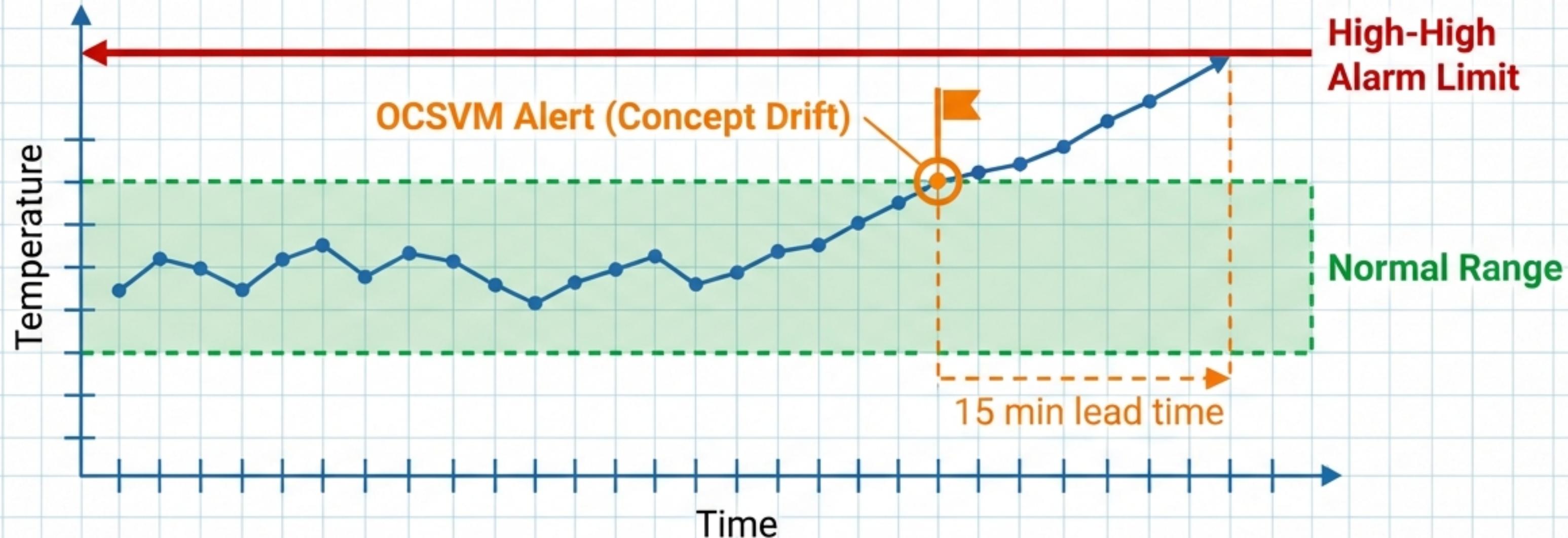


結果 (Results)

設定 $\nu = 0.05$
成功檢測出異常批次 (Recall = 90%)
相比人工抽檢，漏報率降低 50%

應用案例 2：反應器早期預警 (Early Warning System)

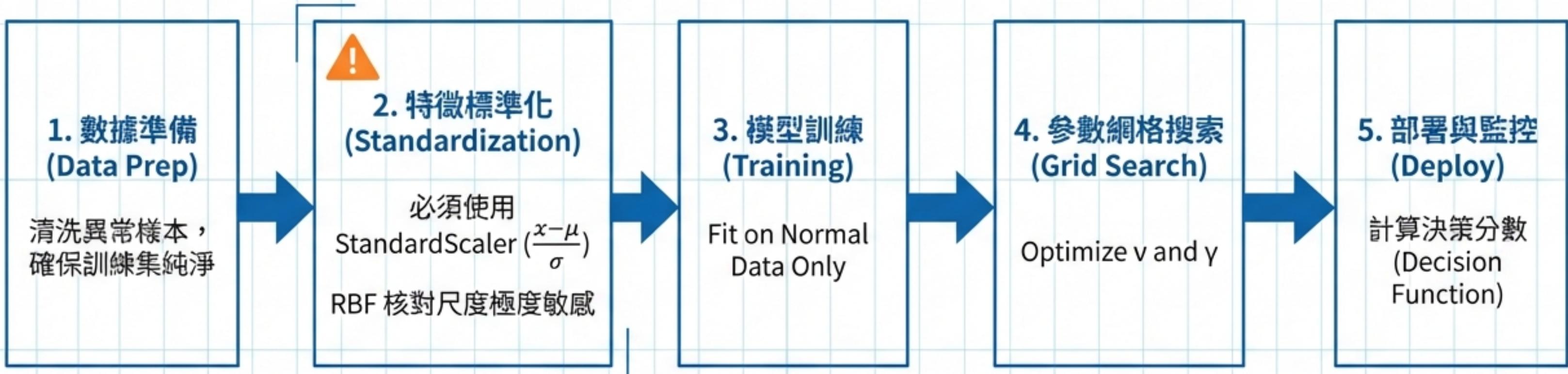
Reactor Temperature



Key Logic Text

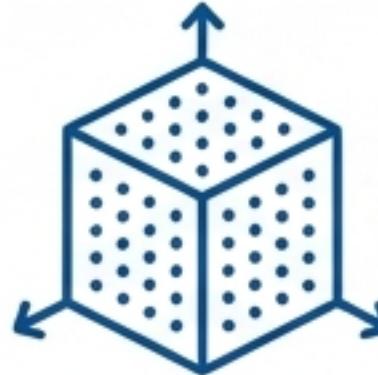
檢測多變數耦合異常：溫度雖未超標，但在當前壓力下已屬異常。

實務工作流程 (Implementation Workflow)



現實世界的挑戰 (Real-World Challenges)

維度災難
(Curse of Dimensionality)



特徵超過 50 維時，RBF 核距離失效。

解決方案 (Solution): →
先進行 PCA 降維。

運算成本
(Computational Cost)



訓練複雜度 $O(n^3)$ 。樣本數 $> 10,000$ 時需謹慎。

解決方案 (Solution): →
使用 Nyström 近似或改用 Isolation Forest。

數據純度
(Data Purity)



訓練數據若混入異常，將導致模型失效 (Garbage In)。

解決方案 (Solution): →
專家預先篩選數據。

總結與下一步 (Summary & Next Steps)

- One-Class SVM 是化工領域處理 **小樣本 (Small Data)** 與 **非線性 (Non-linear)** 異常檢測的黃金標準。
- 成功的關鍵在於：正確的 **標準化 (Scaling)** 與 **參數調校 (Tuning)**。



AI 不會取代工程師，但懂得使用 AI 的工程師將取代不懂的人。

