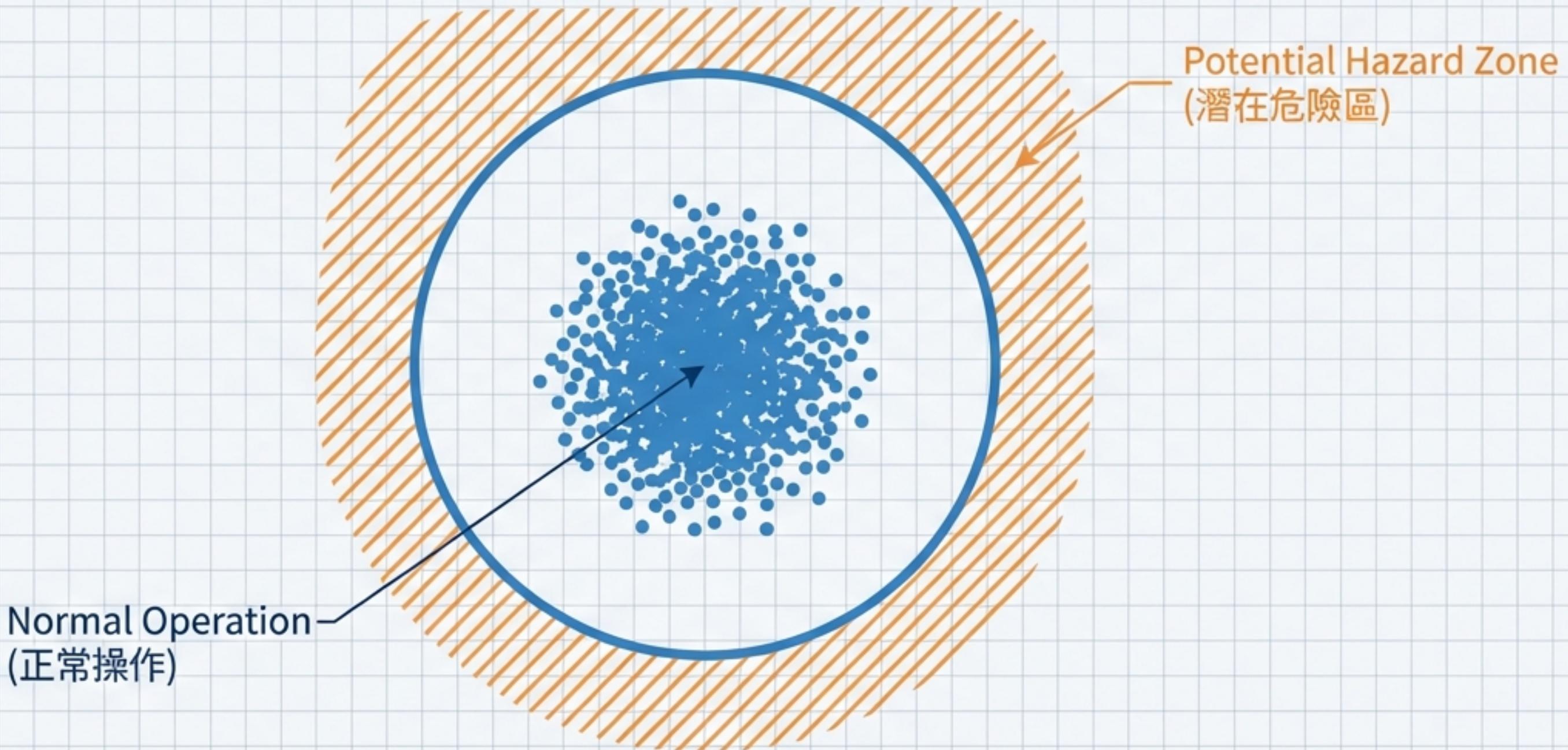


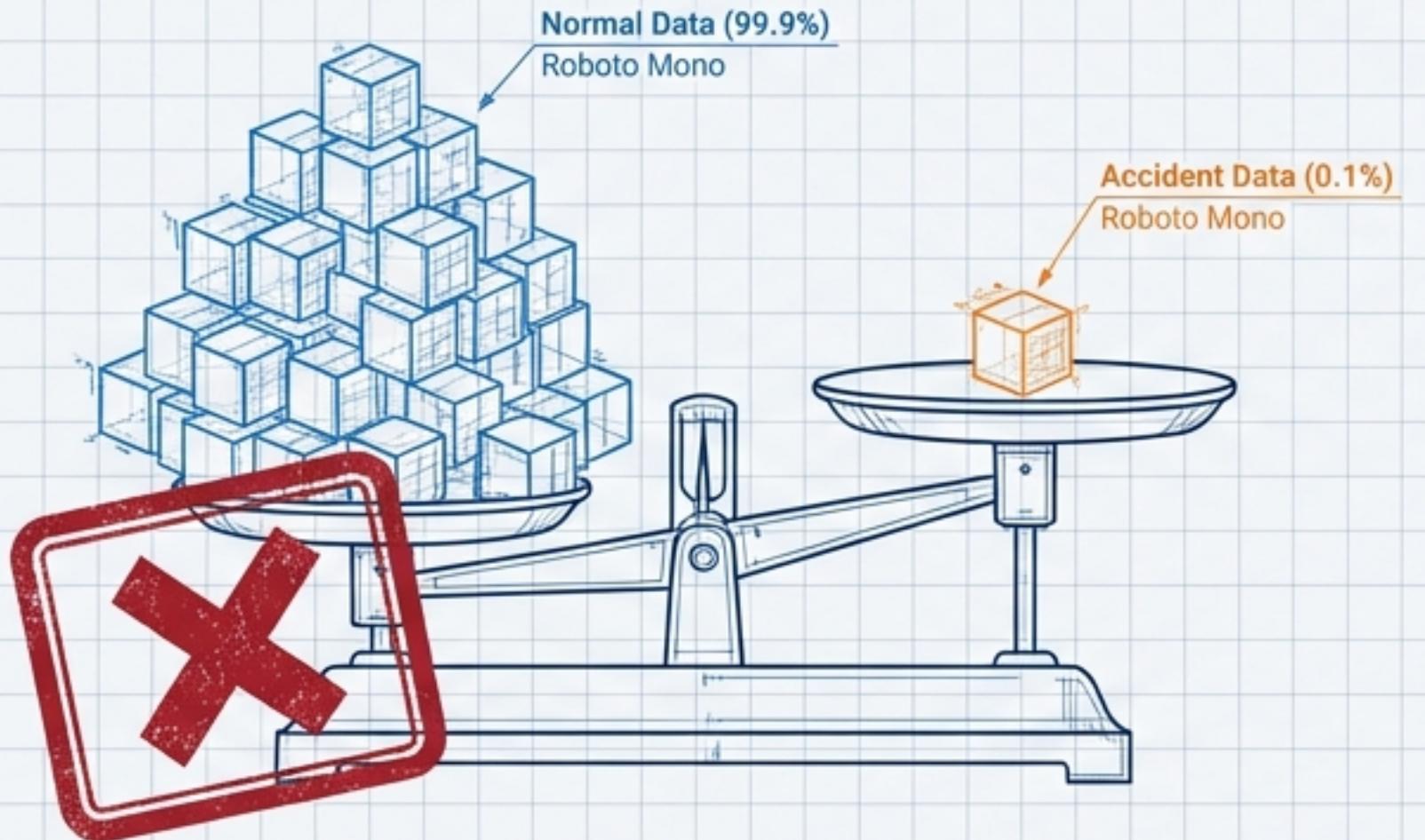
Unit 07: One-Class SVM 單類支持向量機

為化工製程建立精確的異常檢測邊界



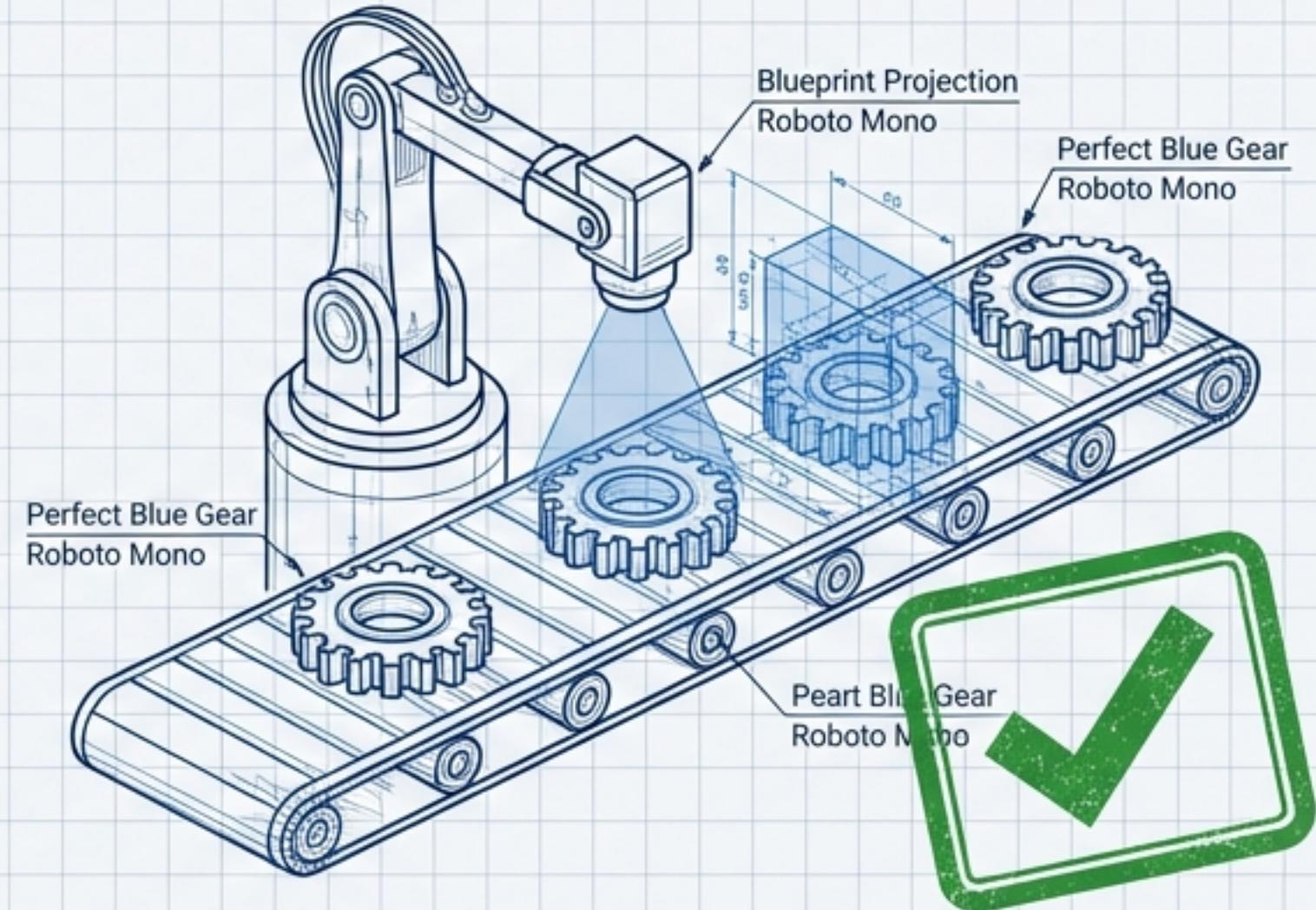
工程師的難題：如何在沒有事故數據時預測事故？

傳統分類器的局限



需要同時擁有正常與異常標籤

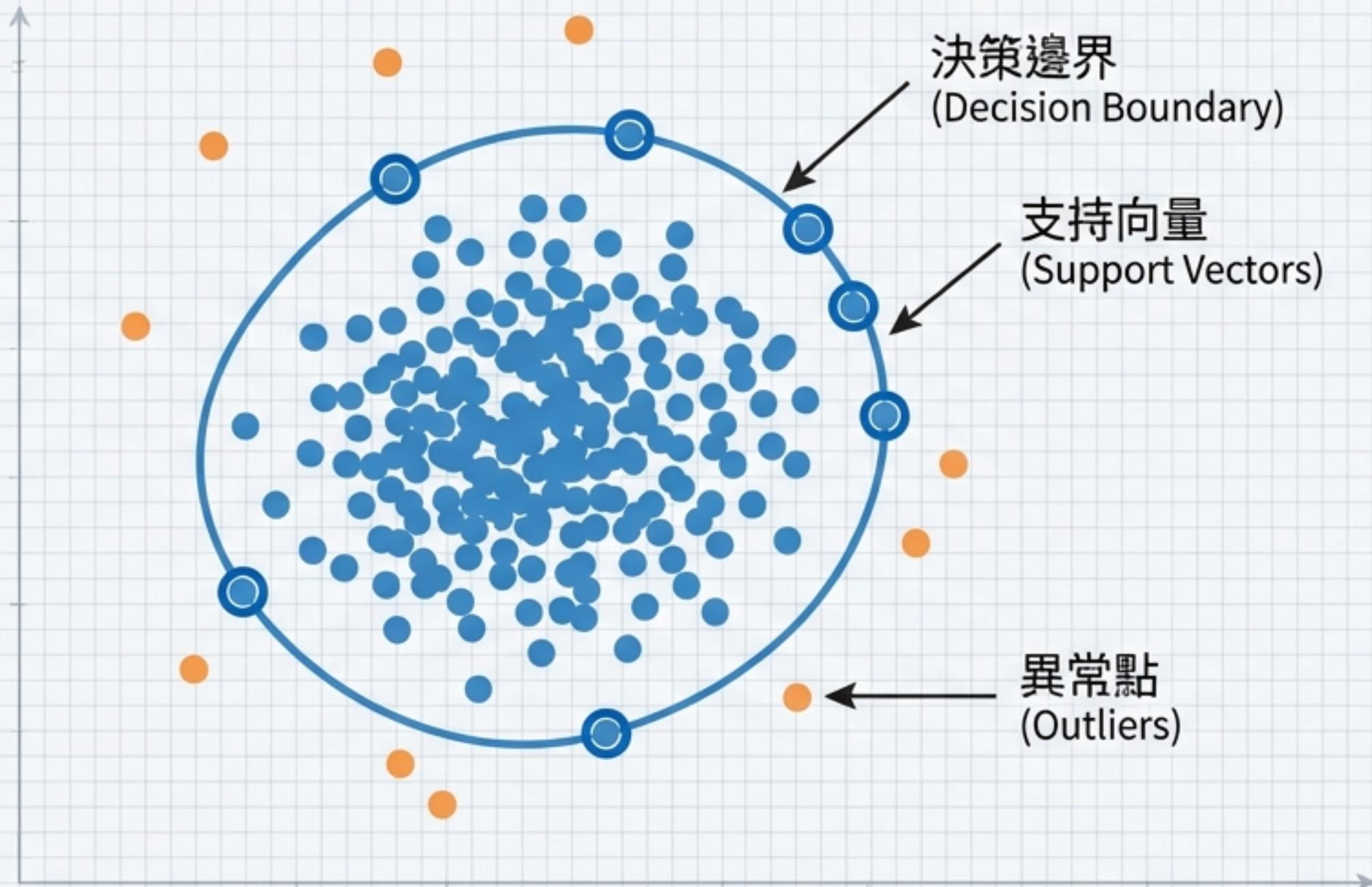
One-Class SVM 的策略



僅需正常數據進行訓練

適用場景：正常數據充足，但異常數據稀少或未知的高純度製程數據。

核心原理：打造數據的「安全容器」



超球面方法 (Hypersphere)

尋找包含數據的最小體積球體
(中心 a ，半徑 R)。

超平面方法 (Hyperplane)

尋找將數據與原點隔開的最大
間隔超平面。

處理非線性操作空間：核技巧 (Kernel Trick)

將低維非線性數據映射到高維空間以實現線性分割。

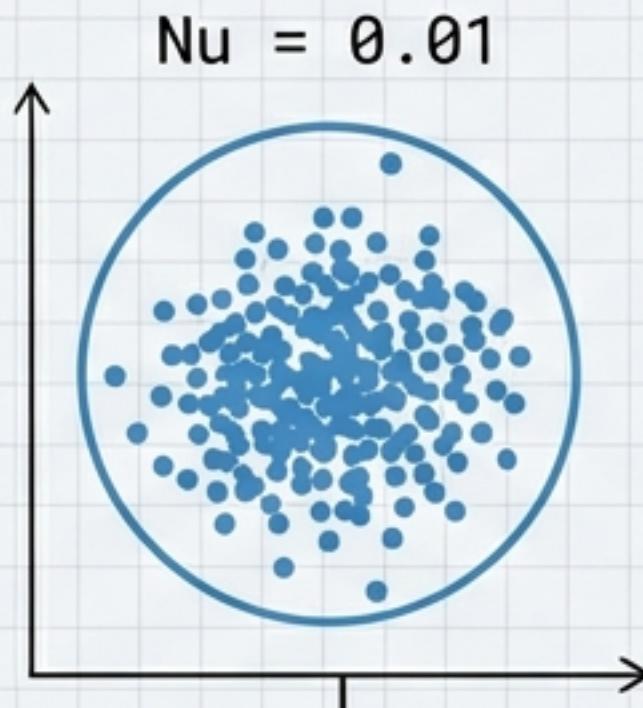
Kernel Type	Visual Concept	Characteristics
Linear (線性核)		簡單快速，適合高維感測器數據 (>50 features)。
RBF (高斯核)		化工首選。能建立複雜曲面邊界，適合反應器操作空間。
Polynomial / Sigmoid		特殊用途，如模擬神經網路。

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma * ||\mathbf{x}_i - \mathbf{x}_j||^2)$$



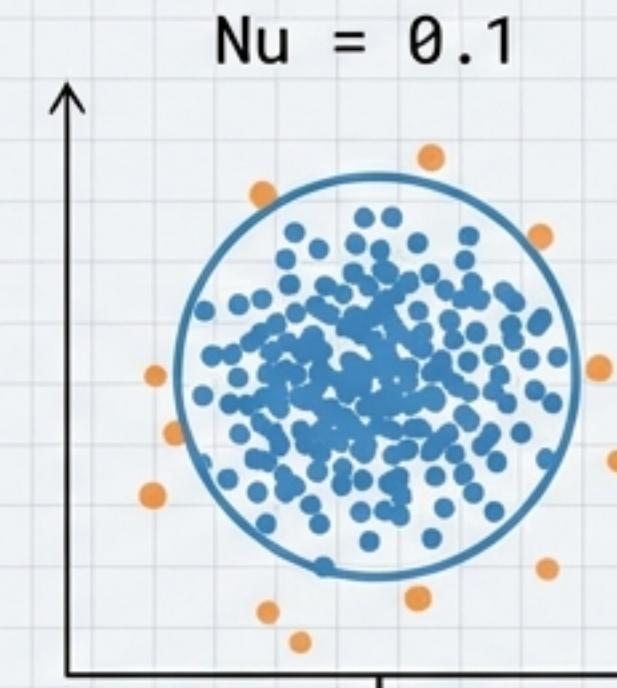
調校旋鈕 1：參數 Nu (v) - 定義容忍度

控制支持向量的數量上限與訓練誤差的下界。



寬鬆邊界 (Loose Fit)

優點：誤報極少
缺點：可能漏報微小異常
適用：連續穩定製程



嚴格邊界 (Tight Fit)

優點：捕捉所有偏差
缺點：誤報率較高
適用：高風險反應器 (Safety First)

Nu Value



Low Nu

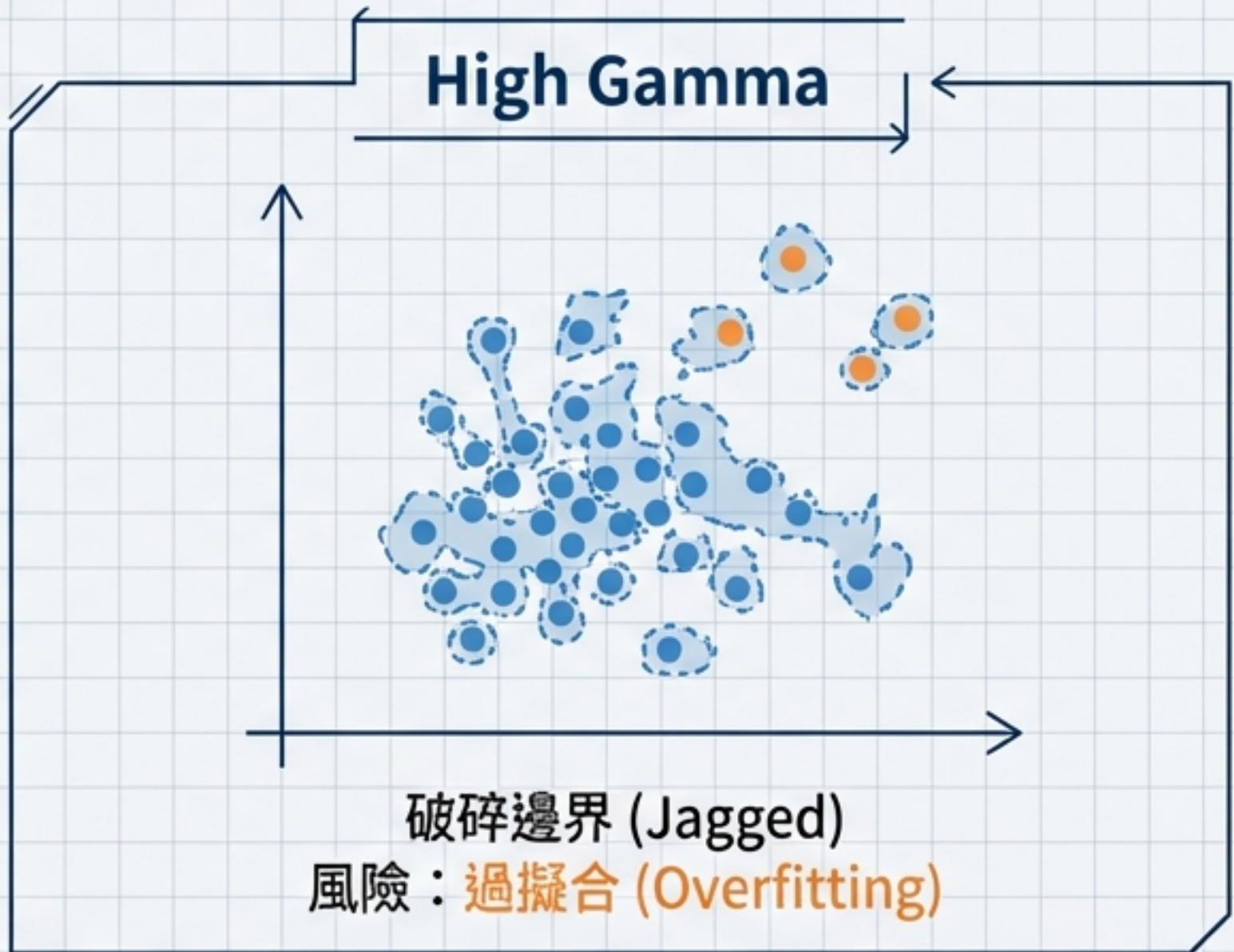
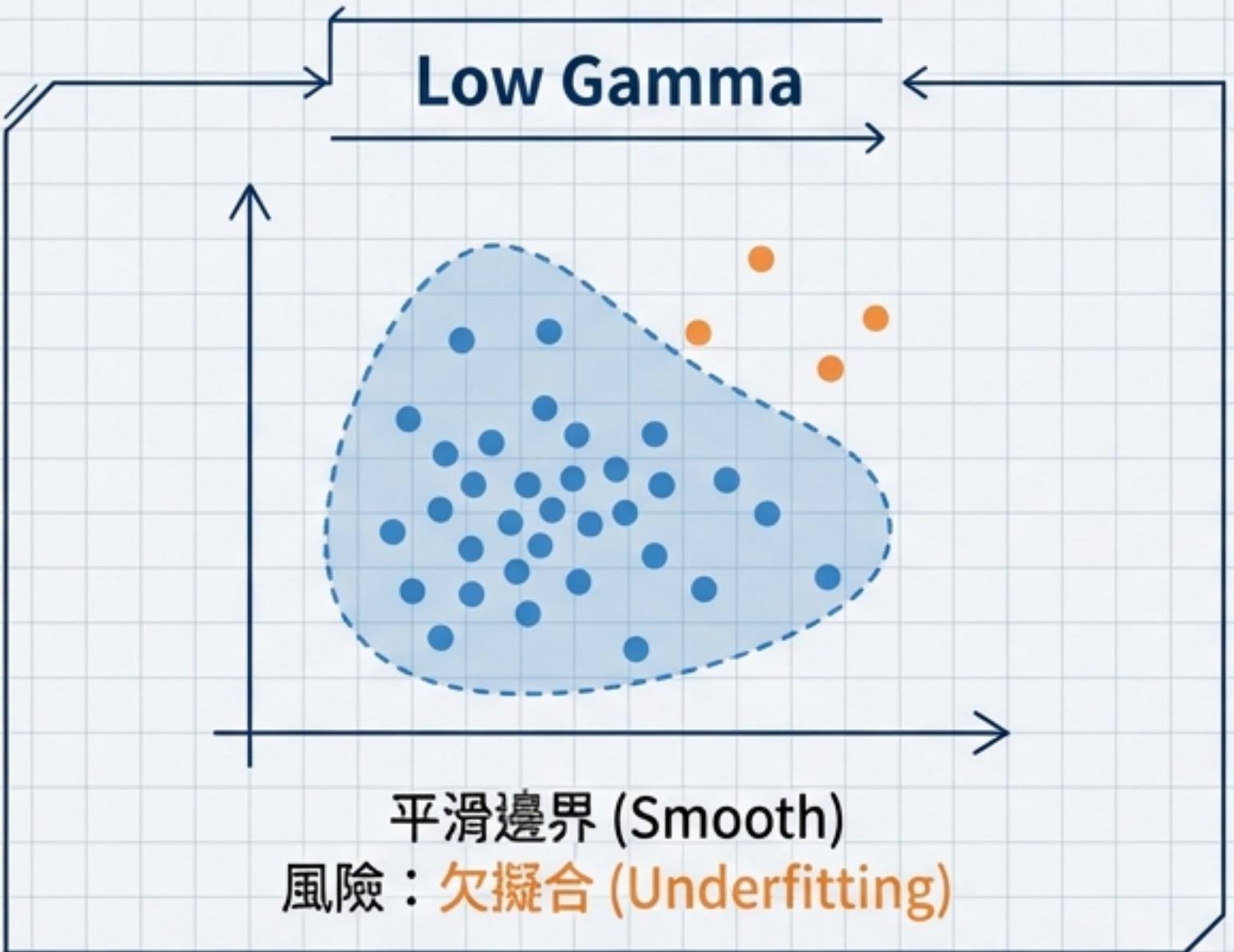
High Nu



經驗法則：將 Nu 設定為預期的異常污染比例（例如 0.05 代表 5% 異常）。

調校旋鈕 2：參數 Gamma (γ) - 控制影響範圍

僅適用於 RBF Kernel，控制單個樣本的影響力半徑。



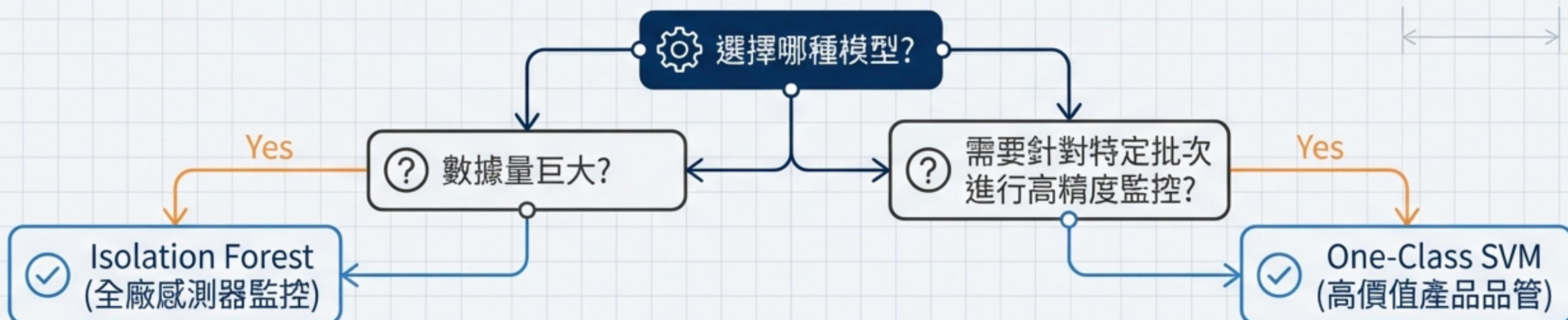
建議設定 (Recommended Strategy)：
使用 `gamma='scale'` 由演算法自動計算。
`gamma = 1 / (n_features * Var(X))`



注意：High Gamma + High Nu = 模型
過於複雜，難以泛化。

系統規格比較：OCSVM vs. Isolation Forest

Feature (特徵)	OCSVM	Isolation Forest (iForest)
Data Size (數據規模)	中小規模 (< 10,000 samples)	大規模 (Massive)
Dimensions (特徵維度)	低維度 (< 50 features)	高維度 (> 100 features)
Precision (精確度)	高 (Deterministic Boundary)	近似 (Stochastic)



品質保證：有標籤數據的評估指標

Predicted (預測)			
		Actual (實際)	
		True Positive (正報) - 準確警報 (Accurate Alarm)	誤報 (False Alarm) - 操作員忽略警報
TP	FP		FP
FN	FN	TN	True Negative (正負) - 正常操作 (Normal Operation)
漏報 (Missed Detection) - 安全事故發生			
Predicted (預測)			



Precision (精確率)

可信度。警報響起時，多大機率是真的？
適用：誤報成本高的場景（避免停機）



Recall (召回率)

安全性。我們抓住了多少比例的異常？
適用：安全關鍵場景（如硝化反應）



F1-Score

兩者的平衡點。

$$F1 = 2 * \frac{(Precision * Recall)}{(Precision + Recall)}$$

盲測驗證：無標籤數據的評估策略

當沒有歷史故障數據時 (Unsupervised) 的應對方法。



應用案例 1：高價值產品品質監控 (Batch Process)

項目細節 (Project Details)



場景：醫藥中間體生產



數據：300 批正常，10 批異常

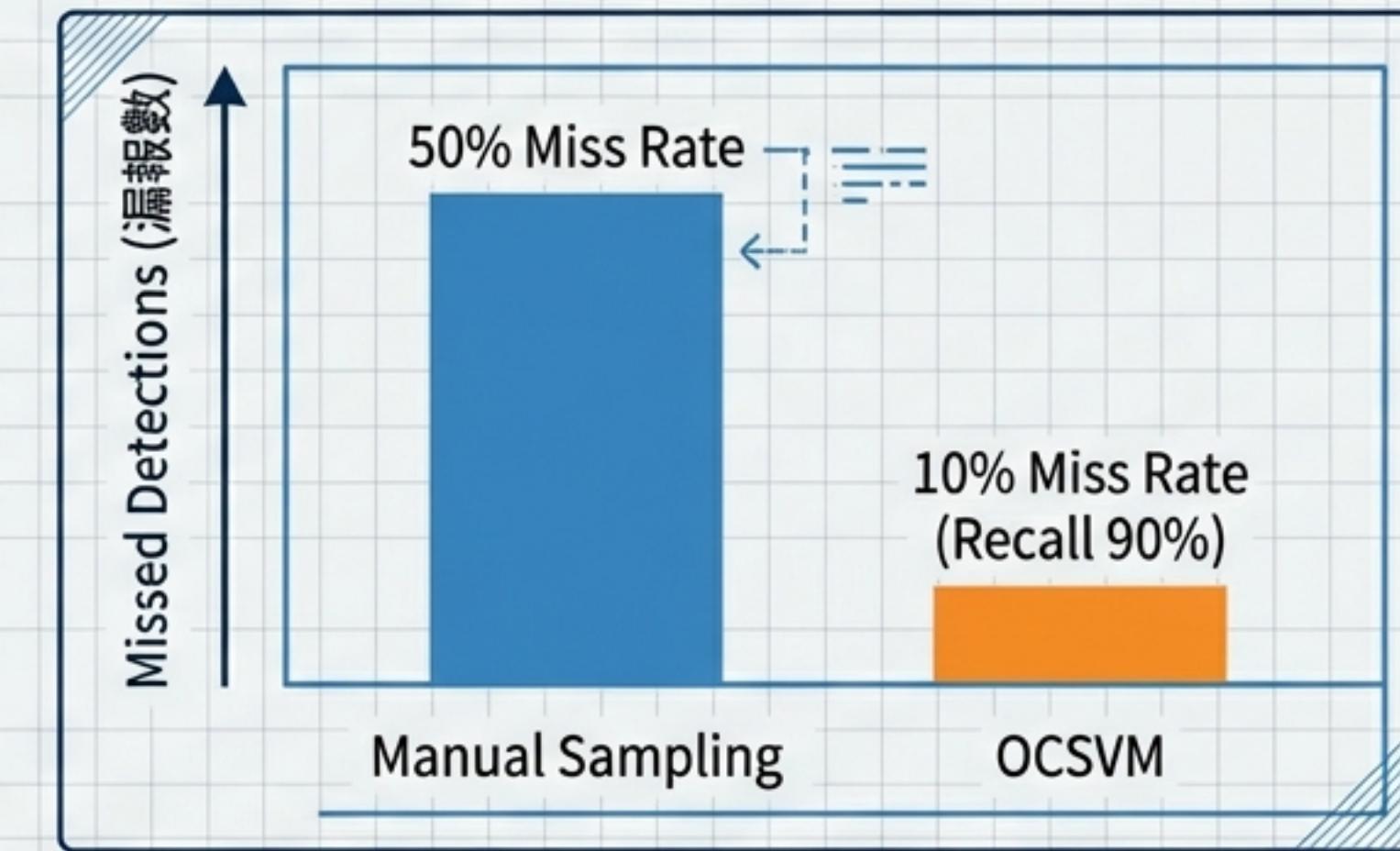


特徵：8 個品質指標 (純度、水分、雜質...)



設定：RBF Kernel, Nu = 0.05

結果可視化 (Results Visual)



成功攔截 9/10 個不良批次，取代昂貴的人工全檢。

結論：顯著提升品質保障能力，降低成本。

應用案例 2：關鍵設備預測性維護 (Continuous Process)

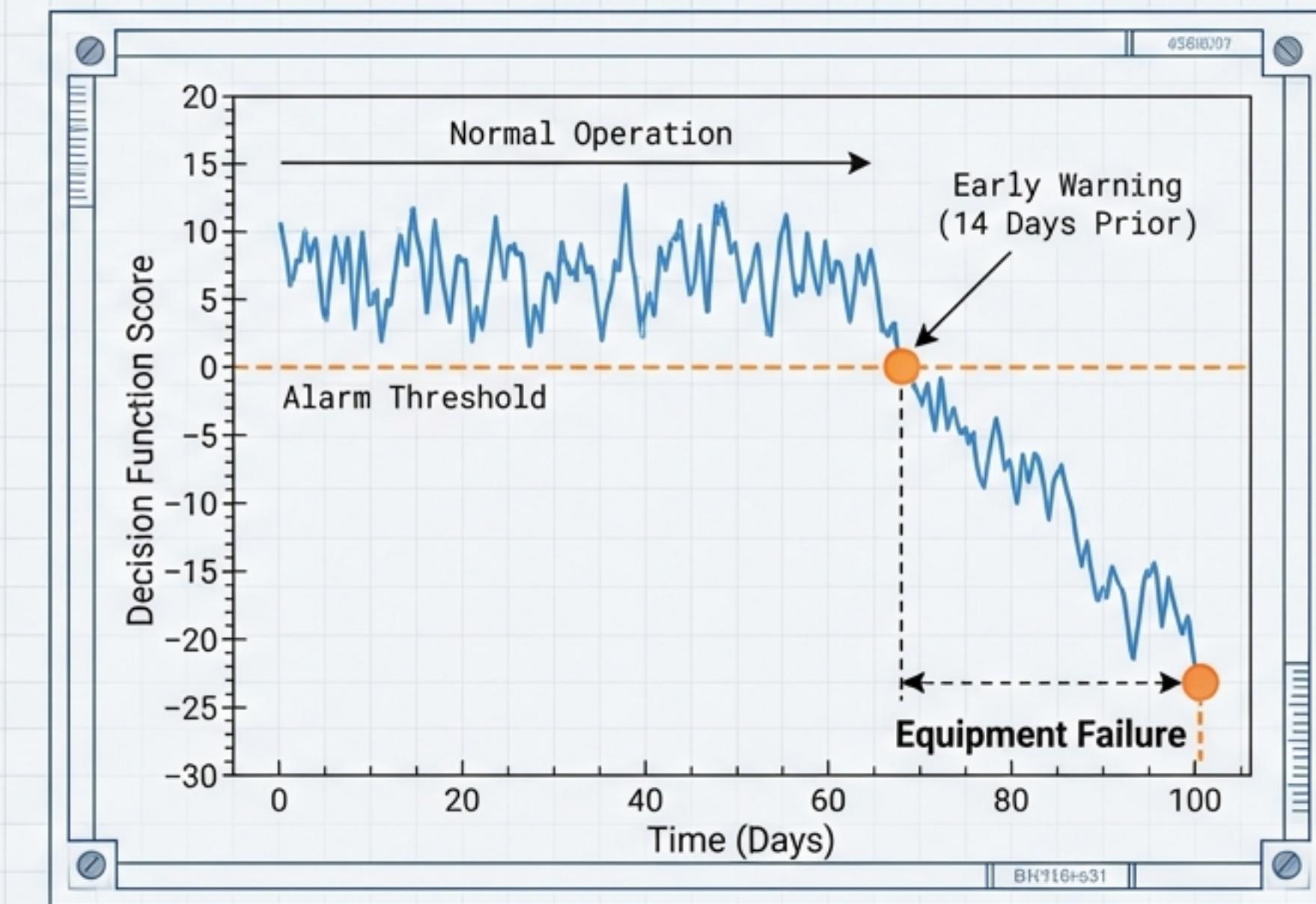
項目細節 (Project Details)

 **場景：**離心泵振動監測
(Pump Vibration)

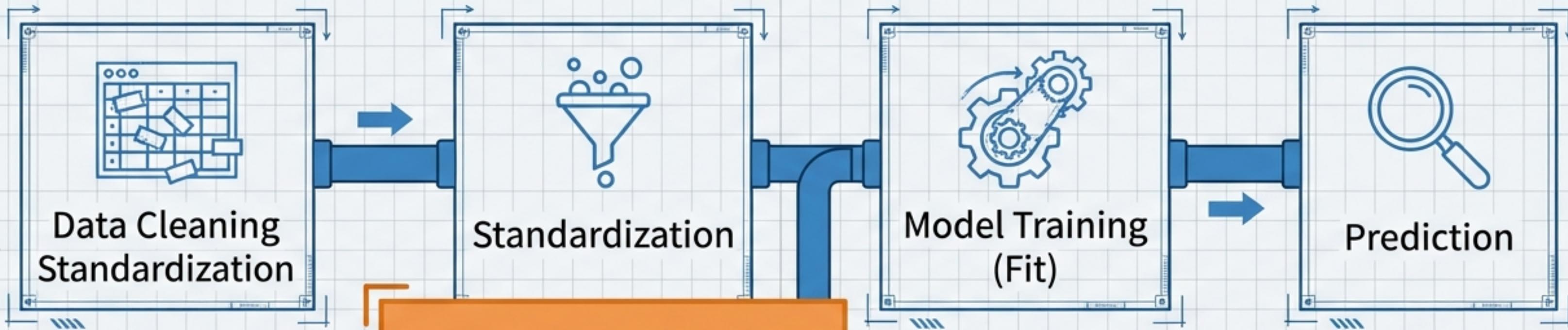
 **策略：**僅使用健康數據訓練

 **效益：**每年節省 200 萬台幣
維護成本

結果可視化 (Results Visual)



實作藍圖：標準化是成功的關鍵



CRITICAL: StandardScaler
(Mean=0, Std=1)

為什麼？RBF Kernel 依賴歐式距離。若不標準化，壓力 (100000 Pa) 的數值影響力將完全蓋過溫度 (100 C)。

Why? Unscaled large values dominate the distance calculation.

程式碼實作：Scikit-learn 核心語法

```
from sklearn.svm import OneClassSVM  
from sklearn.preprocessing import StandardScaler
```

Industrial Blueprint v1
Font: Roboto Mono

```
# 1. Preprocessing (Must do!)  
scaler = StandardScaler()  
X_train_scaled = scaler.fit_transform(X_train)
```

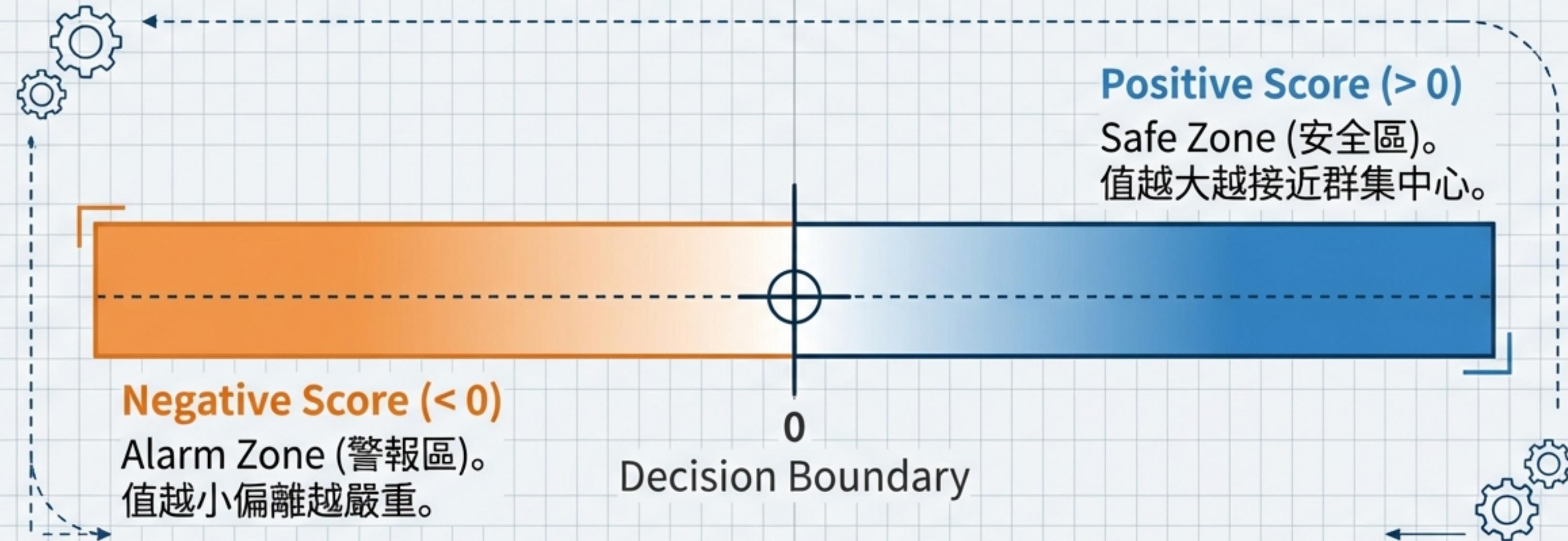
```
# 2. Model Definition  
# nu=0.05 implies ~5% outliers allowed  
model = OneClassSVM(nu=0.05, kernel="rbf", gamma="scale")
```

```
# 3. Training  
model.fit(X_train_scaled)
```

```
# 4. Prediction (1 = Normal, -1 = Anomaly)  
y_pred = model.predict(scaler.transform(X_test))
```



解讀模型輸出：決策分數 (Decision Function)



- Score = -2.5 (Severe Anomaly - Emergency Stop)
- Score = -0.1 (Minor Deviation - Warning)

總結與操作檢查表 (Operational Checklist)

- Data Check:** 訓練數據是否乾淨且為「正常」狀態？
- Scaling:** 是否已應用 StandardScaler？
- Parameter Nu:** 是否根據預期異常率設定 Nu (e.g., 0.05)？
- Kernel:** 非線性製程是否使用 RBF Kernel？
- Validation:** 偵測到的異常是否經過專家確認？

下一步 (Next Step): 開啟 Jupyter Notebook
`'Unit07_OneClass_SVM.ipynb'` 進行實作演練。