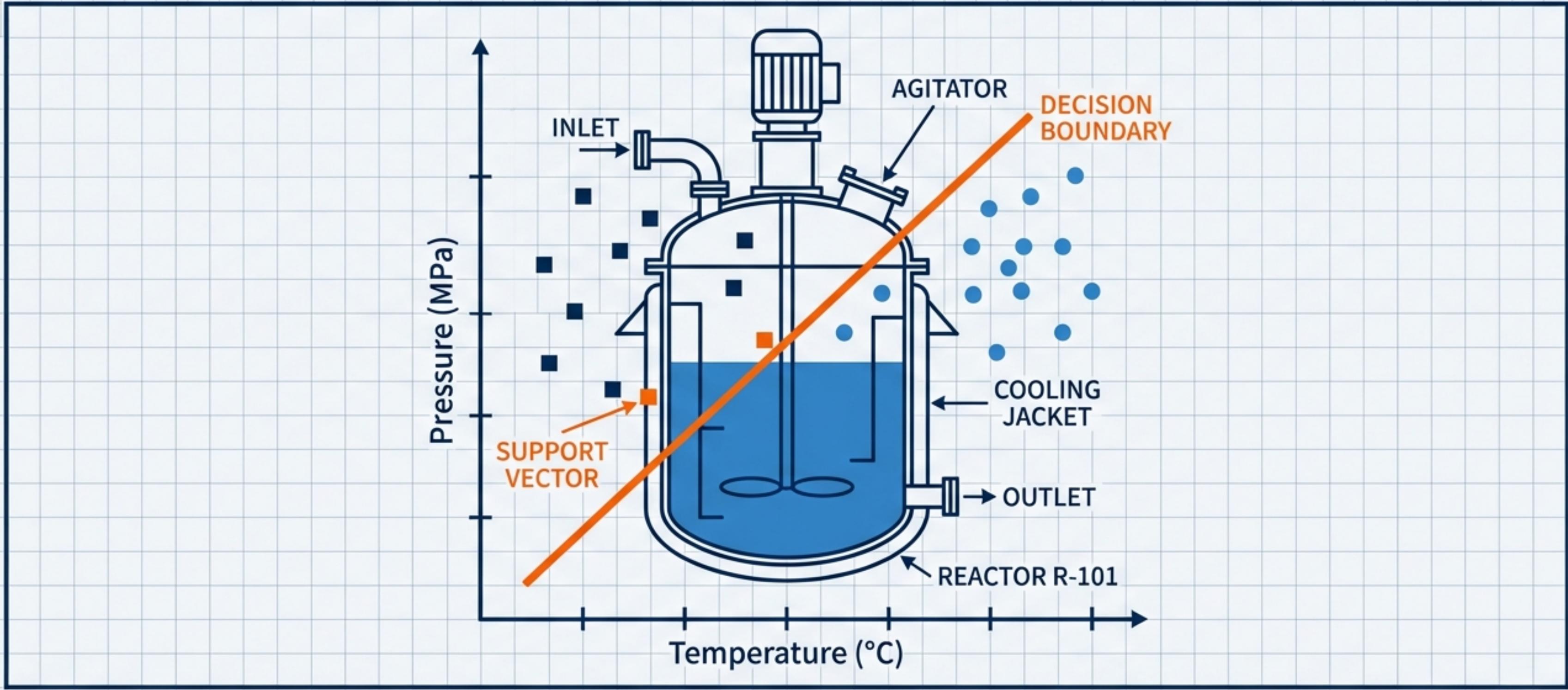
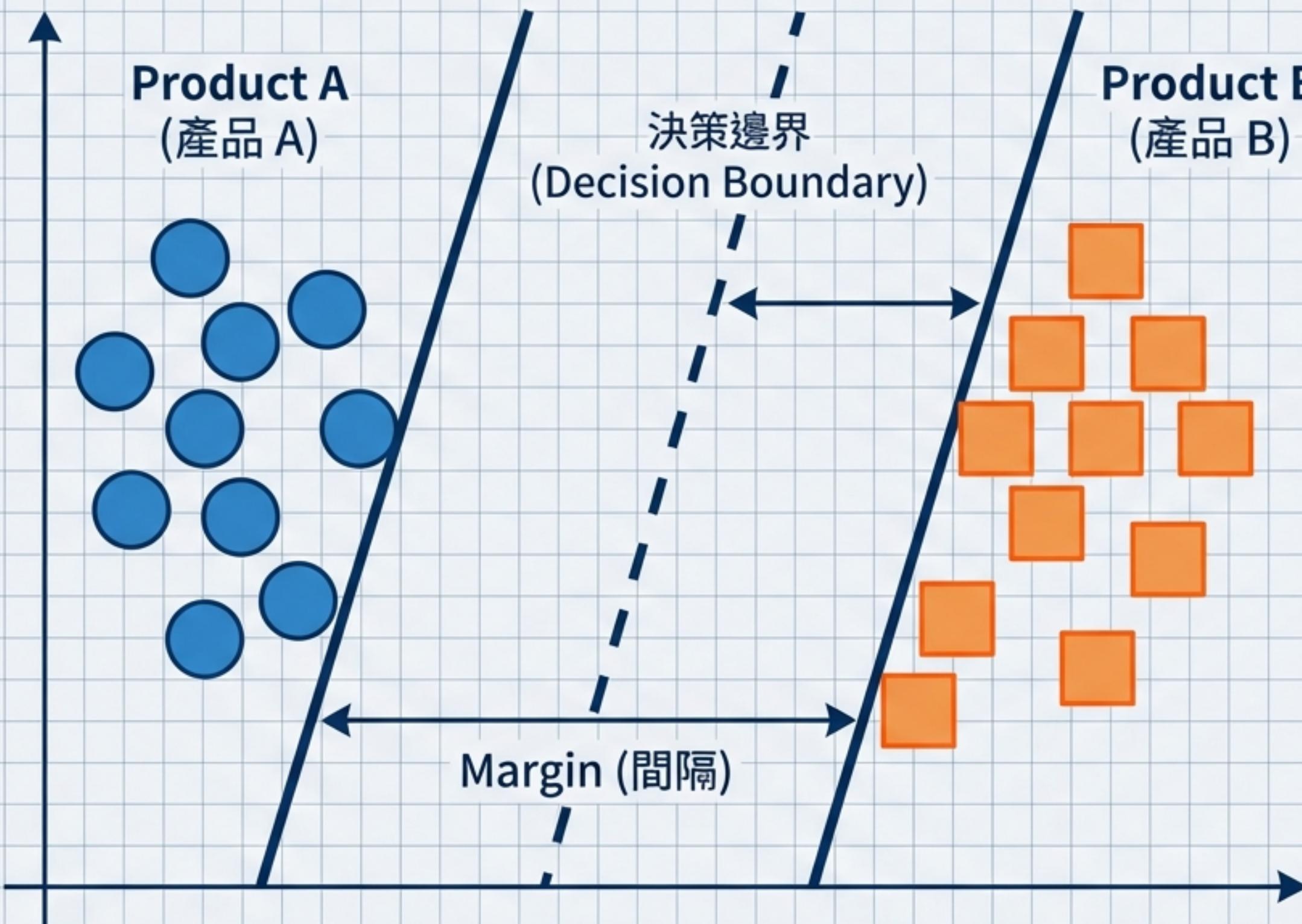


Unit 12: 支持向量分類 (SVC)

為化學製程建構數位安全護欄



核心原理：尋找最寬的『安全緩衝區』

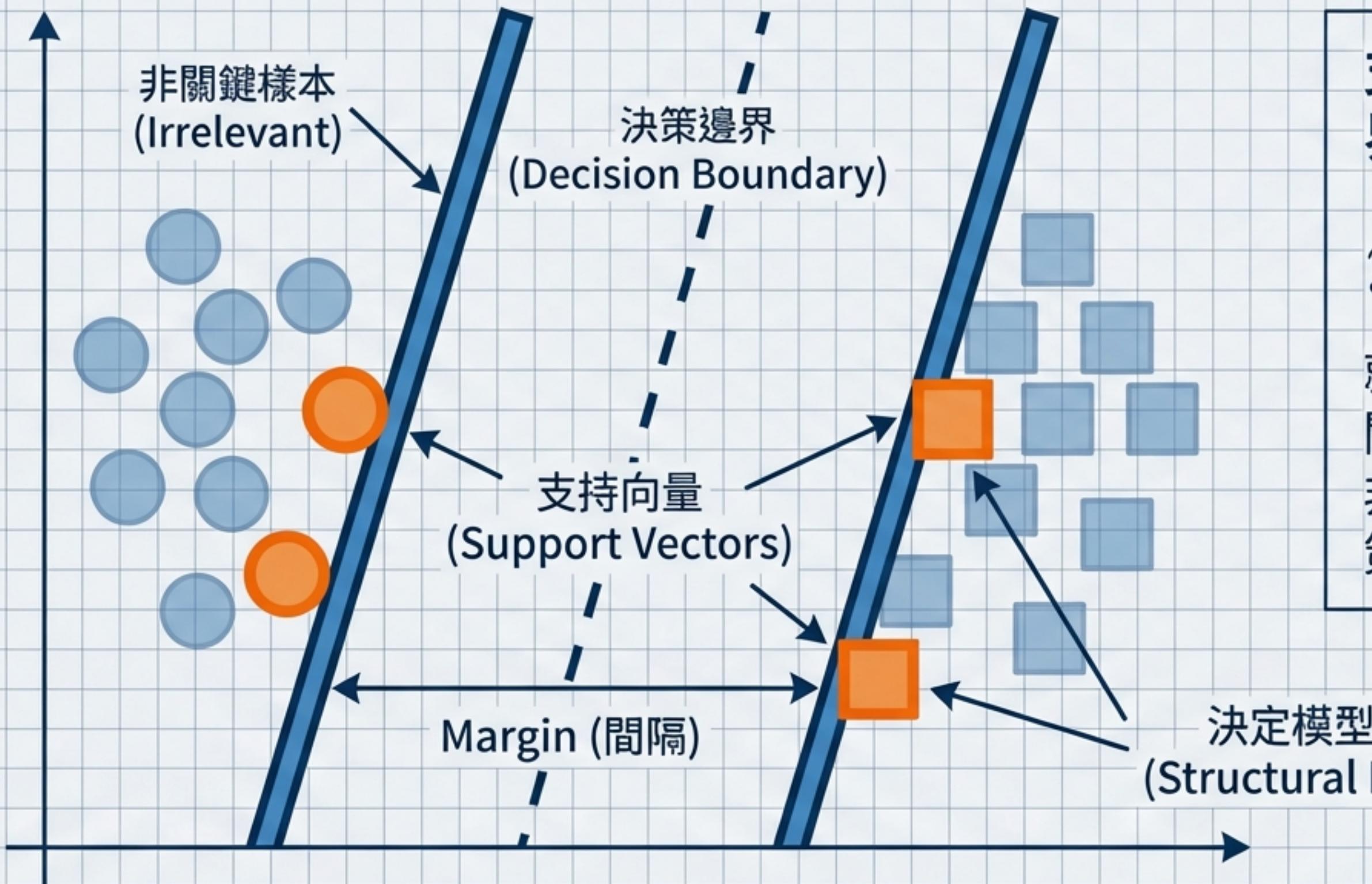


最優超平面 (Optimal Hyperplane)

SVC 的目標不僅是區分數據，而是建立最大的安全緩衝距離。在化工安全中，這相當於設定最保守的操作邊界。

$$\text{Margin} = \frac{2}{\|w\|}$$

結構力學：支持向量的作用



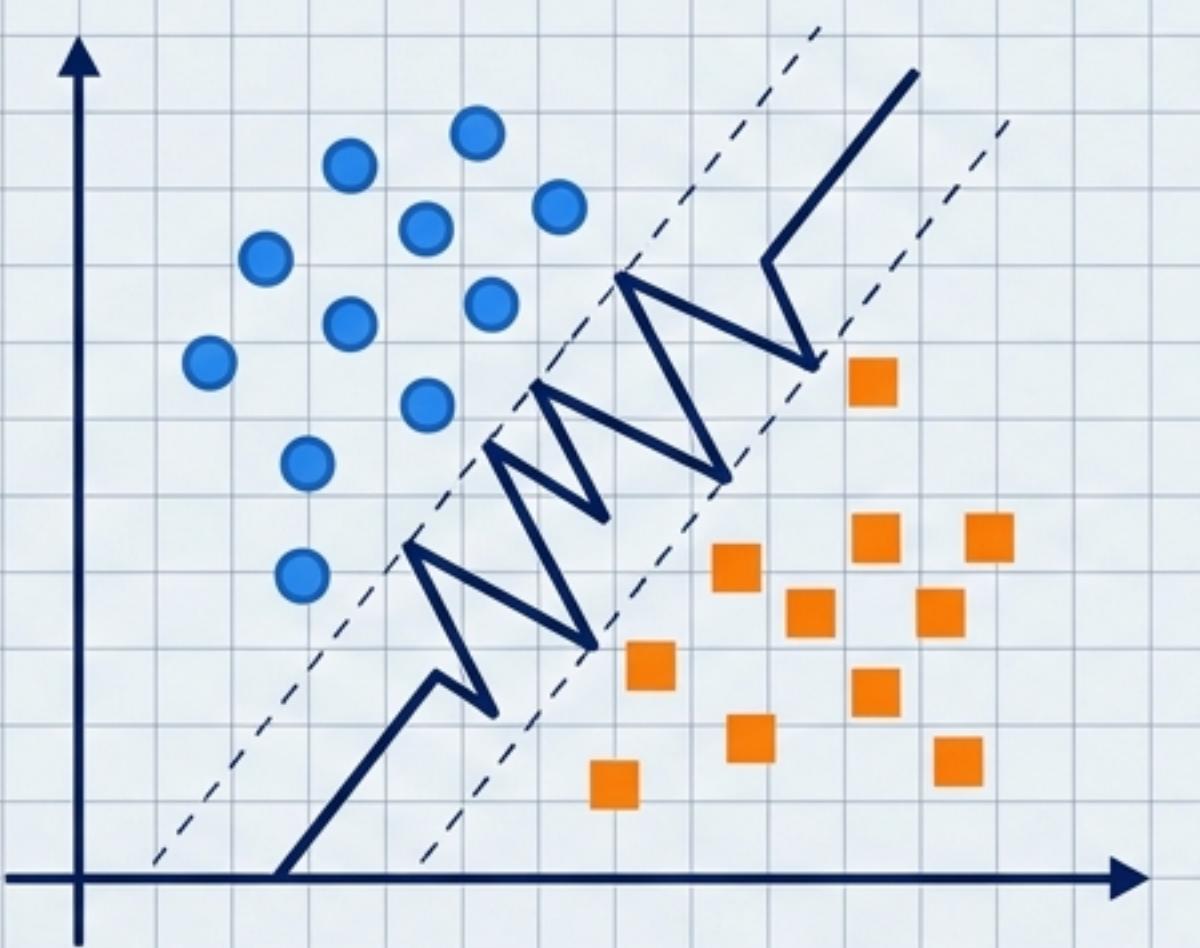
支持向量是位於間隔邊界上的關鍵數據點。

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) = 1$$

就像橋樑的承重柱，只有關鍵點在受力。移除其他非支持向量的數據點，決策邊界完全不會改變。

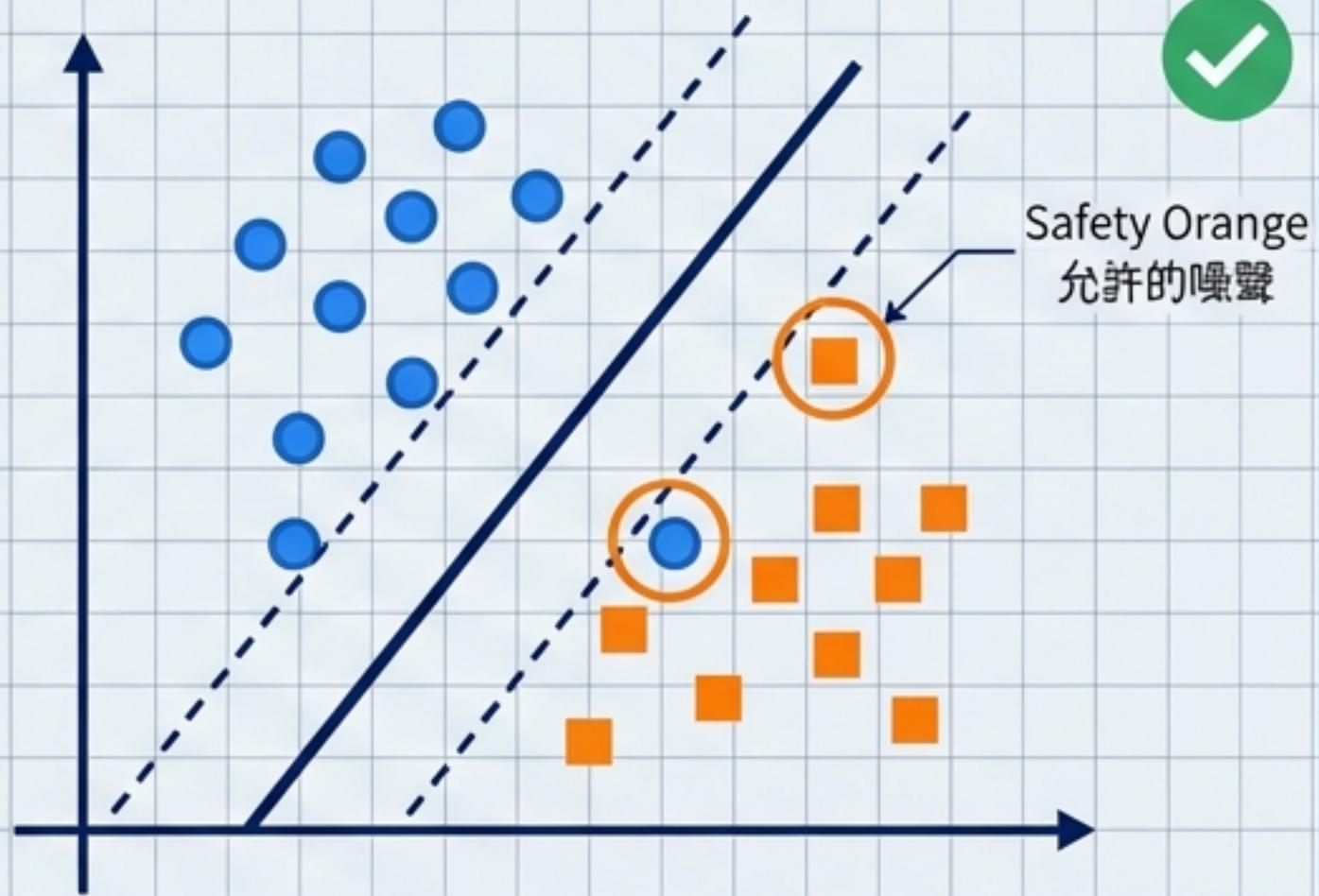
容錯設定：硬間隔 vs. 軟間隔

硬間隔 (Hard Margin)



理論理想狀態。C 值極大。對噪聲極度敏感
容易過擬合 (Overfitting)。

軟間隔 (Soft Margin)

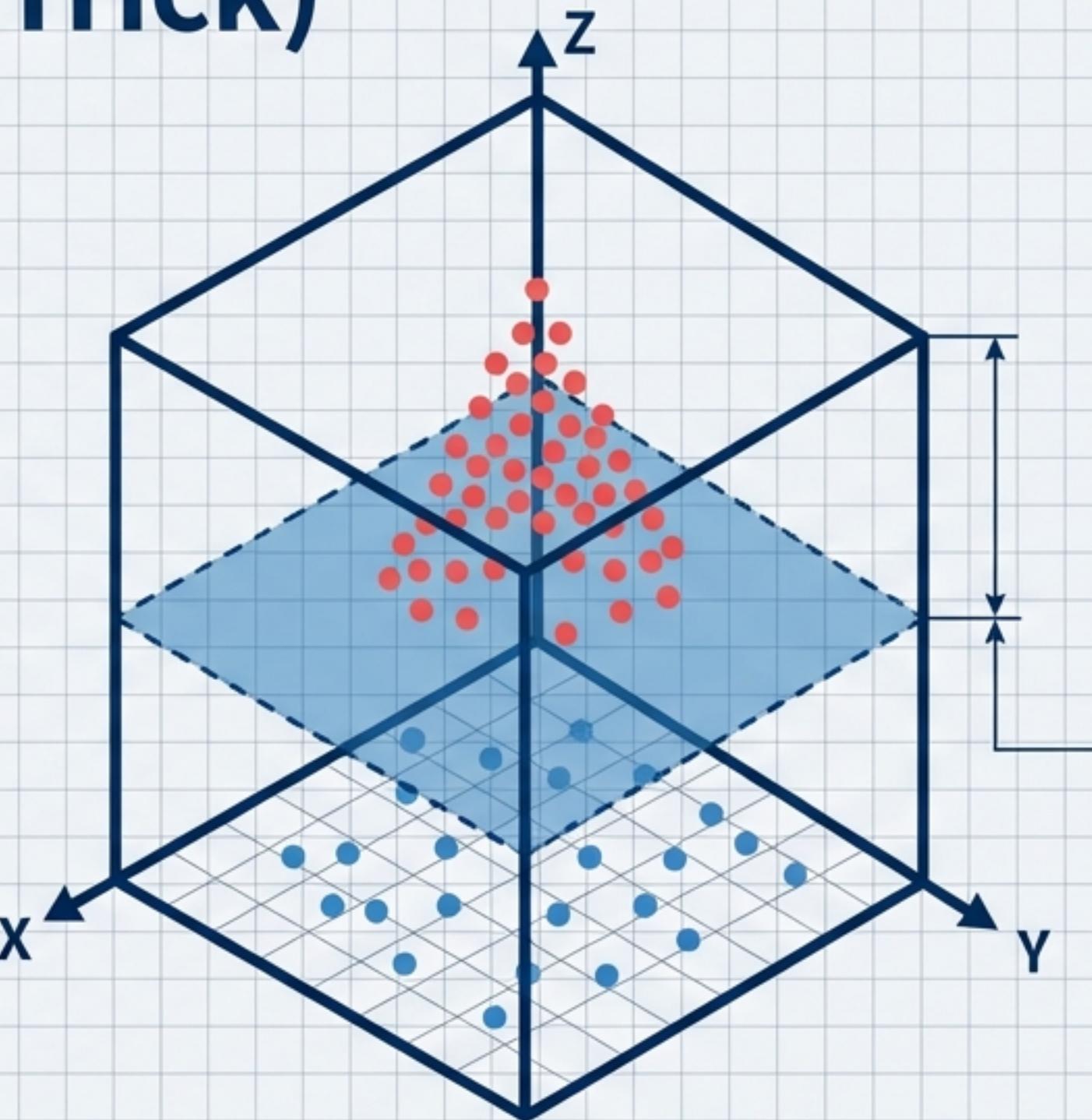
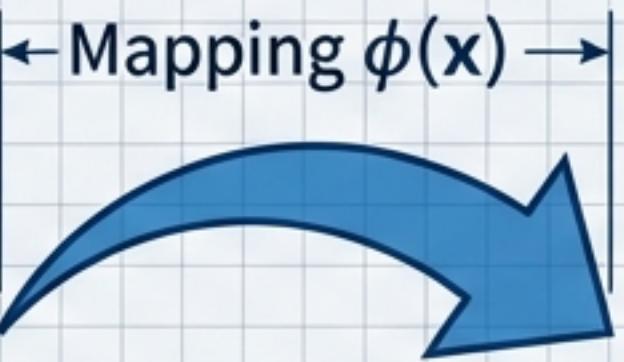
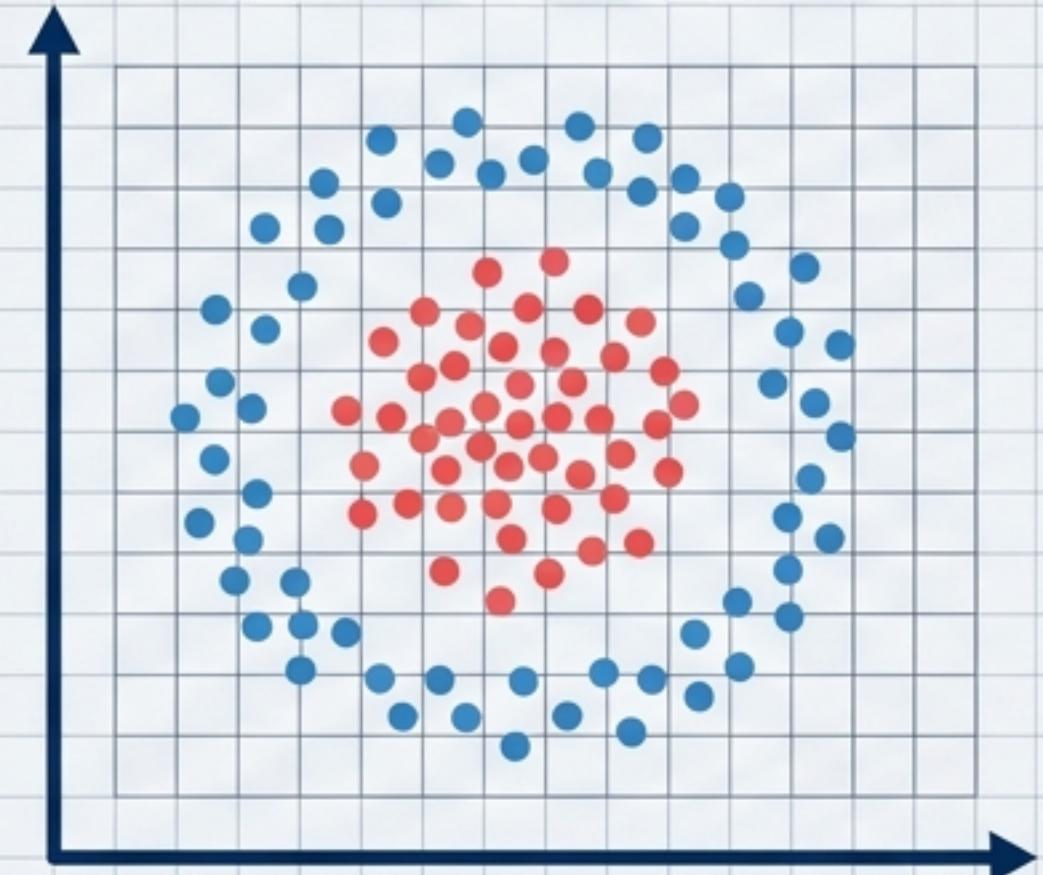


工程實際狀態。C 值適中。引入鬆弛變數 (ξ_i)
，允許少量誤差以換取更寬的安全邊界。

參數 C 控制對誤差的『懲罰力度』：C 越大，執法越嚴格 (間隔窄)；C 越小，越寬容 (間隔寬)。

處理非線性：核技巧 (Kernel Trick)

2D 原始空間：無法線性分割



Analogy Box

就像將油水混合物導入立體槽體中：在二維表面難以
區分，但在三維空間中利用密度差異自然分層。

核函數 $K(x, x')$ 讓我們能在高維空間計算距離，而無需實際進行繁重的座標轉換運算。

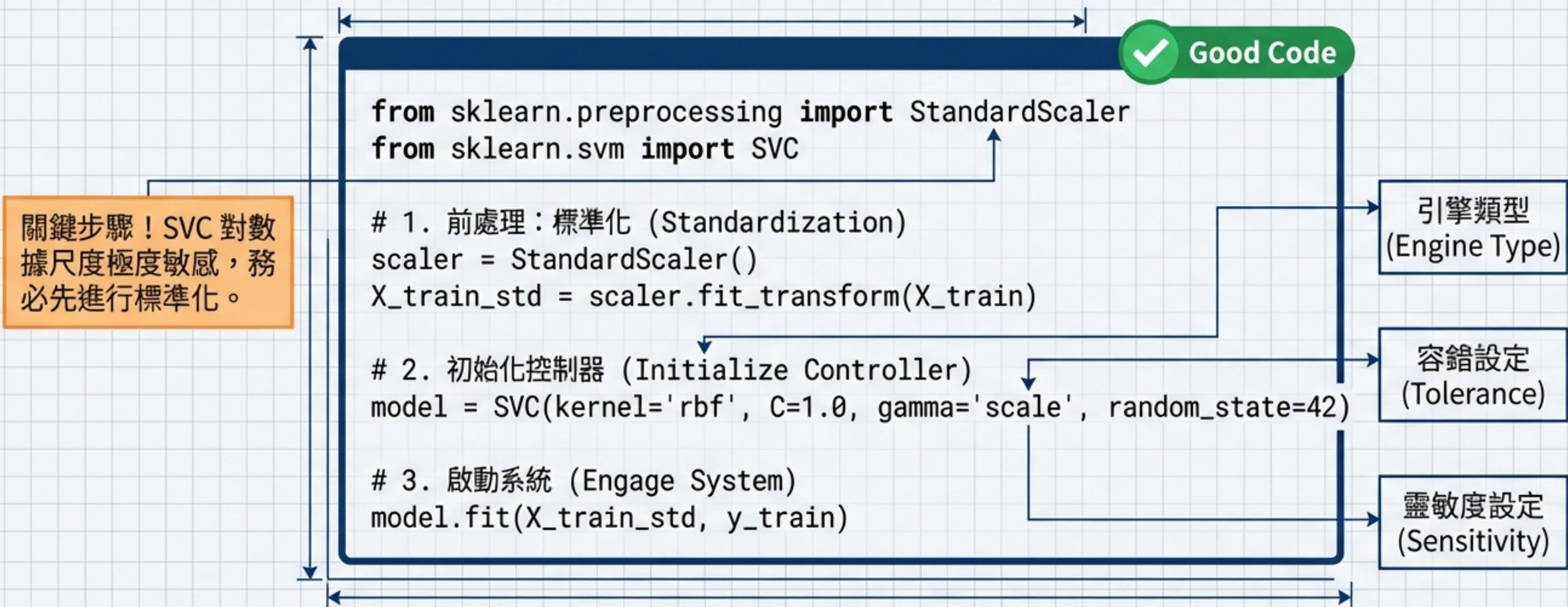
工具箱選擇指南：核函數類型

核函數類型 (Kernel Type)	圖示	適用場景 (Scenario)	優缺點 (Pros/Cons)
Linear Kernel (線性核)		線性可分數據，或特徵數量巨大 (文字探勘)。	計算最快，不易過擬合。
RBF Kernel (高斯徑向基核)		複雜非線性關係，未知數據分佈。	工業標準預設值。適應性最強。
Polynomial Kernel (多項式核)		明確知道數據符合特定物理定律 (如二次方關係)	參數繁多，計算不穩定。

工程實務法則 (Engineering Rule of Thumb)

不確定時，優先選用 RBF 核。就像在幫浦選型時，離心幫浦通常是首選。

控制面板 : Scikit-Learn 實作



關鍵調校參數：Gamma 與 C

Gamma (影響範圍)



High Gamma = 局部視角，邊界崎嶇 (Overfitting Risk)

C (懲罰強度)

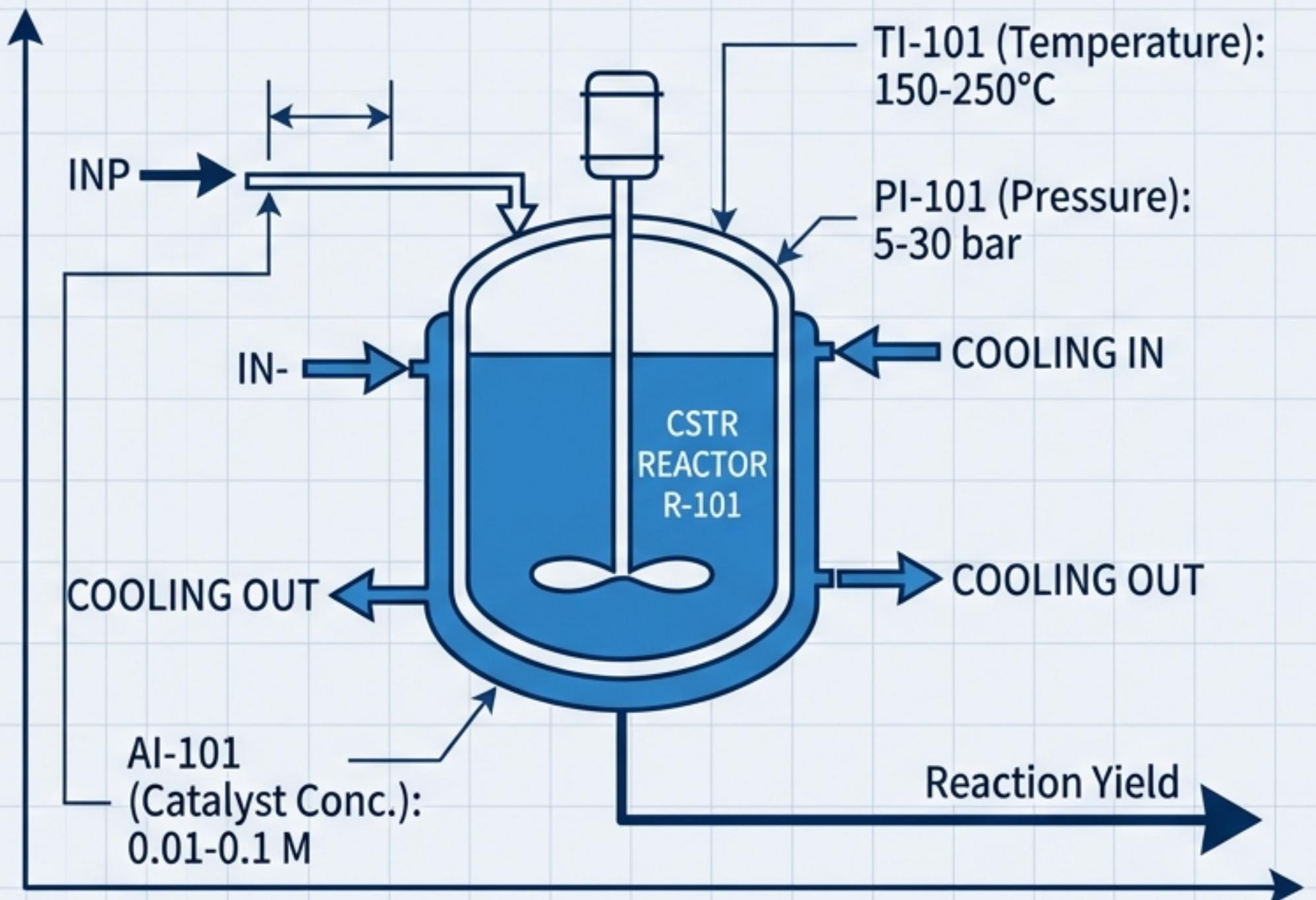


High C = 嚴格執法，模型複雜 (Low Bias)



危險組合：高 Gamma + 高 C = 高度複雜模型，極易過擬合。

實戰案例：化學反應條件分類



Mission:

預測反應是否成功 (Success/Failure)

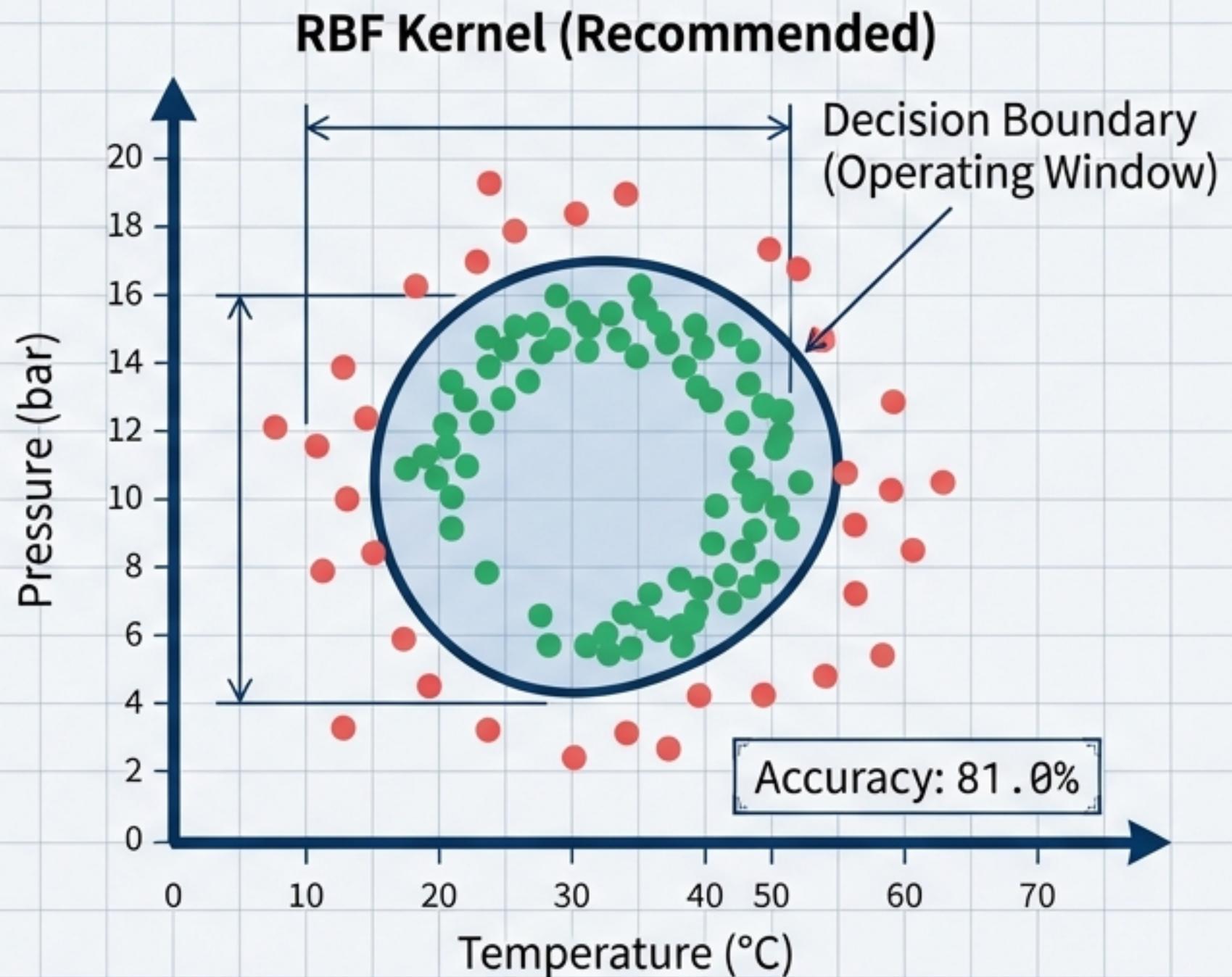
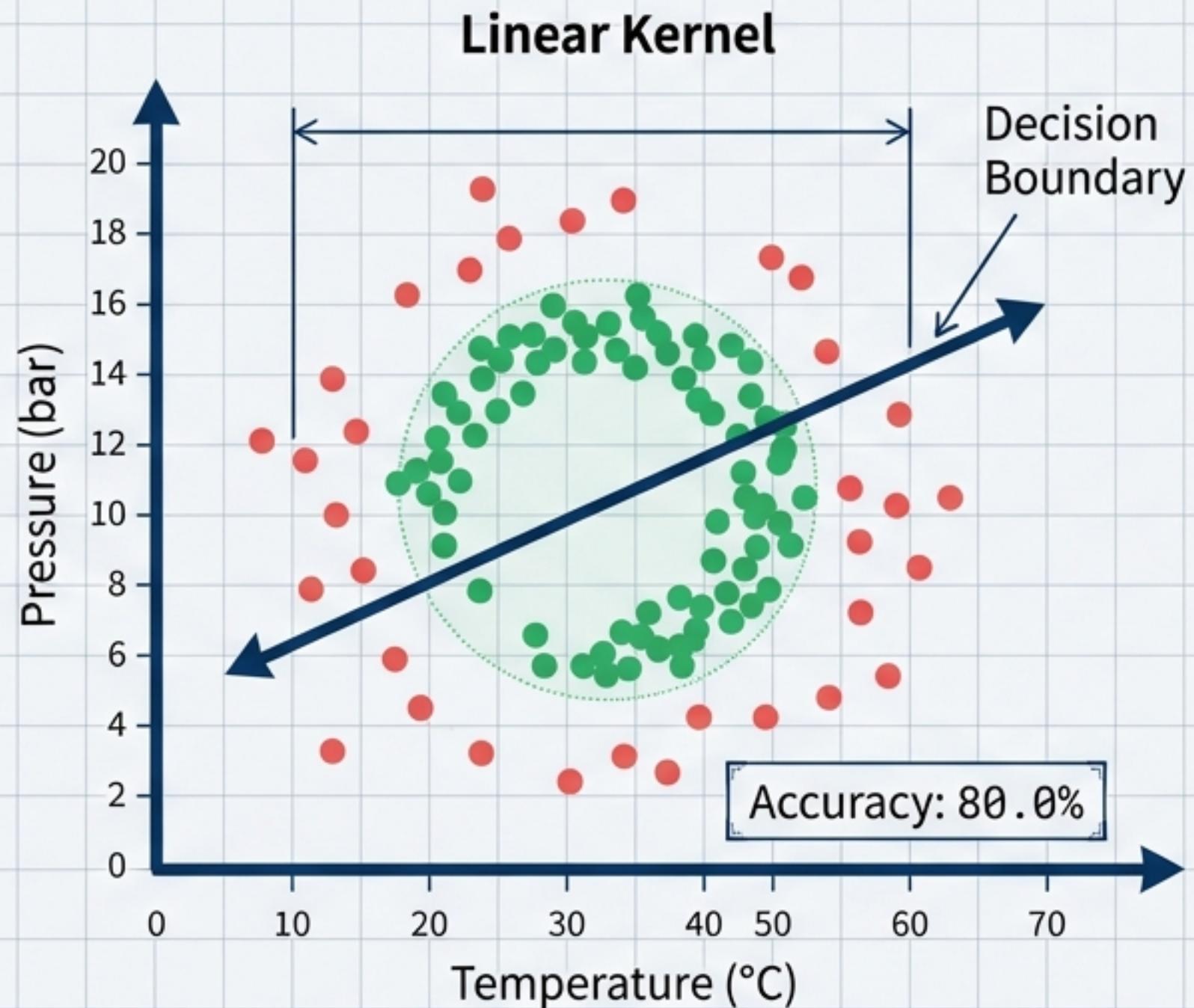
Class Imbalance Alert:

Success (1): 365 筆 (73%)

Failure (0): 135 筆 (27%)

⚠ 注意：類別不平衡
(Imbalanced)

模型效能比較：線性 vs. RBF



化學反應通常存在特定的『操作視窗』(Operating Window)，而非簡單的線性關係。
RBF 核能精確描繪這個幾何形狀。

隱藏的危險：準確率的陷阱

Metric Highlight

Overall Accuracy: 81%

Looks Good

Failure Recall: 41%

我們漏檢了 59% 的失敗反應！

	Predict Failure	Predict Success
Actual Failure	誤報 (FP)	漏報 (False Negative) - 危險
Actual Success	正報 (TP/TN)	預測會成功，實際卻失敗 → 原料浪費、潛在安危。

化工代價是不對稱的。漏報事故 (FN) 的代價通常遠高於誤報 (FP)。

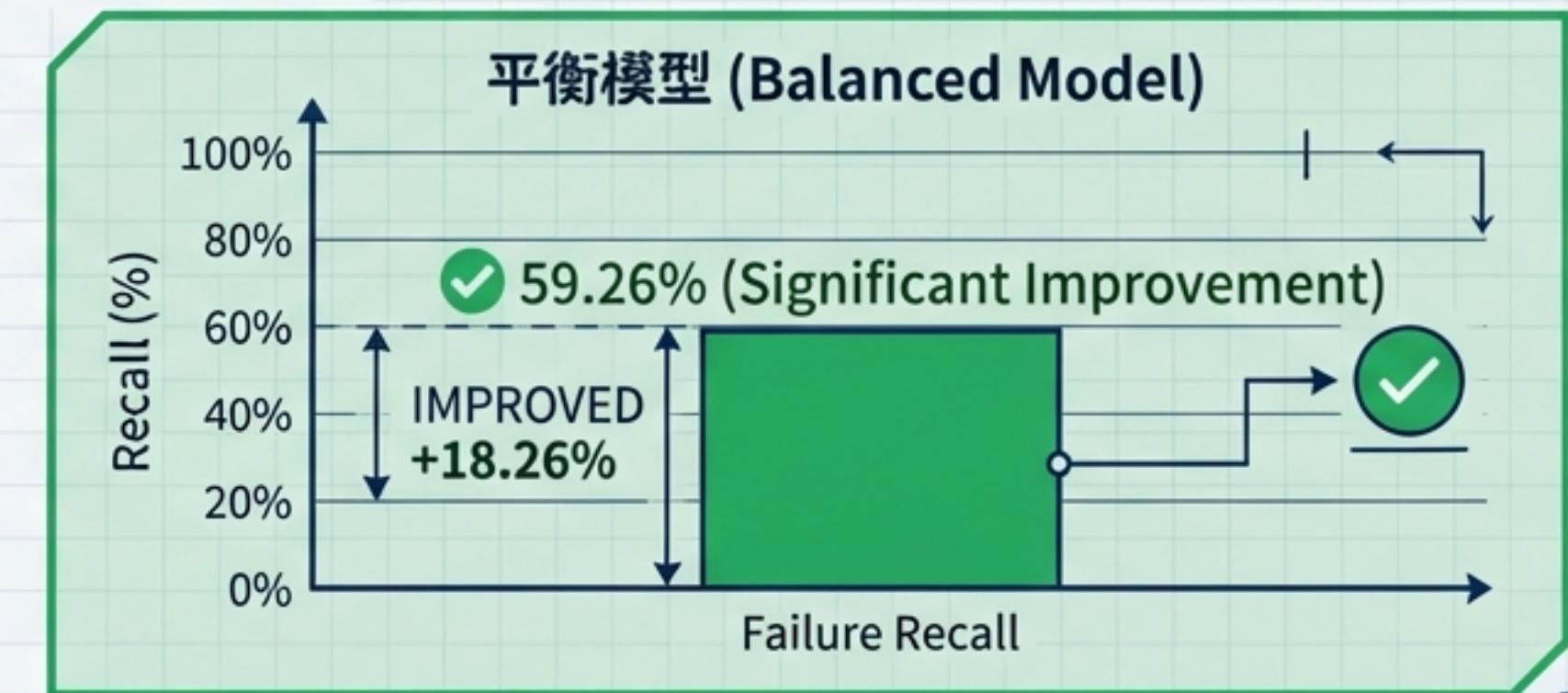
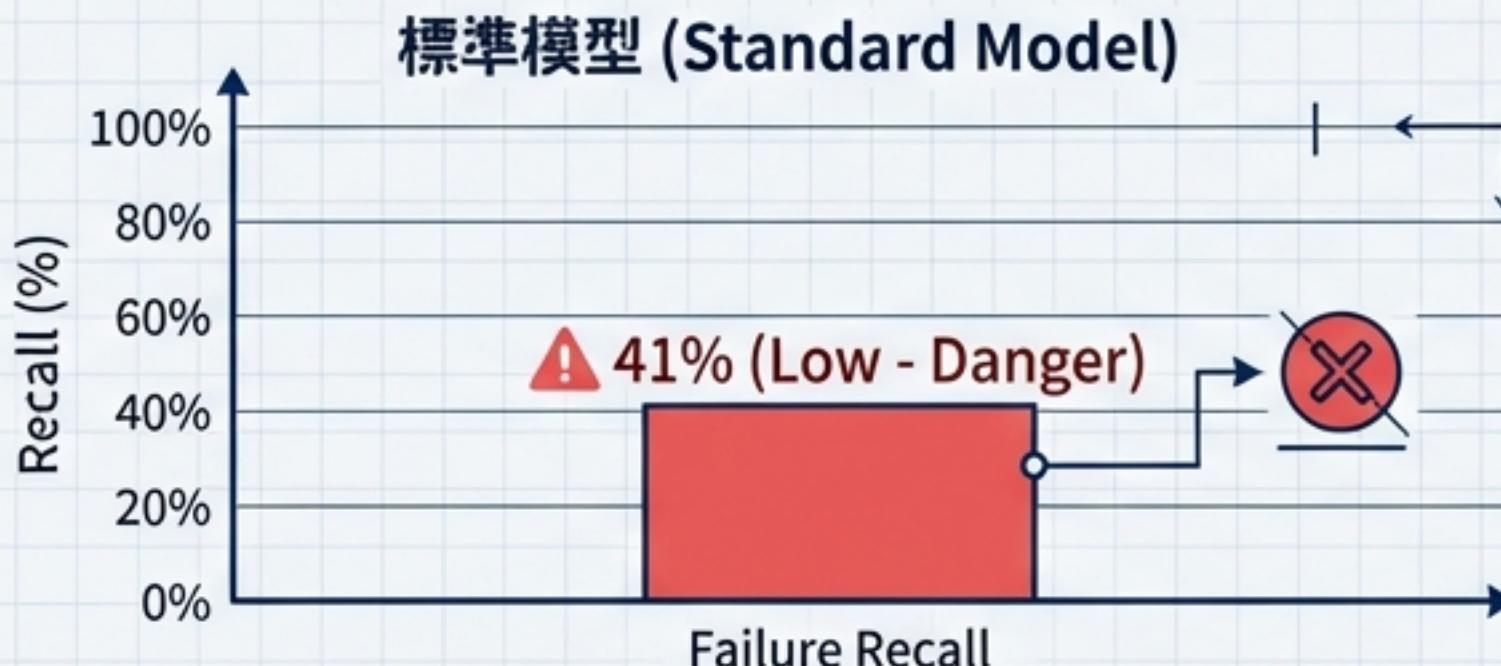
安全校準：處理類別不平衡

SOLUTION CODE

```
model = SVC(class_weight='balanced')
```

自動調高少數類別 (Failure) 的懲罰權重。

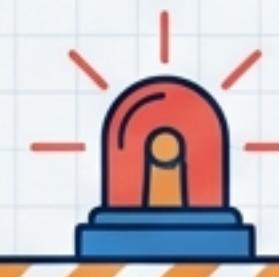
BEFORE / AFTER COMPARISON: FAILURE RECALL IMPROVEMENT



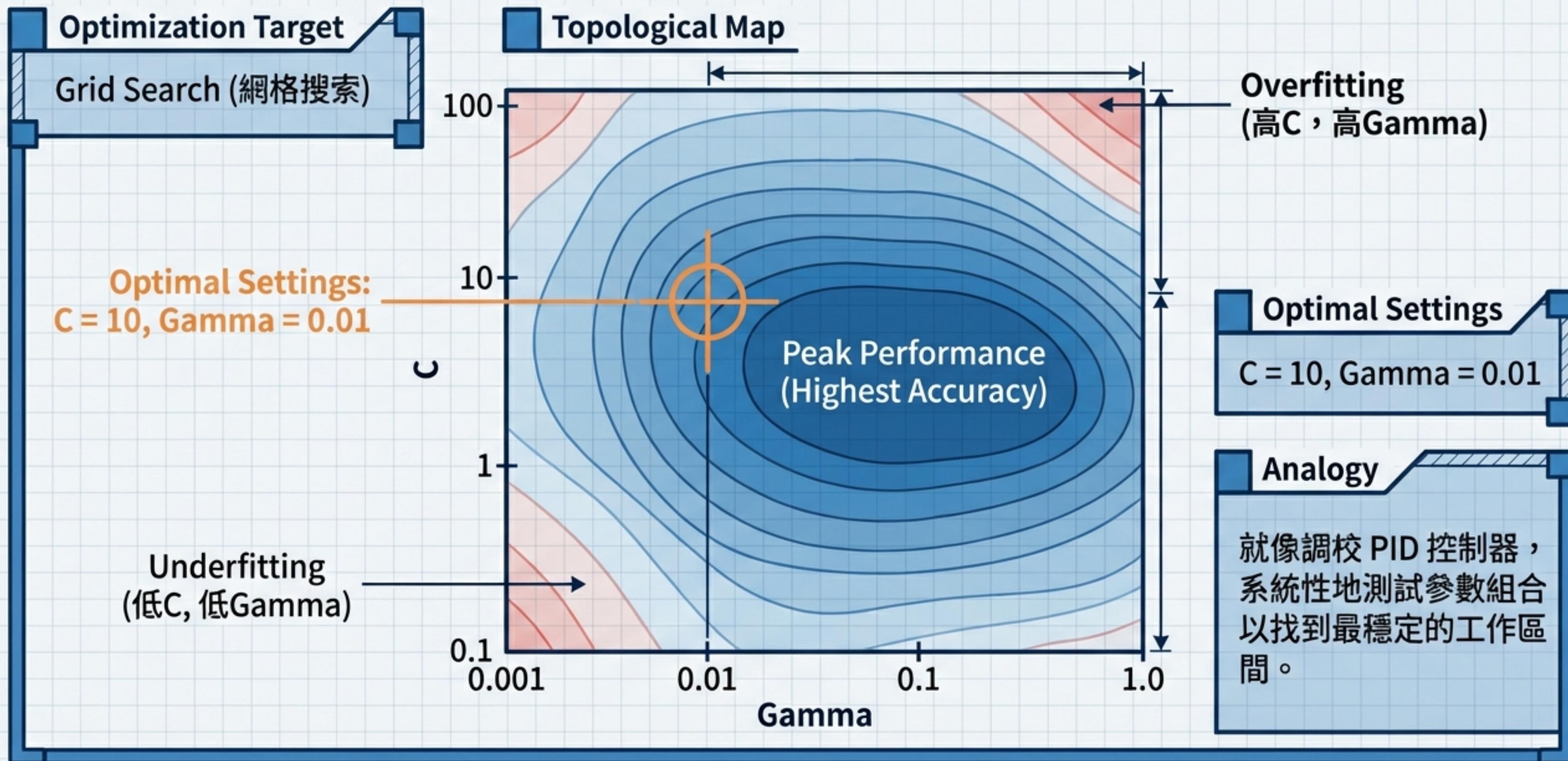
ENGINEERING TRADE-OFF: SAFETY OVER ACCURACY

- Failure Recall: 41% → 59.3% Safety Improved (優先考量)
- Overall Accuracy: 81% → 79% Acceptable Cost (可接受代價)

寧可誤報 (False Alarm)，不可漏報 (Missed Alarm)。

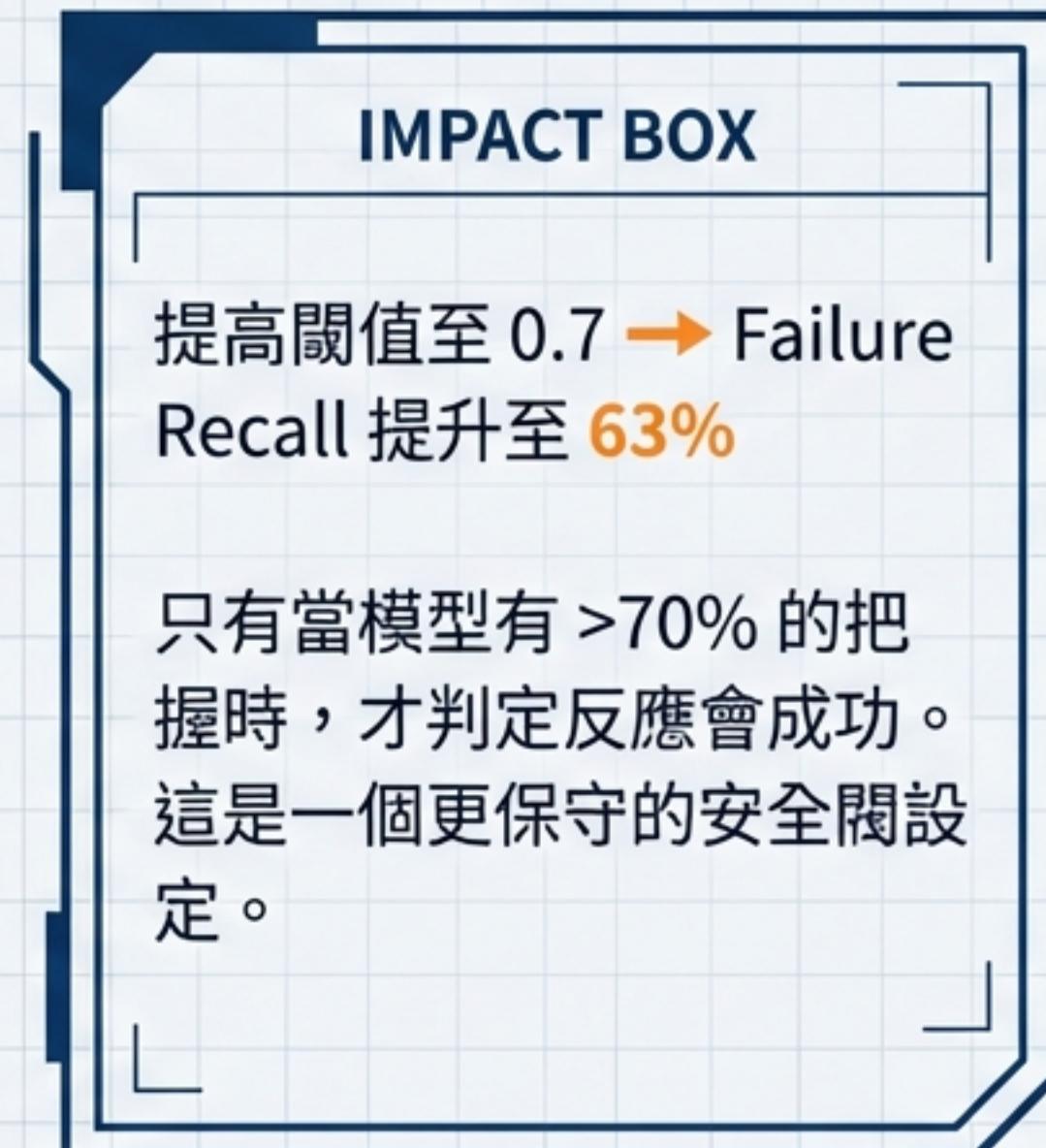
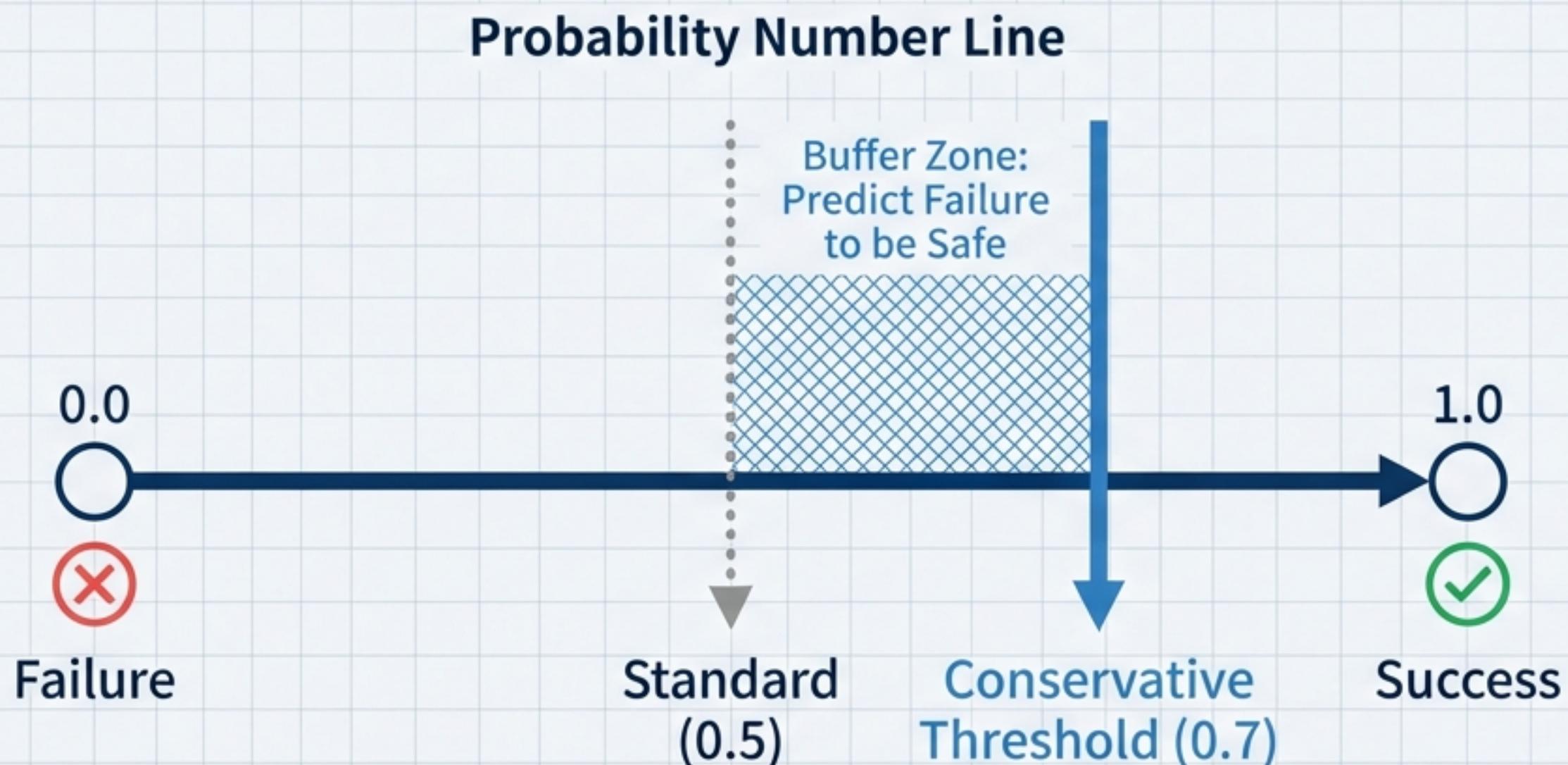


尋找最佳工作點：參數網格搜索

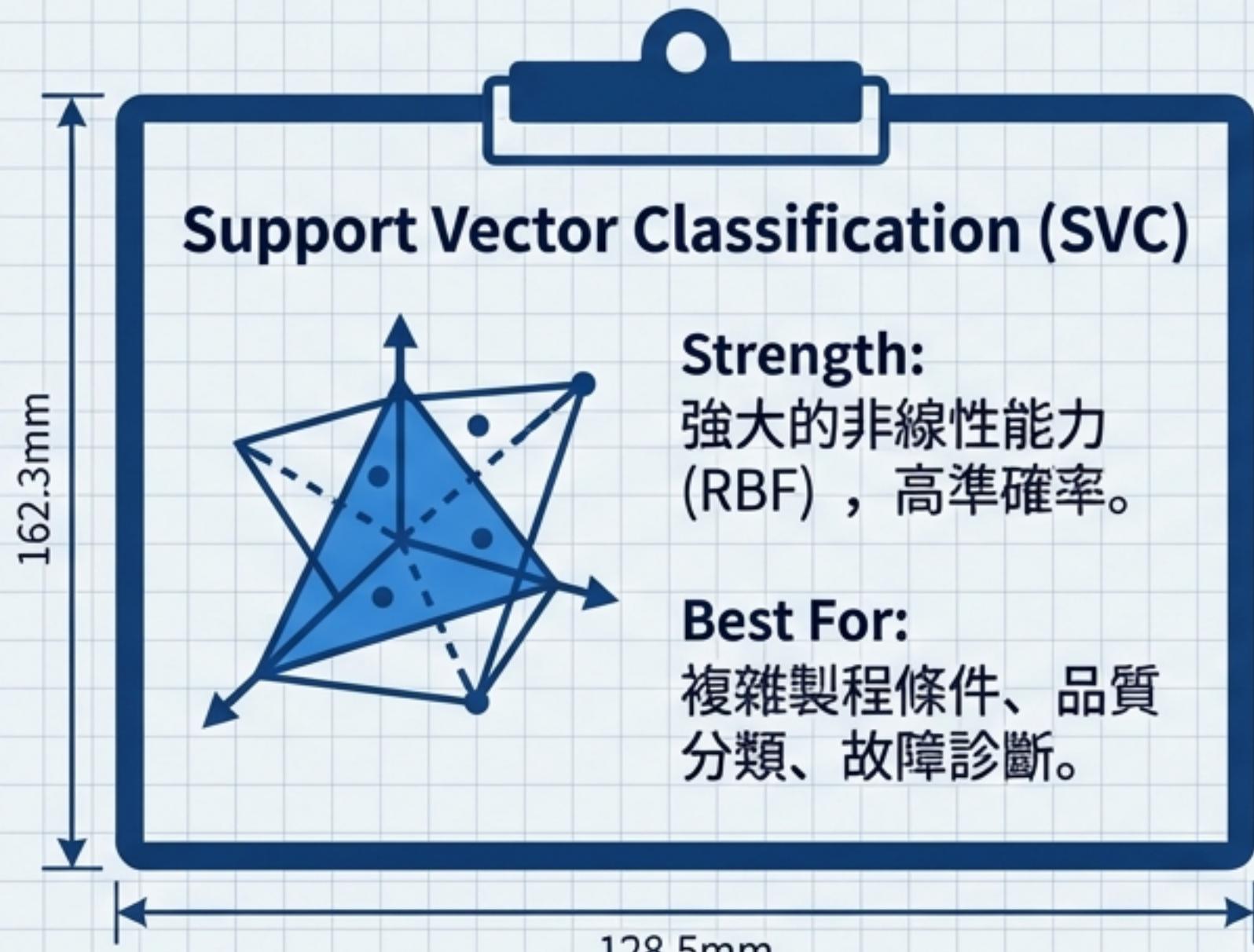


進階控制：機率與閾值調整

啟用 `probability=True`，將二元判斷轉為信心指數。



總結與選型指南



Final Checklist

- 數據標準化 (StandardScaler)
- 檢查不平衡 (Class Weight)
- 網格搜索調優 (C & Gamma)

SVC 是您定義複雜製程安全視窗的精密儀器。