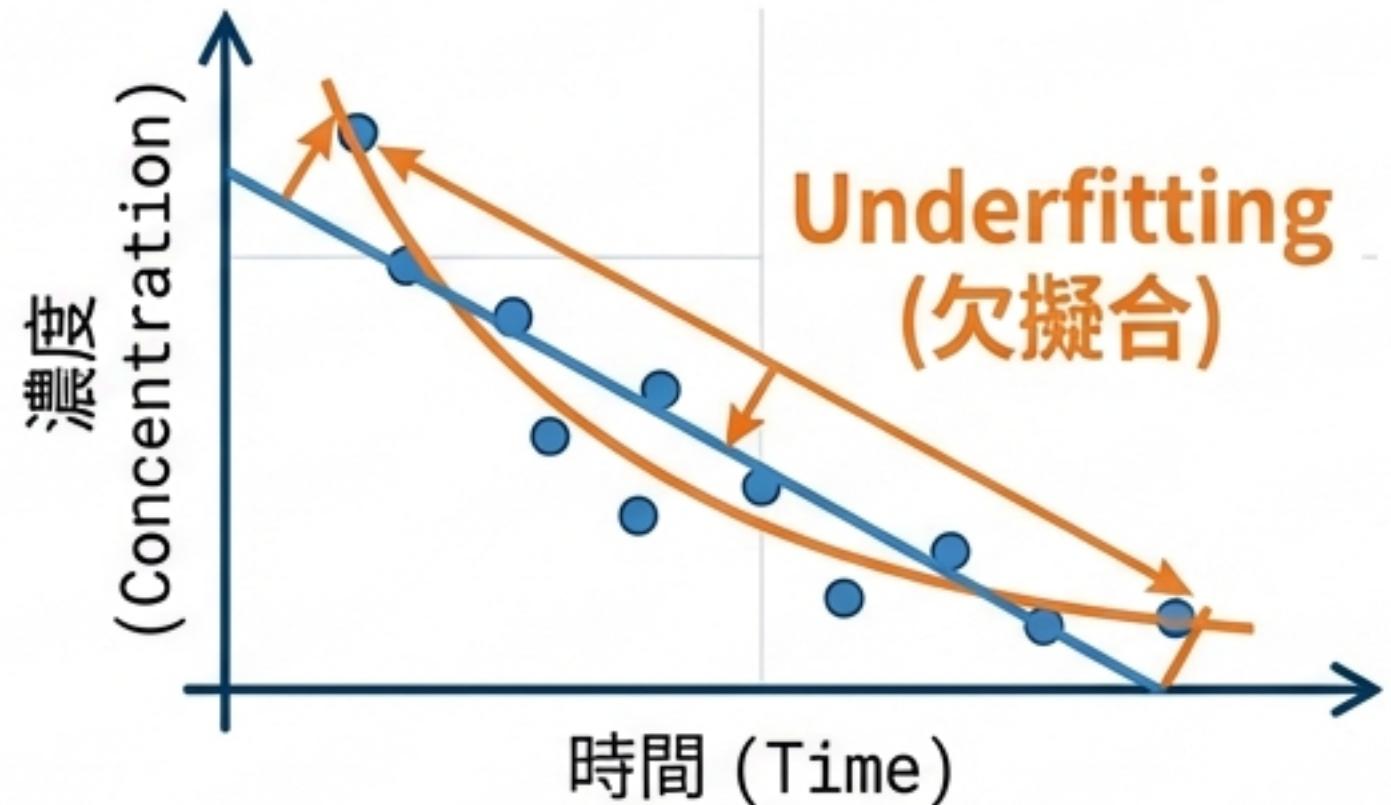


Unit 11: 支持向量機回歸 (SVR)

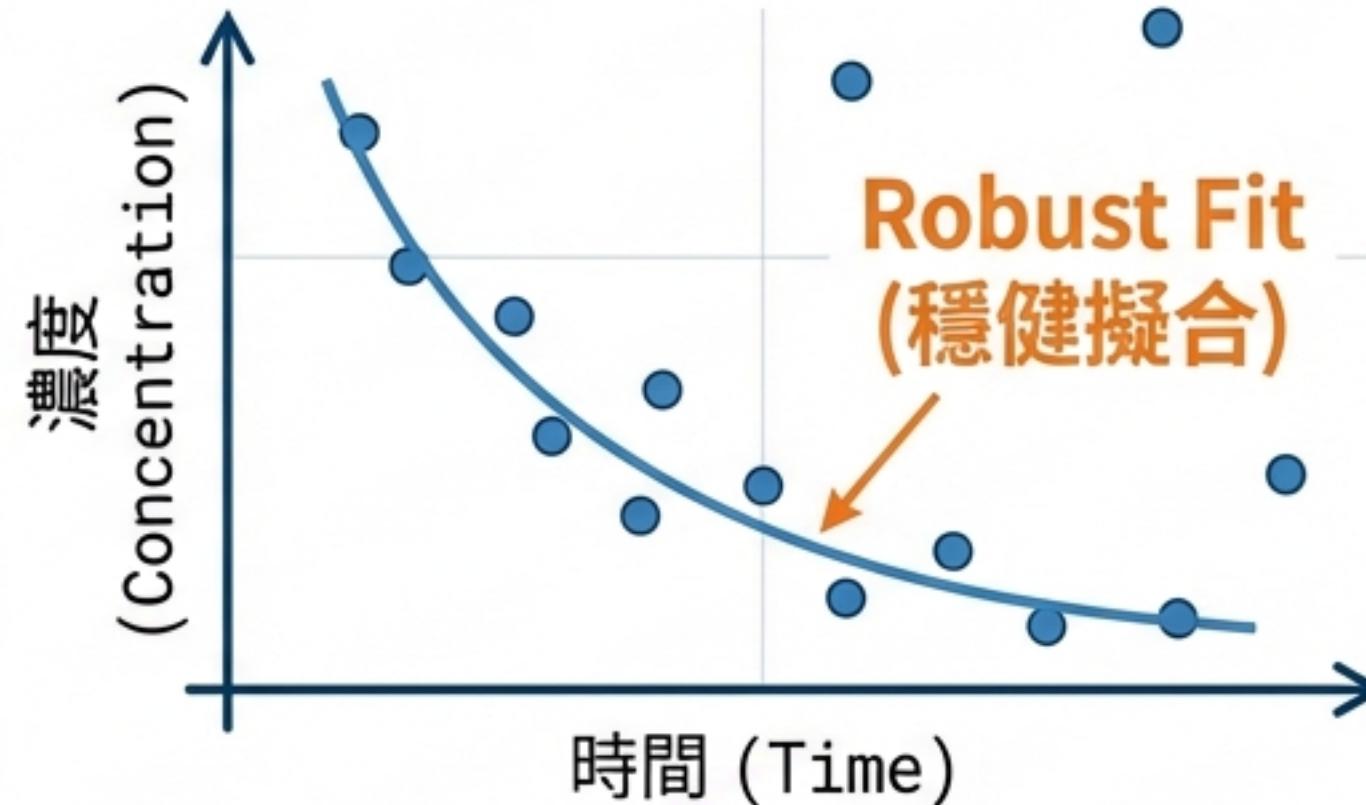
建構化工製程的強健非線性模型

工程挑戰：為何傳統線性回歸不足以應對化工製程？

傳統線性回歸 (Linear Regression)



支持向量機回歸 (SVR)



非線性關係 (Non-Linearity)

化學反應動力學、熱力學平衡
通常是非線性的。

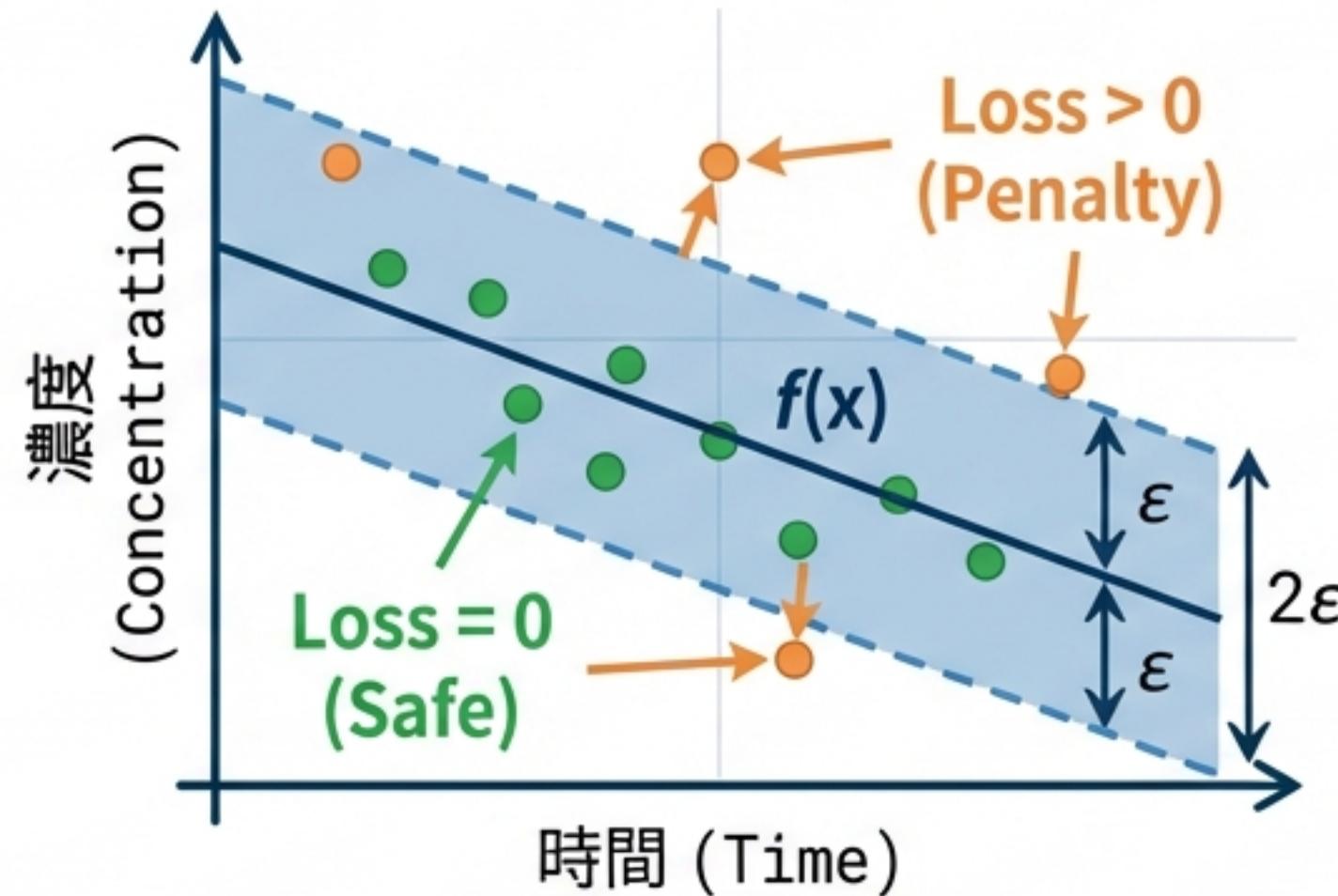
抗噪聲需求 (Robustness)

工廠感測器數據常含有異常值
(Outliers)。

小樣本高維度

實驗數據稀少，但變數（溫
度、壓力、濃度）繁多。

SVR 的核心機制：建立「誤差容忍管道」(ε -Tube)



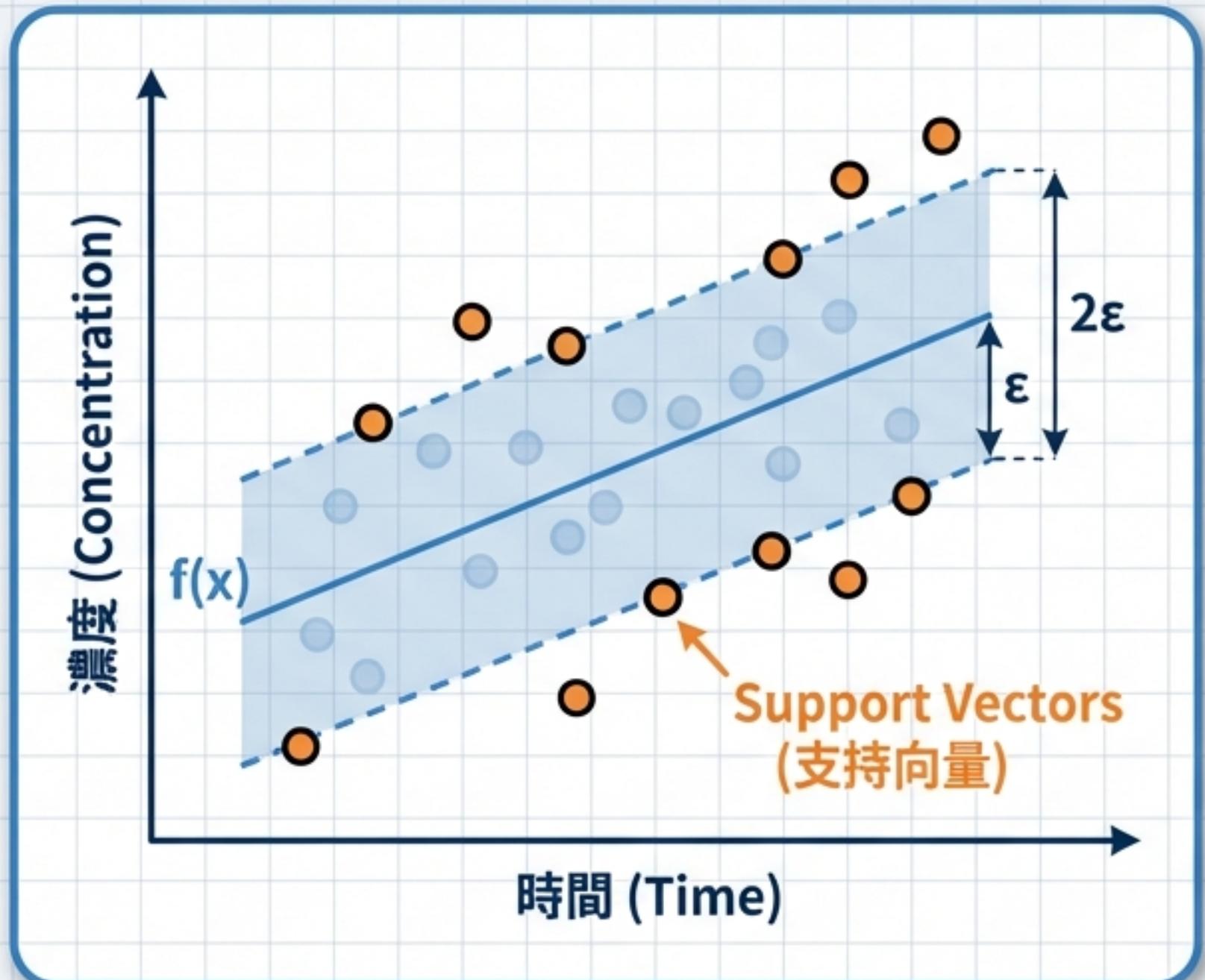
ε -Insensitive Loss Function:

$$L_\varepsilon(y, f(x)) = \max(0, |y - f(x)| - \varepsilon)$$

Engineering Translation

- ε (Epsilon) : 管道的半徑。在此範圍內的預測被視為「夠準確」，不計入損失。
- 類似概念：工業規格中的公差 (Tolerance) – 只要在規格內，就是合格產品。

結構力學：支持向量 (Support Vectors) 的角色



定義與特性

- 定義：只有落在 ϵ -tube 邊界上或邊界外的樣本點（誤差 > 0 ）才會影響模型。
- 稀疏解 (Sparsity)：落在管道內的大多數數據點被模型「忽略」。這使得 SVR 計算高效且穩健。

優化目標 (Optimization Goal)

$$\text{Minimize: } \frac{1}{2} \|\mathbf{w}\|^2 + C \sum \xi$$

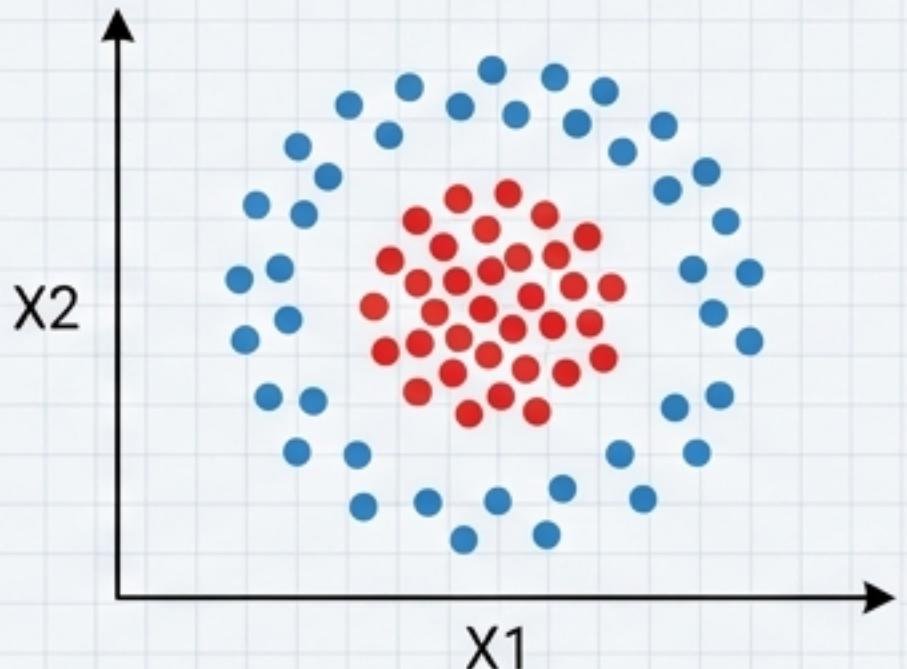
尋找最平坦的函數 ($\|\mathbf{w}\|^2$)，同時限制誤差總和 ($\sum \xi$)。

工程隱喻：就像建築結構，只有關鍵的「支柱」承載重量，其他的牆面只是填充。

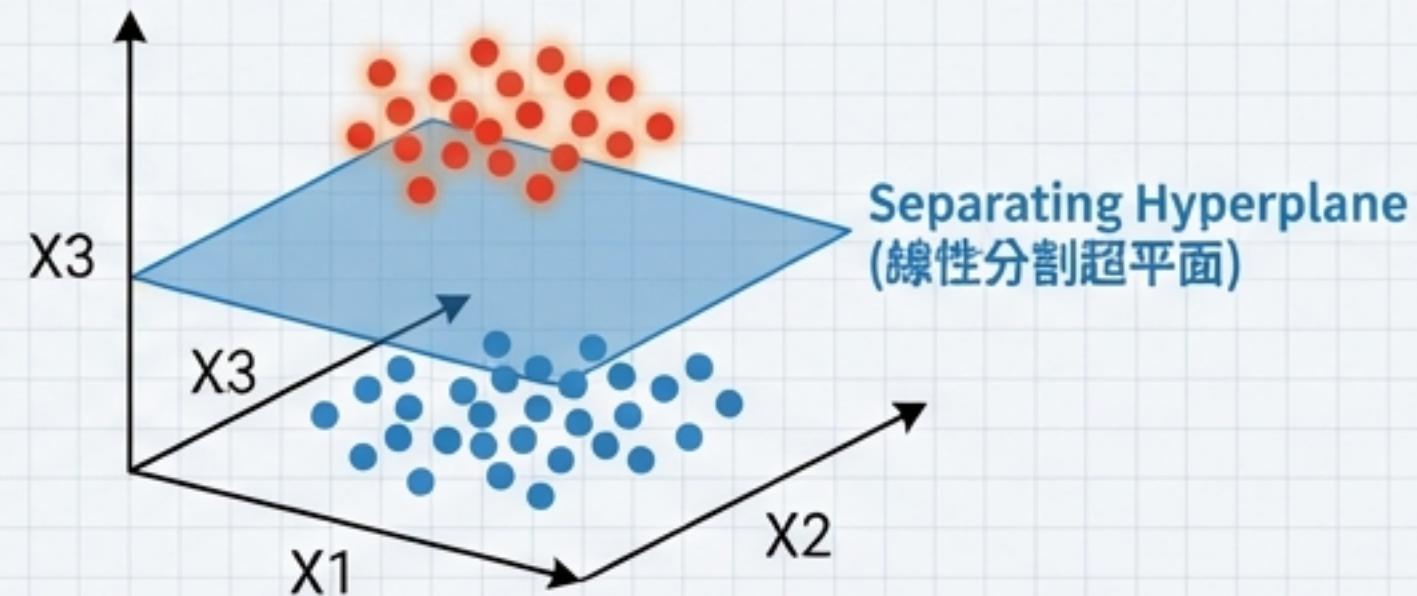
核函數技巧 (Kernel Trick)：高維空間的映射機制

核函數映射示意 (Kernel Mapping Diagram)

Input Space (Low Dimension): Non-Linear

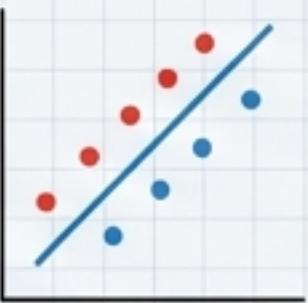


Feature Space (High Dimension): Linear



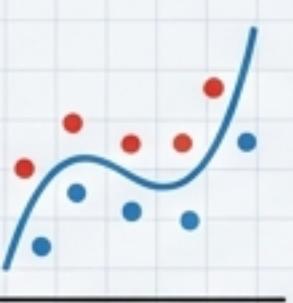
常用核函數工具箱 (Kernel Toolkit)

Linear (線性)



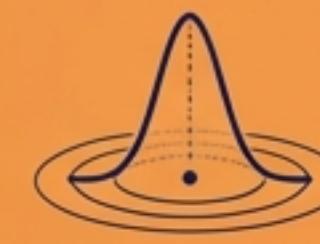
適用於線性問題，速度快。

Poly (多項式)



適合具有多項式交互作用的特徵。

RBF (徑向基函數)

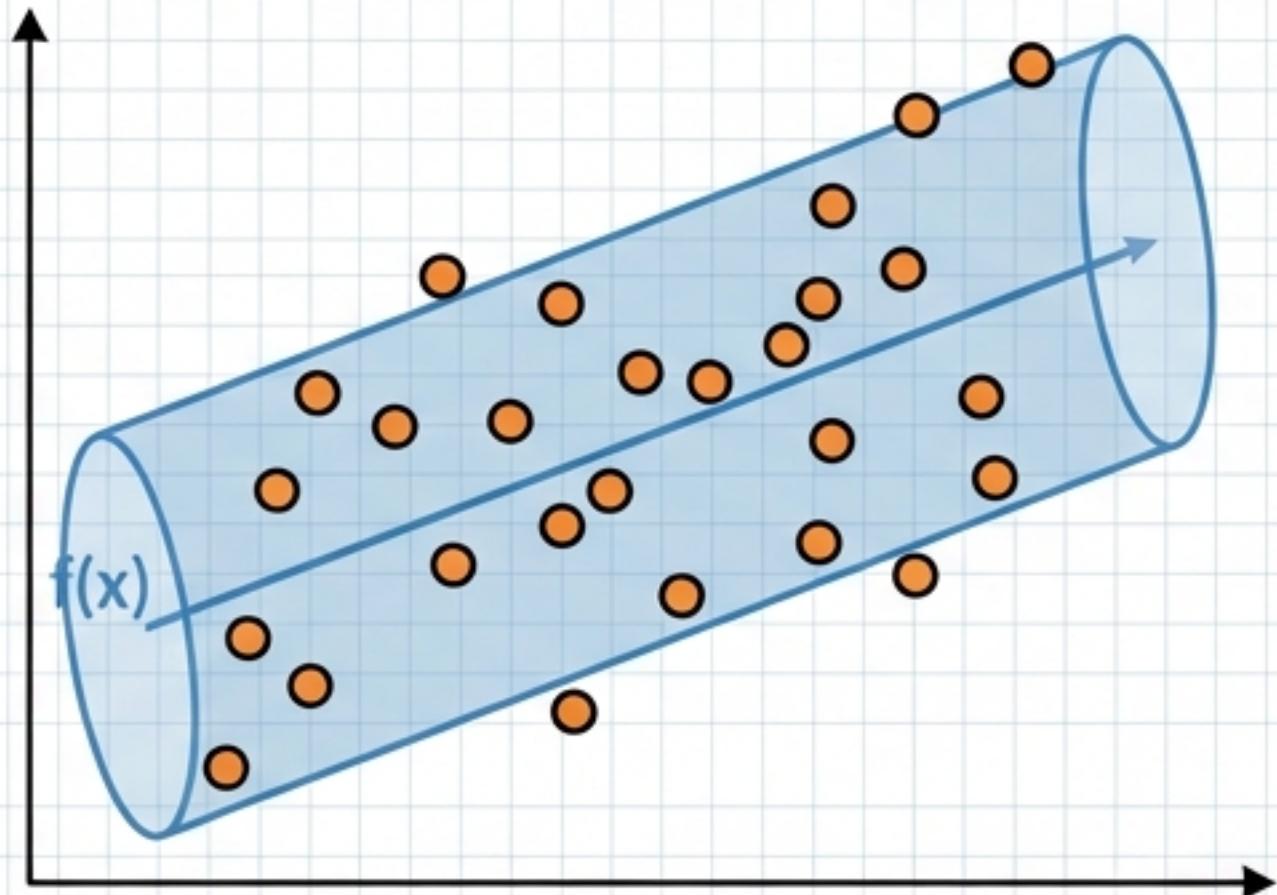


預設且最強大。如同高斯分佈，基於距離衡量相似度，能處理任意非線性。

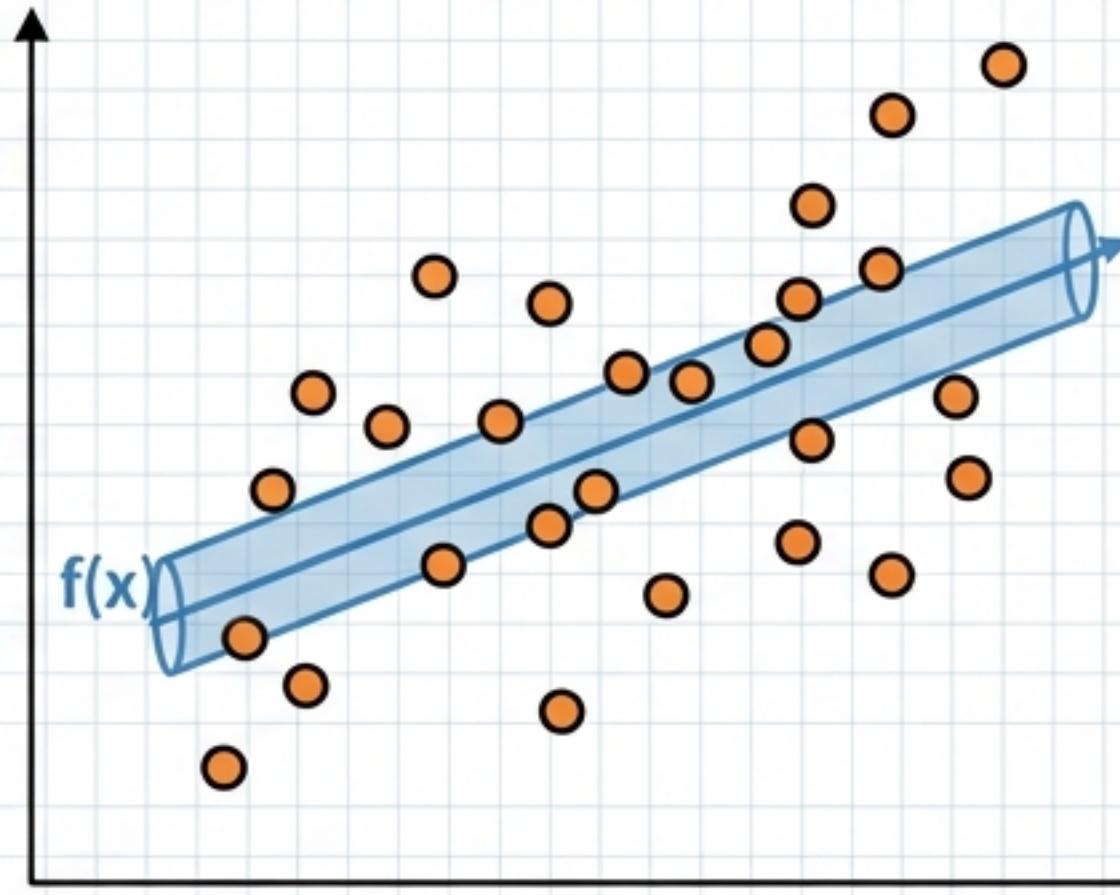
控制閥門 I：Epsilon (ε) — 定義精確度

Metaphor: 管道寬度 (Tube Width)

Large ε (Wide Tube)



Small ε (Narrow Tube)



粗略估計 (Rough Estimate)。支持向量很少，模型簡單。風險：欠擬合 (Underfitting)。

精確擬合 (Precision Fit)。支持向量很多，模型複雜。風險：過擬合 (Overfitting)。

調參策略：通常設定為目標變數標準差的 0.1 - 0.5 倍。預設值：0.1。

控制閥門 II：Penalty (C) — 誤差懲罰強度

Metaphor: 壓力閥 (Pressure Gauge)

寬容模式 (Soft Margin)。

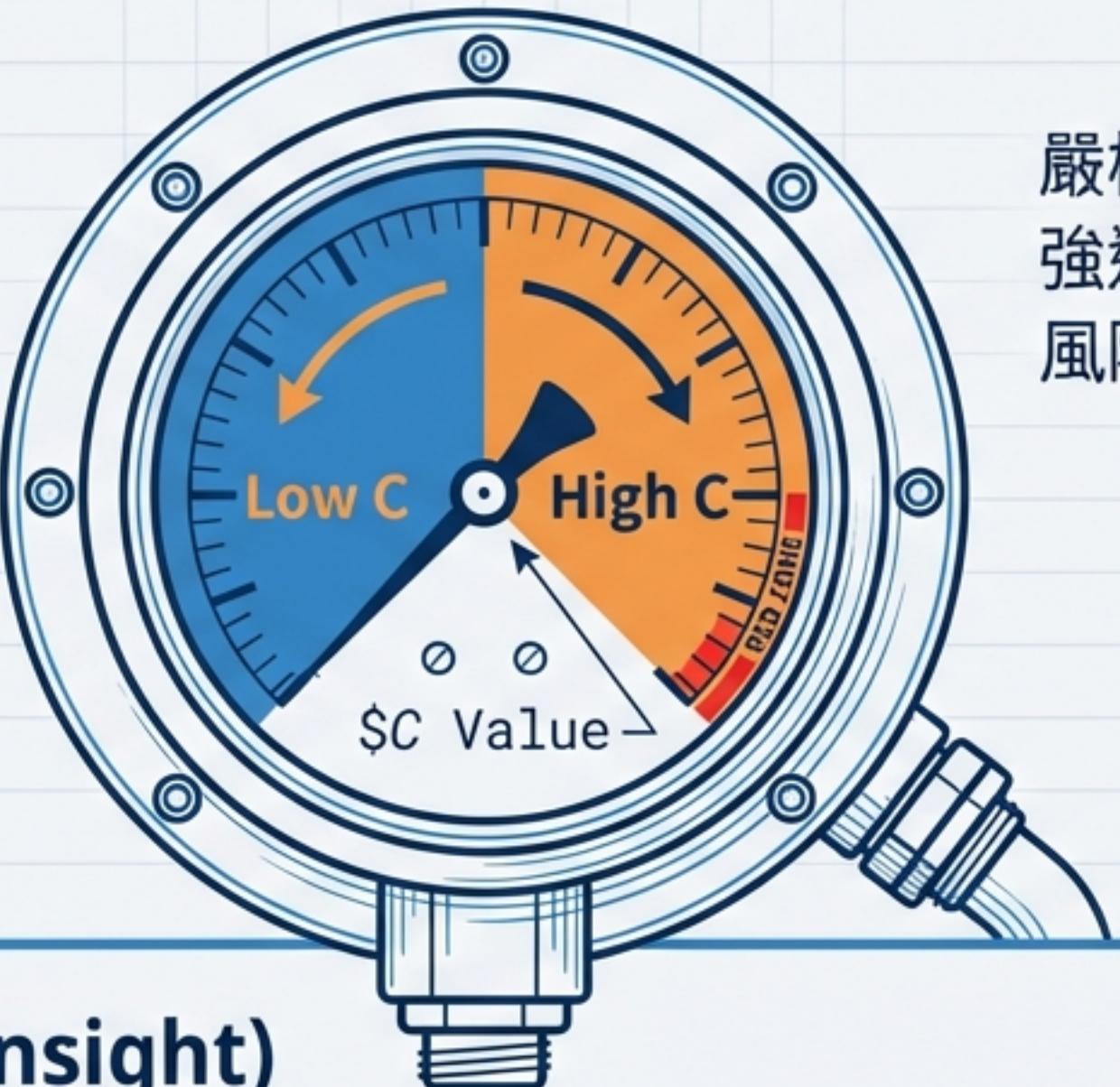
允許更多誤差以換取平滑模型。

風險：欠擬合 (Underfitting)。

嚴格模式 (Hard Margin)。

強迫模型擬合每一個點。

風險：過擬合 (Overfitting)。



工程觀點 (Engineering Insight)

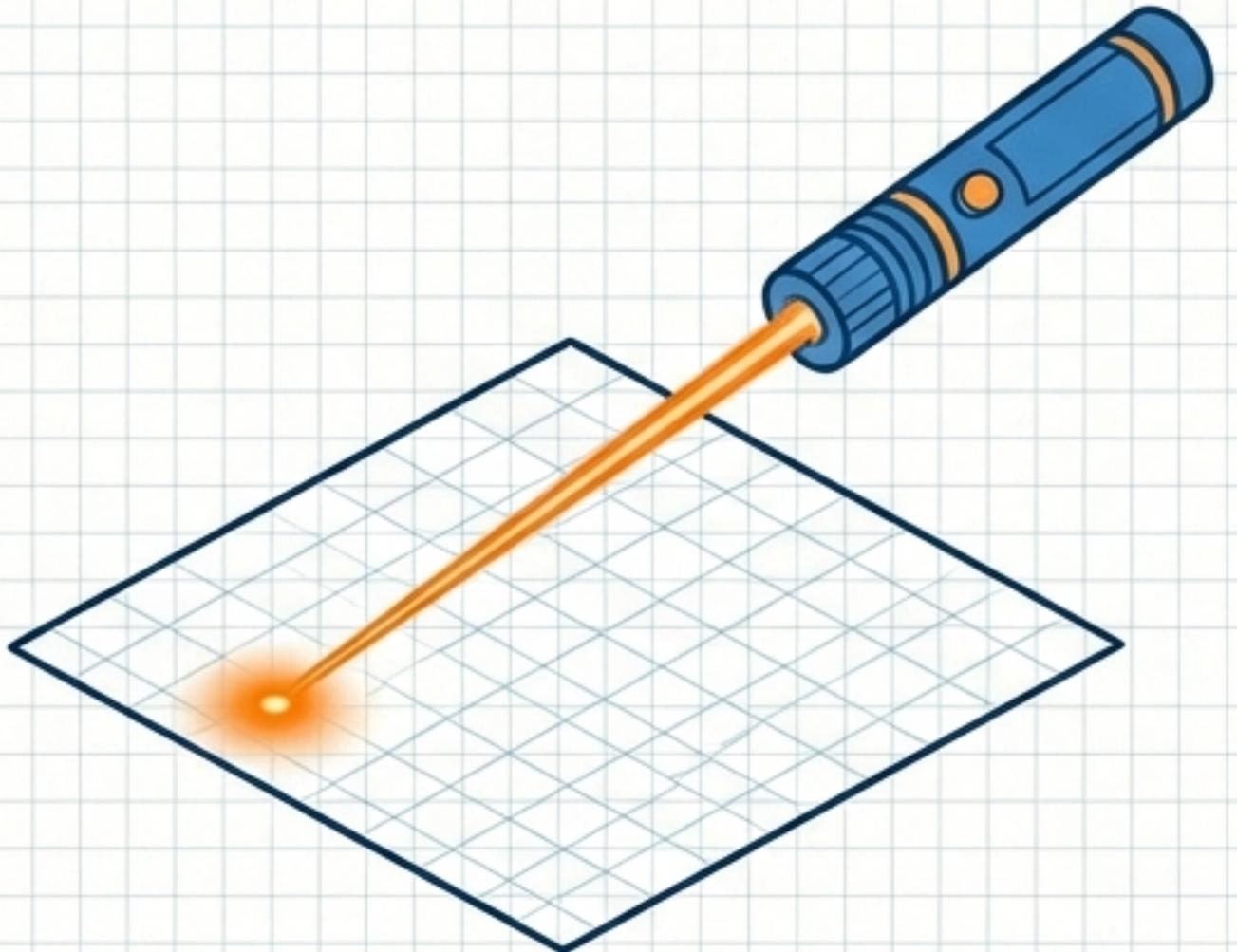
當數據噪聲很大時，應降低 C 值以避免模型去擬合噪聲 (Overfitting to noise)。

控制閥門 III : Gamma (γ) — 影響範圍 (僅限 RBF 核)

Metaphor: 聚光燈 (Spotlight) 的聚焦程度

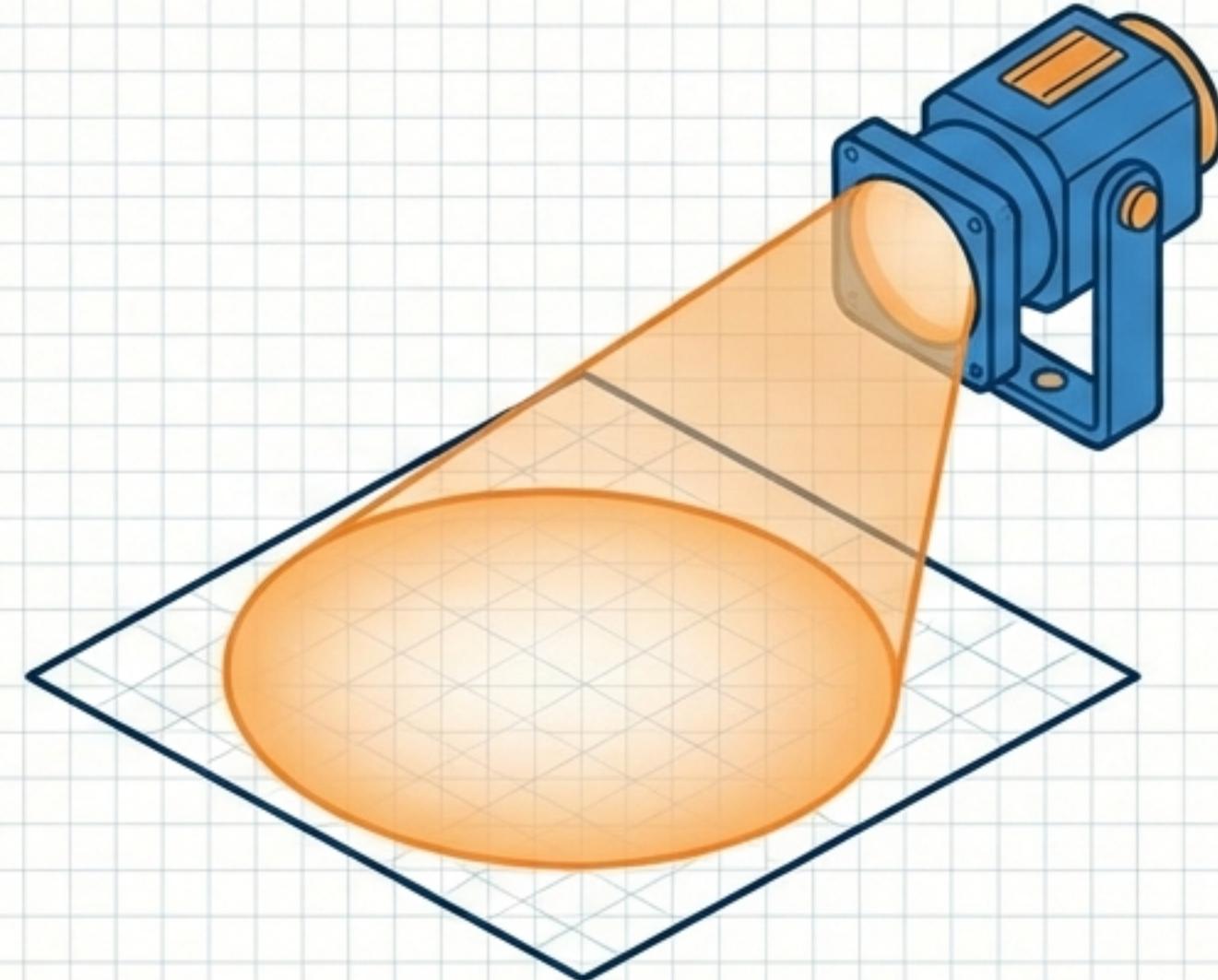
$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$$

High γ (Large Value)



局部影響 (Local)。每個樣本只影響極小範圍。
模型邊界呈鋸齒狀 (Spiky)。易過擬合。

Low γ (Small Value)



全域影響 (Global)。單一樣本影響範圍大。
模型邊界平滑 (Smooth)。易欠擬合。

關鍵標準作業程序 (SOP)：資料標準化

SVR (特別是 RBF 核) 基於歐氏距離。若特徵尺度差異過大，模型將失效。

未處理數據 (Raw Data)

Reactant Temp: 300 - 500 K



Pressure: 1 - 10 bar



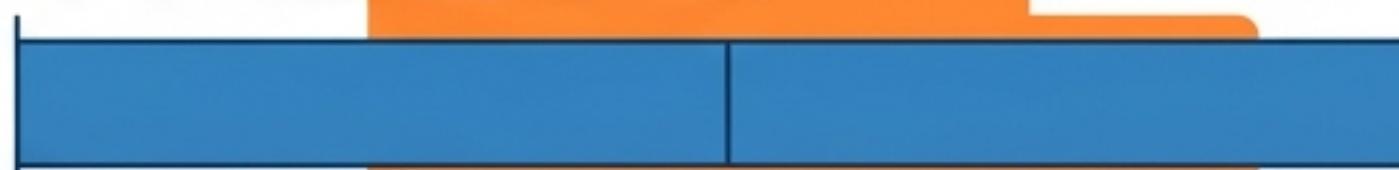
大數值特徵完全主導距離計算。

標準化數據 (Scaled Data)

Reactant Temp: -1.5 to +1.5



Pressure: -1.5 to +1.5



所有特徵權重相等。

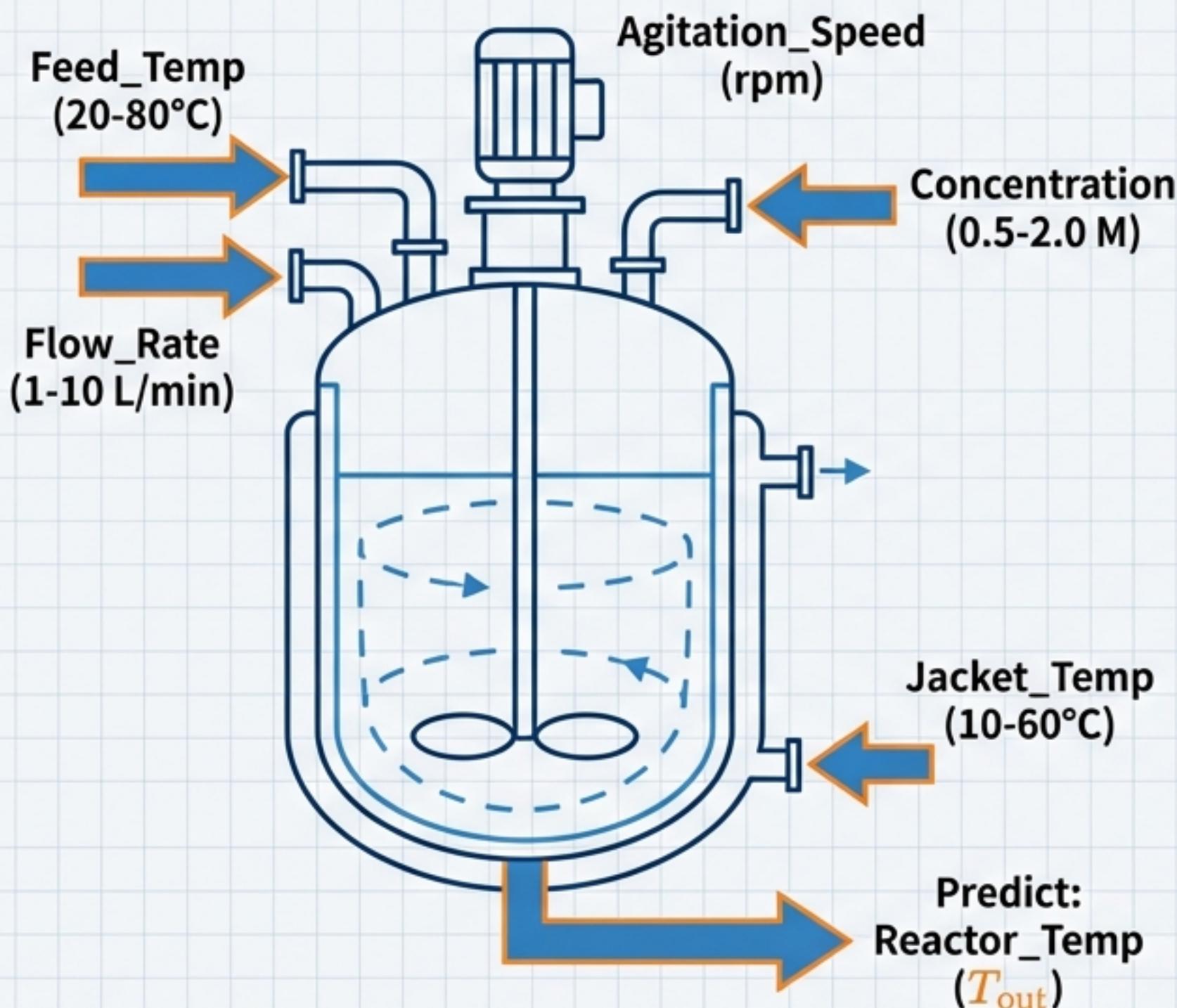


解決方案：使用 StandardScaler 將所有特徵轉換為均值=0，變異數=1 的分佈。

Caution

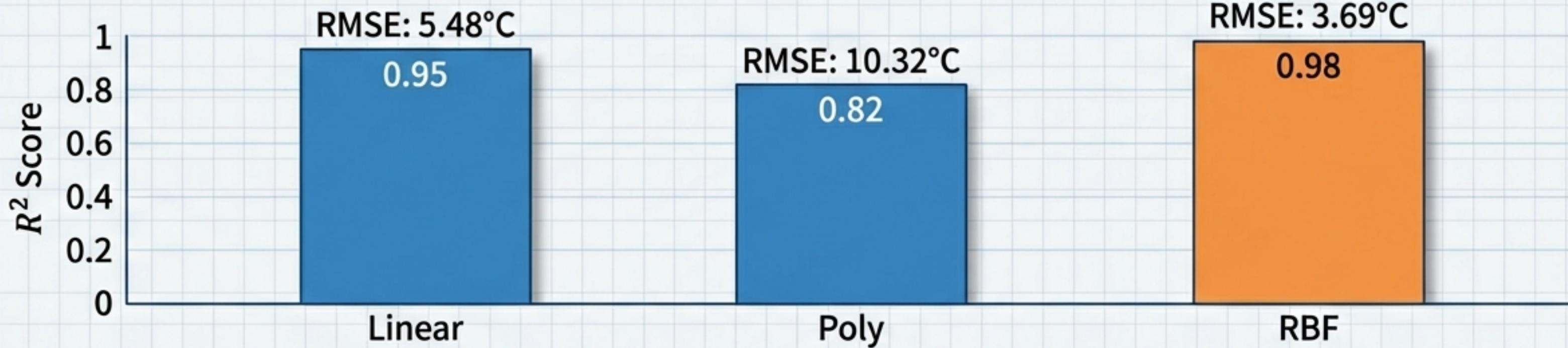


實戰案例：CSTR 反應器出口溫度預測



- **目標 (Objective)**：根據操作條件預測 T_{out} 。
- **數據集 (Dataset)**：300 筆模擬數據 (70% 訓練 / 30% 測試)。
- **物理特性**: 放熱反應，具有非線性熱平衡特徵。

案例結果 I：核函數性能比較



Kernel	R^2	RMSE	結論
RBF	0.9775	3.69	最佳 (Best Fit)
Linear	0.9505	5.48	尚可，細節不足
Poly	0.8242	10.32	過於複雜，表現最差

誤論：CSTR 的熱平衡是非線性的，Linear 核只能做到近似，而 Poly 核在此案例中過度擬合且不穩定。RBF 是首選。

案例結果 II：超參數優化 (Grid Search)

Grid Search (C, γ, ϵ) → 5-Fold CV → Best Parameters

Best Parameters Found:

- $C = 100$ (較高的懲罰，嚴格擬合)
- $\gamma = 0.01$ (較廣的影響範圍，平滑邊界)
- $\epsilon = 0.1$ (適度的誤差容忍)

Default RBF

RMSE: 3.69°C

誤差降低 23%

Optimized RBF

RMSE: 2.84°C

誤差 $\pm 2.8^{\circ}\text{C}$ 已滿足工業控制需求。

SVR 的工程適用性分析

優勢 (Pros)

-  **強魯棒性**： ϵ -loss 機制有效對抗工廠數據噪聲。
-  **高維處理**：適合變數多、樣本少的化工實驗數據。
-  **泛化能力**：結構風險最小化原則，避免過擬合。

限制 (Cons)

-  **計算成本**：訓練時間隨樣本數平方增長 ($O(N^2)$)，不適合大數據 ($N > 10,000$)。
-  **解釋性低**：黑盒子模型，不如決策樹直觀。
-  **需預處理**：對數據尺度極度敏感，必須標準化。

工程師的實踐清單 (Best Practices)

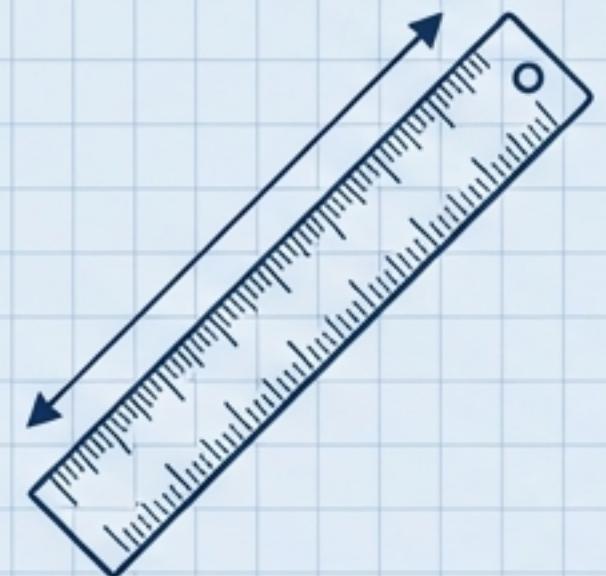


SVR Implementation Checklist

- 數據標準化：務必使用 StandardScaler。
- 處理異常值：SVR 雖穩健，但極端值仍需檢查。
- 首選 RBF 核：通用性最強，適合大多數化工問題。
- 網格搜尋 (Grid Search)：使用對數尺度搜尋 C ($10^{-1} \dots 10^2$) 和 γ ($10^{-3} \dots 10^0$)。
- 監控支持向量比例：理想應在 30-70% 之間。

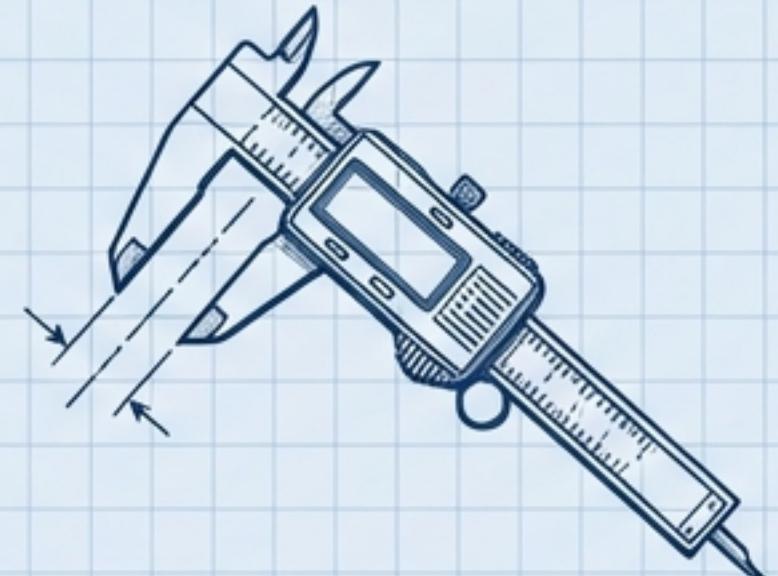
“Code is read much more often than it is written.”

結論：您的數位工具箱



線性回歸 = 直尺

用於簡單、線性問題的快速近似與基準比較。



SVR = 精密卡尺

雖然慢，但在非線性與高維度問題上提供卓越精確度。



實作

將理論轉化為實際的工程解決方案，掌握工具。

→ 下一步 (Next Step): 開啟 Jupyter Notebook `Unit11_Support_Vector_Machine.ipynb` 進行實作。
課後作業: 比較 SVR 與 Random Forest 在相同數據集上的表現。