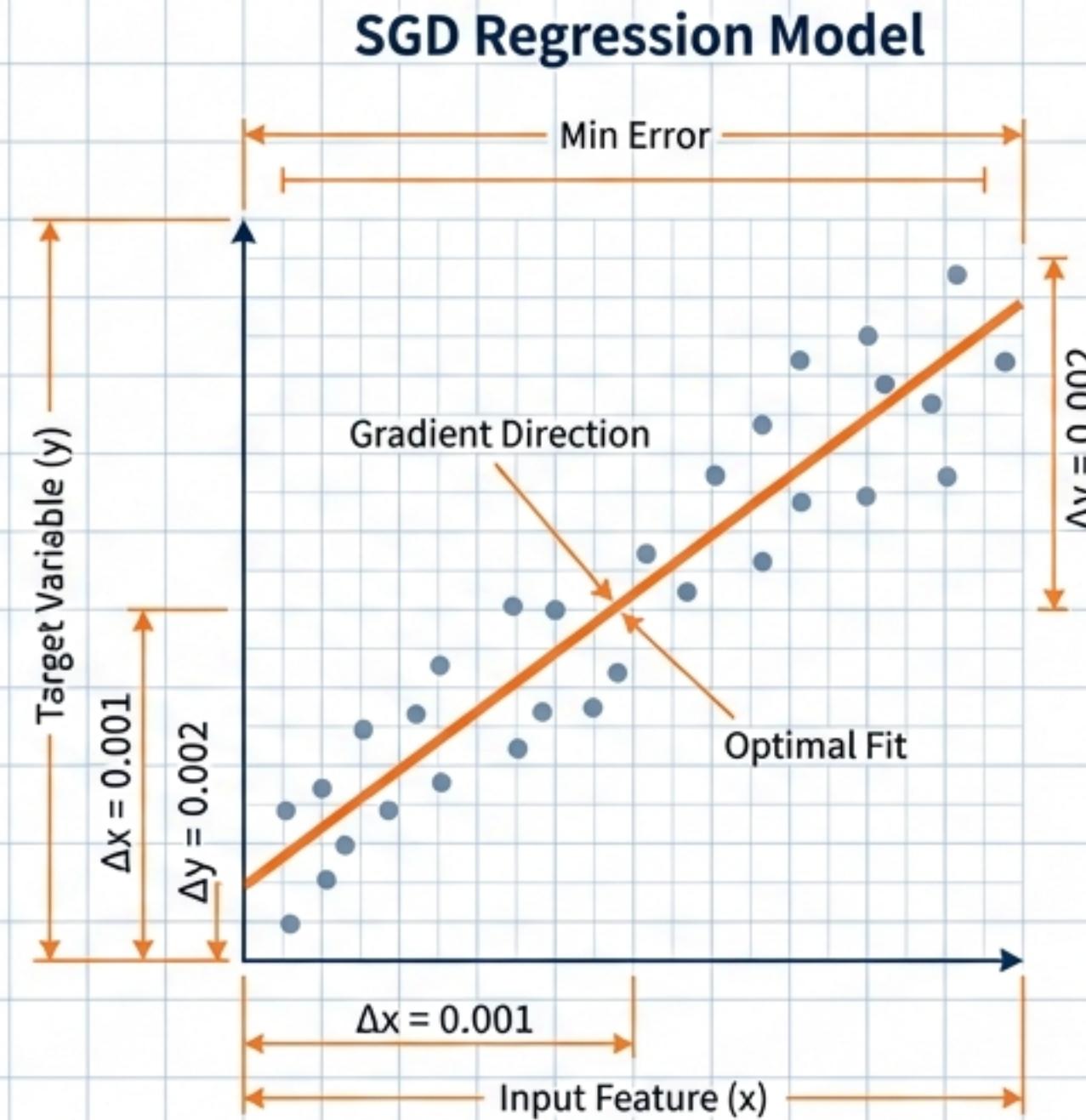
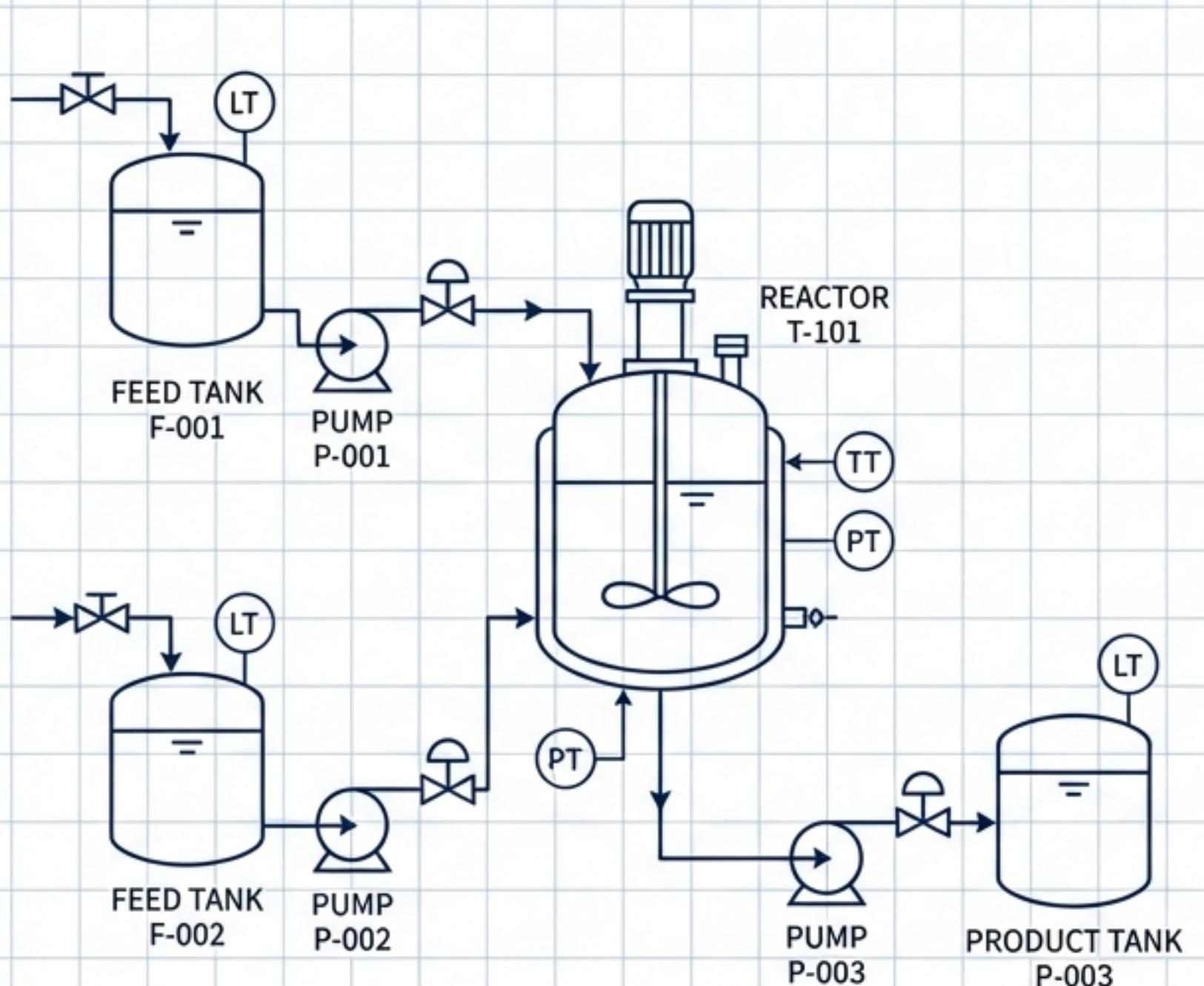


Unit 10: 隨機梯度下降回歸 (SGD Regression)

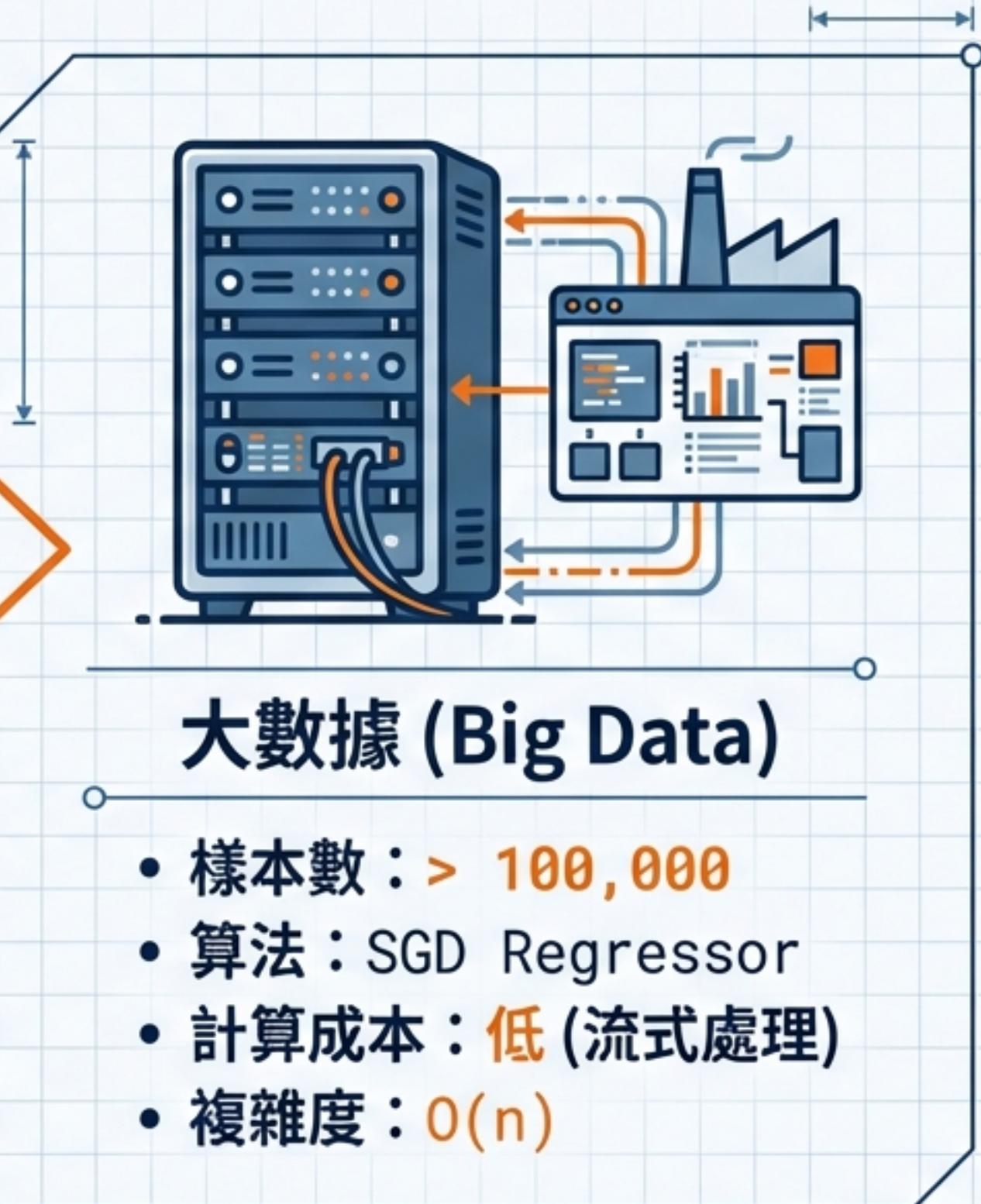
應對大規模化工數據與在線學習的利器



為什麼需要 SGD？從實驗室到化工廠的範式轉移

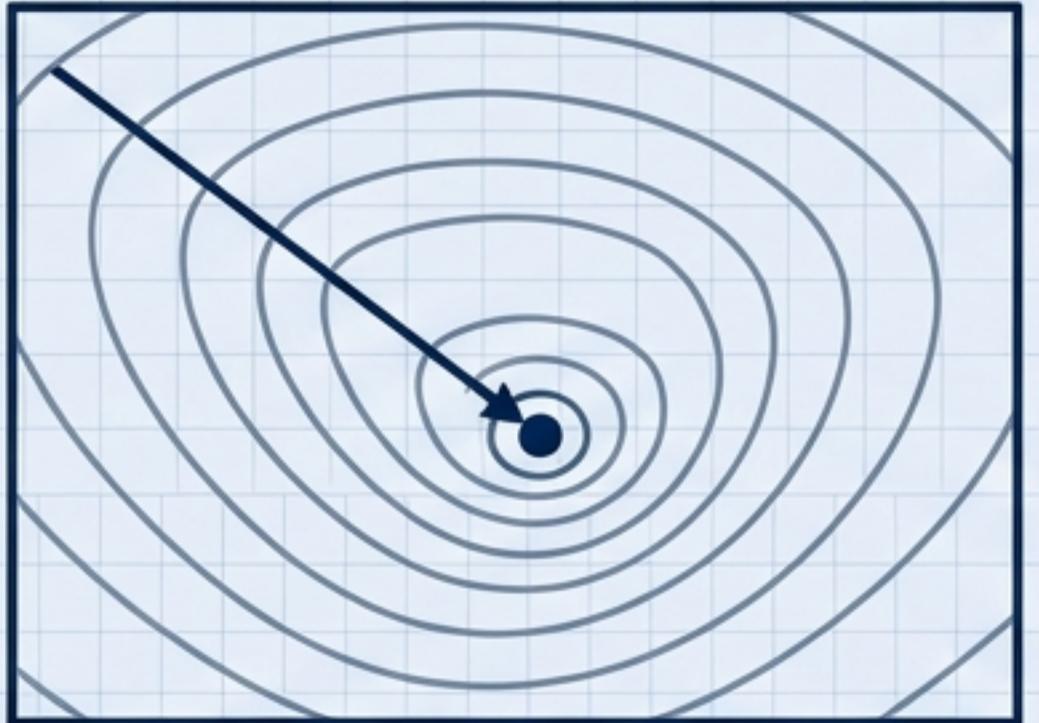


典範轉移
(Paradigm Shift)



梯度下降策略比較：速度與穩定性的權衡

Batch Gradient Descent (BGD)



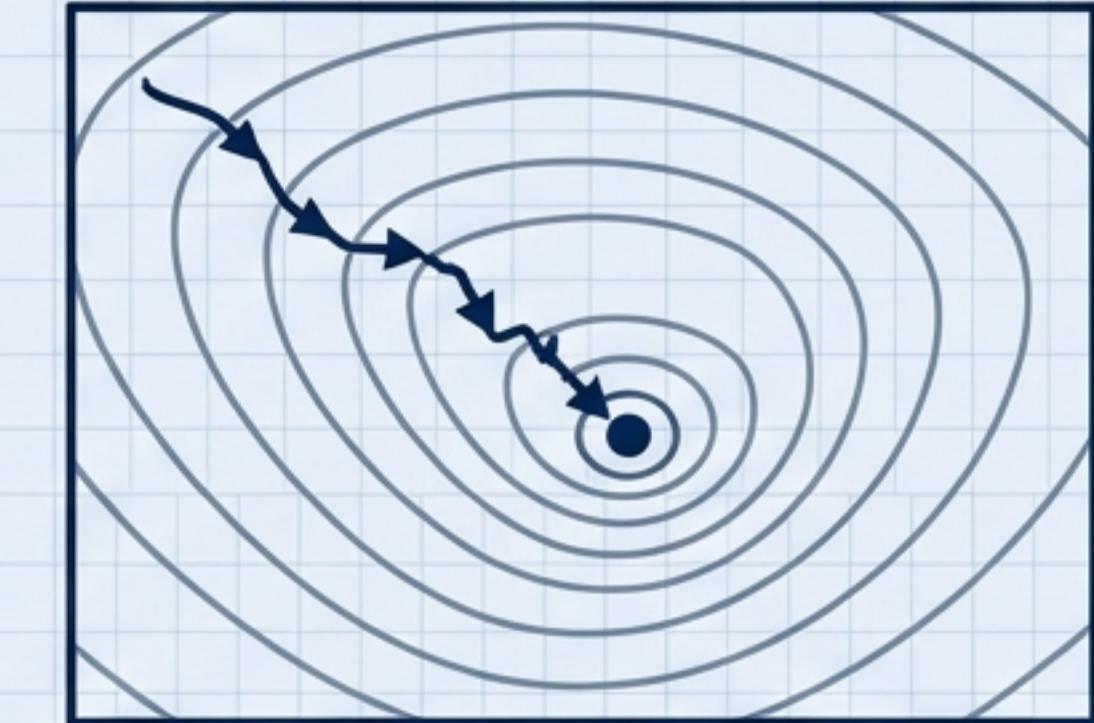
使用全部樣本 (Stable but Slow)

Stochastic Gradient Descent (SGD)



使用 1 個樣本 (Fast but Noisy)

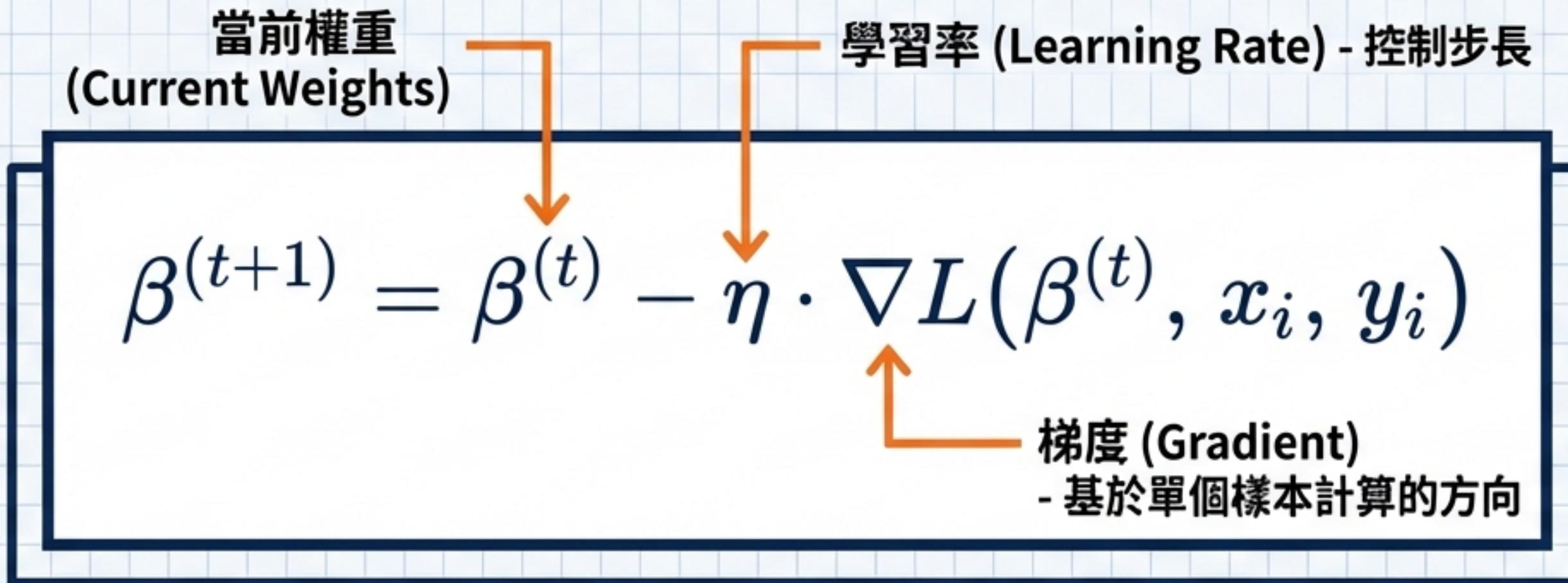
Mini-Batch Gradient Descent (MBGD)



使用 m 個樣本 (Balanced)

	計算成本 (Computational Cost)	收斂性 (Convergence)	適用場景 (Suitability)
BGD	高 $O(n)$	收斂穩定	適合小數據
SGD	低 $O(1)$	收斂波動大	適合大數據/在線學習
MBGD	中等	相對穩定	適合 GPU 加速

數學原理：參數更新機制 (The Update Rule)

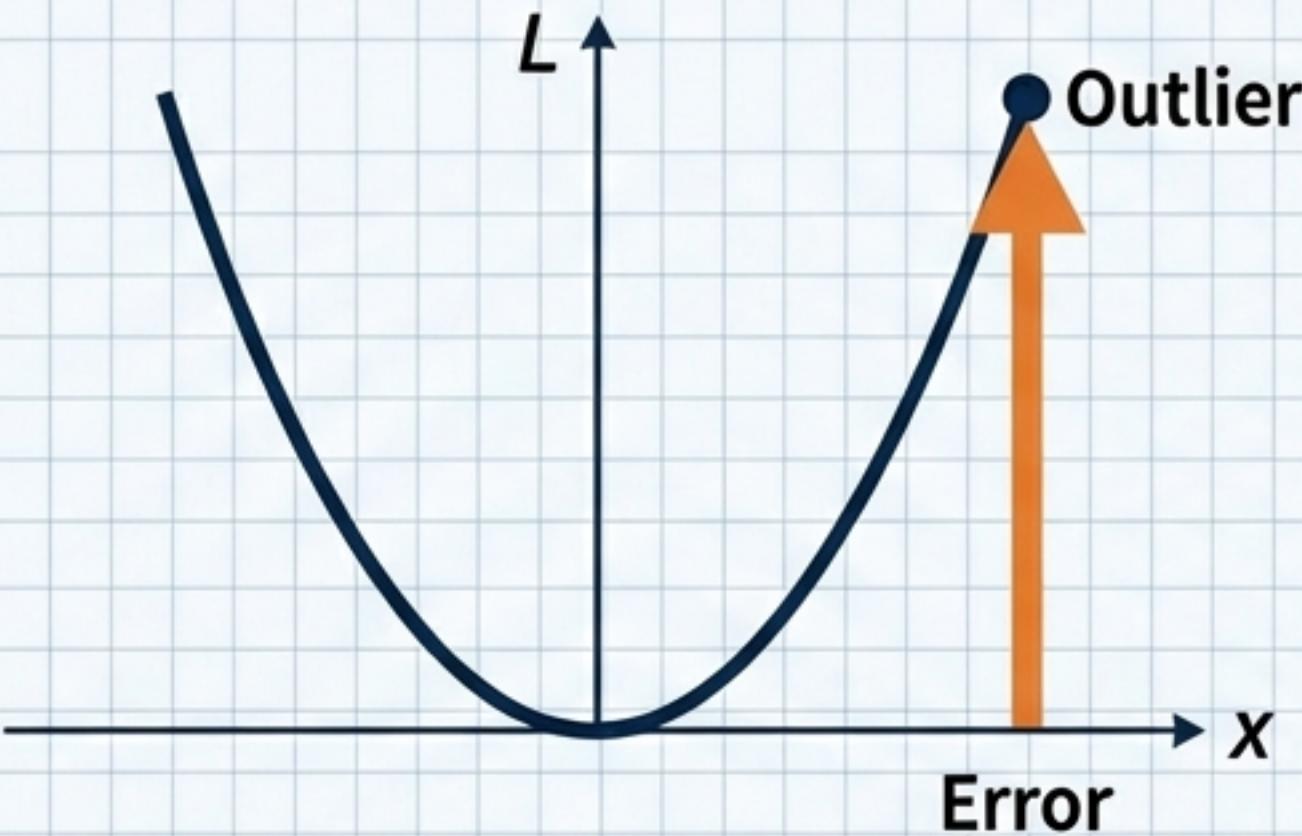


核心機制：

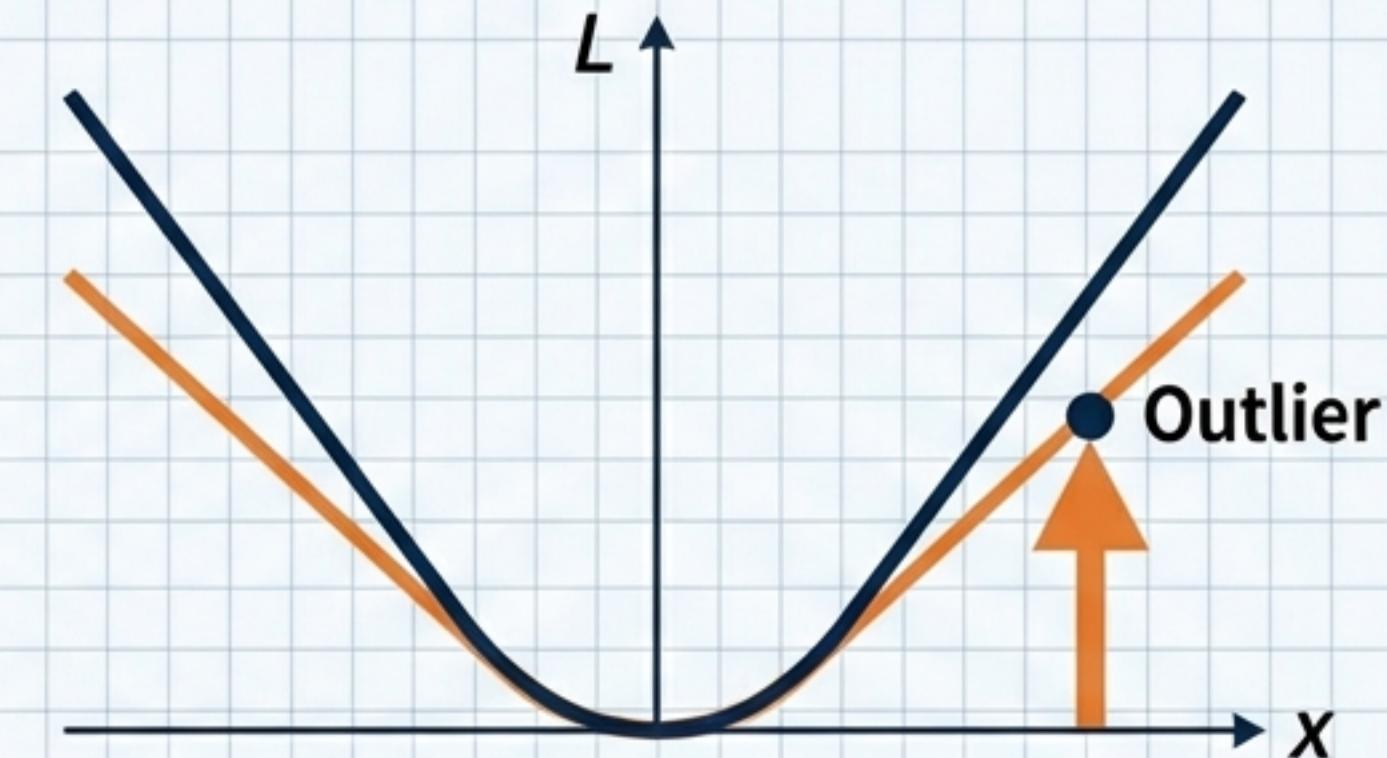
1. 每次迭代僅隨機抽取一個樣本 (x_i, y_i) 。
2. 計算該樣本的梯度並更新權重。
3. 若加入正則化 (α)，公式增加權重衰減項 $(1 - \eta\alpha)$ 。

損失函數 (Loss Functions)：應對現實世界的噪聲

Squared Loss (OLS)



Huber Loss

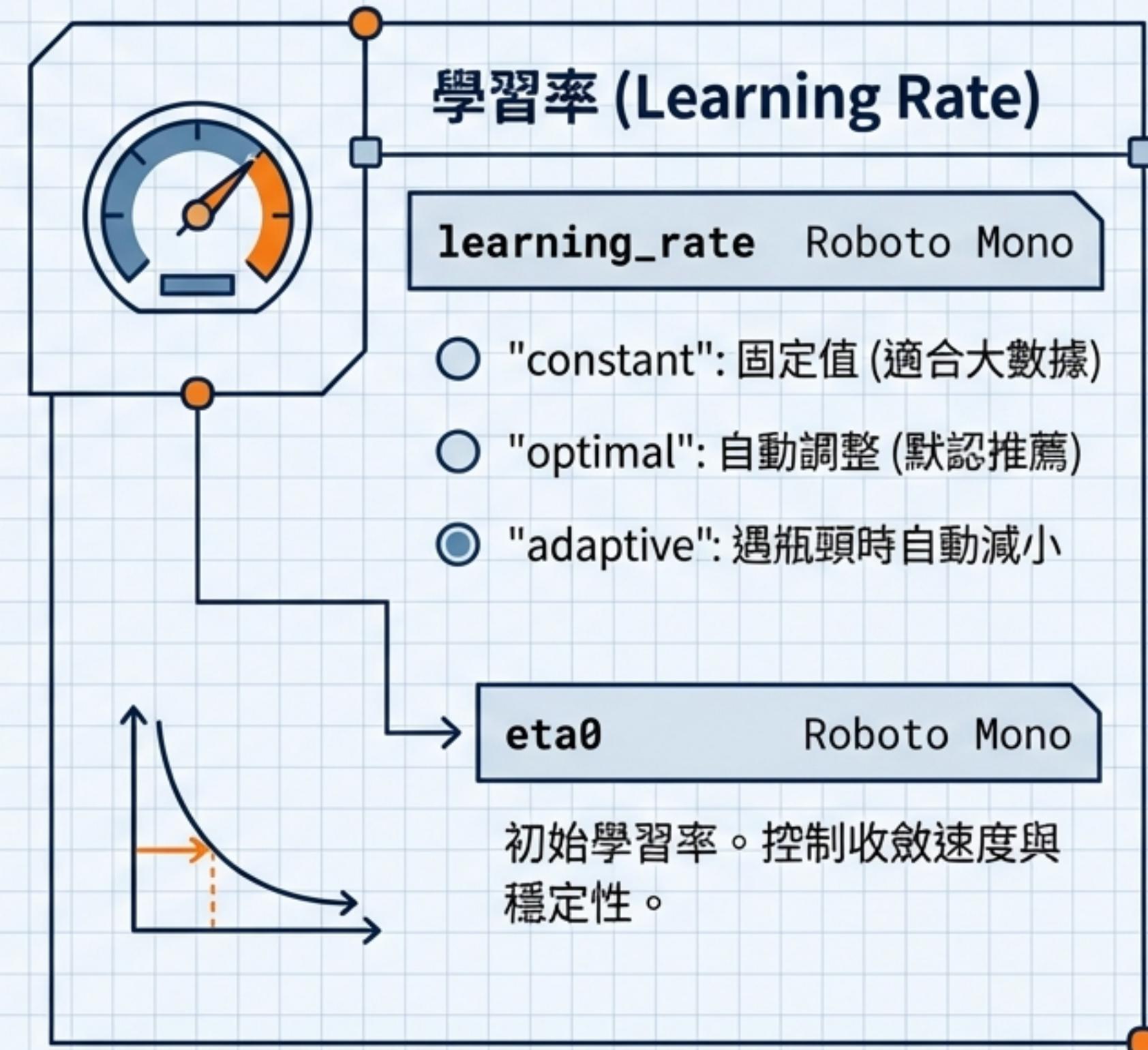
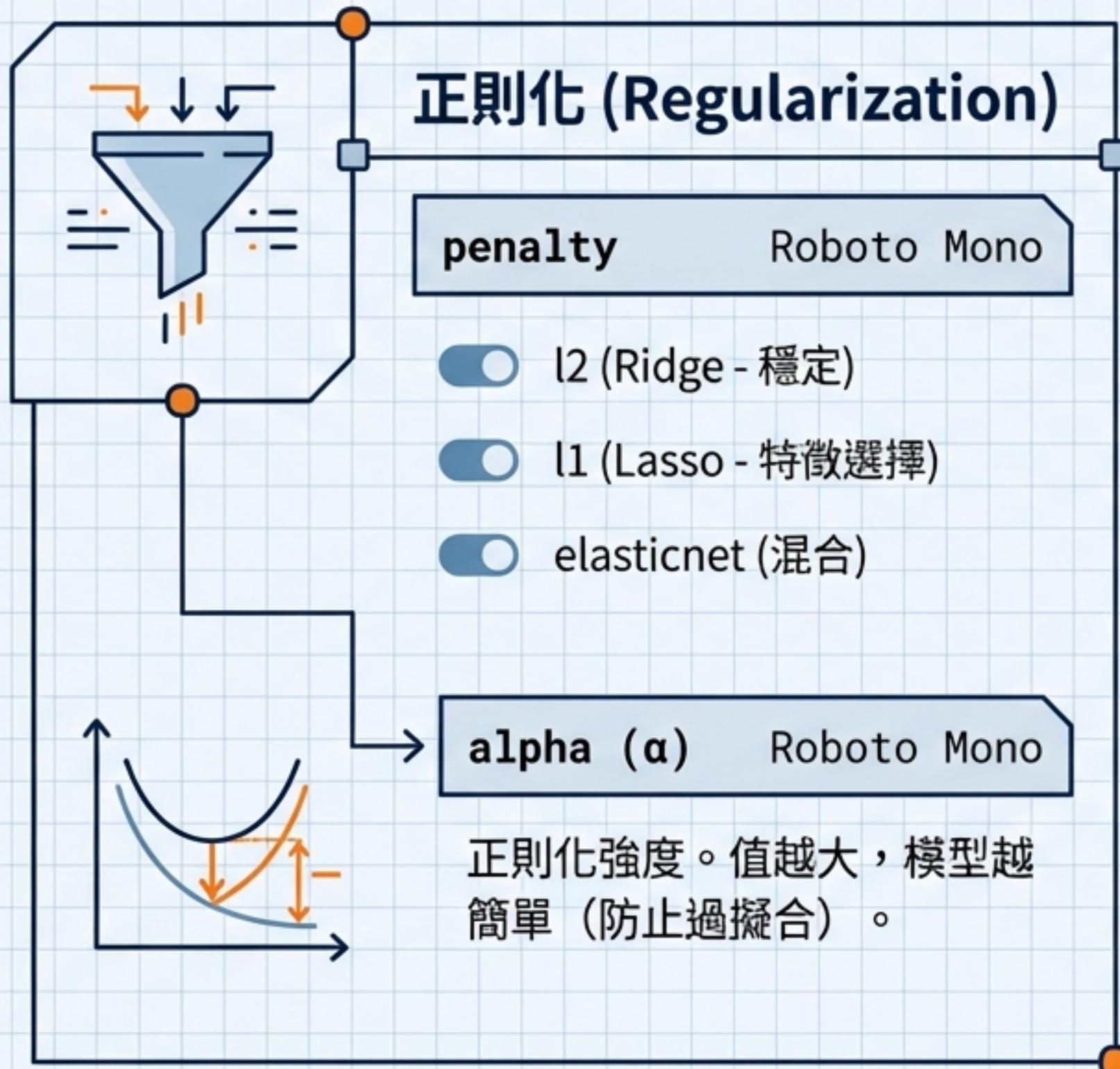


對異常值敏感 (Sensitive)。
默認選擇，計算快。

對異常值魯棒 (Robust)。
適合傳感器故障或噪聲數據。

化工應用場景：當數據包含傳感器瞬間突波 (Spikes) 時，Huber Loss 能防止模型被異常值誤導。

關鍵超參數調控 (Hyperparameter Tuning)



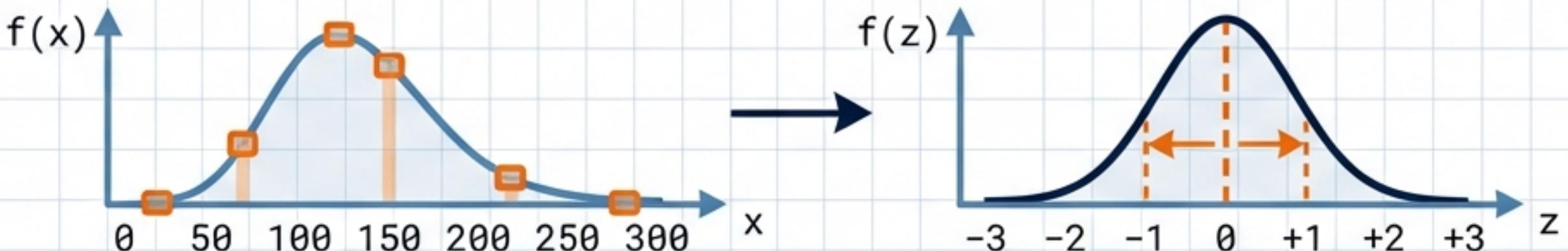
⚠ 致命陷阱：數據標準化 (Standardization)

SGD 對特徵尺度極其敏感 (Extremely Sensitive to Scaling)

如果特徵尺度不同（例如：溫度 300 K vs 壓力 1 bar），梯度更新路徑會被拉伸，導致模型收斂極慢甚至發散。

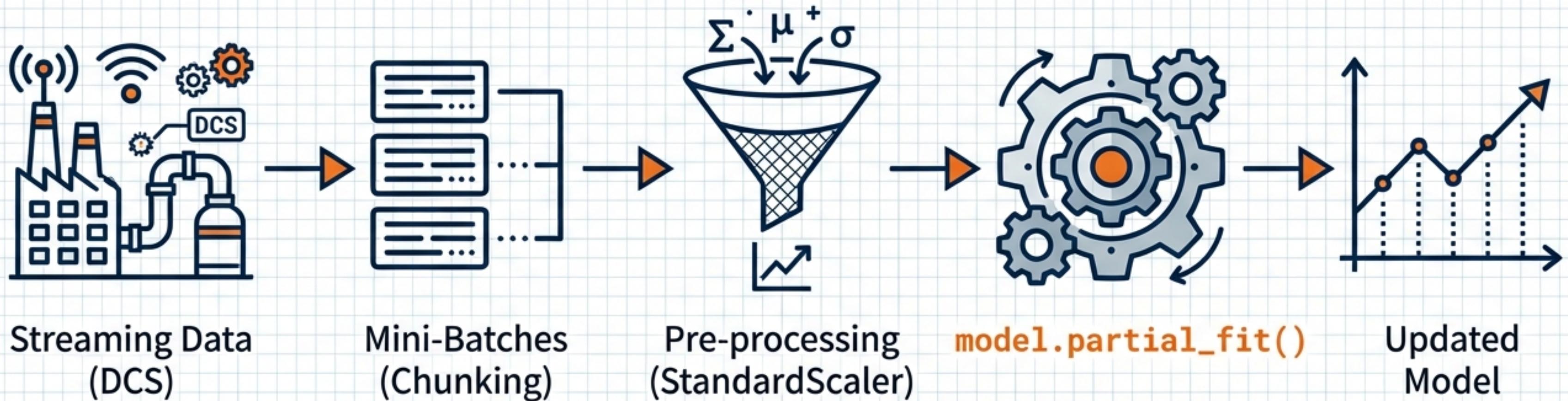
解決方案：必須使用 StandardScaler

$$z = \frac{x - \mu}{\sigma}$$



```
scaler.fit(X_train)  
X_train_scaled = scaler.transform(X_train)
```

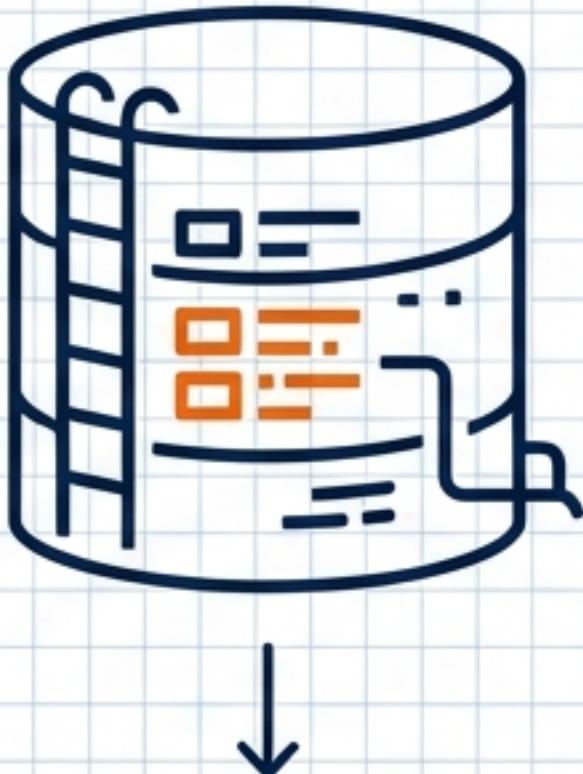
工程工作流程：從批次到流式處理



核心概念比較：

- `fit(X, y)` → 重置模型，重新訓練 (Reset & Retrain)
- `partial_fit(X, y)` → 增量更新 (Incremental Update)，保留權重，適應新數據

化工領域應用場景 (ChemE Applications)



大規模生產建模

DCS 系統累積數百萬筆歷史數據。用於反應器產率預測 (**Yield Prediction**)，無需載入全部記憶體。



在線學習與監控

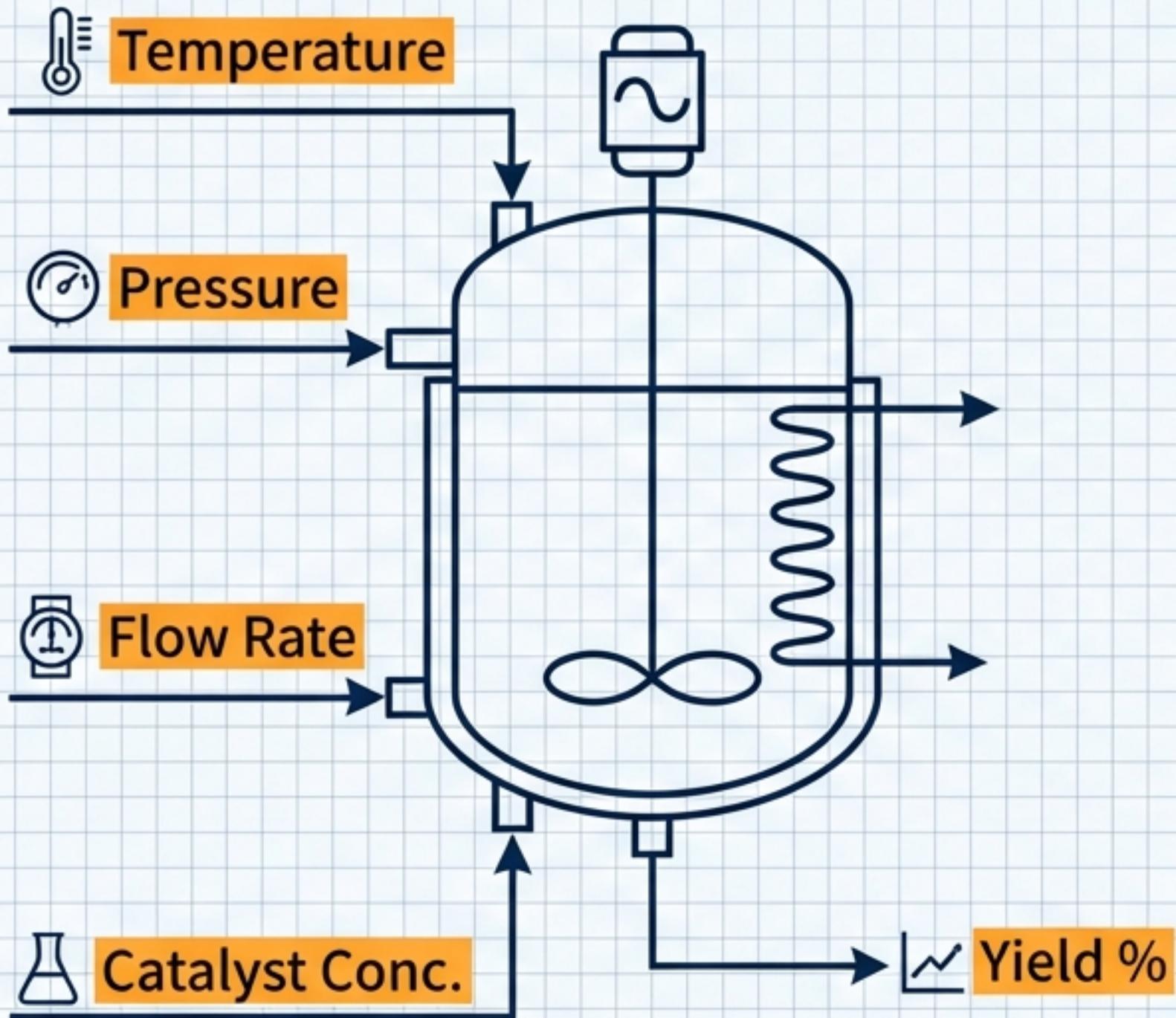
應對**概念漂移 (Concept Drift)**。當操作條件隨時間變化時，使用 `partial_fit` 即時更新軟感測器 (**Soft Sensors**)。



魯棒異常檢測

利用 **Huber Loss** 過濾傳感器噪聲。在含雜訊的數據中準確預測產品純度。

實戰案例：化工反應器產率預測



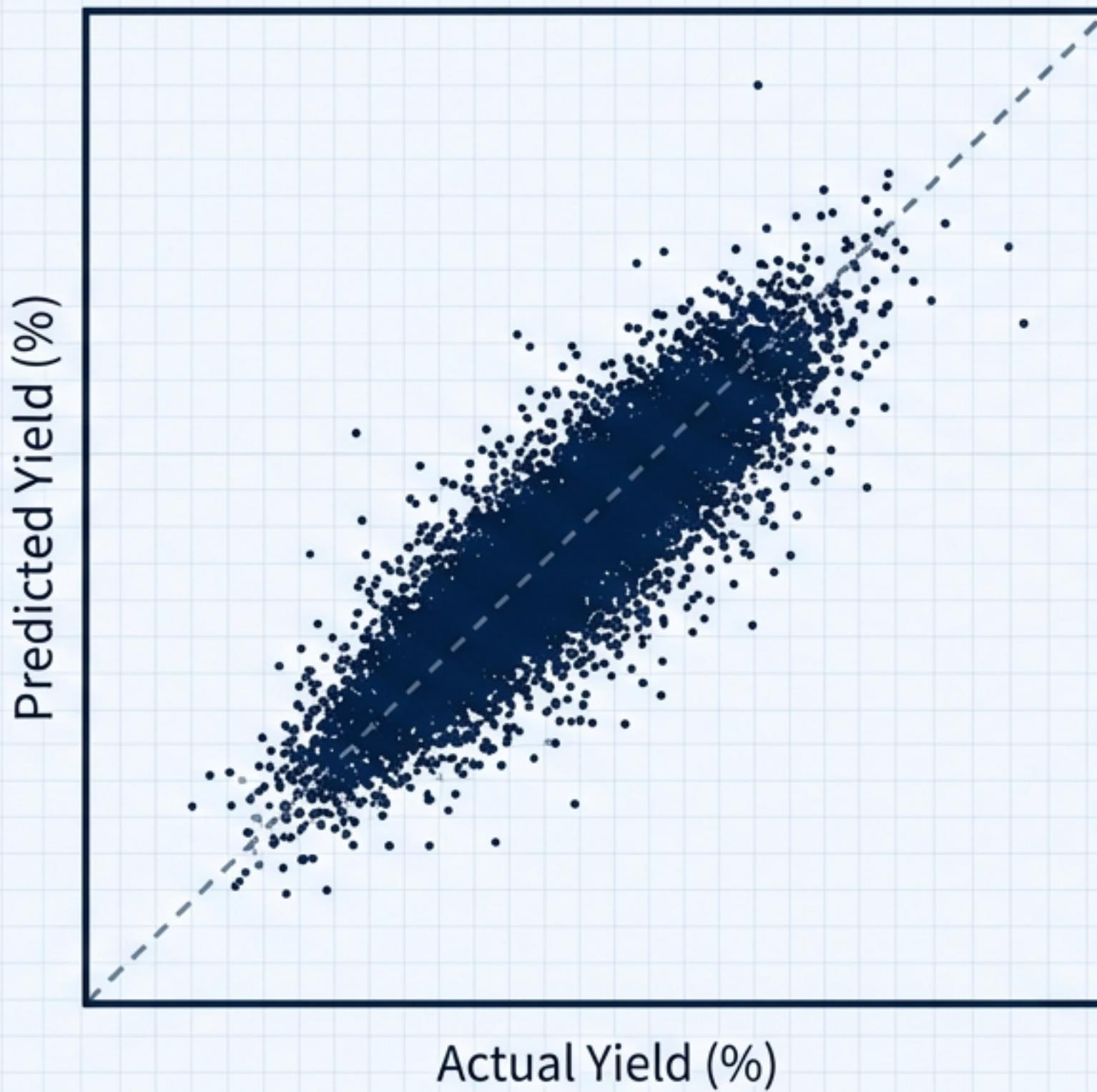
數據規格 (Data Specs):

- 樣本數 (N): 50,000 筆操作數據
- 特徵 (Features): 10 個變數 (Temp, Pressure, etc.)
- 噪聲 (Noise): 含 5% 異常值 (模擬傳感器故障)
- 目標 (Objective): 預測產品產率 (Yield)

關鍵挑戰：

數據量大，且包含現實世界的測量噪聲。需建立魯棒且快速的回歸模型。

模型訓練結果與評估



Performance Metrics

R² Score: 0.8083

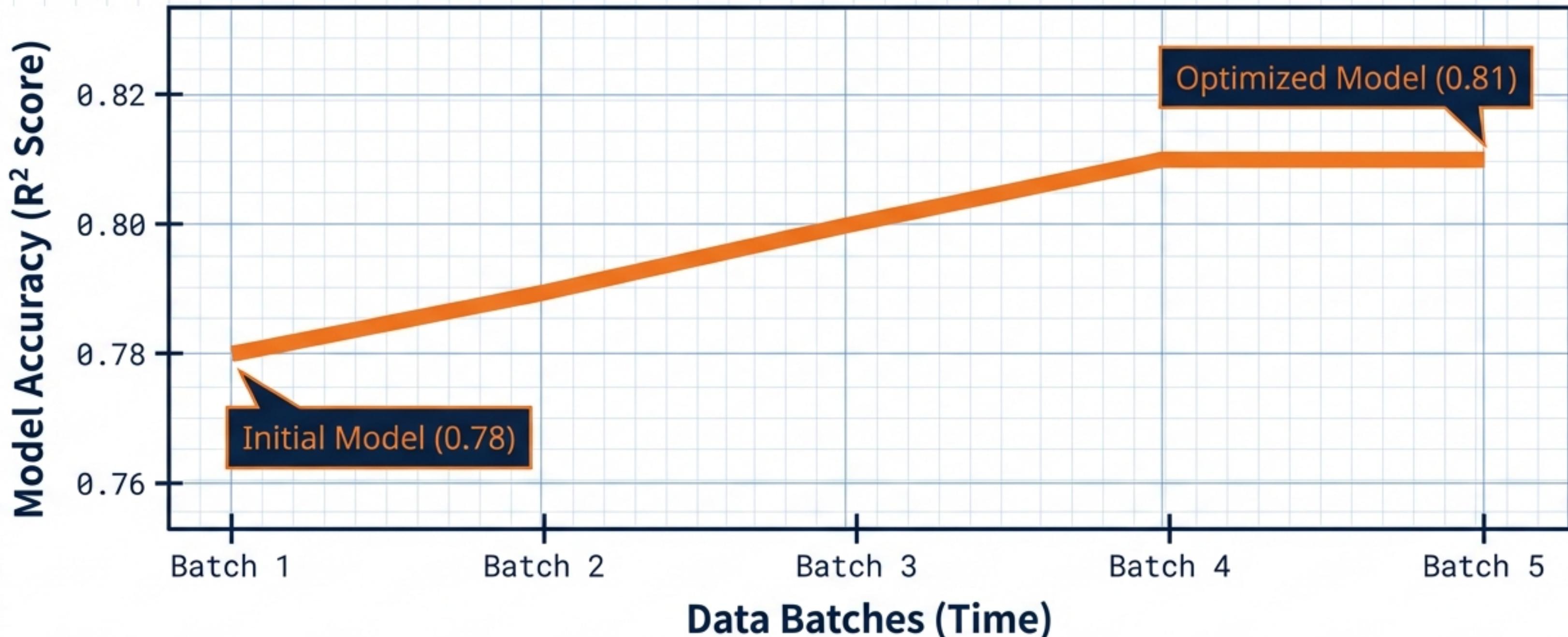
RMSE: 8.2%

Convergence: 91 Epochs (Fast!)

特徵重要性 (Feature Importance)

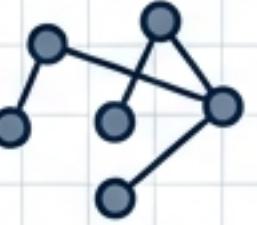
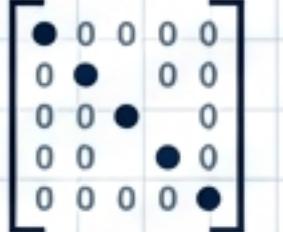
- Temperature (+): 正相關，最關鍵因素
- Feed Rate (-): 負相關，流速過快導致反應不完全
- Conclusion: SGD 成功從含噪聲的大數據中提取了物理規律。

在線學習演示 (Online Learning in Action)



- 使用 `partial_fit()` 模擬流式數據到達。
- 價值：無需重新訓練即可適應新數據，實現「持續改善」。

模型選擇指南：何時使用 SGD？

場景與數據特徵	決策與建議	理由與依據
小數據 (< 10k 樣本)	 → 避免 SGD Ridge / Lasso / ElasticNet (更穩定)	模型容易過擬合，傳統線性回歸模型表現更佳
大數據 <td> → 推薦 SGD Regressor</td> <td>計算效率高，線性時間複雜度 $O(n)$</td>	 → 推薦 SGD Regressor	計算效率高，線性時間複雜度 $O(n)$
流式數據 / 在線監控	 → 推薦 SGD Regressor	唯一支持 <code>partial_fit</code> 的線性模型，適合實時更新
特徵稀疏或高維度	 → 推薦 SGD Regressor	處理稀疏數據效率極高，有效利用內存

常見問題與故障排查 (Troubleshooting)



模型發散或 $R^2 < 0$ ？

👉 檢查是否使用了 StandardScaler。
調整學習率 (learning_rate)。



訓練過慢？

👉 啟用早停法 (early_stopping=True)。
減少迭代次數 (max_iter)。



過擬合 (Overfitting)？

👉 增大正則化參數 alpha。嘗試 L1 或 ElasticNet。



受異常值影響？

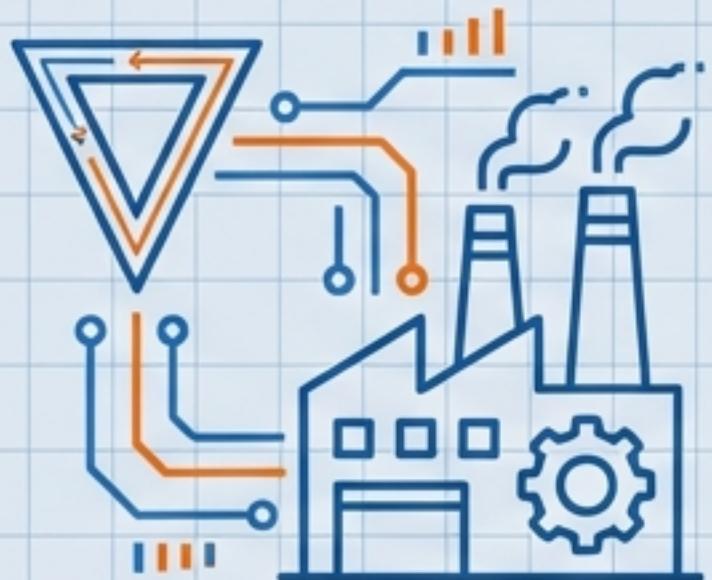
👉 將 loss 參數改為 'huber'。



結果無法重現？

👉 固定 random_state。

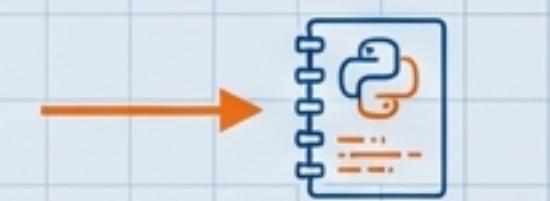
結語：連接統計學與大數據的橋樑



SGD 是現代機器學習的基石，連結了傳統統計回歸與深度學習優化。掌握 **標準化**、**損失函數** 與 **學習率** 是成功的關鍵。

下一步 (Next Step) :

開啟 **Unit10_SGD_Regression.ipynb** 進行實作練習。



不要懼怕大數據。利用 SGD，讓您的模型隨工廠數據一同成長。

