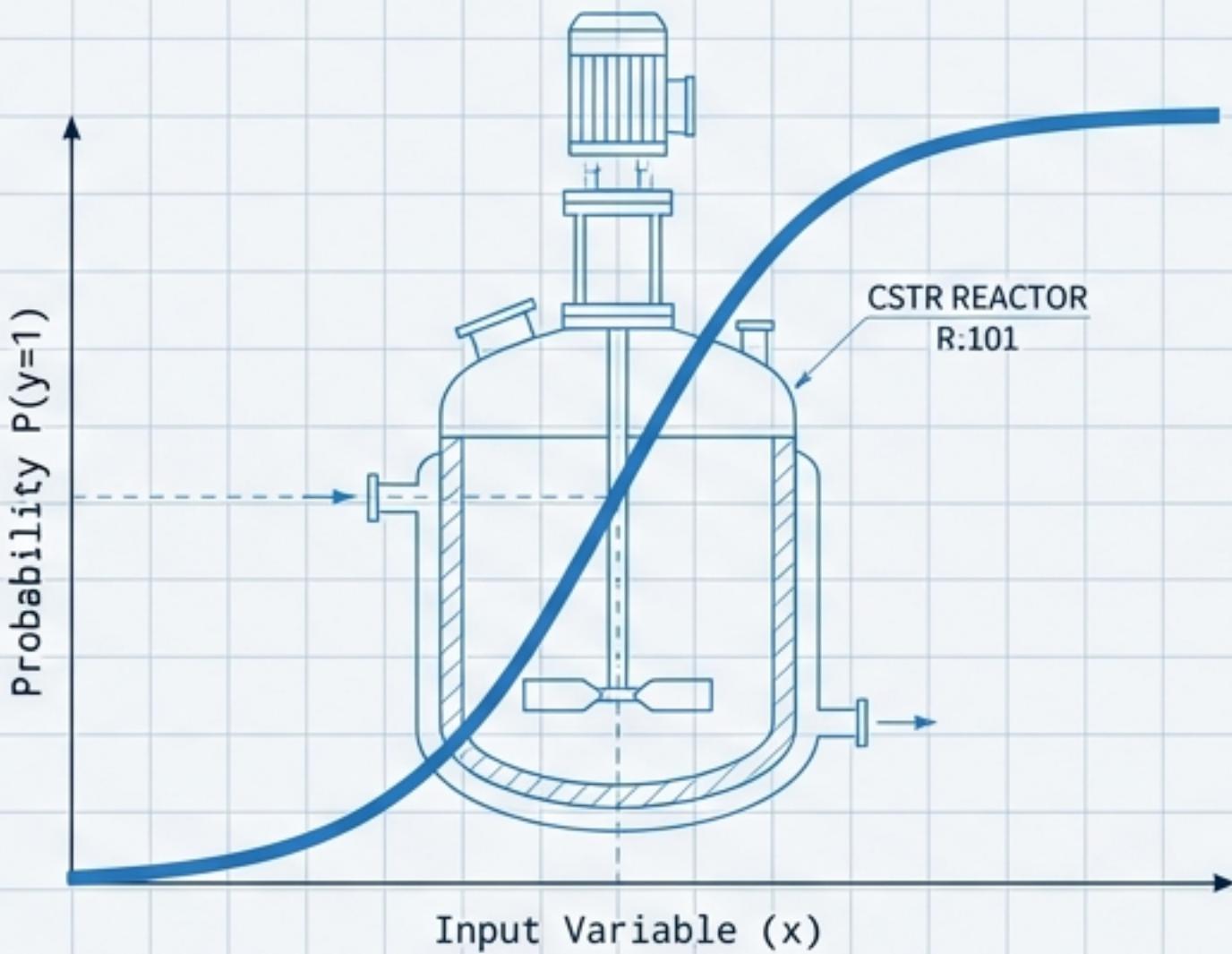


# Unit 12: 邏輯迴歸 (Logistic Regression)

## 機器學習中的二元分類與化工應用



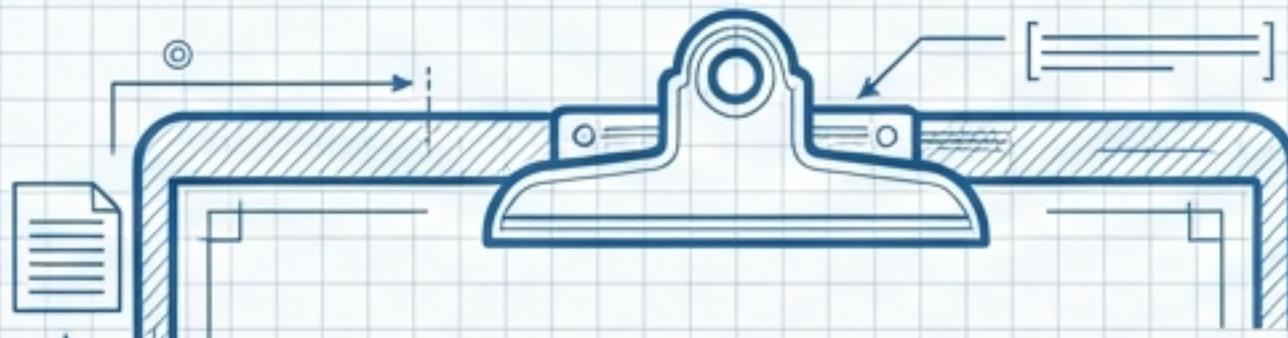
課程：AI在化工上之應用

製作：逢甲大學 化工系 智慧程序系統工程實驗室

指導：莊曜楨 助理教授

日期：2026-01-28

# 本單元學習目標 (Spec Sheet)

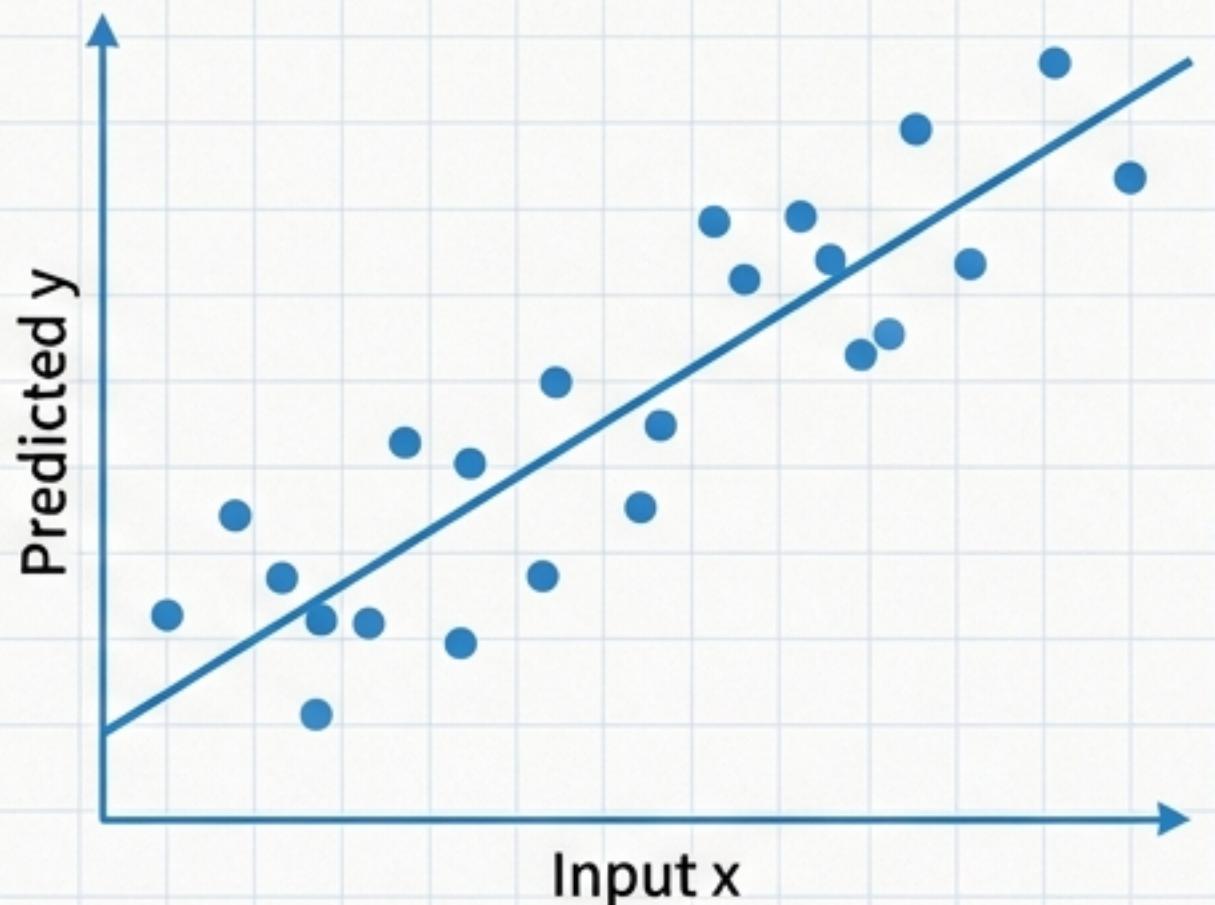


- ✓ **原理理解**：理解 Sigmoid 函數將線性輸出轉換為機率的數學機制。
- ✓ **核心運算**：掌握對數損失 (Log Loss) 函數的推導與優化原理。
- ✓ **工具實作**：熟練 `sklearn` LogisticRegression`` 的參數設定 (C, penalty, solver)。
- ✓ **化工應用**：應用於產品合格判定、反應成功預測與異常檢測。
- ✓ **模型評估**：解讀混淆矩陣、ROC 曲線與 AUC 指標。



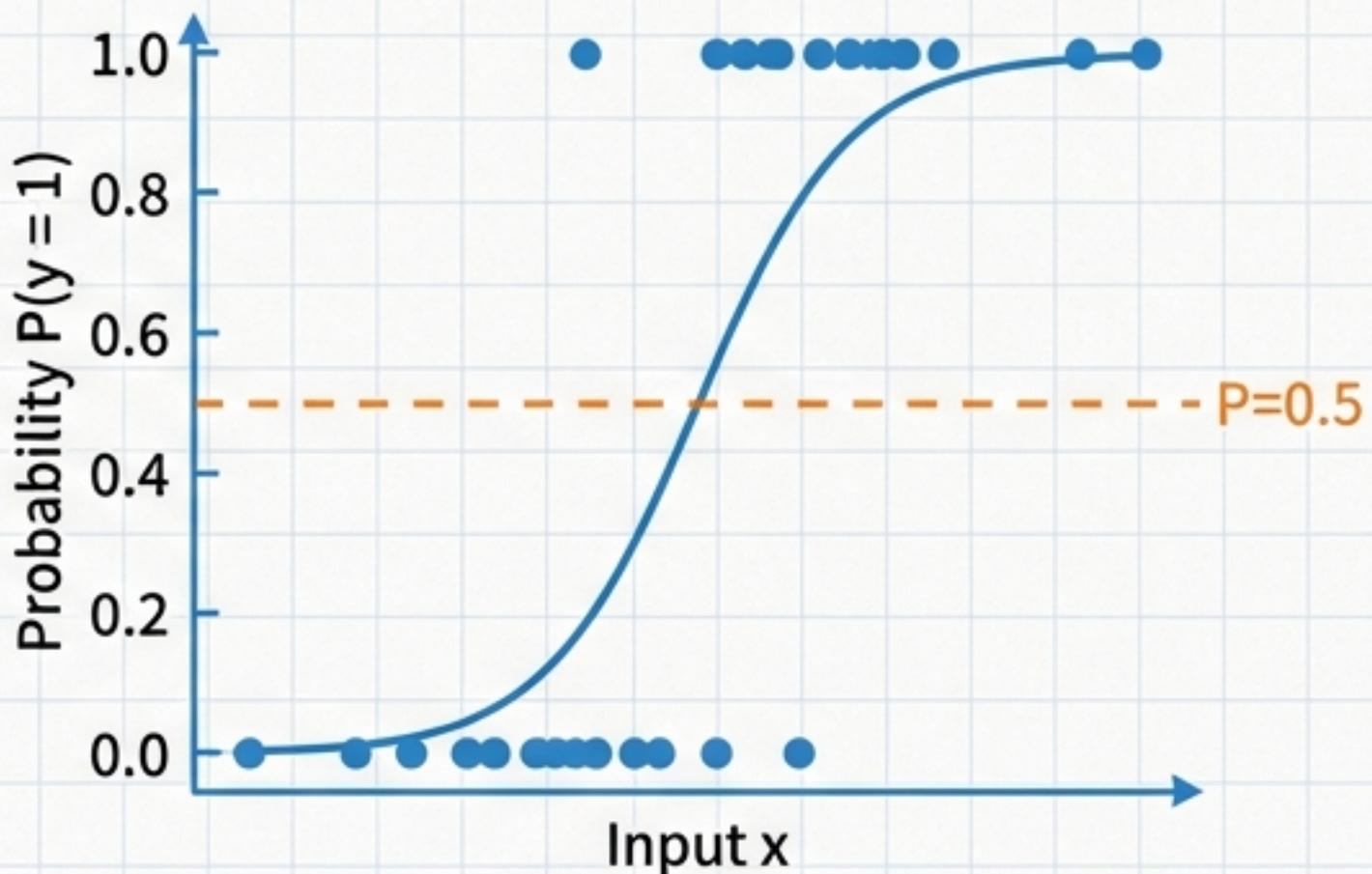
# 核心概念：雖名為「迴歸」，實為「分類」

## 線性迴歸 (Linear Regression)



$$\hat{y} = \mathbf{w}^T \mathbf{x} + b \text{ (預測連續值)}$$

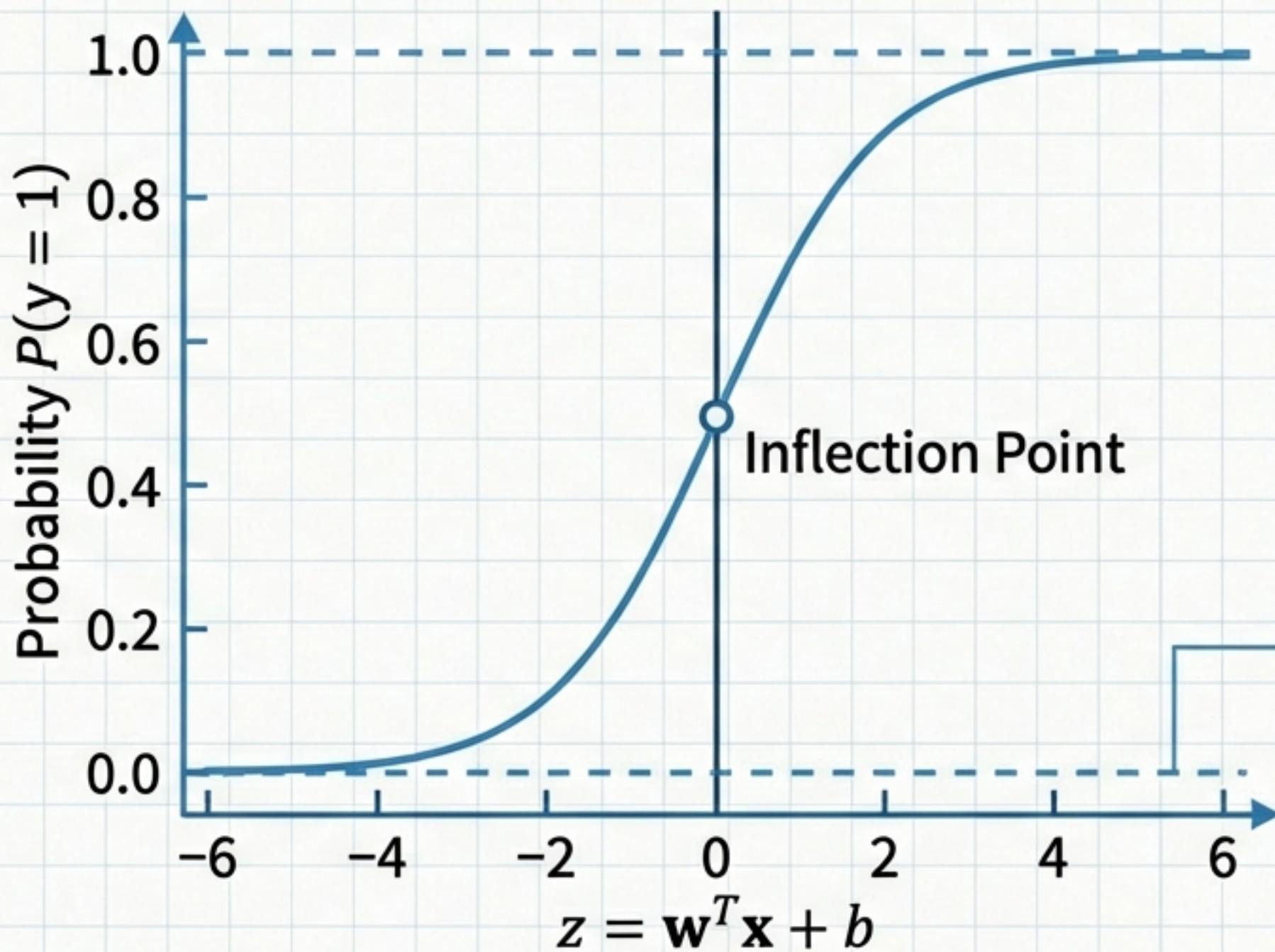
## 邏輯迴歸 (Logistic Regression)



$$P(y = 1) = \sigma(\mathbf{w}^T \mathbf{x} + b) \text{ (預測機率 [0, 1])}$$

**決策邊界 (Decision Boundary)**：當機率  $P = 0.5$  時， $\mathbf{w}^T \mathbf{x} + b = 0$  定義了分類的線性邊界。  
判定規則：機率  $\geq 0.5$  判為正類 (1)；機率  $< 0.5$  判為負類 (0)。

# 核心機制：Sigmoid 函數 (The Activation Valve)



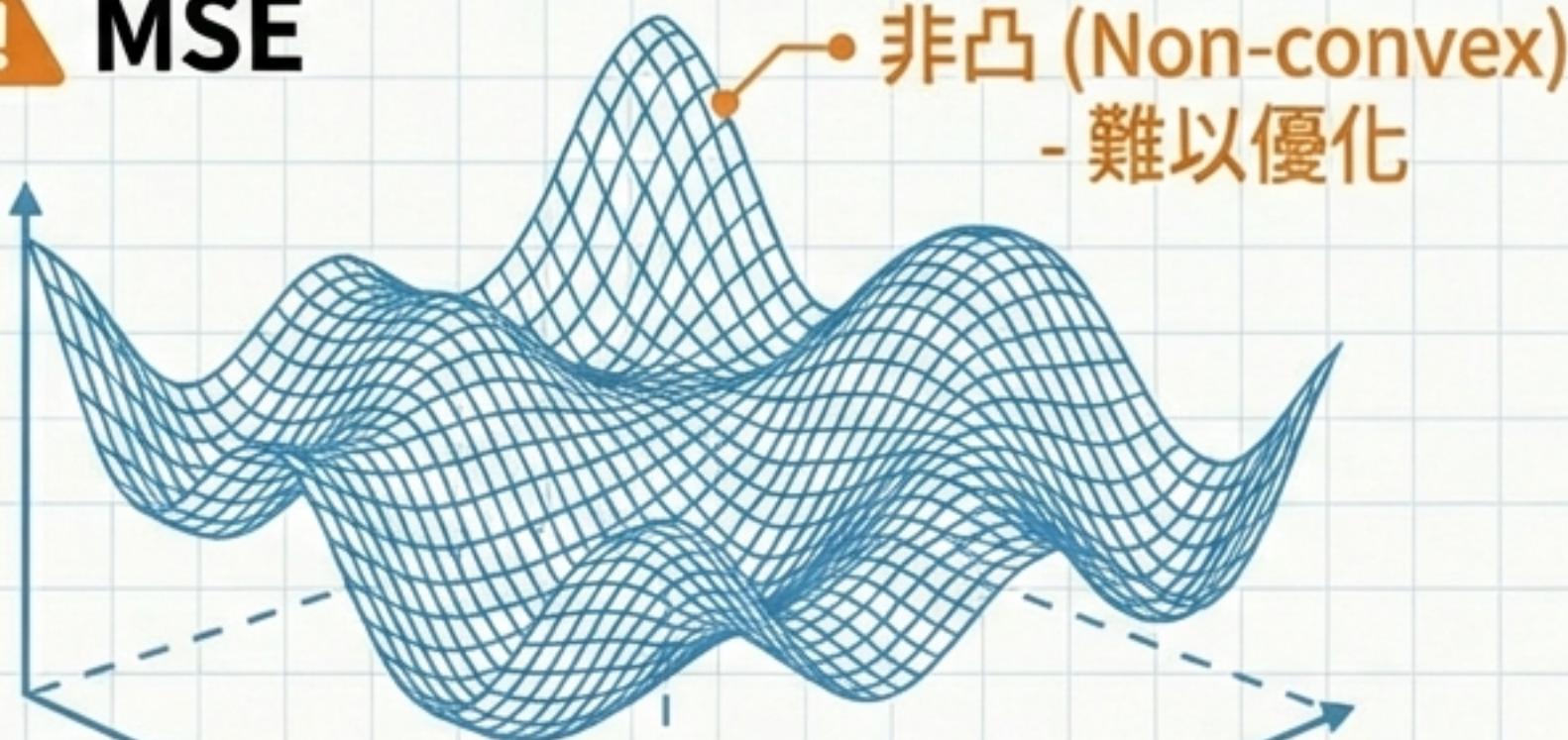
$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

- **壓縮特性**：將任意實數  $z$  映射至  $(0, 1)$  區間，完美對應機率定義。
- **對稱性**： $\sigma(-z) = 1 - \sigma(z)$ 。
- **導數優雅**： $\sigma'(z) = \sigma(z)(1 - \sigma(z))$ ，利於反向傳播運算。

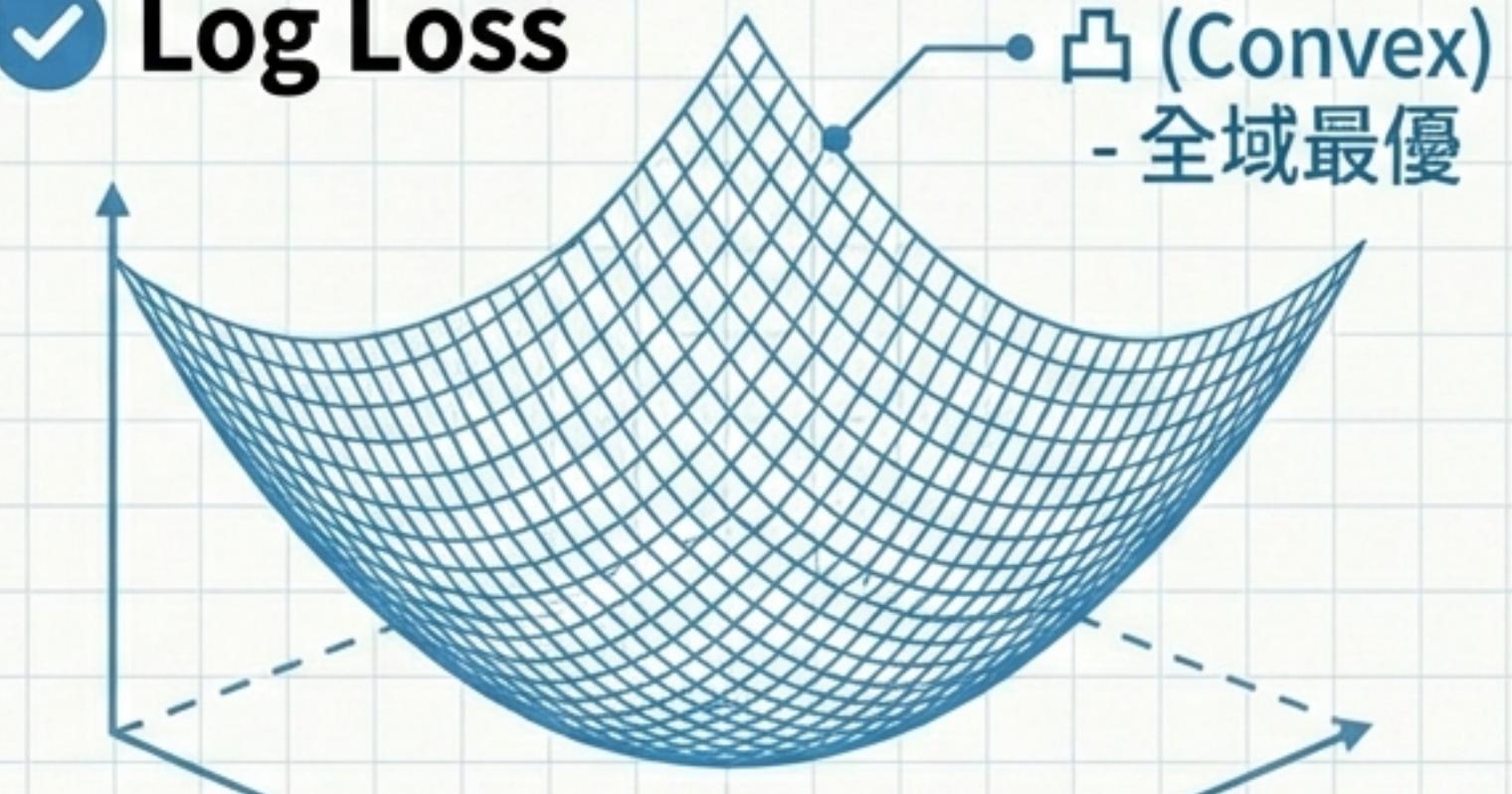
Concept : 這是模型的「軟開關」，將剛性的線性數值轉化為柔性的機率評估。

# 損失函數：為何棄用 MSE 改用 Log Loss ?

⚠ MSE



✓ Log Loss



$$J(\mathbf{w}, b) = -\frac{1}{m} \sum [y_i \log(\hat{p}_i) + (1 - y_i) \log(1 - \hat{p}_i)]$$

## 直觀理解 (The Penalty)

⇒ 若  $y = 1$  但預測  $\hat{p} \rightarrow 0 \Rightarrow$  損失趨向  $+\infty$ 。

⚠ 結論：損失函數會「嚴厲懲罰」那些置信度高但錯誤的預測。

# Sklearn 實作工具箱 (The Control Panel)

## Containment Area

```
from sklearn.linear_model import LogisticRegression  
  
# 實例化模型  
model = LogisticRegression(  
    penalty='l2',  
    C=1.0,  
    solver='lbfgs',  
    class_weight=None  
)  
model.fit(X_train, y_train)
```

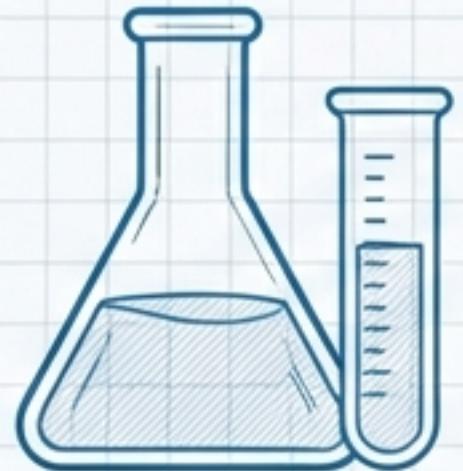
**C**：正則化強度的倒數。數值越小，模型越簡單 (防止過擬合)；數值越大，擬合越強。

**class\_weight**：設定為‘balanced’可自動調整權重，解決化工數據常見的類別不平衡 (Imbalance) 問題。

**solver**：小型數據用‘liblinear’，大型數據用‘saga’。

# 化工領域應用場景 (Process Applications)

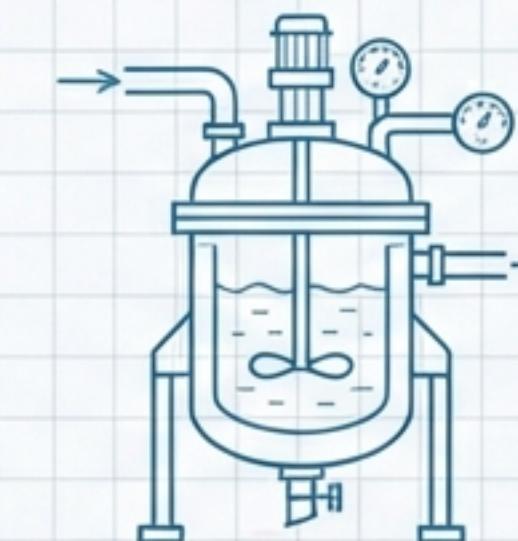
## 1. 產品品質分類 (Product Quality)



輸入：反應溫度、原料純度。

輸出：合格 (1) / 不合格 (0)。

## 2. 反應成功預測 (Reaction Success)



輸入：催化劑濃度、壓力、時間。

輸出：達到目標轉化率 (1) / 失敗 (0)。

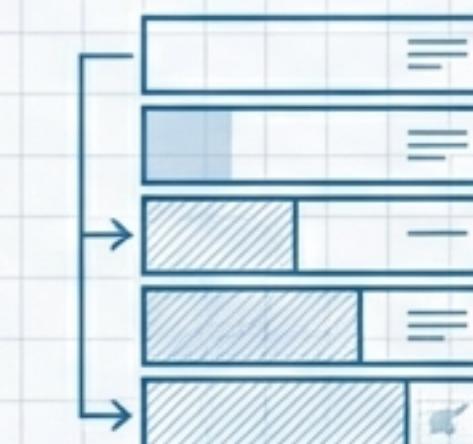
## 3. 設備異常檢測 (Anomaly Detection)



輸入：振動頻率、能耗、噪音。

輸出：異常報警 (1) / 正常運行 (0)。

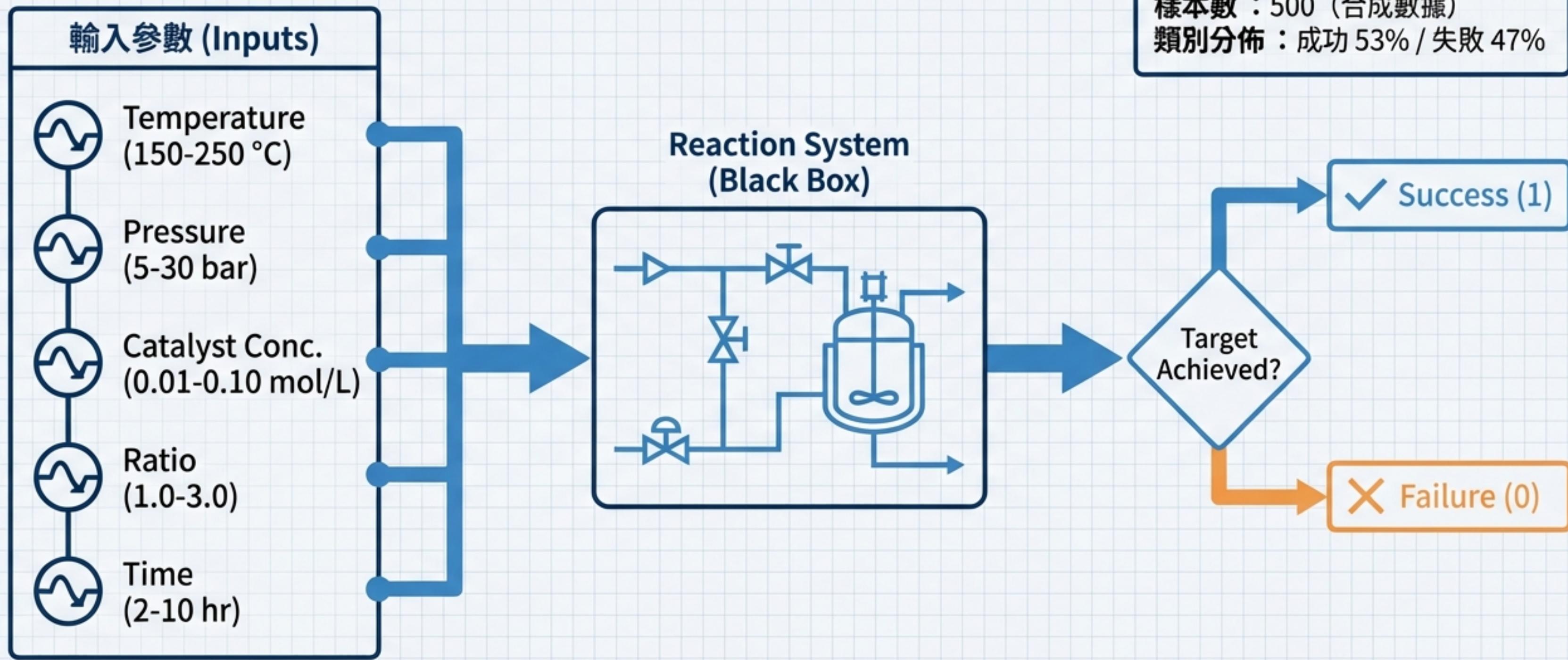
## 4. 批次質量分級 (Batch Grading)



輸入：批次製程曲線數據。

輸出：A級品 / 非A級品。

# 實作案例：化學反應成功預測



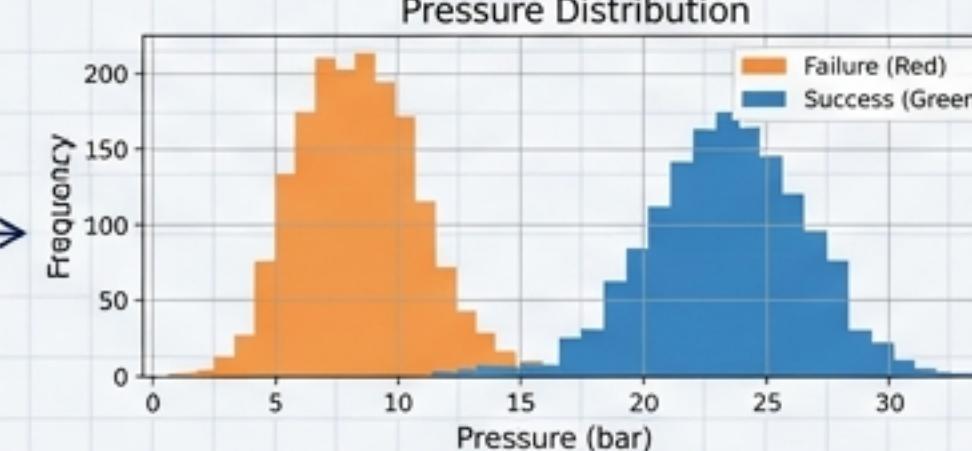
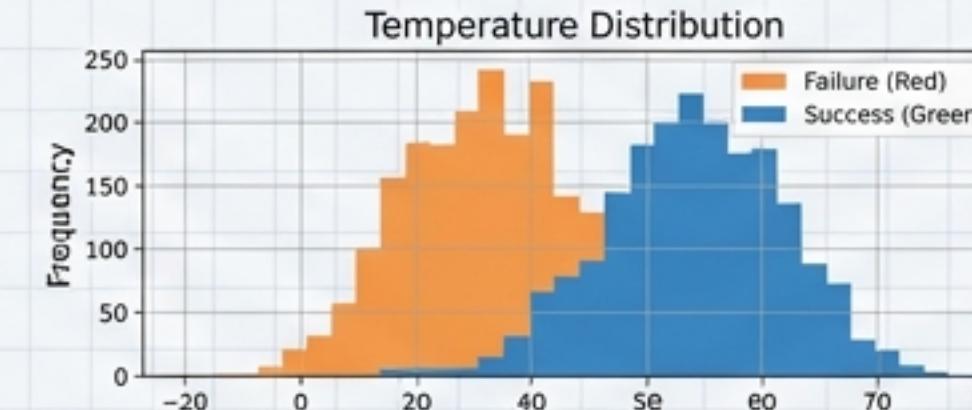
# 數據探索與分析 (EDA)

## Correlation Heatmap

	Temp	Pressure	Cat. Conc.	Ratio	Time
Temp		-0.05	0.21	0.08	0.12
Pressure	-0.05		0.13	0.29	0.06
Cat. Conc.	0.21	0.18		0.13	-0.05
Ratio	0.08	-0.03	0.18		0.08
Time	0.12	-0.05	0.05	0.34	

→ 無多重共線性 ( $|r| < 0.3$ )

## Feature Distribution (Temp & Pressure)



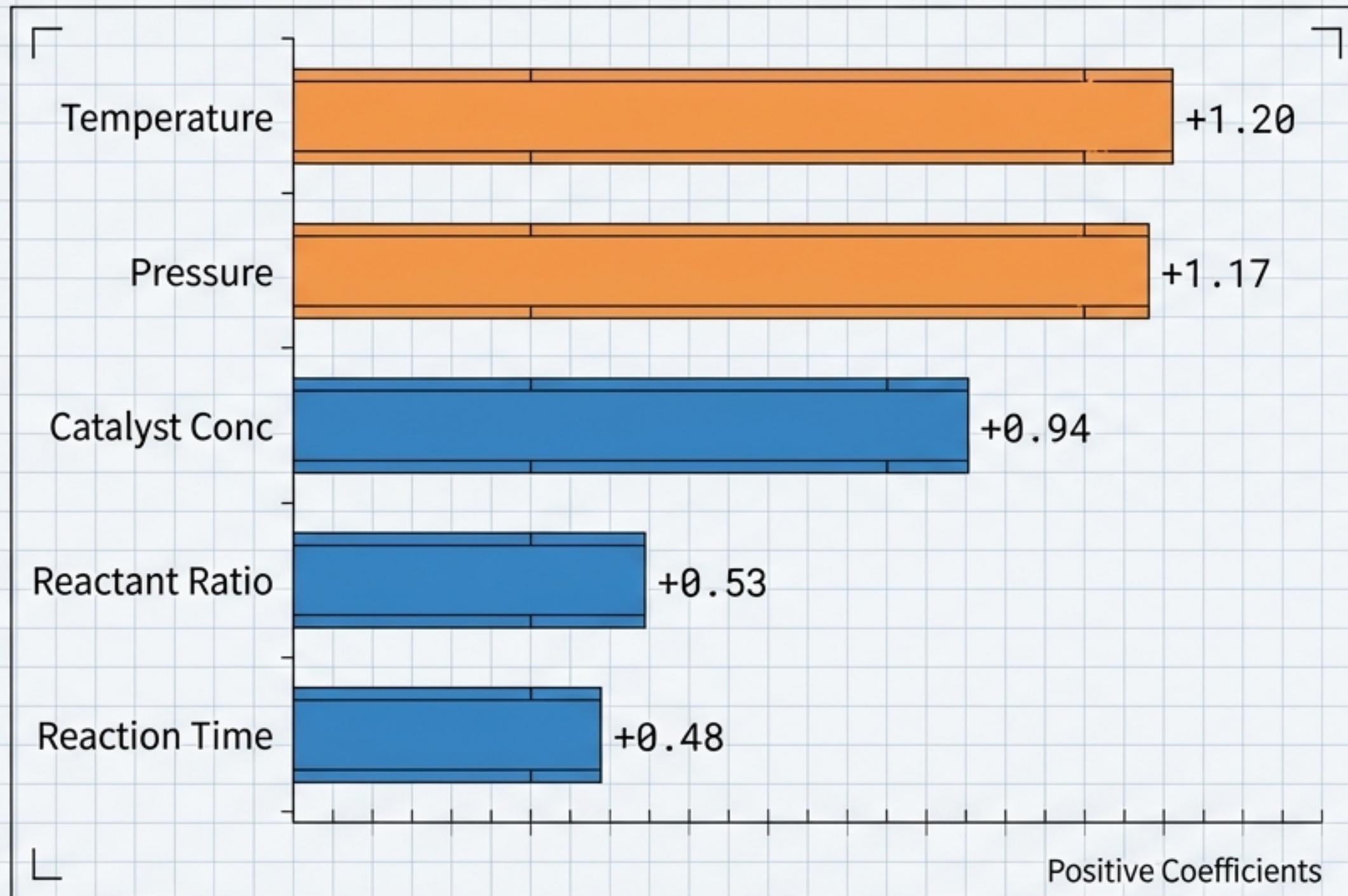
溫度與壓力顯示出明顯的類別區分能力



**Takeaway:** 數據乾淨且特徵獨立，適合邏輯迴歸模型。



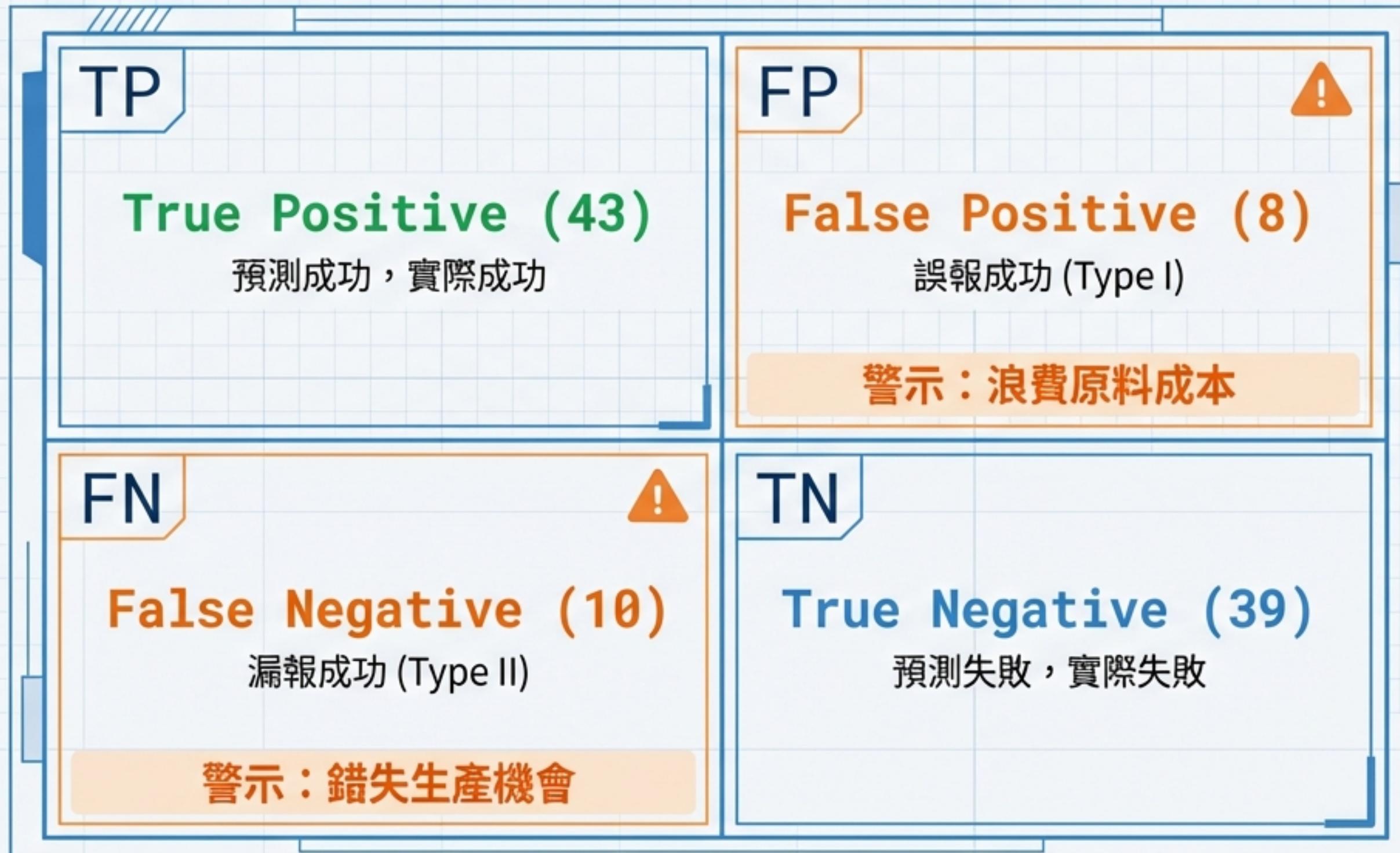
# 模型解密：特徵重要性分析 (Coefficients)



## 工程觀點解析

1. **正係數**：所有特徵皆為正值，符合化學動力學原理 (增加能量/濃度有利於反應)。
2. **關鍵因子**：溫度與壓力是主導變數 (Drivers)。若要提高成功率，優先調整溫壓條件，而非單純延長反應時間。

# 效能評估：混淆矩陣 (Confusion Matrix)



## Metrics

Accuracy

82%

Noto Sans TC

Precision (Success)

84%

Noto Sans TC

Recall (Success)

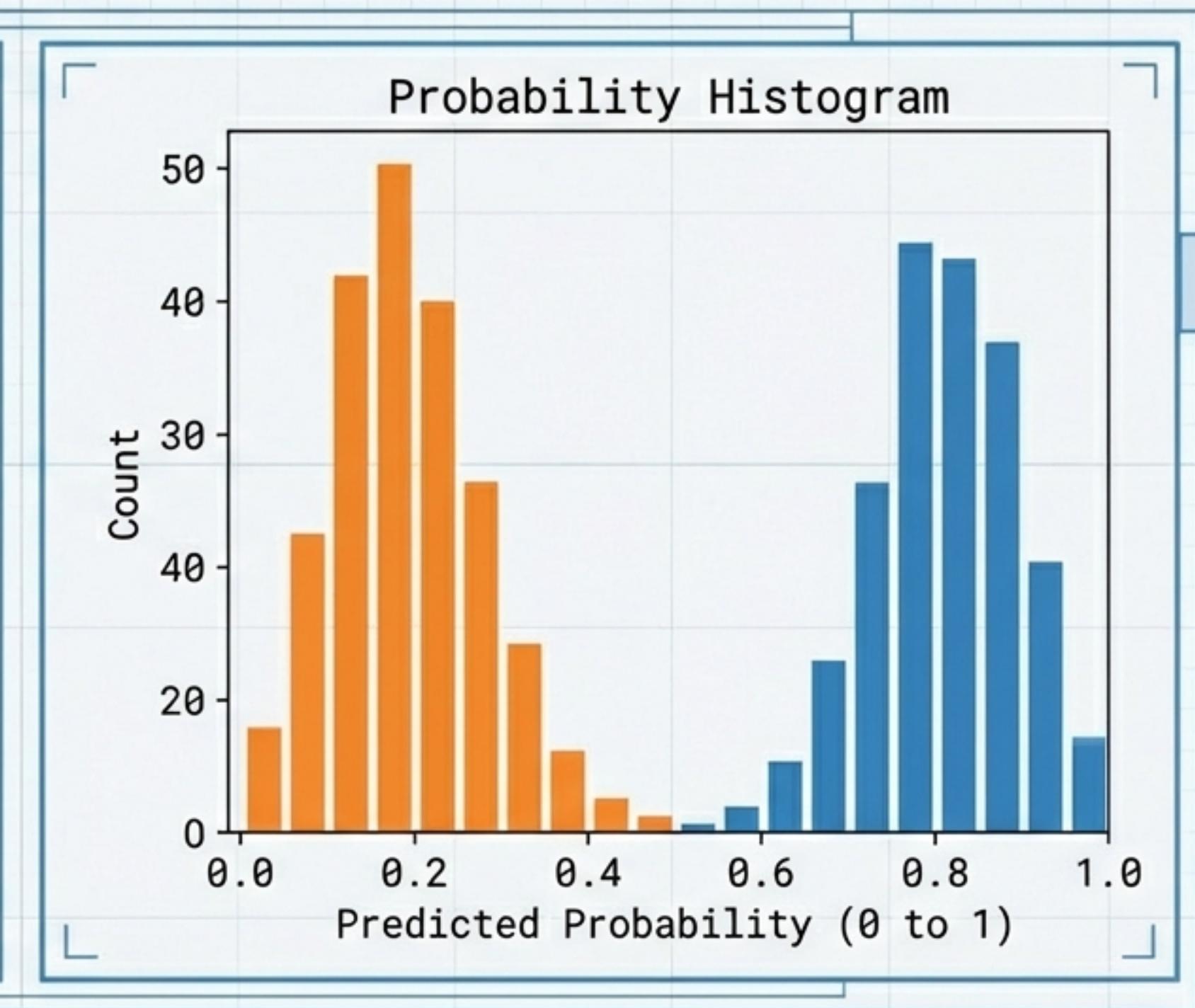
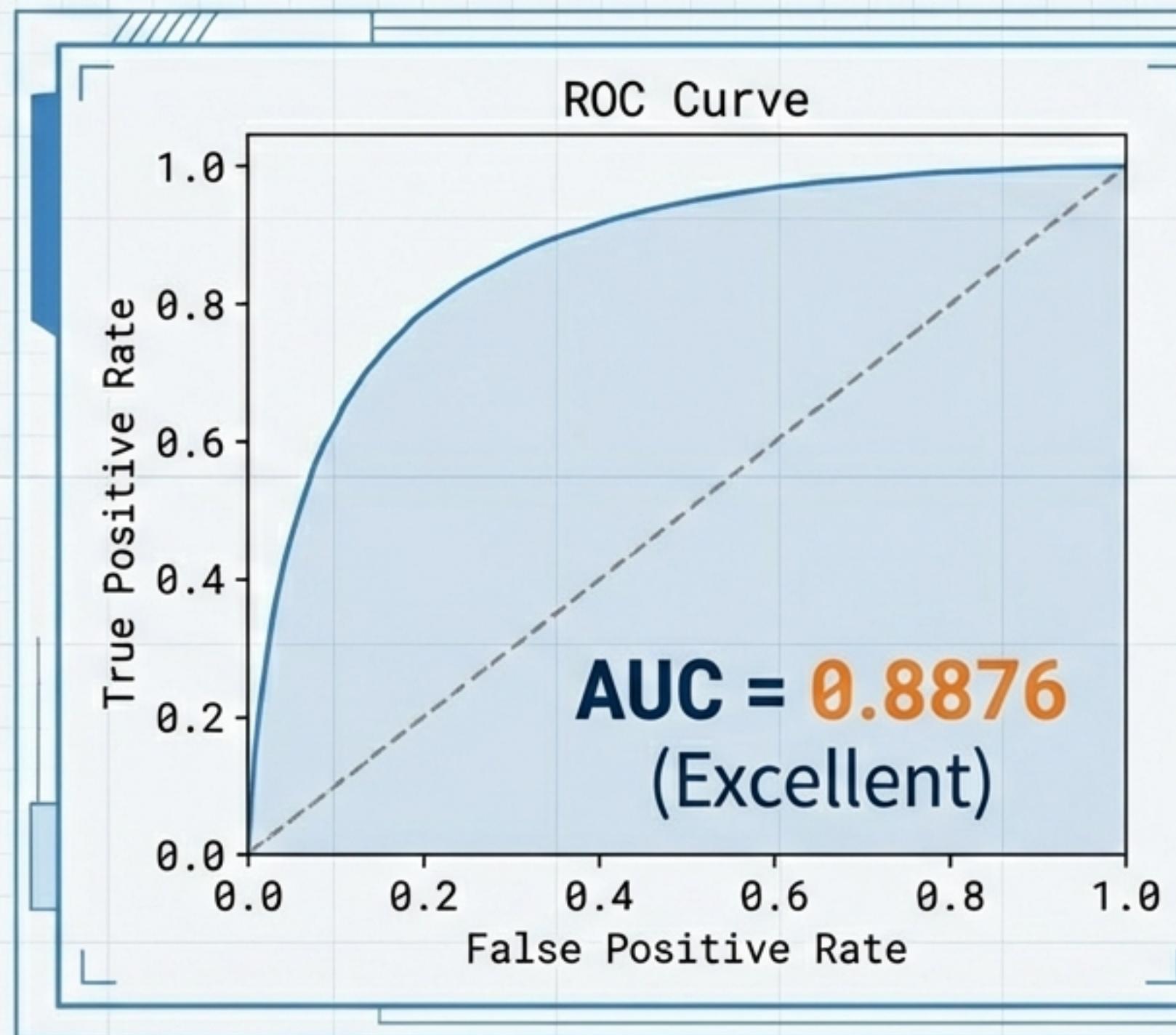
81%

Noto Sans TC

Swiss Style meets Industrial Blueprint

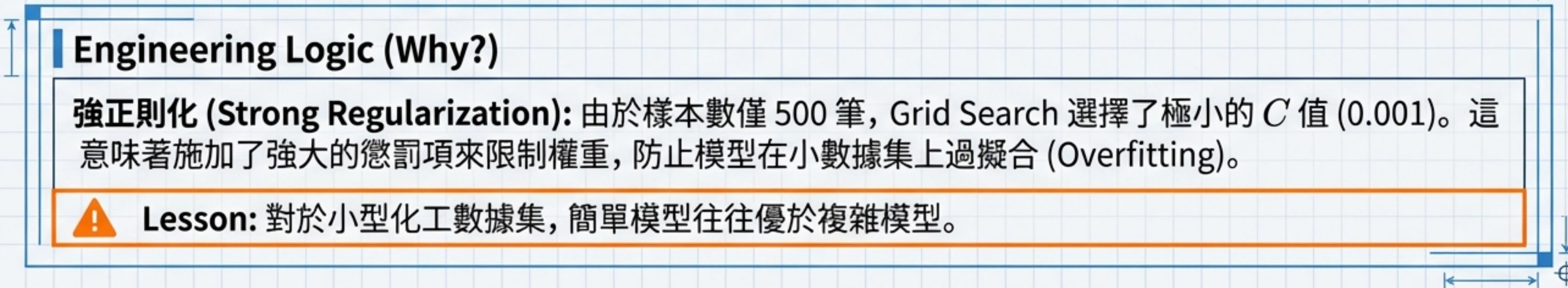
Unit 02 結束 | 前往 Unit 03

# 進階評估：ROC 曲線與機率分佈



AUC 接近 0.9 代表模型具有極高的區分能力，且機率輸出具有明確的置信度 (High Confidence)。

# 模型優化：超參數調整 (Hyperparameter Tuning)



# 總結：邏輯迴歸的優勢與限制

## ✓ 優勢 (Pros)

- ✓ 可解釋性強：係數直接對應物理意義（如溫度影響力 > 時間）。
- ✓ 機率輸出：提供決策置信度，不僅僅是 Yes/No。
- ✓ 計算極快：適合即時監控系統。

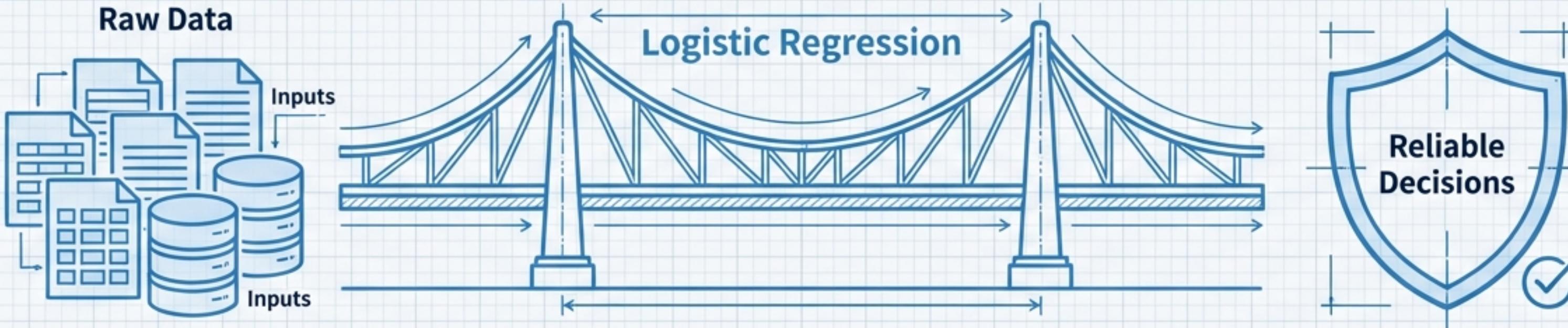
## ⚠ 限制 (Cons)

- ⚠ 線性邊界：無法處理複雜的非線性關係（如太極圖分佈）。
- ⚠ 特徵獨立假設：對多重共線性敏感。

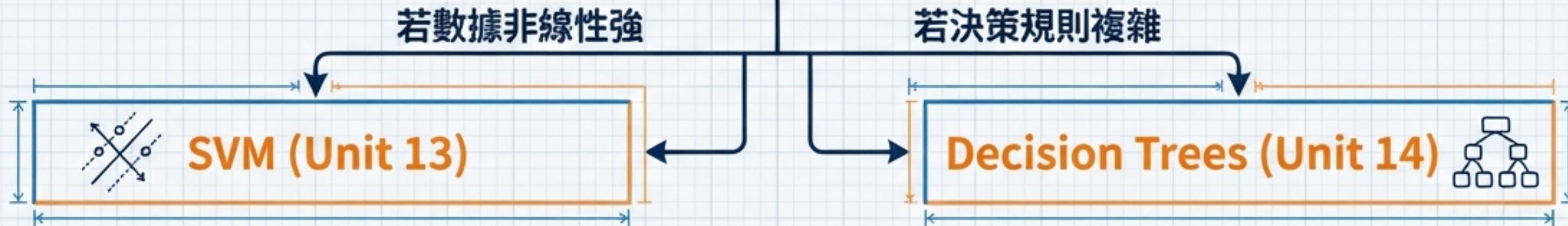
## Conclusion

邏輯迴歸是所有分類問題的基準 (Baseline)。先用它建立標準，再嘗試更複雜的模型。

# 結論：化工AI的「第一草稿」模型



## Next Steps



*"Model interpretability is often more important than pure accuracy in engineering safety."*