

Gaussian Naive Bayes Classifier

高斯樸素貝氏分類器：應用於化工反應預測的機率模型

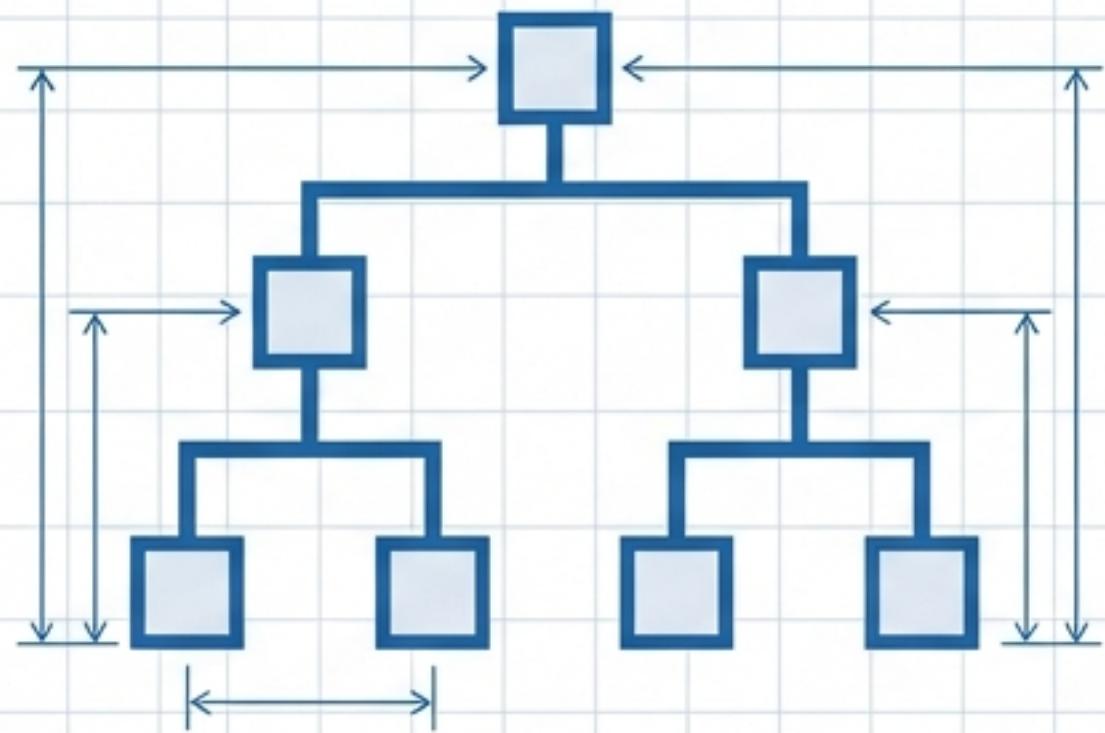
課程製作：逢甲大學 化工系 智慧程序系統工程實驗室

授課教師：莊曜楨 助理教授

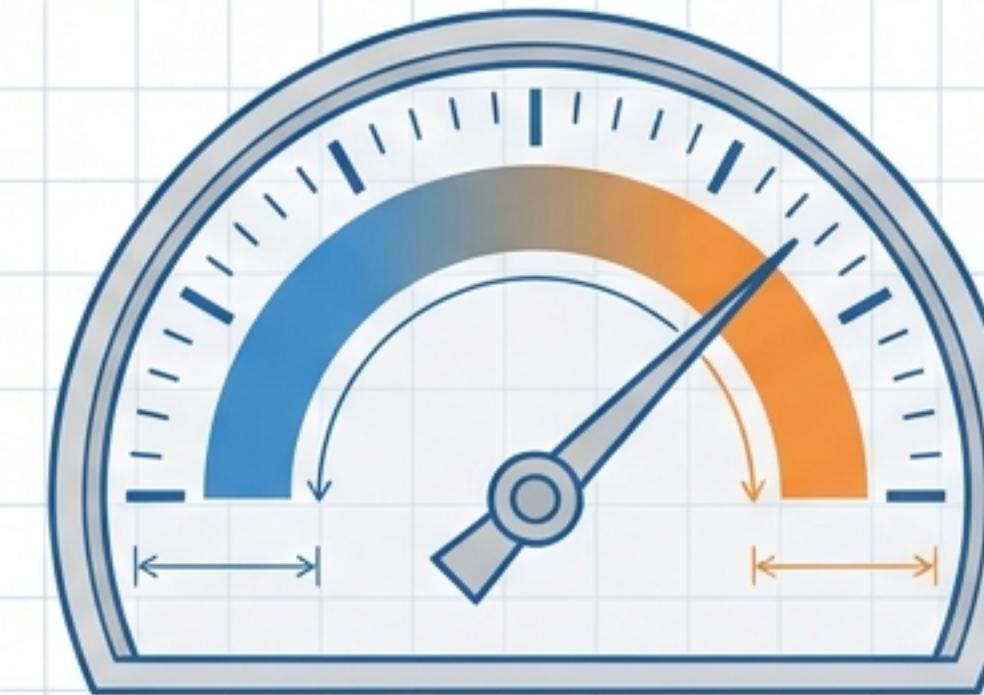
更新日期：2026-01-28

機率視角的診斷工具

決策樹 (硬性規則)



樸素貝氏 (機率推論)



適合模糊的趨勢評估 (e.g., 80% chance of failure)



快速篩選 (Rapid Screening) :
適合即時判斷反應條件是否適當。



初步診斷 (Initial Diagnosis) :
快速識別製程異常狀態。



小樣本學習 (Small Sample Learning) : 數據量有限 (幾十筆)
即可訓練，優於深度學習。

貝氏定理：從證據更新信念

後驗機率 (Posterior)
最終診斷結果 (Diagnosis)

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

似然度 (Likelihood)

在特定狀態下出現該數據的機率 (Sensor Evidence)

先驗機率 (Prior)

根據歷史數據的經驗法則
(Historical Rate)

邊際機率 (Evidence)

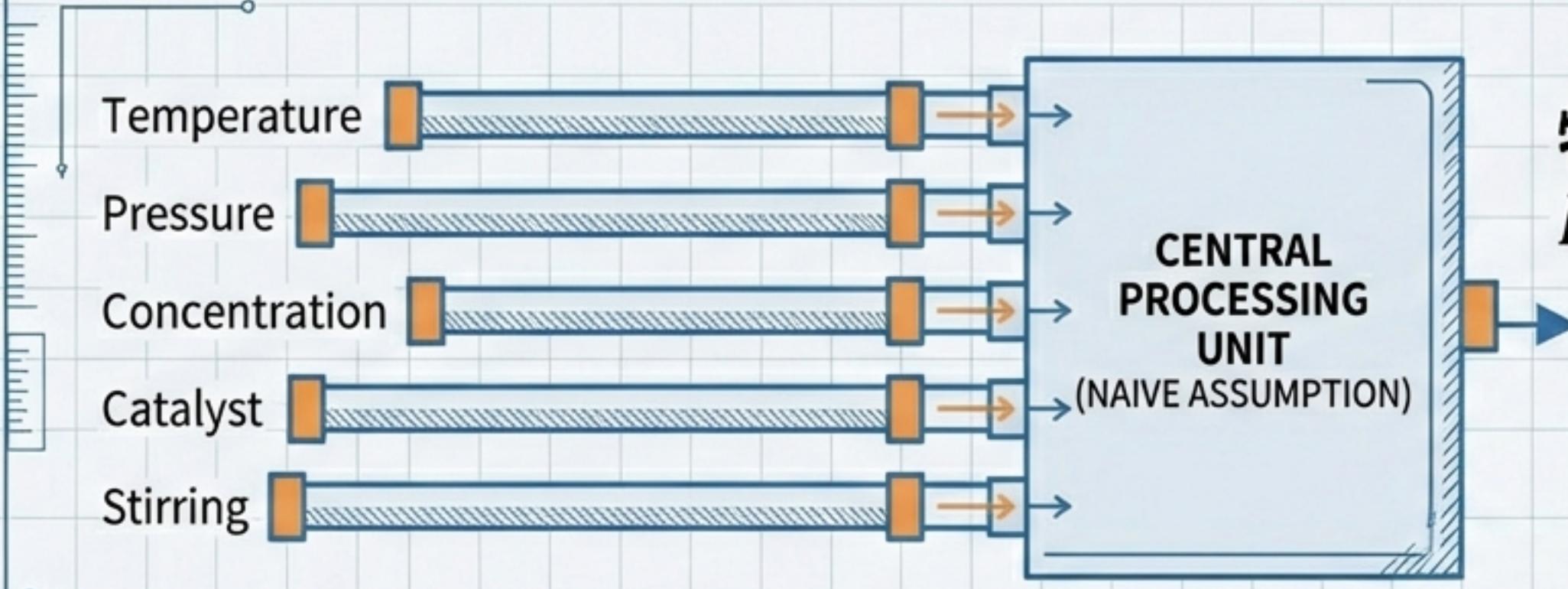
觀測數據的總機率 (Normalization)



»» 證據 (Sensor Data) + 先驗知識 (History) = 診斷 (Diagnosis) »»

描述在已知某些證據 (如溫度升高) 的情況下，如何更新對假設 (反應將會失敗) 的信念。

「樸素」的代價與優勢



特徵條件獨立假設 - 假設

$$P(x_i, x_j | C_k) = P(x_i | C_k) \cdot P(x_j | C_k)$$

化工現實：變數通常高度相關（例如：理想氣體方程式 $PV = nRT$ ，溫度與壓力連動）。

模型假設：視為完全獨立。

Why it works (Engineering Pragmatism)



降維效果：將複雜的聯合機率簡化為獨立機率乘積。

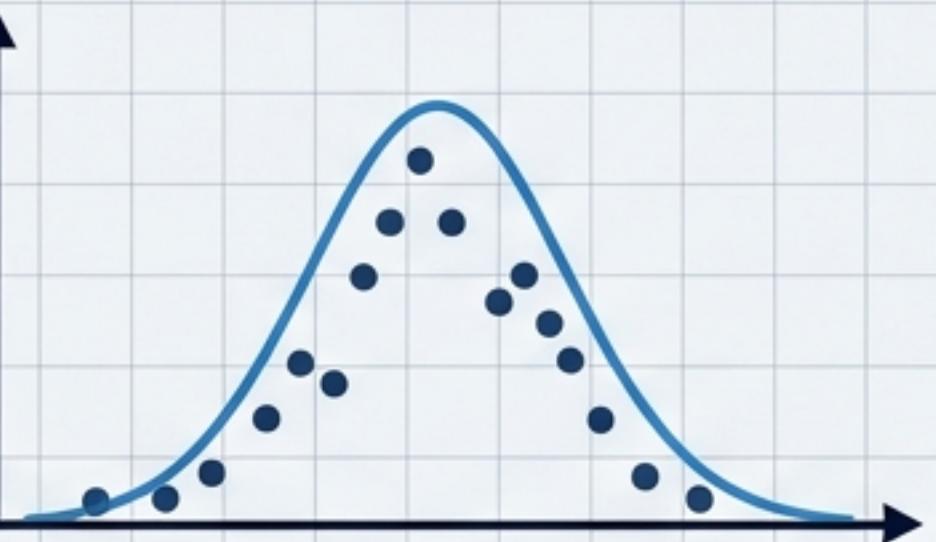


抗干擾性：對特徵間的弱相關具有容忍度。

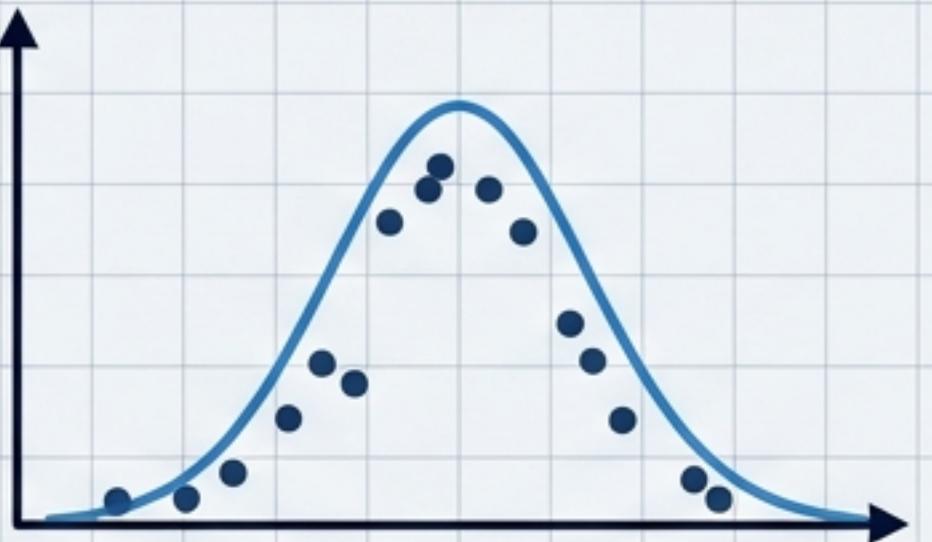


計算效率：線性時間複雜度 $O(n)$ ，極適合即時系統。

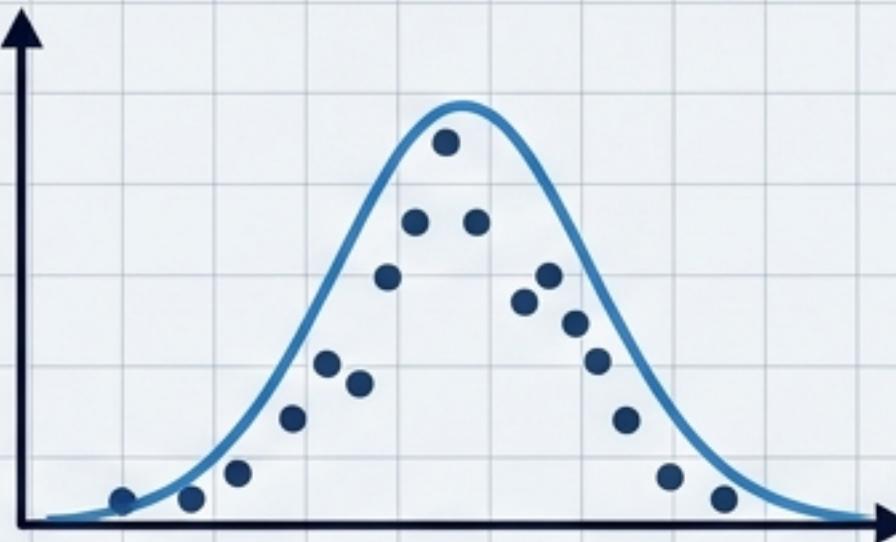
處理連續數值：高斯分佈假設



Temperature (T)



Pressure (P)



Flow (F)

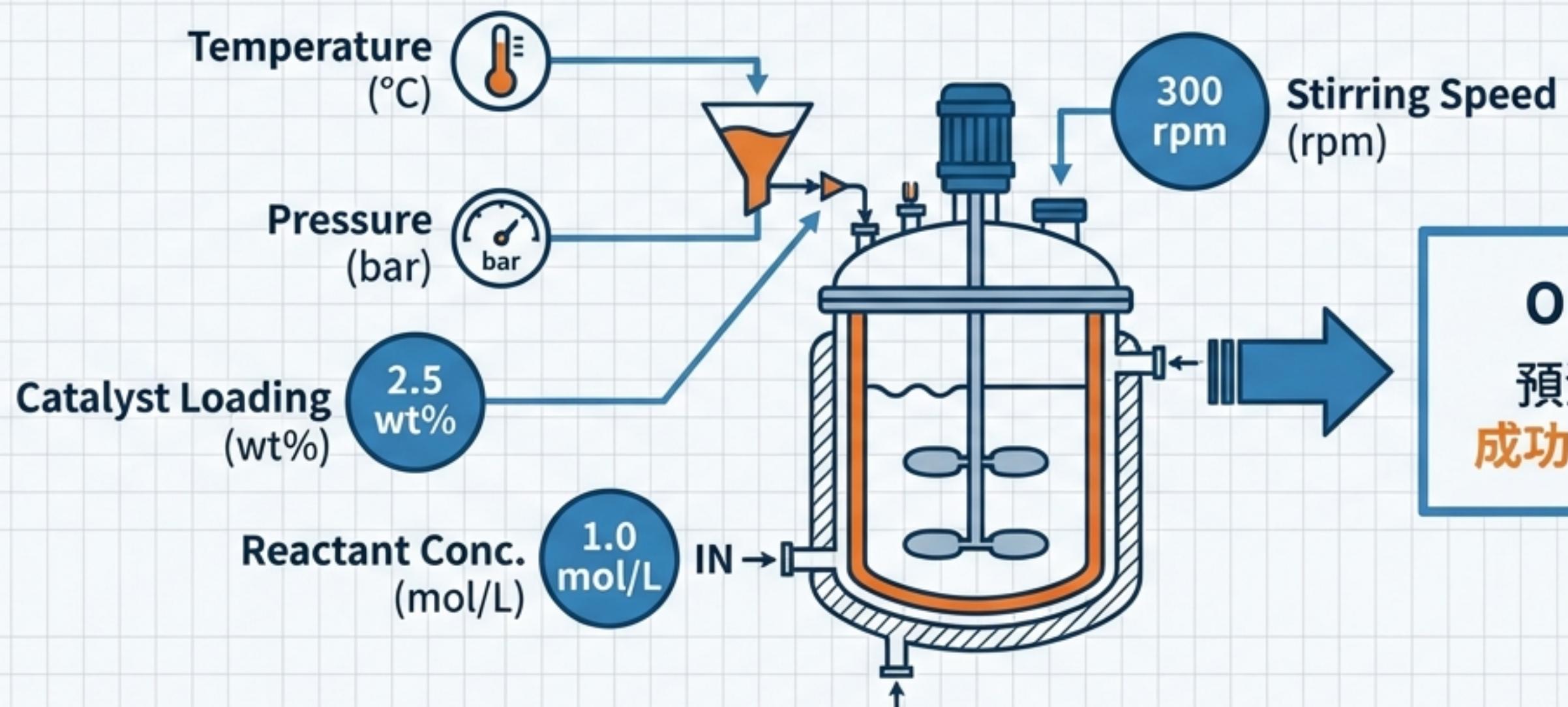
$$P(x_i | C_k) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right)$$

由兩個參數決定：

- 平均值 (μ) $\underline{\mu}$
- 變異數 (σ^2) $\underline{\sigma^2}$

工程背景：大多數化工製程的感測器雜訊與操作變異，自然傾向於常態分佈，使得此假設在實務上相當合理。

案例研究：反應成功率預測



Data Context

1000 筆模擬數據，存在特徵協同效應（違反獨立性假設），測試模型極限。

執行結果：超出預期的效能

準確率 (Accuracy)
83.00%
(超出預期 +18%)



訓練時間 (Training Time)
0.0010 秒
(極速)

過擬合 (Overfitting)
無
(訓練/測試差異僅 0.57%)

INSIGHT

儘管違反獨立性假設，但由於「溫度」與「壓力」具有強大的單變量區分力，
加上特徵標準化 (Standardization) 的輔助，模型表現優異。

速度與精度的權衡



Random Forest	94.67% 準確度	慢 (Slow)	複雜度高
Gaussian NB	83.00% 準確度	極快 (Fast)	記憶體極省

Takeaway:

對於需要毫秒級響應的 **即時監控系統 (Real-time Monitoring)**
或 **嵌入式裝置**，GNB 是最佳選擇。

保守的代價：高安全性與低召回率



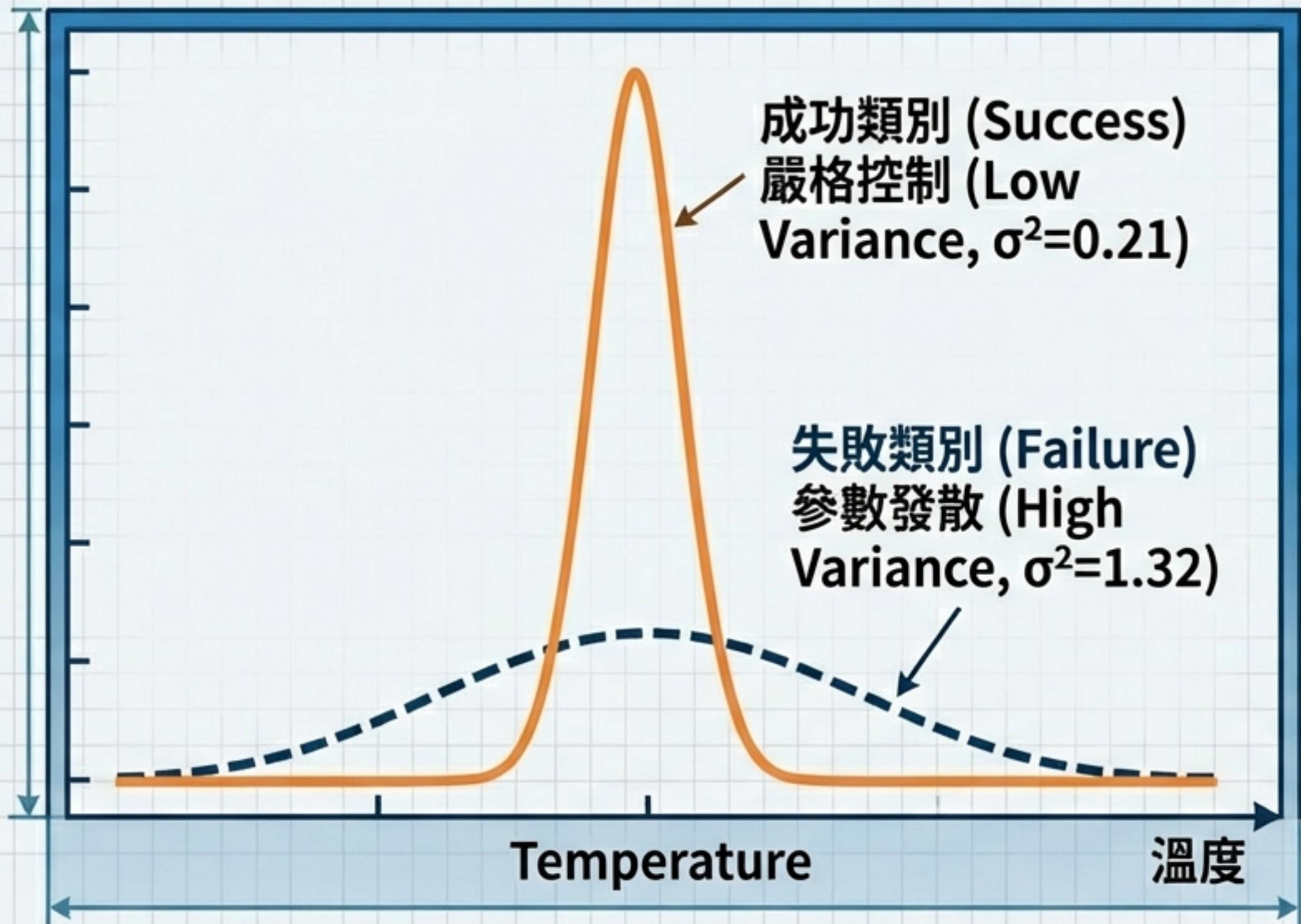
極度保守 (Paranoid Model)

寧可錯過 50 個好反應，也不願執行 1 個壞反應。

應用場景：

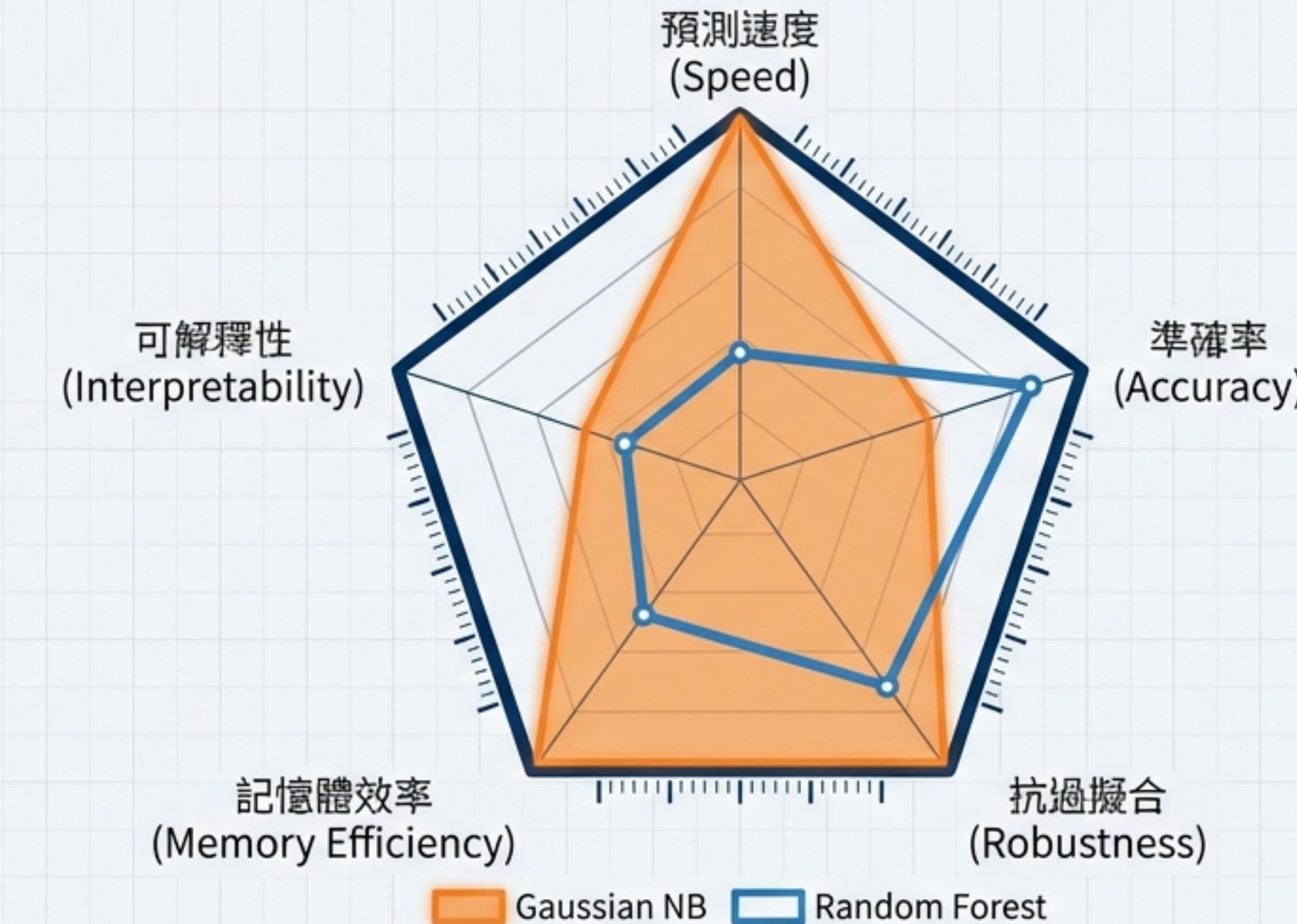
適合高風險、高成本的製程，
要求「絕對確信」才執行操作。

數據背後的物理意義



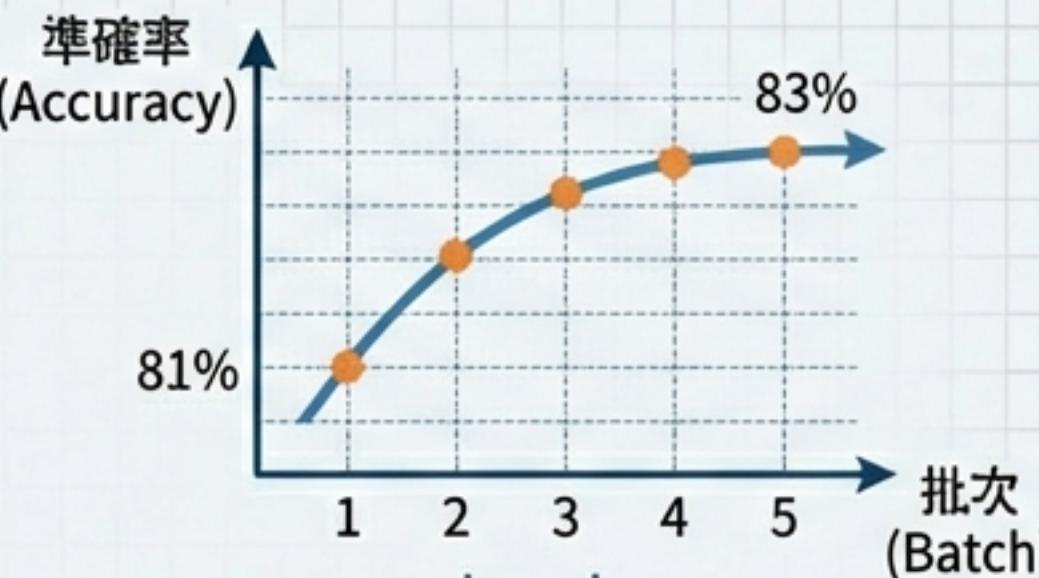
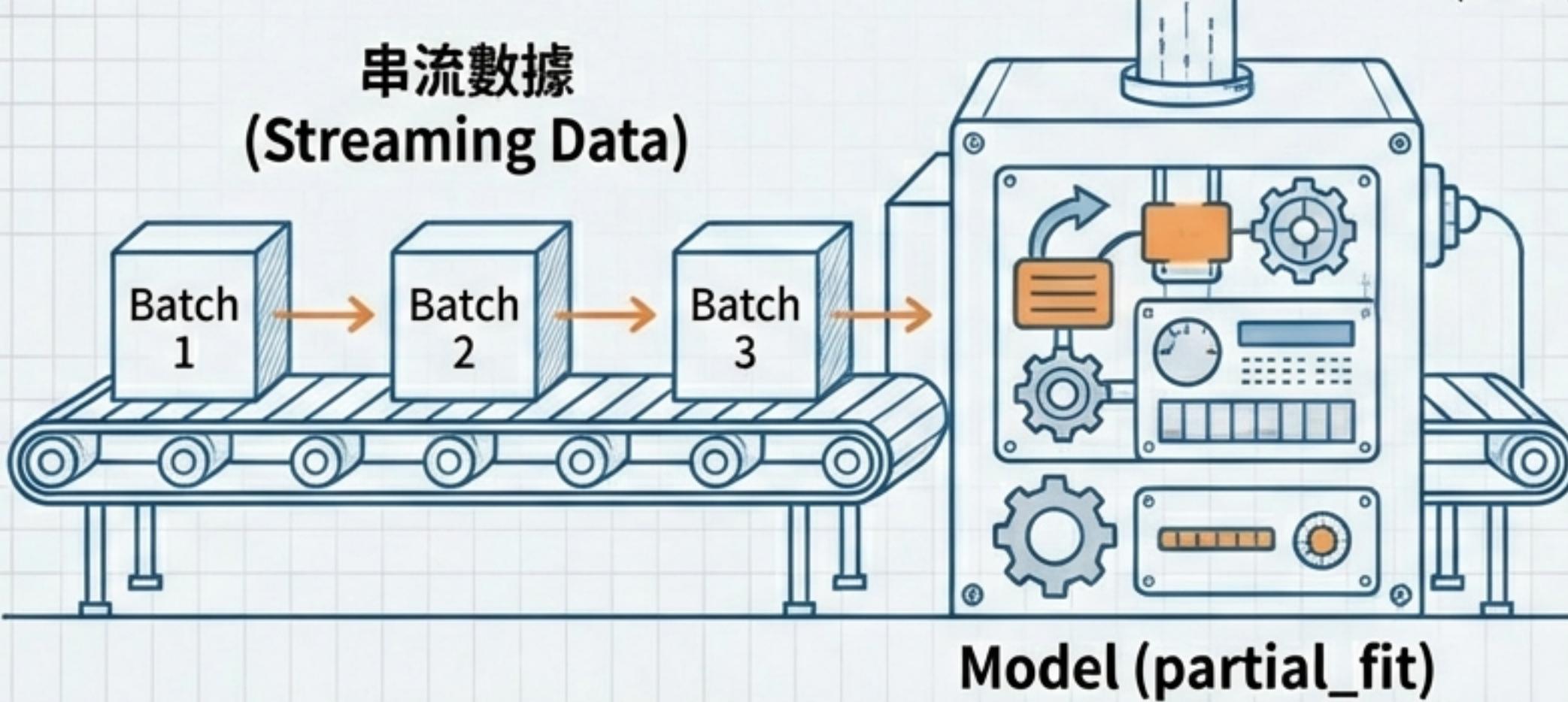
- 成功反應需要將溫度控制在狹窄區間。
- 失敗可以在廣泛的參數範圍內發生。
- 由於成功類別的分佈極其緊湊，模型僅在數據落入此狹窄區間時才敢預測為「成功」，導致了保守傾向。

模型特性雷達圖



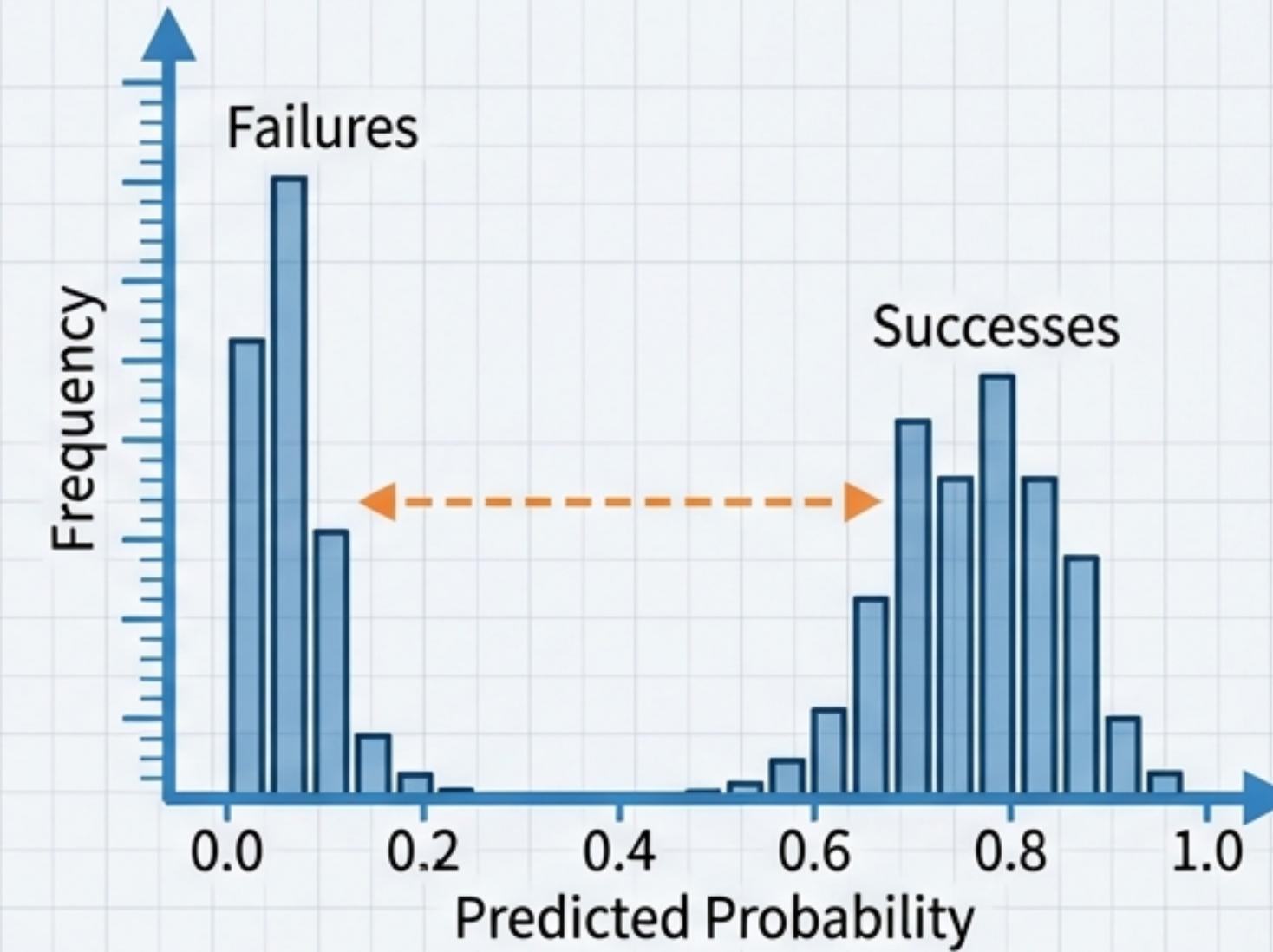
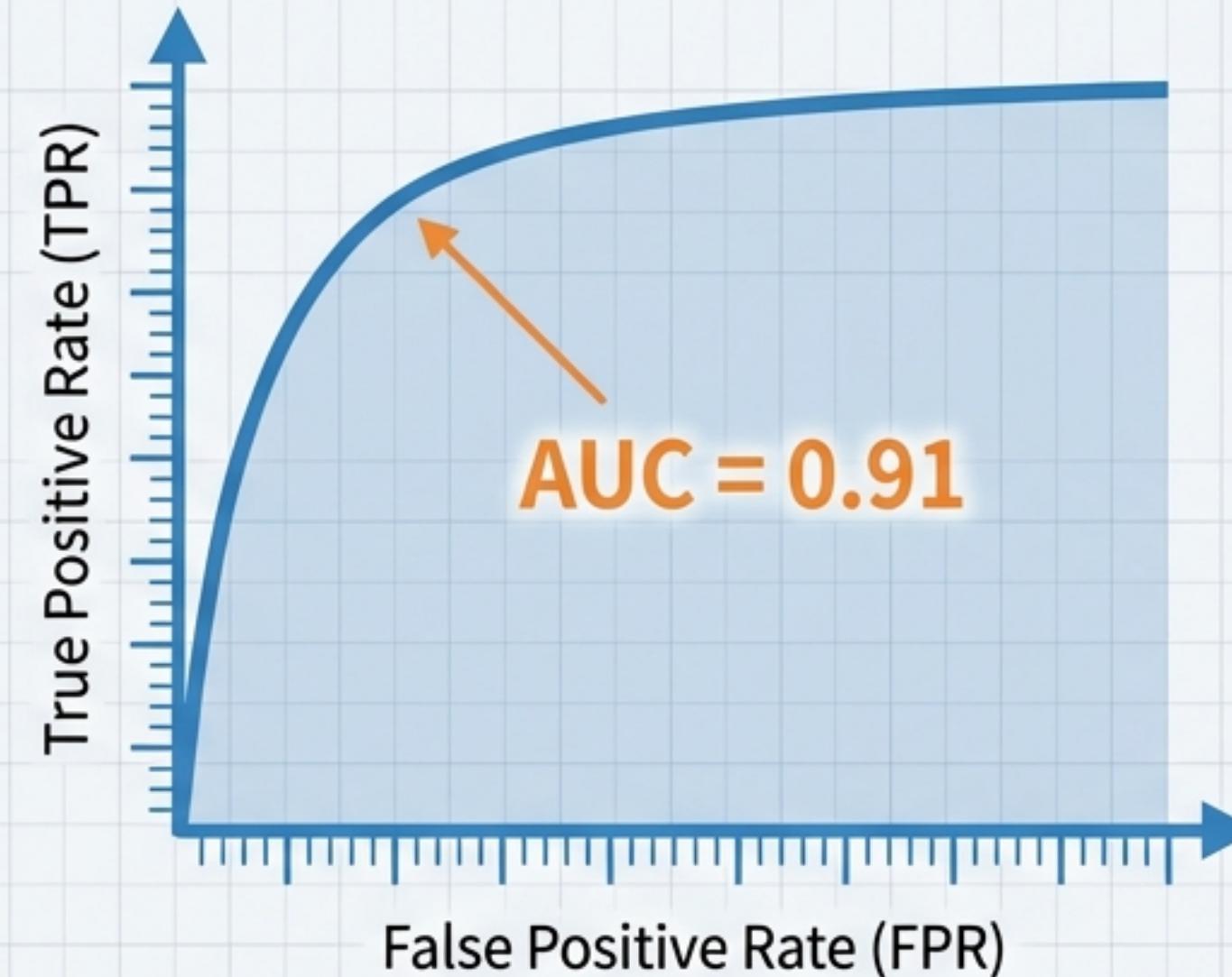
Conclusion: GNB 犧牲了處理複雜邊界的能力，換取了極致的速度與資源效率。

增量學習：適應串流數據



- **partial_fit()**：支援分批訓練，無需將所有數據載入記憶體。
- 線上更新 (Online Update)：模型可隨程式運作持續持續更新，適應概念漂移 (Concept Drift)。
- 應用：IoT 感測器串流數據分析。

機率校準與閾值調整



Implication :

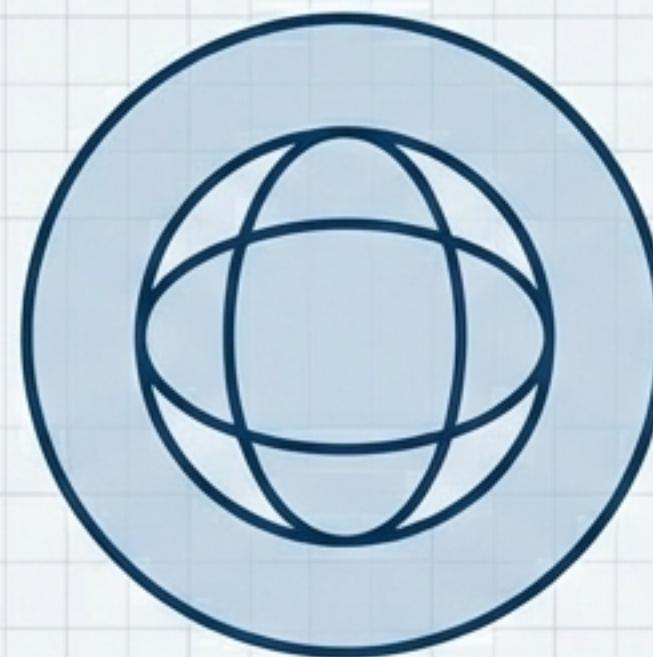
- AUC 0.91 表示模型排序能力極佳。
- **決策靈活性**：雖然預設準確率為 83%，但透過調整 **決策閾值 (Threshold)**，我們可以根據需求用精確率交換召回率 (**Trade Precision for Recall**)。

實務應用指南

CHECKLIST

-  **資源受限環境**：嵌入式系統、邊緣運算 (Edge AI)。
-  **即時反應需求**：毫秒級預測 (High-frequency Monitoring)。
-  **建立基準 (Baseline)**：專案初期快速評估數據可行性。
-  **高安全性需求**：需要極高的精確率 (High Precision)。
-  **非線性複雜邊界**：若需捕捉精細的分類邊界，請改用 Random Forest。

總結：簡單、快速、安全



Simple (樸素假設)

模型輕量，參數極少。



Fast (極致速度)

比傳統模型快百倍，即時運算首選。



Safe (保守策略)

98% 精確率，為高風險製程提供安全保障。

Next Step: 開啟 Unit12_Gaussian_Naive_Bayes.ipynb 開始實作