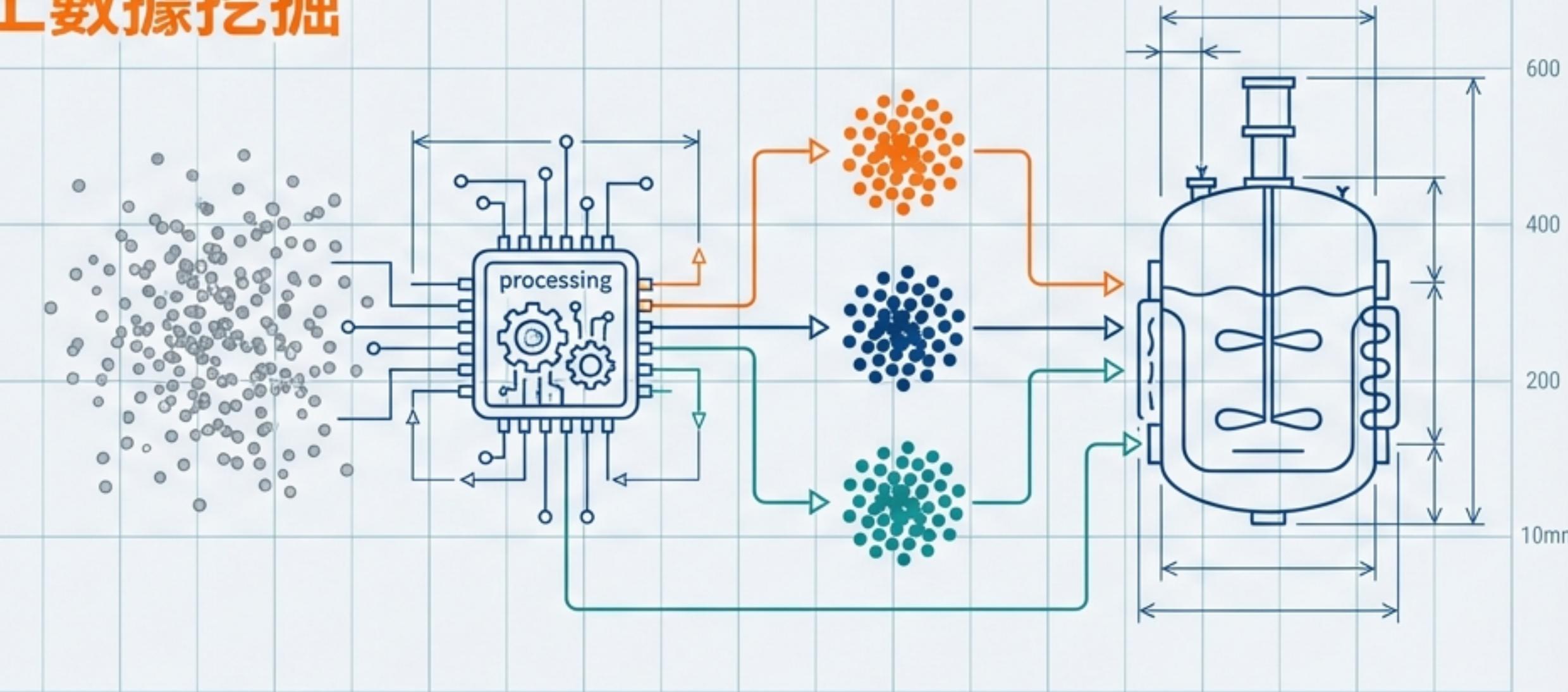


AI 在化工上的應用：從原理到實踐

# Unit 05 分群分析總覽

非監督式學習與化工數據挖掘



# 什麼是分群分析？(What is Clustering?)

- 一種非監督式學習 (Unsupervised Learning) 方法
- 不需要預先標記 (No labels required)
- 目標：自動發現數據中的內在結構

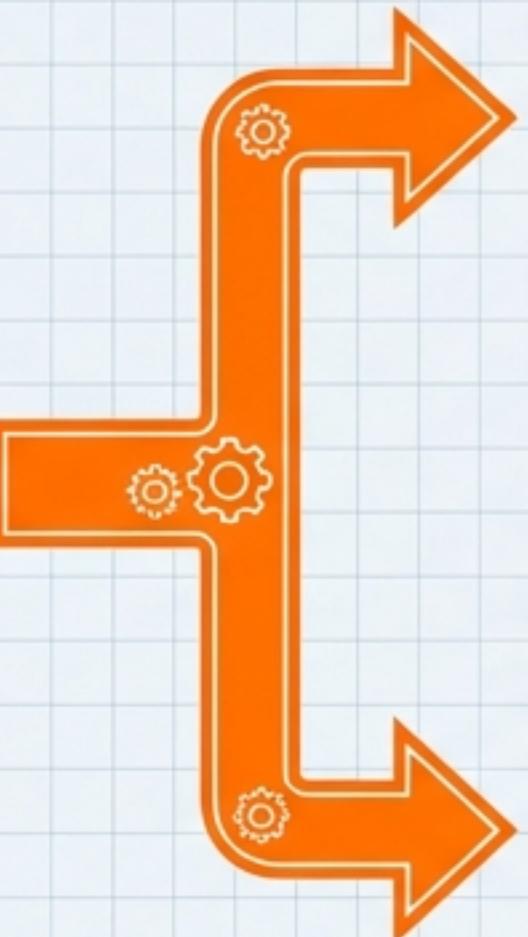
## Key Metrics



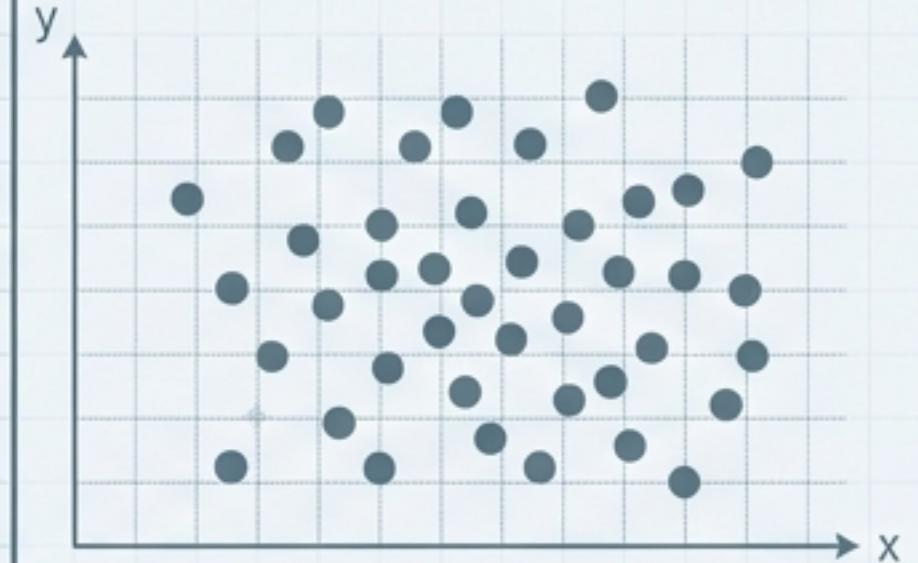
高內聚性 (High Cohesion)：  
同群集內數據點相似度高



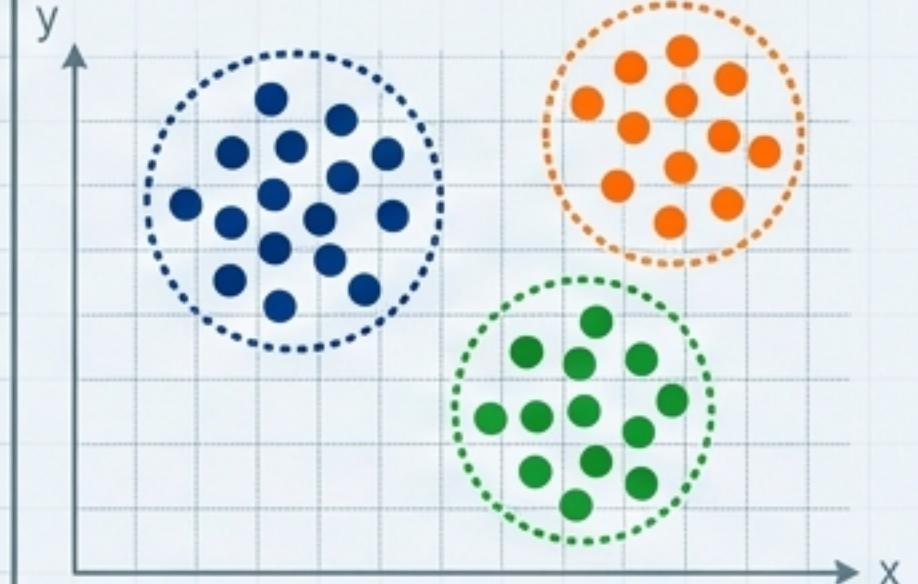
低耦合性 (Low Coupling)：  
不同群集間數據點明顯分離



## 原始數據 (Raw Data)



## 分群結果 (Clustered Data)



距離度量 (Distance)：歐幾里得距離

$$d(x, y) = \sqrt{\sum (x_i - y_i)^2}$$

應用：適用於連續變數

# 化工領域的核心應用目標 (Core Objectives in Chemical Engineering)



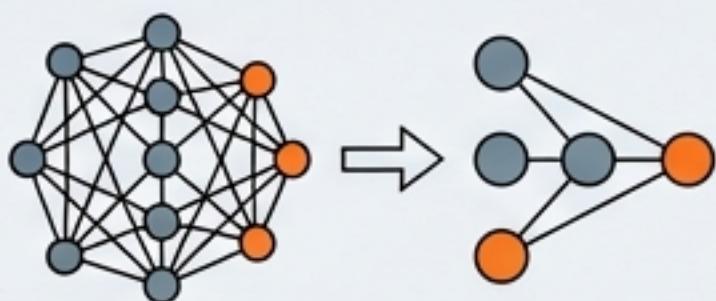
## 模式識別 (Pattern Recognition)

識別製程中的不同操作模式 (Start-up, Normal, Shutdown) 或產品類型。



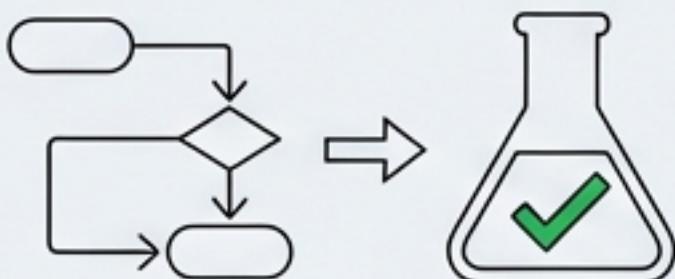
## 異常檢測 (Anomaly Detection)

發現偏離正常操作的異常模式 (Faults, Deviations)。



## 數據探索 (Data Exploration)

理解高維數據結構，將相似操作條件歸類。



## 決策支持 (Decision Support)

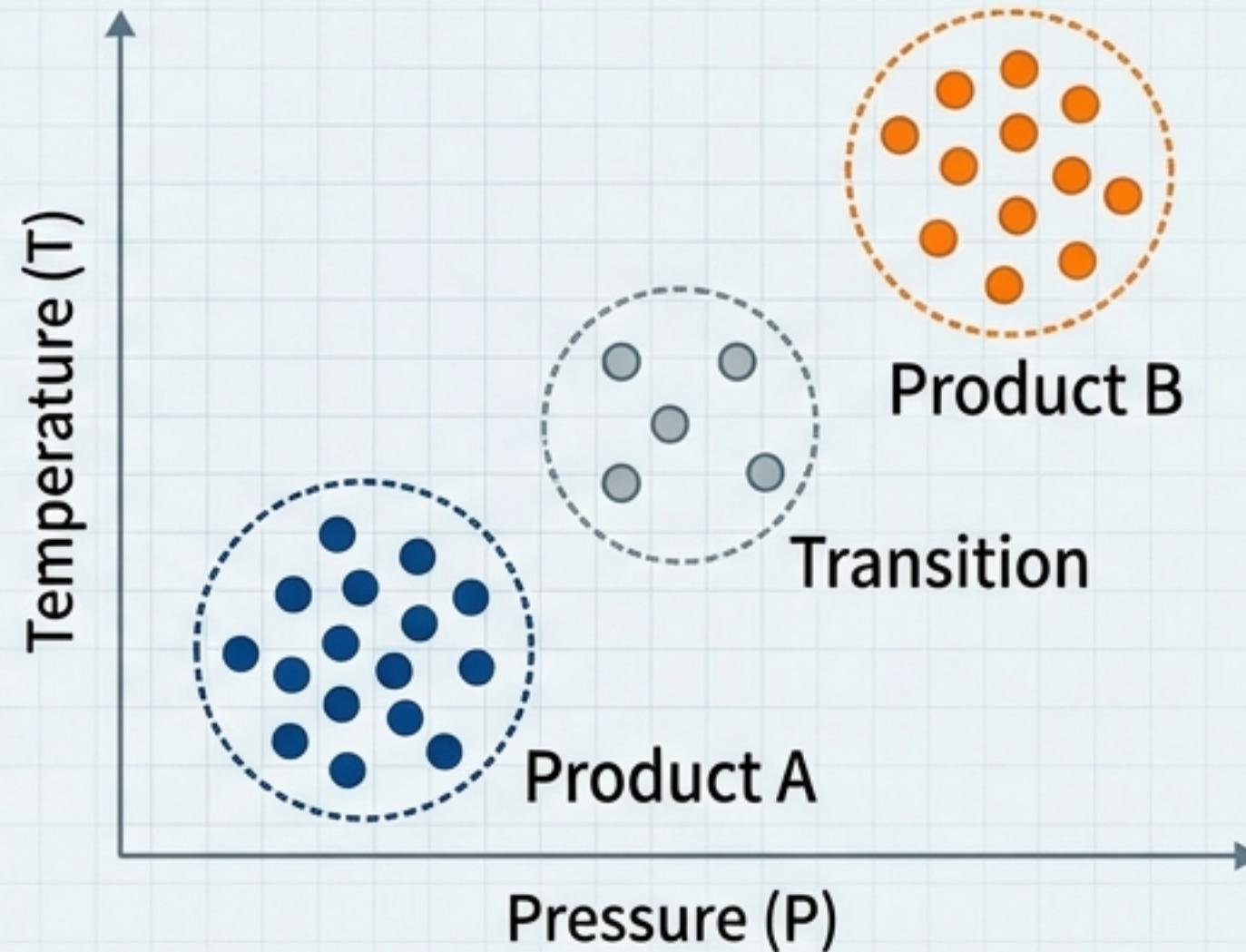
溶劑與配方篩選，縮小實驗範圍。

# 演算法 1：K-平均演算法 (K-Means Clustering)

## 原理與機制

- 機制：基於中心點 (Centroid-based) 的迭代算法
- 目標函數：最小化群集內距離總和 (Minimize WCSS)
- 步驟：
  1. 隨機初始化 K 個中心
  2. 分配數據點至最近中心
  3. 更新中心位置
  4. 重複直至收斂
- 優缺點：快速簡單，但需預先指定 K 值且假設為球形分佈

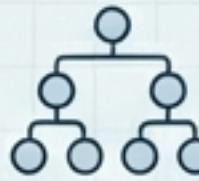
## ChemE Example: 操作模式識別



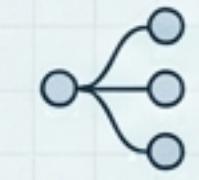
自動識別反應器的穩定狀態 (Steady States)

# 演算法 2：階層式分群 (Hierarchical Clustering)

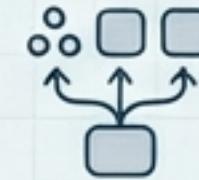
## 原理與機制



- 機制：建立樹狀結構 (Tree Structure)



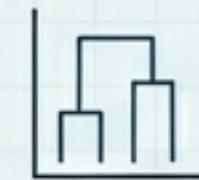
- 特點：不需預先指定群集數



- 兩種類型：

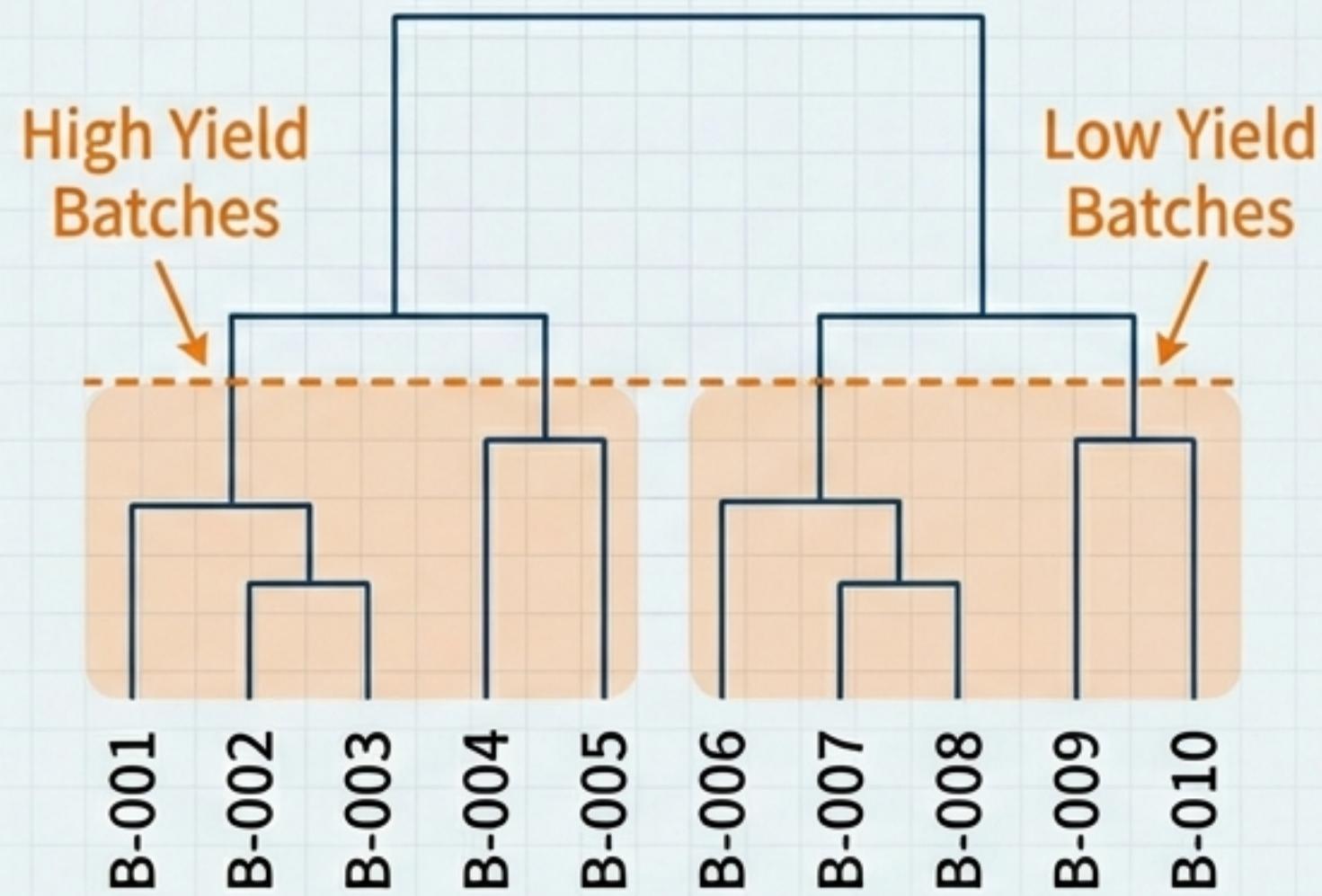
- 凝聚式 (Agglomerative)：  
由下而上合併

- 分裂式 (Divisive)：  
由上而下分割



- 核心視覺化工具：樹狀圖 (Dendrogram)

## ChemE Example: 批次製程分析



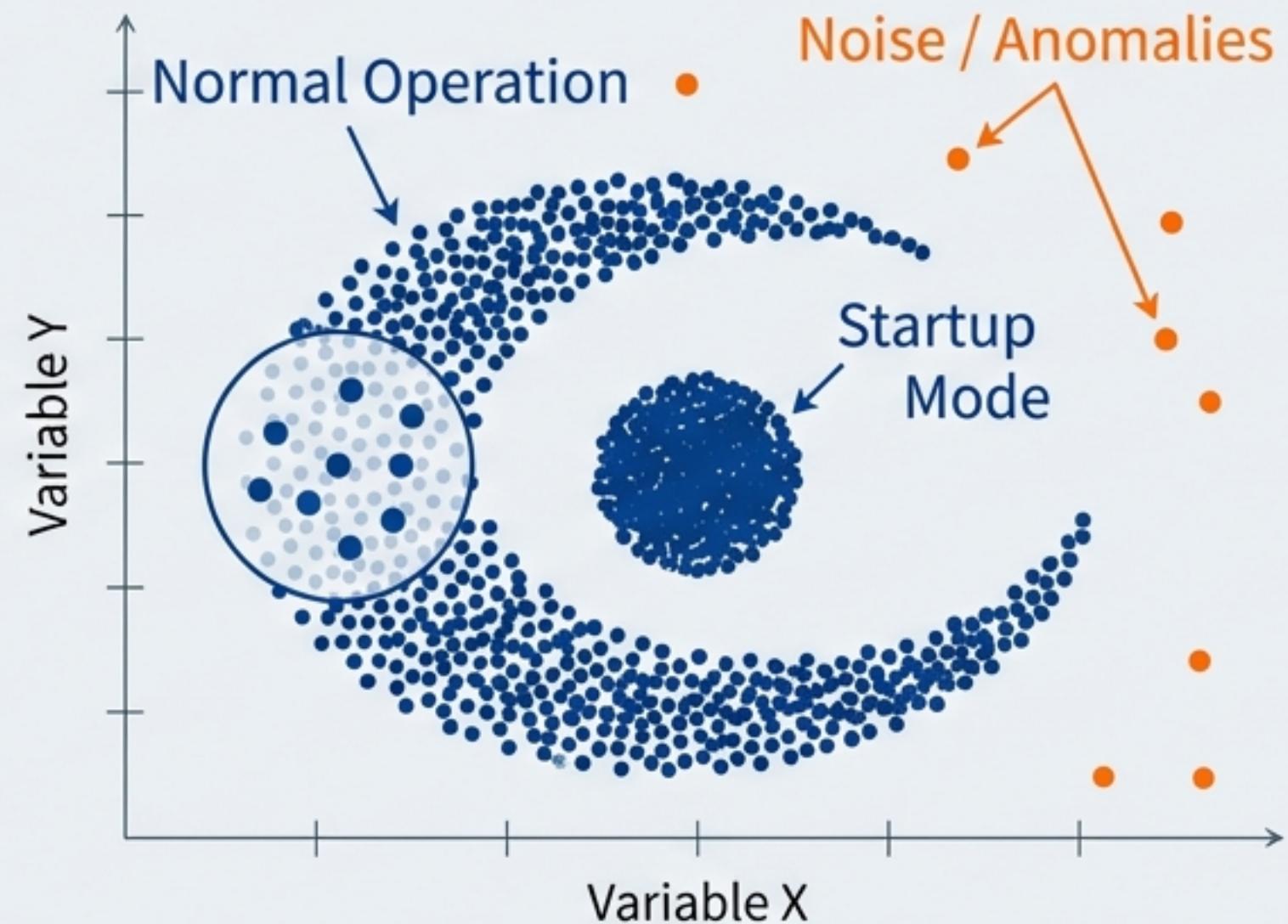
理解批次之間的相似性階層關係

# 演算法 3：基於密度的分群 (DBSCAN)

## 原理與機制

- 機制：密度導向 (Density-based)
- 特點：能發現任意形狀群集，自動排除噪音
- 關鍵參數：
  - $\text{eps} (\varepsilon)$ ：鄰域半徑
  - $\text{min\_samples}$ ：最小鄰居數
- 點的分類：核心點 (Core)、邊界點 (Border)、噪音點 (Noise)

## ChemE Example: 異常檢測 (Fault Detection)



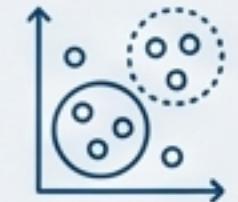
識別不屬於正常操作分佈的孤立點 (Faults)

# 演算法 4：高斯混合模型 (Gaussian Mixture Models, GMM)

## 原理與機制



**機制：**假設數據由多個高斯分佈混合而成



**核心概念：**軟分群 (Soft Clustering)



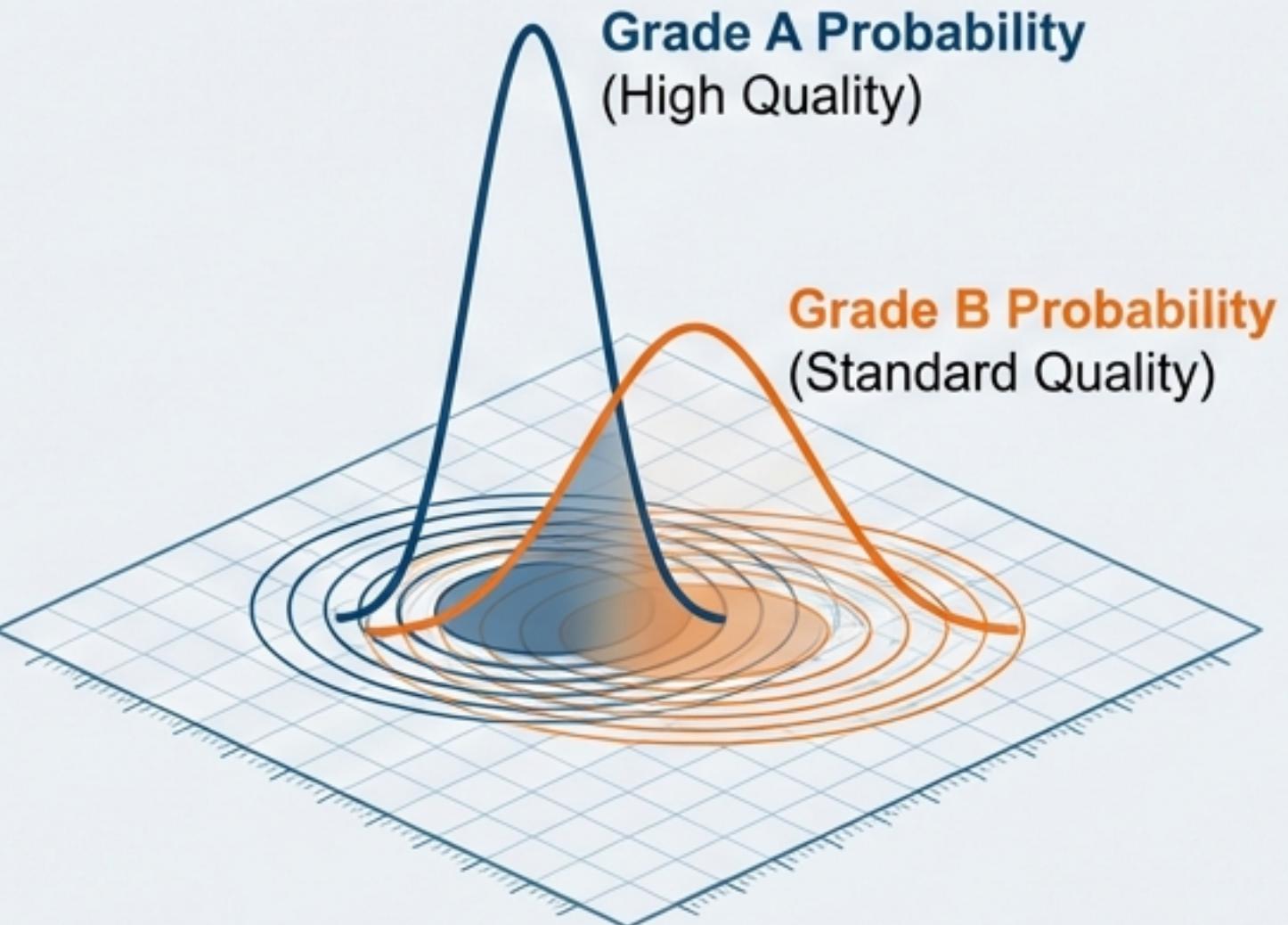
**數學基礎：**計算點屬於某群集的機率

$$p(x) = \sum_k \pi_k \mathcal{N}(x | \mu_k, \Sigma_k)$$



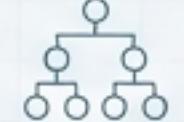
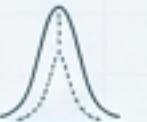
**優點：**可處理橢圓形群集，提供不確定性估計

## ChemE Example: 產品品質分佈建模



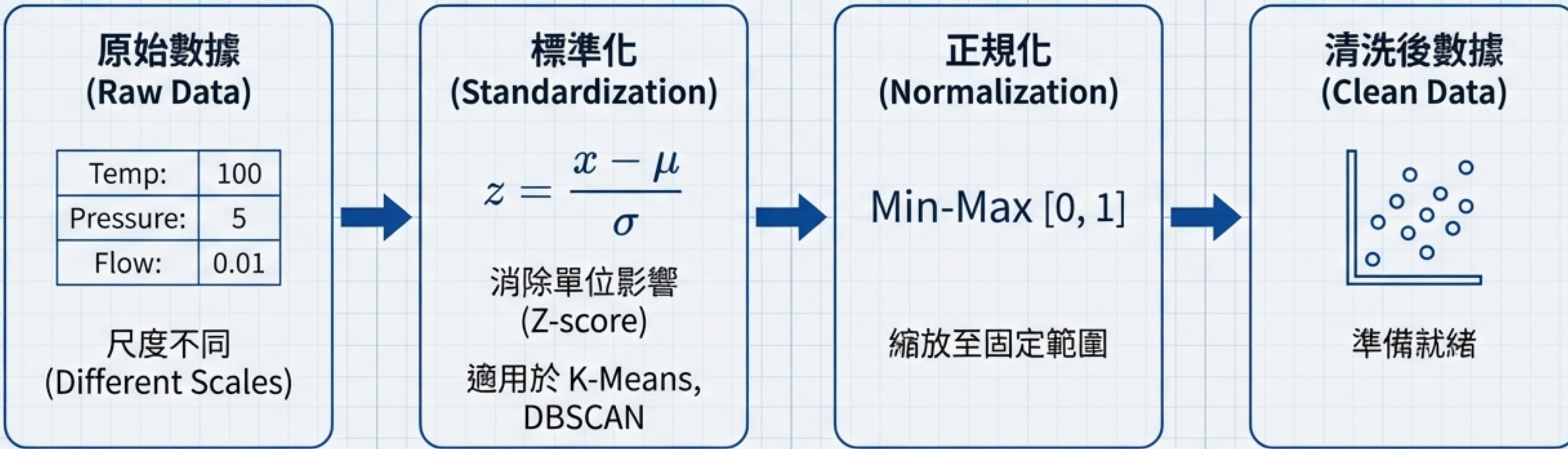
當邊界模糊時，給出屬於特定等級的機率

# 決策指南：如何選擇正確的演算法？

| 演算法<br>(Algorithm)   | 核心特點<br>(Features) | 計算效率<br>(Efficiency)  | 優點<br>(Pros) | 化工適用場景<br>(ChemE Use Case) |
|--|--------------------|-----------------------|--------------|----------------------------|
| K-Means         | 球形分佈，需指定 K         | 高 ↑<br>(High)         | 快速簡單         | 操作模式識別<br>(穩定狀態)           |
| Hierarchical  | 樹狀結構，無需 K          | 低 ↓<br>(Low)          | 視覺化階層關係      | 溶劑/配方分類體系                  |
| DBSCAN        | 任意形狀，抗噪聲           | 中 └<br>(Medium)       | 自動識別異常       | 異常/故障檢測                    |
| GMM           | 機率模型，軟分群           | 中/低 ↗<br>(Medium/Low) | 不確定性估計       | 品質機率評估                     |

💡 選擇策略：先評估數據形狀與噪音存在與否，再決定演算法。

# 關鍵前處理：Garbage In, Garbage Out

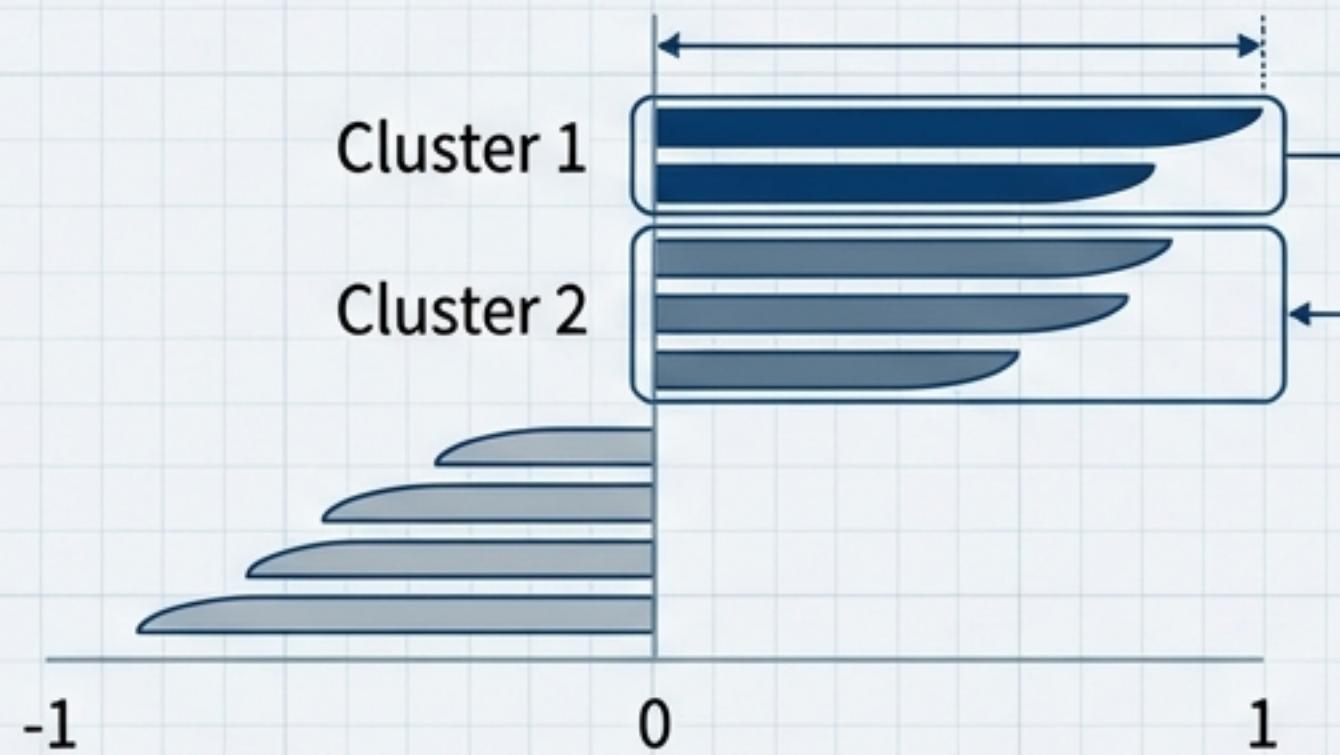


重要提示：忽略標準化會導致大數值特徵（如壓力）主導距離計算，導致錯誤分群。

# 模型評估：沒有答案時，如何知道結果好壞？(Internal Evaluation)

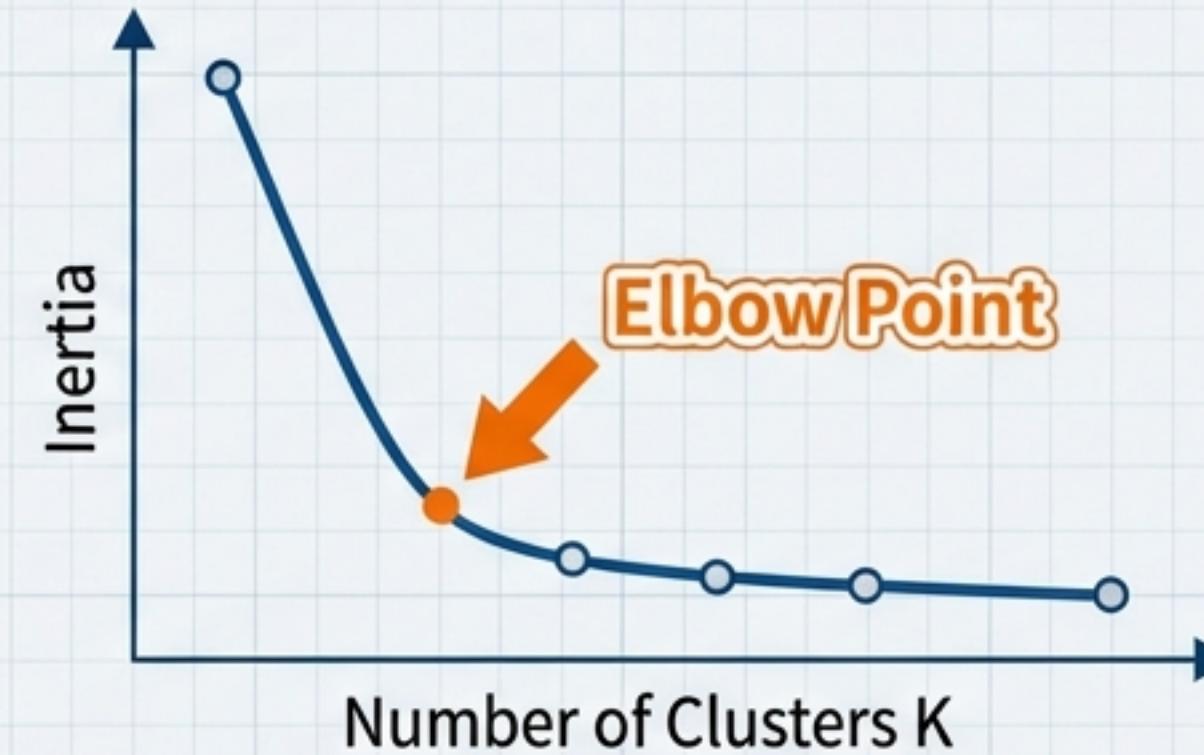
## 輪廓係數 (Silhouette Score)

衡量指標：內聚度 (Cohesion) vs 分離度 (Separation)  
範圍：[-1, 1] (越接近 1 越好)



## 手肘法 (Elbow Method)

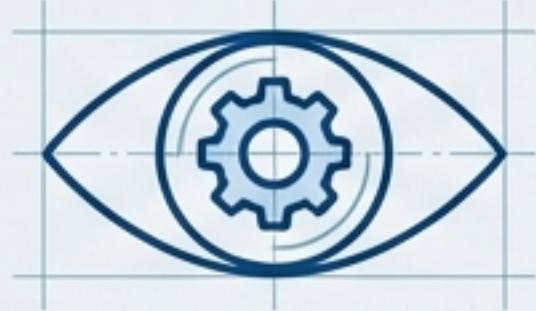
用途：決定最佳的 K 值 (K-Means)  
原理：尋找邊際效益遞減點



⚠ 注意：這些指標僅供參考，最終需結合化工領域知識驗證。

# 工程觀點的驗證 (Engineering Interpretation)

化工廠不是 Kaggle 競賽，物理意義至關重要。



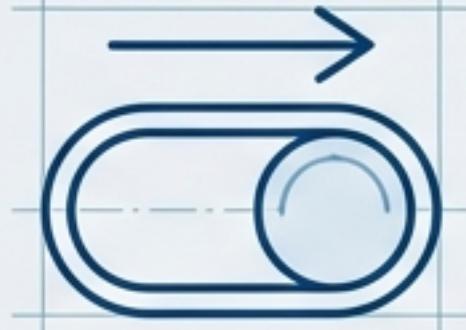
## 物理可解釋性 (Interpretability)

每個群集是否對應明確的操作狀態？  
(e.g., Is Cluster 1 truly 'Startup'?)



## 物理限制 (Physics Constraints)

結果是否違反質量守恆  
(Mass Balance) 或熱力學  
定律？



## 操作合理性 (Feasibility)

模式切換是否過於頻繁？  
(Rapid switching is often  
physically impossible)

結論：數據科學家必須與製程工程師 (Domain Expert) 緊密合作。

# 實務應用流程 (Implementation Workflow)

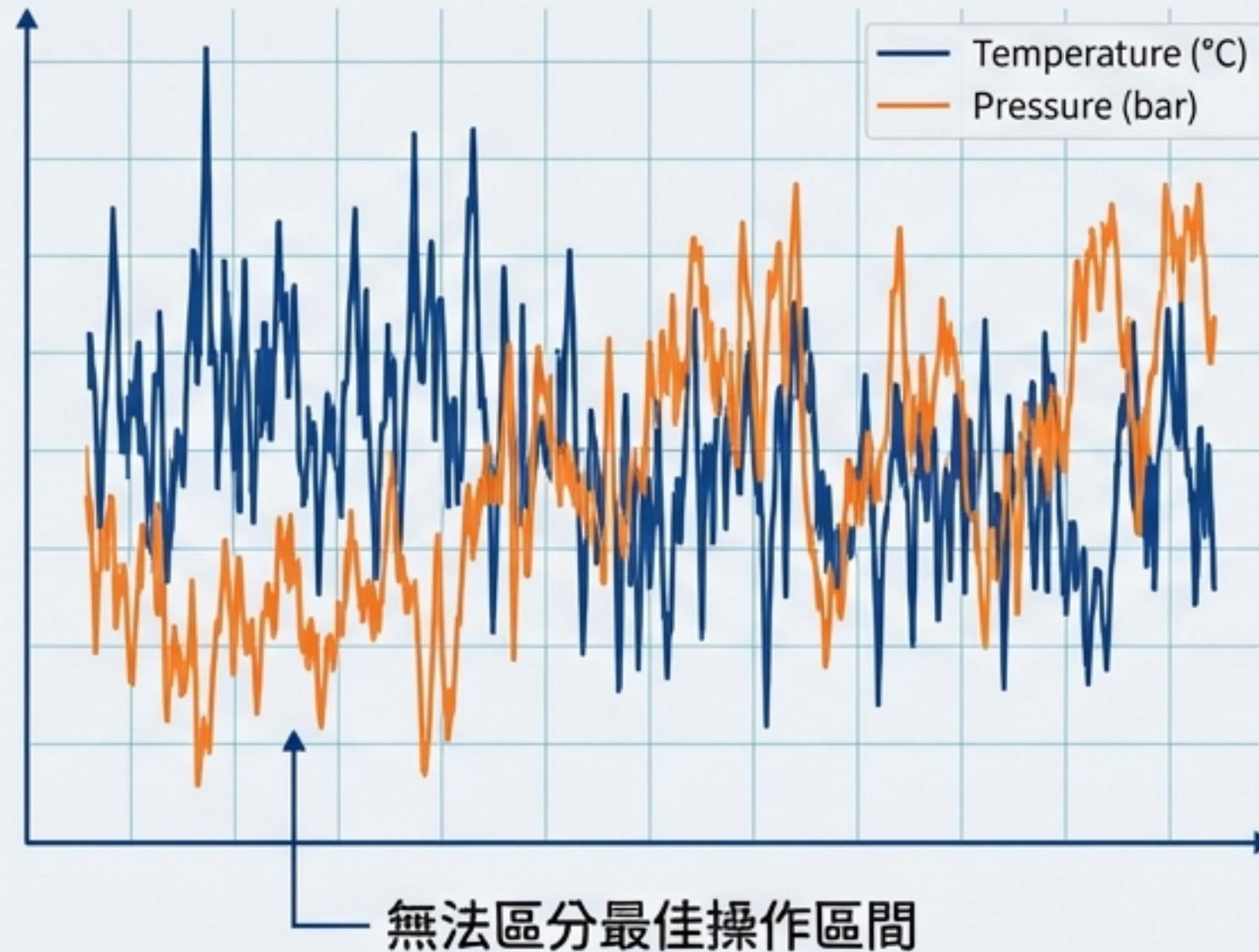


10mm

# 案例研究：反應器多模式操作識別

Case Study: Reactor T-101 Mode Identification

## 原始信號 (Raw Signals)



## 分群結果 (Clustered Results)



成效：建立即時監控系統，當操作點移出 Cluster 2 時自動發出警報。

# 重點回顧 (Key Takeaways)



## 智慧化操作 (Intelligent Operations)

### 核心概念 (Concept)

分群是發現數據隱藏結構的強大非監督工具。

### 工具選擇 (Tools)

K-Means (快速),  
DBSCAN (抗噪),  
GMM (機率)。  
沒有萬能演算法。

### 工程實踐 (Practice)

數據標準化是前提。  
物理可解釋性是最終標準。

10mm

# 下一步：處理高維數據的挑戰

當變數從 3 個變成 300 個時，分群效果會變差（維度詛咒）。

## Unit 05 Clustering

## Unit 06 Dimensionality Reduction (PCA, t-SNE)

AI 不會取代化工工程師；但懂得使用 AI 的化工工程師將取代不懂的人。

(AI won't replace ChemEs; but ChemEs who use AI will replace those who don't.)