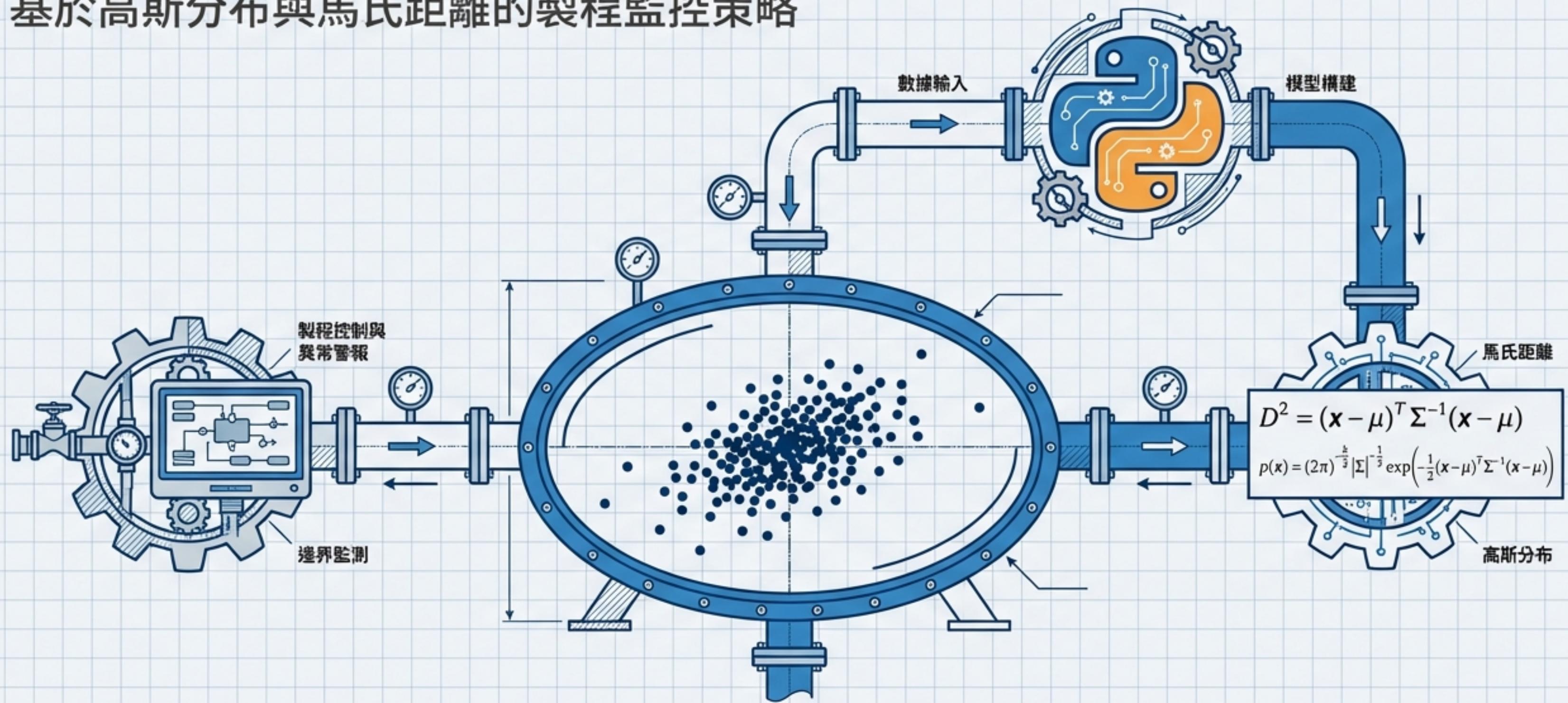
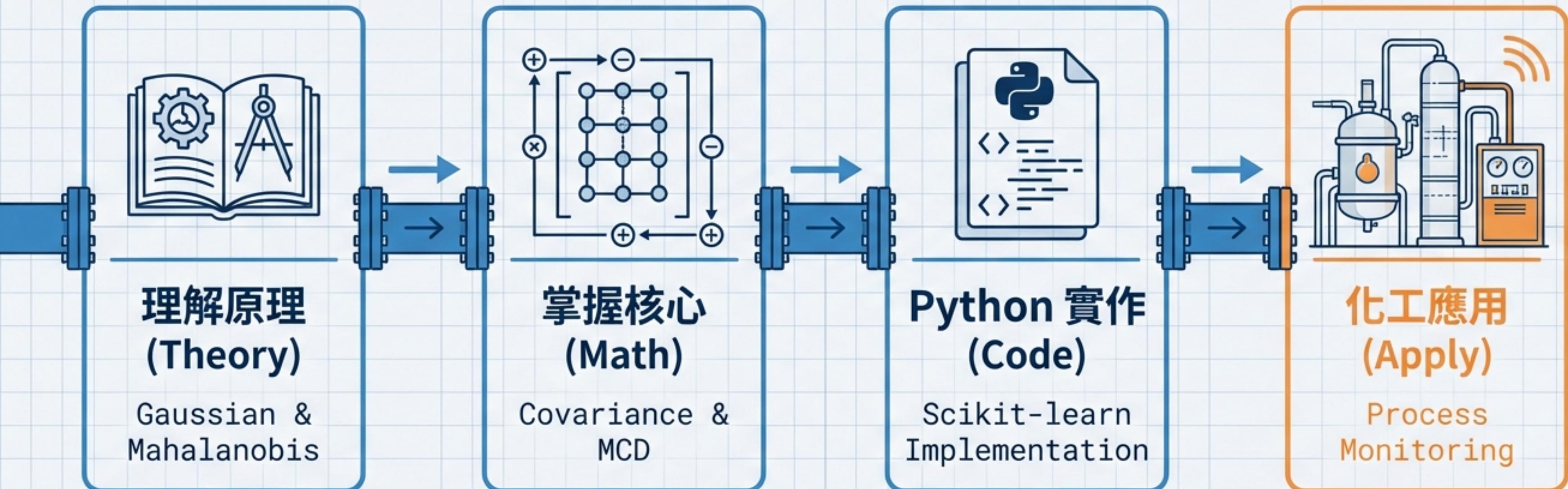


Unit 07: 橢圓包絡 (Elliptic Envelope) 異常檢測

基於高斯分布與馬氏距離的製程監控策略



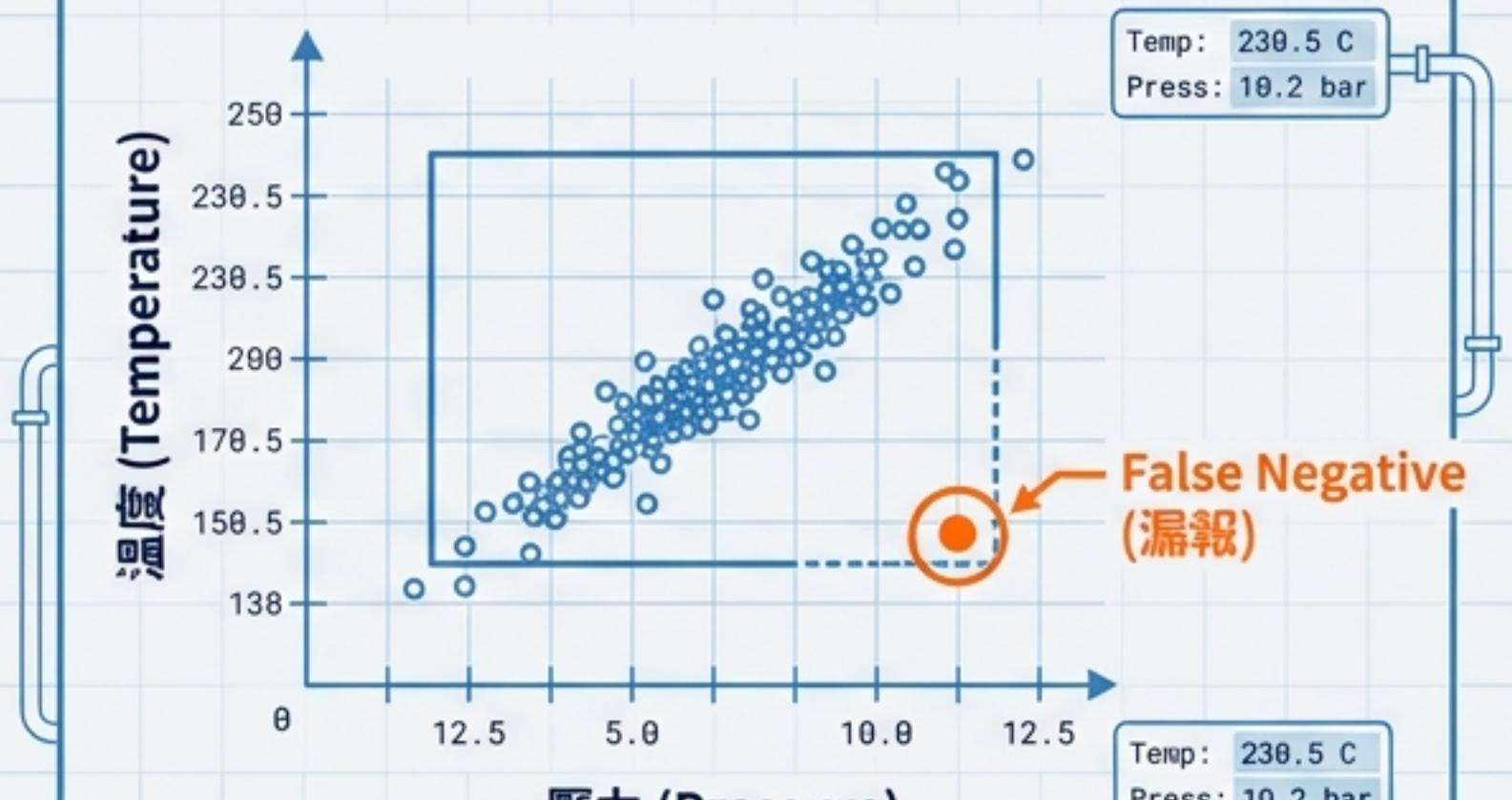
本單元學習目標



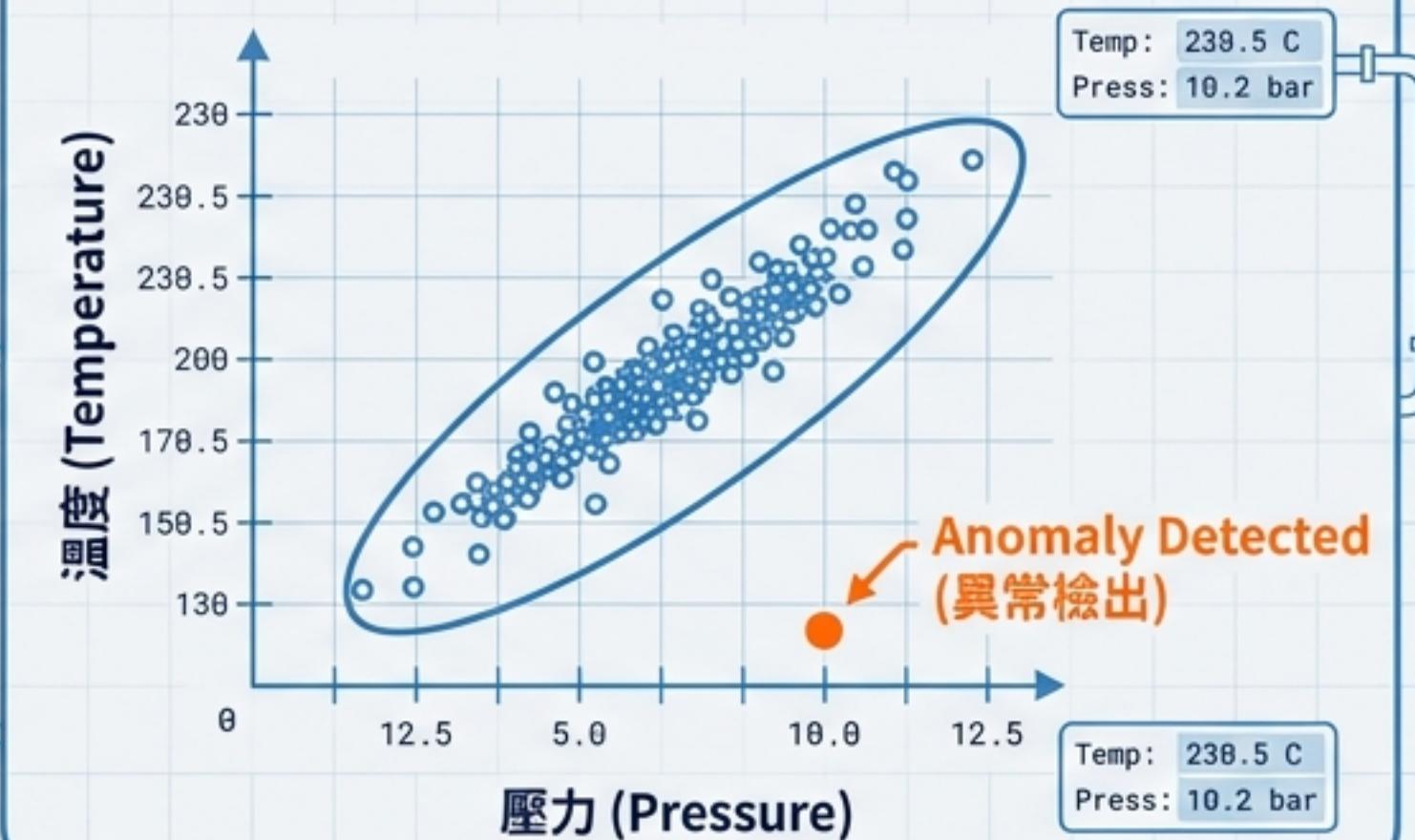
從數學理論到數位化工廠的實際部署

為什麼需要橢圓包絡？(Why Elliptic Envelope?)

傳統上下限 (Univariate Limits)

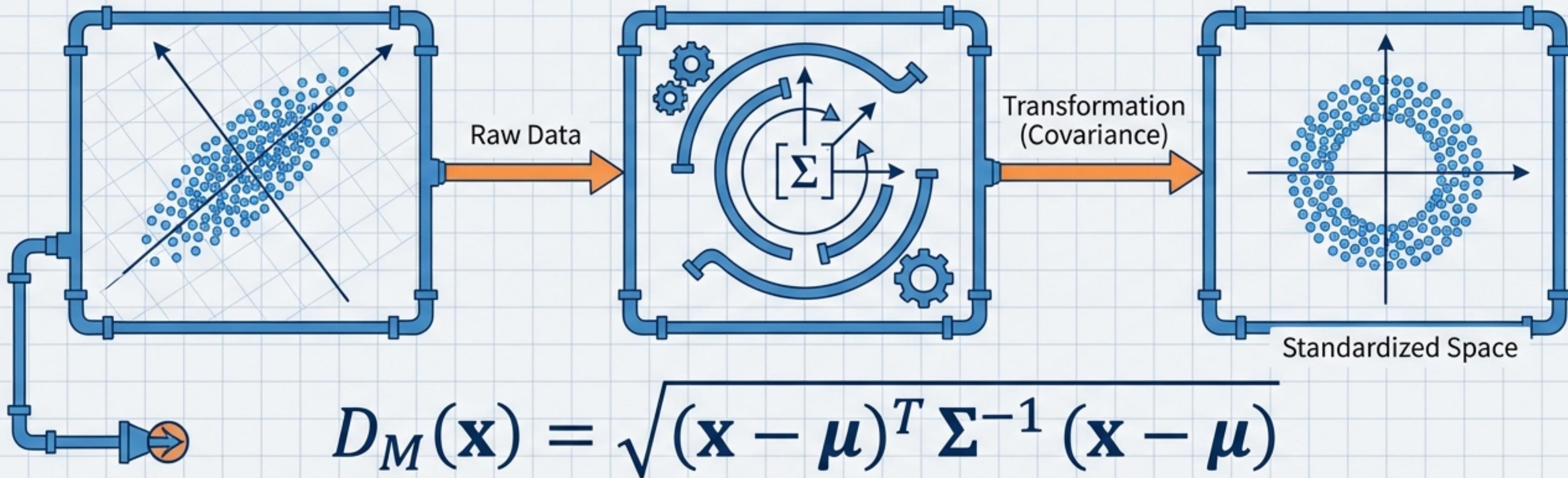


多變量包絡 (Multivariate Envelope)



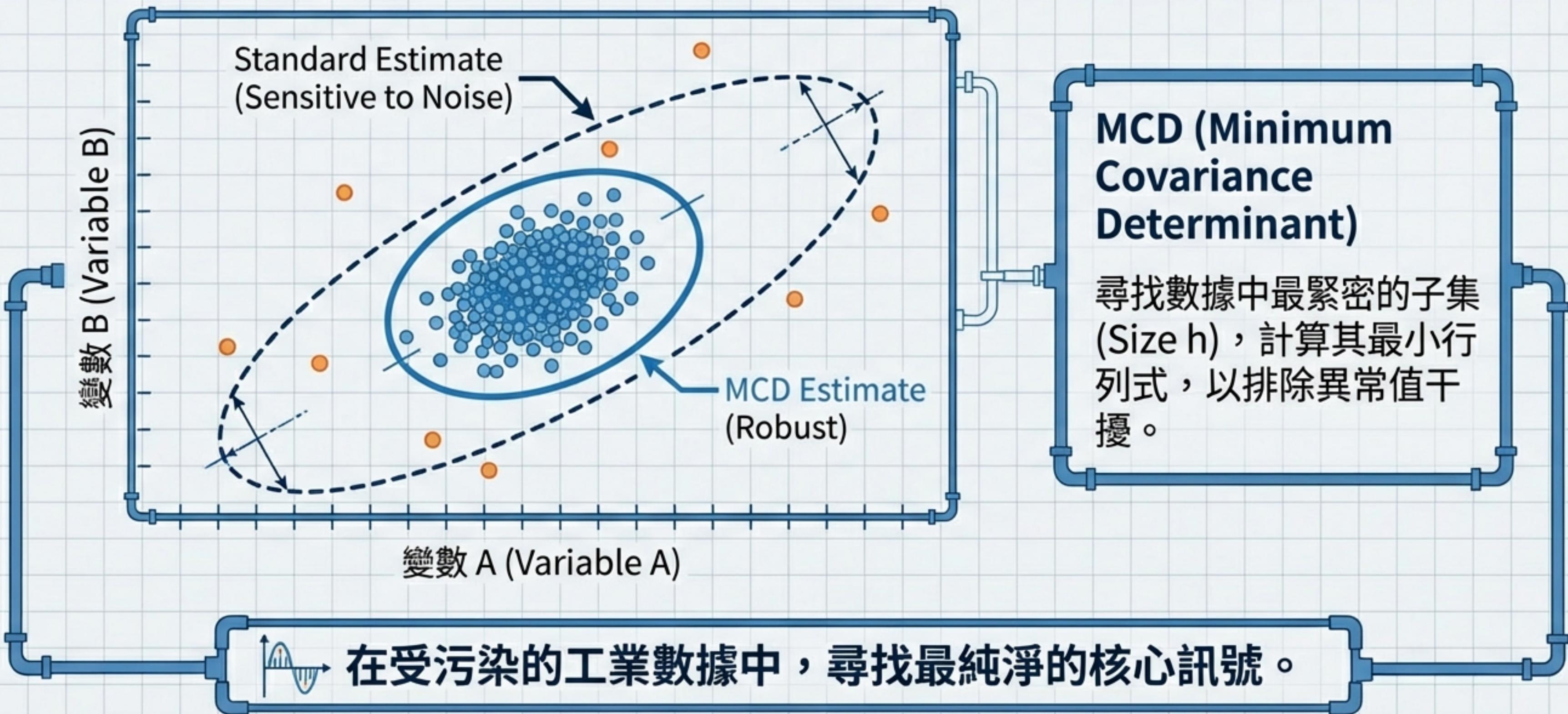
- 製程穩定性：變數圍繞設定點波動
- 變數相關性：溫度與壓力非獨立變數
- 核心假設：正常數據服從多變量高斯分布

核心原理：馬氏距離 (Mahalanobis Distance)

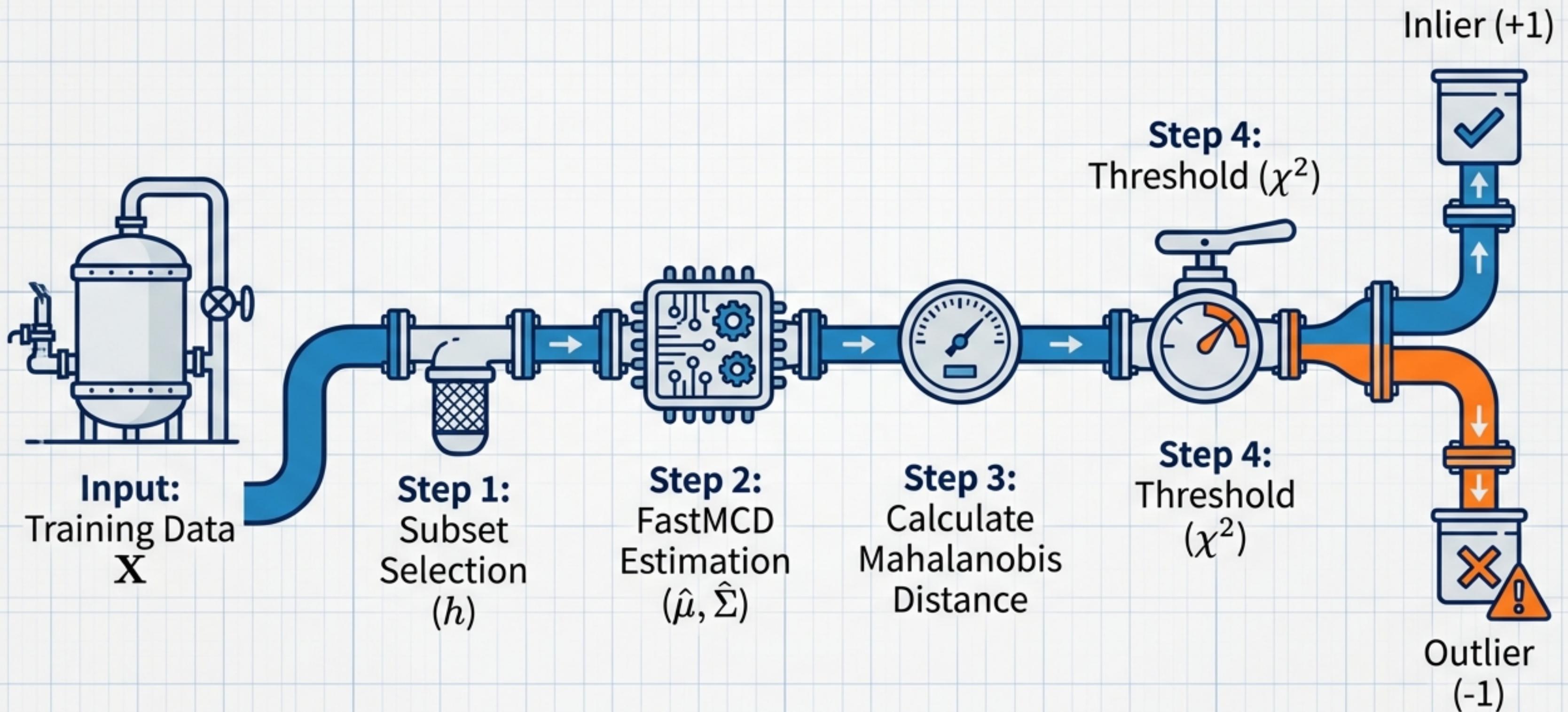


- 標準化 (Standardization)：自動處理不同單位 (如 °C vs Bar)
- 相關性 (Correlation)：考慮變數間的共變異數矩陣 Σ
- 直觀意義：多變量 Z 分數 (Multivariate Z-Score)

穩健估計：MCD 的作用



演算法運作流程 (Algorithm Workflow)



Python 實作範例 (Implementation Example)

```
from sklearn.covariance import EllipticEnvelope  
  
# 1. 設定模型 (Configure)  
# contamination: 預期異常比例 (0.1 = 10%)  
model = EllipticEnvelope(contamination=0.1, random_state=42)  
  
# 2. 擬合數據 (Fit) - 學習高斯分布形狀  
model.fit(X_train)  
  
# 3. 預測結果 (Predict) - 判斷是否異常  
# 1 = 正常, -1 = 異常  
y_pred = model.predict(X_test)
```

定義異常偵測器

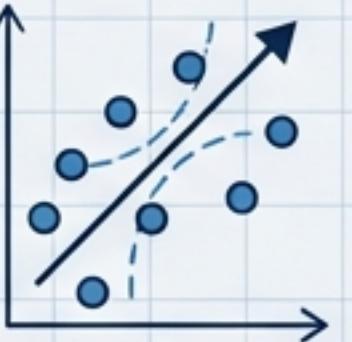
MCD 穩健估計 (μ, Σ)

關鍵參數設定 (Key Hyperparameters)

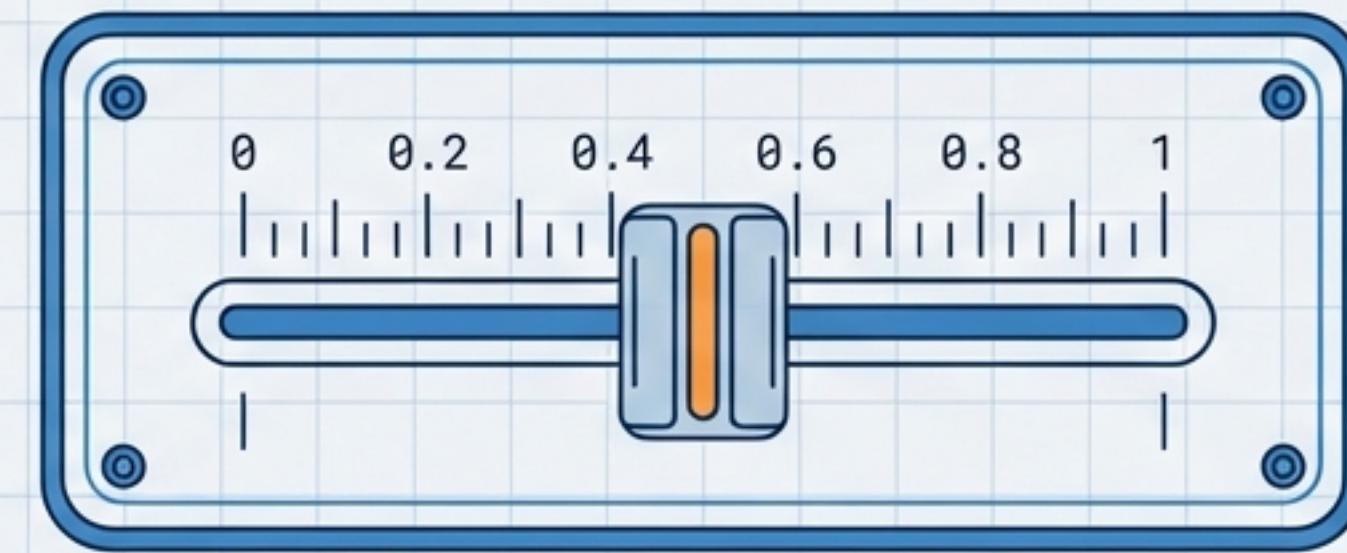
Contamination



決定邊界的鬆緊 (Sensitivity)。
數值越大 = 邊界越緊 (Strict)，
更多點被視為異常。



Support Fraction

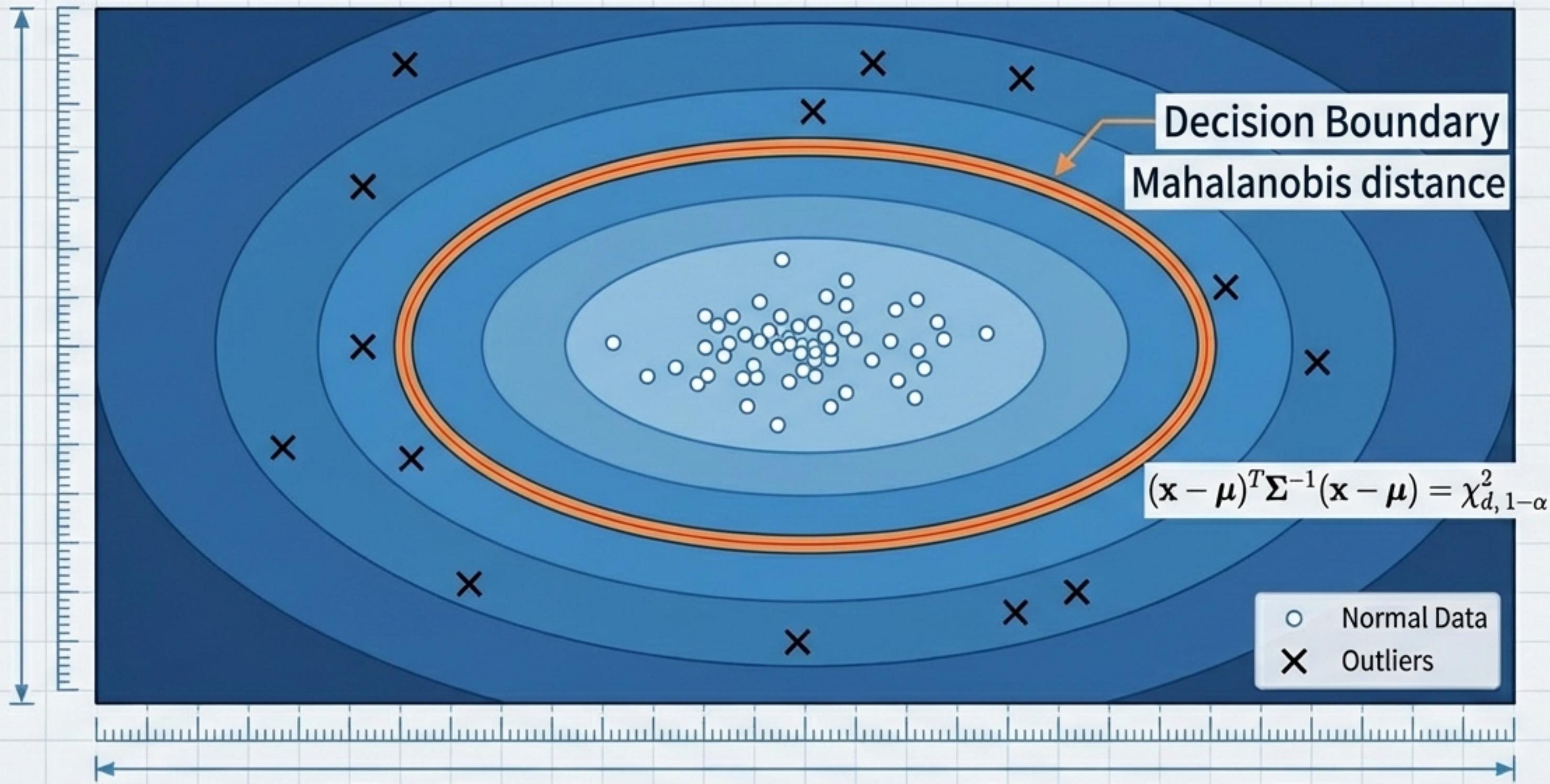


決定估計的穩健度 (Robustness)。
數值越小 = 抗噪能力越強，
但效率較低。



Contamination 決定警報靈敏度；Support Fraction 決定模型抵抗汙染的能力。

決策邊界可視化 (Visualizing the Decision Boundary)



紅色橢圓線即為根據 Contamination 設定的安全操作界限。

橢圓包絡的優勢與限制 (Pros & Cons)

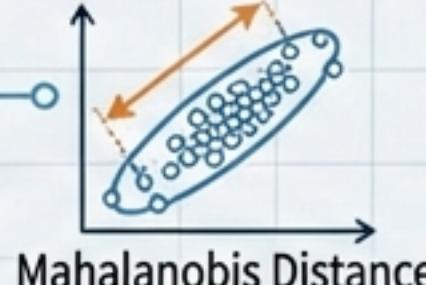


優勢 (Pros)

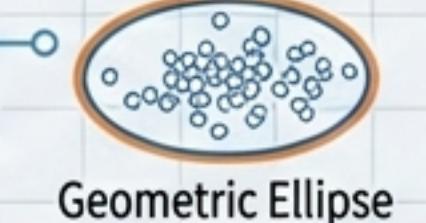
- 數學嚴謹 (Rigorous)
 - 基於高斯分布理論



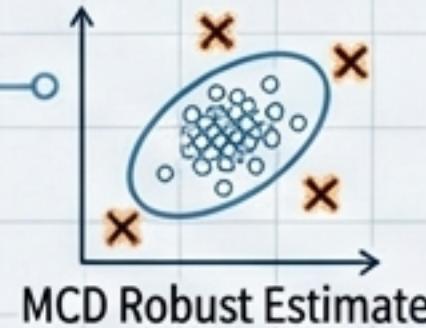
- 處理相關性 (Correlation)
 - 馬氏距離有效解決變數耦合



- 可解釋性 (Explainable)
 - 邊界為幾何橢圓，易於理解

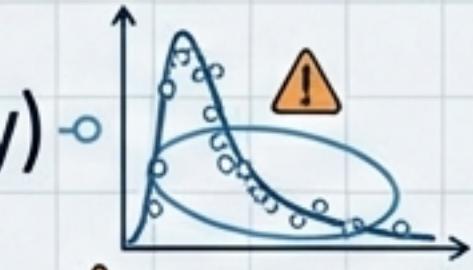


- 穩健抗噪 (Robust)
 - MCD 抵抗離群值干擾

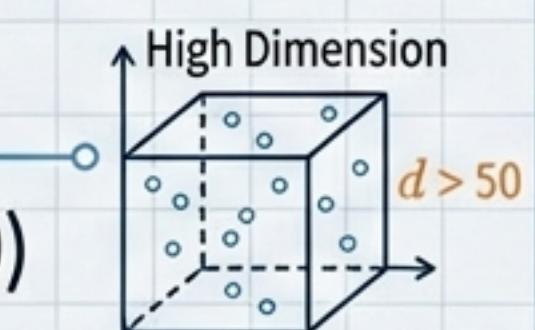


限制 (Cons)

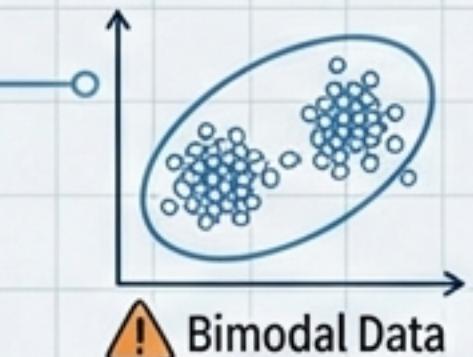
- 高斯假設 (Gaussian Only)
 - 數據必須接近常態分布



- 維度限制 (Dimension)
 - 不適合高維數據 ($d > 50$)



- 單峰限制 (Unimodal)
 - 無法處理多模態分布



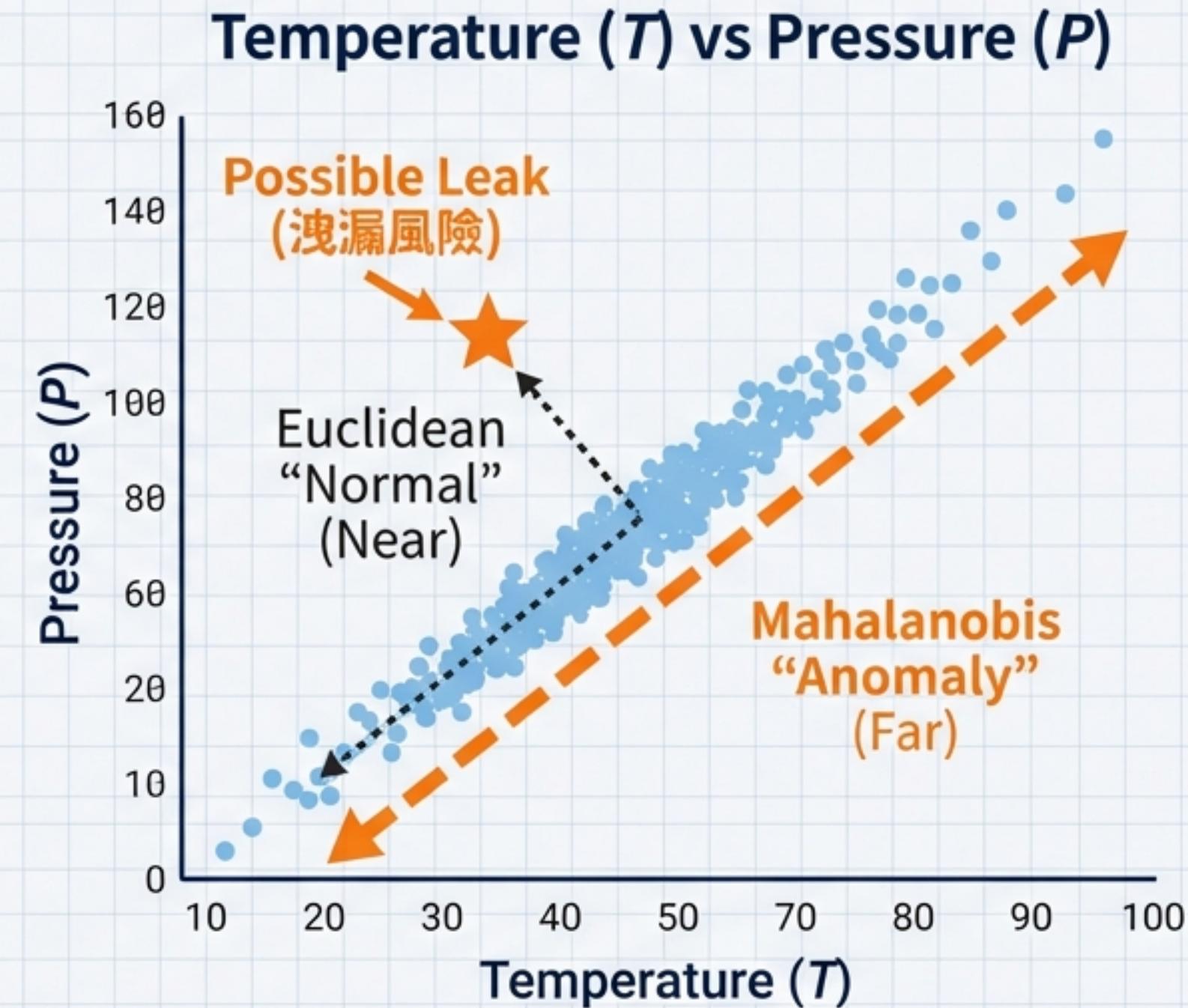
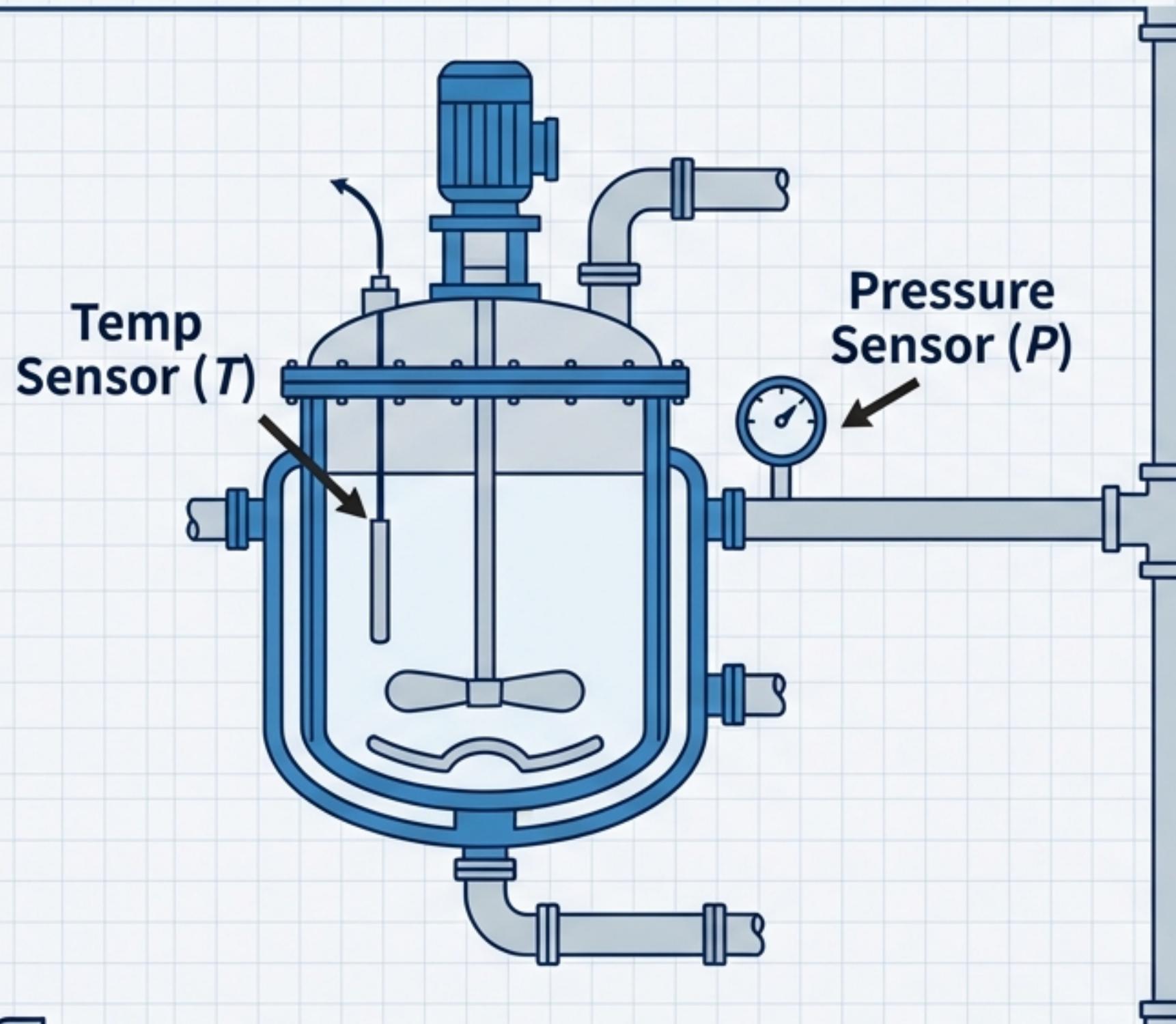
異常檢測演算法對決 (Algorithm Comparison)

特性 (Feature)	橢圓包絡 (Elliptic Envelope)	Isolation Forest	LOF
分布假設 (Assumption)	Gaussian (High)	None	None
可解釋性 (Explainability)	High (Geometric)	Low (Black Box)	Medium
高維數據 (High Dim)	Poor	Good	Poor
多模態 (Multi-modal)	Poor	Good	Good

選擇 Elliptic Envelope → 當數據符合物理高斯分布且需要解釋性。

選擇 Isolation Forest → 當數據複雜、高維且無明顯分布。

化工應用案例：反應器監控 (CSTR Monitoring)

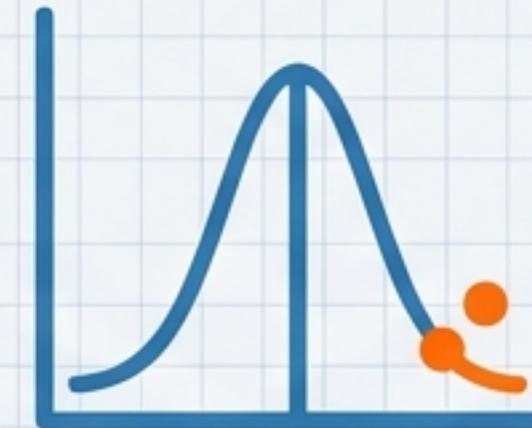


馬氏距離能檢測出『違反物理相關性』的危險訊號 (如高溫卻低壓)，這是單變數監控無法發現的。

實務應用檢查清單 (Best Practices Checklist)



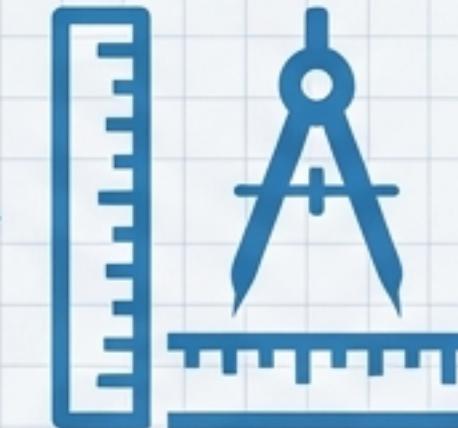
總結：您的數位監控守門員



Gaussian Assumption



MCD Robustness



Mahalanobis Distance

橢圓包絡為穩態製程提供了一個基於統計學的嚴謹邊界，是品質控制與異常偵測的強力工具。

下一步 (Next Step)：
開啟 Unit07_Elliptic_Envelope.ipynb
現在開始您的實作練習！

Unit 02 結束 | 前往 Unit 03