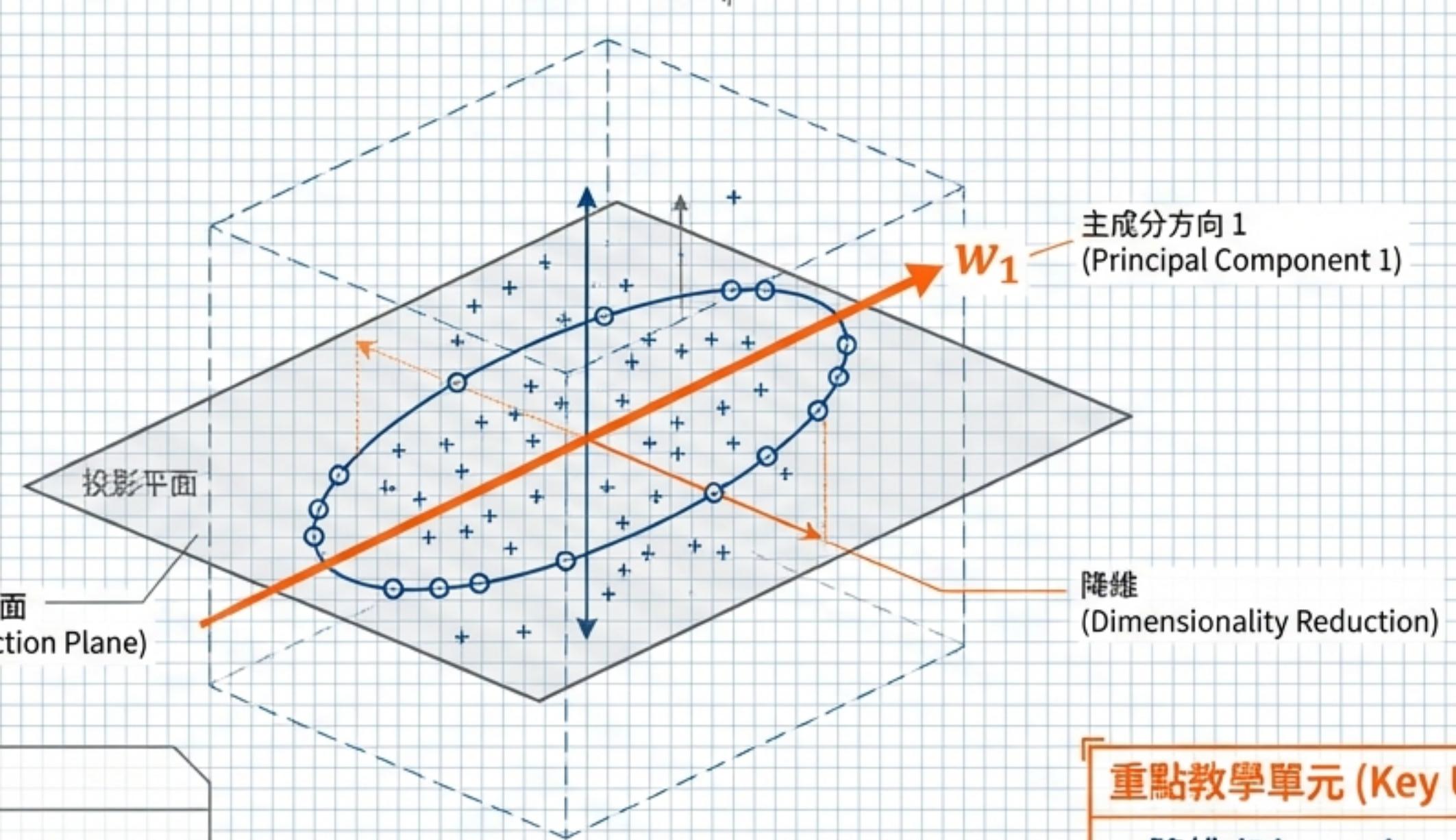


# AI 在化工上的應用：從原理到實踐

## Unit 06 主成分分析 (PCA)：降維與製程監控



### 課程資訊 (Course Metadata)

授課教師	莊曜楨 助理教授
學期	114學年度第2學期
所屬單位	逢甲大學 化工系 智慧程序系統工程實驗室

PCA 原理圖示 (PCA Schematic)

### 重點教學單元 (Key Unit)

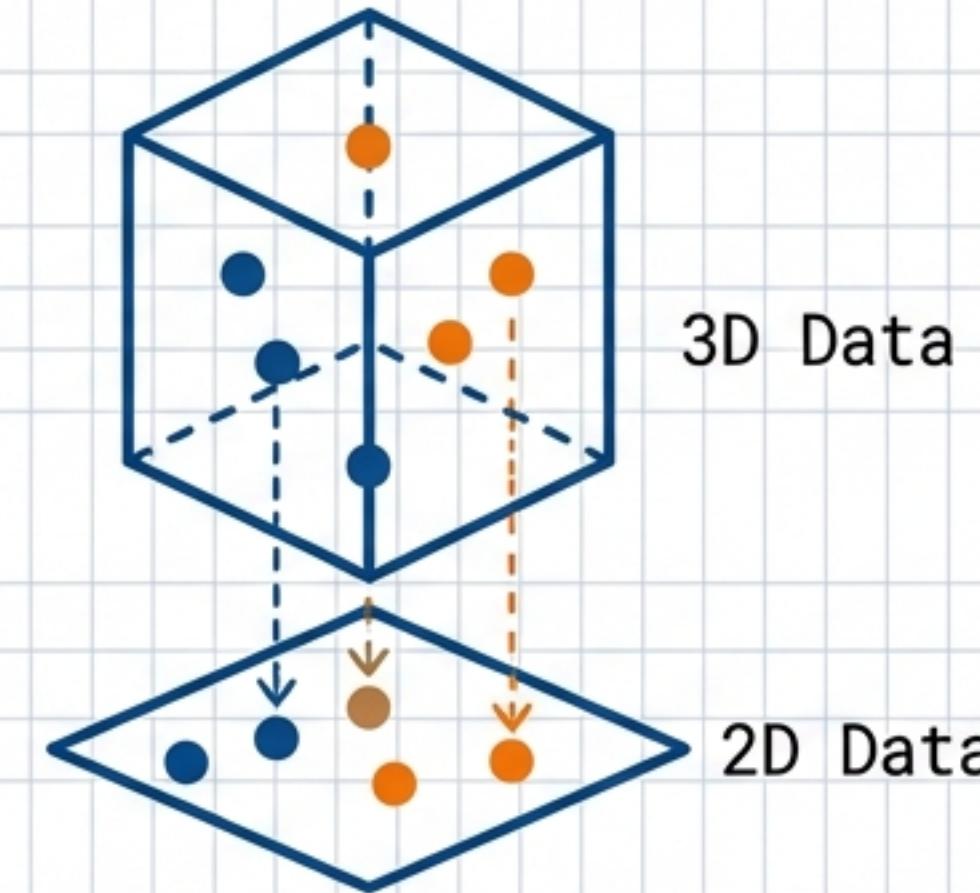
- 降維 (Dimensionality Reduction)
- 特徵提取 (Feature Extraction)
- 多變數統計製程控制 (MSPC)

# 為什麼需要 PCA？化繁為簡的智慧



## 化工挑戰 (The Challenge)

- 現代化工廠擁有成千上萬個感測器 (Sensors)
- 數據量大 (High-Dimensional)
- 變數間高度相關 (Multicollinearity)
- 核心問題：如何從 100+ 個變數中，快速看清反應器的真實狀態？



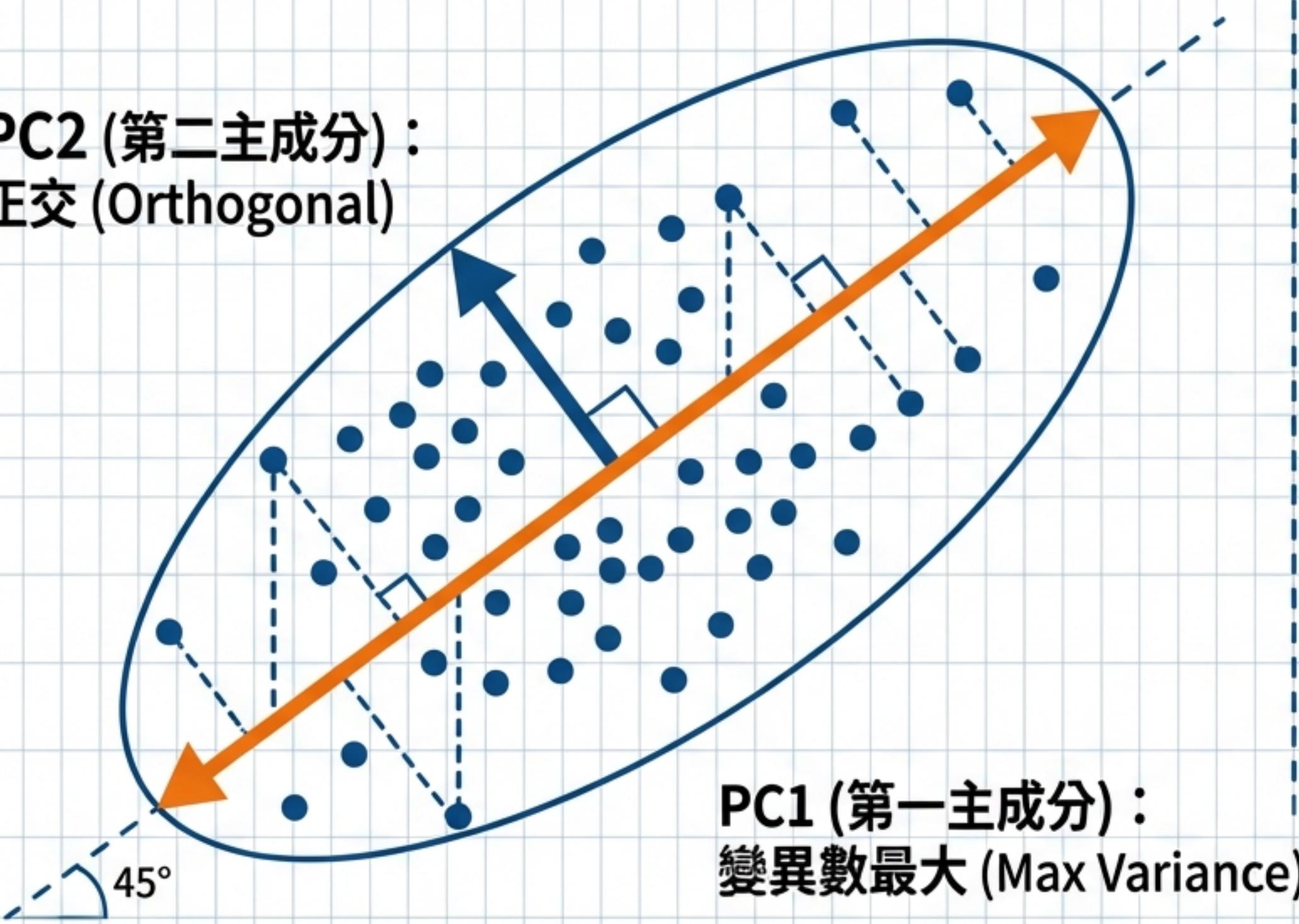
## 核心目標 (Course Objectives)

1. 理解原理 (Understand)：掌握變異數最大化與正交投影
2. 學會實作 (Master)：使用 Python scikit-learn 建立模型
3. 解決問題 (Apply)：應用於製程監控 (Process Monitoring) 與故障診斷

無監督式學習 (Unsupervised Learning)：輸入只有  $X$  (無標籤)，目標是發現數據內部的隱藏結構。

# PCA 核心原理：尋找最佳投影視角

PC2 (第二主成分)：  
正交 (Orthogonal)



## 幾何觀點

找到一個投影方向，使得數據在該方向上的「影子」最寬（保留最多資訊）。

## 降維魔法 (The Magic)

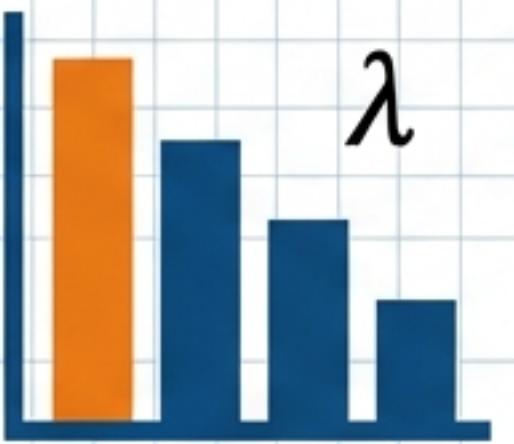
1. 旋轉 (Rotation)：將原本糾纏的變數轉換為新的座標系。
2. 正交 (Orthogonality)：新的主成分彼此不相關，消除共線性。
3. 排序 (Ranking)：資訊量  $PC1 > PC2 > PC3 \dots > PCk$ 。

# 解密 PCA 術語：理解層次結構

$$X \approx ZW^T$$

(Data  $\approx$  Scores  $\times$  Loadings)

## 1. 特徵值 (Eigenvalues, $\lambda$ )



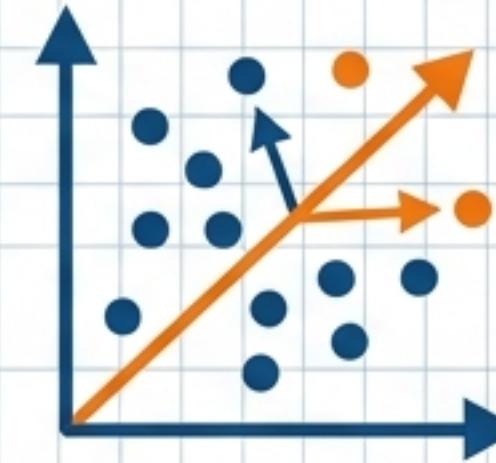
- 定義：每個主成分的**變異數**大小。
- 意義：代表該主成分的「**重要性**」。 $\lambda$  越大，包含資訊越多。

## 2. 載荷 (Loadings, $w$ )



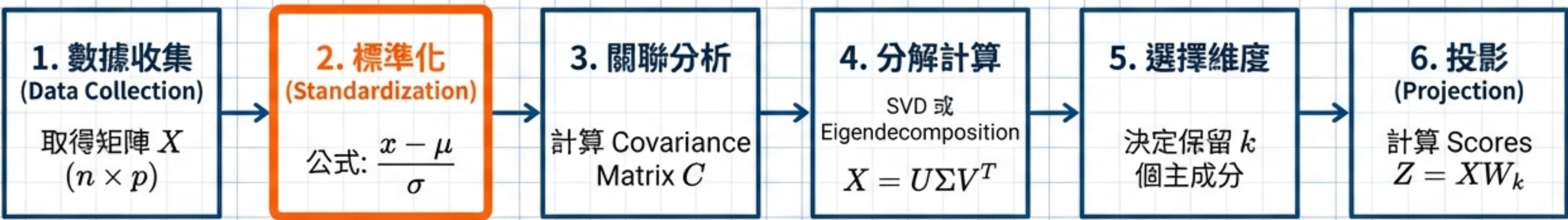
- 定義：原始變數對主成分的**貢獻係數**(權重)。
- 意義：主成分的「**配方表**」。
- Example:  $PC1 = 0.7 \times Temp + 0.5 \times Pressure$

## 3. 得分 (Scores, $z$ )



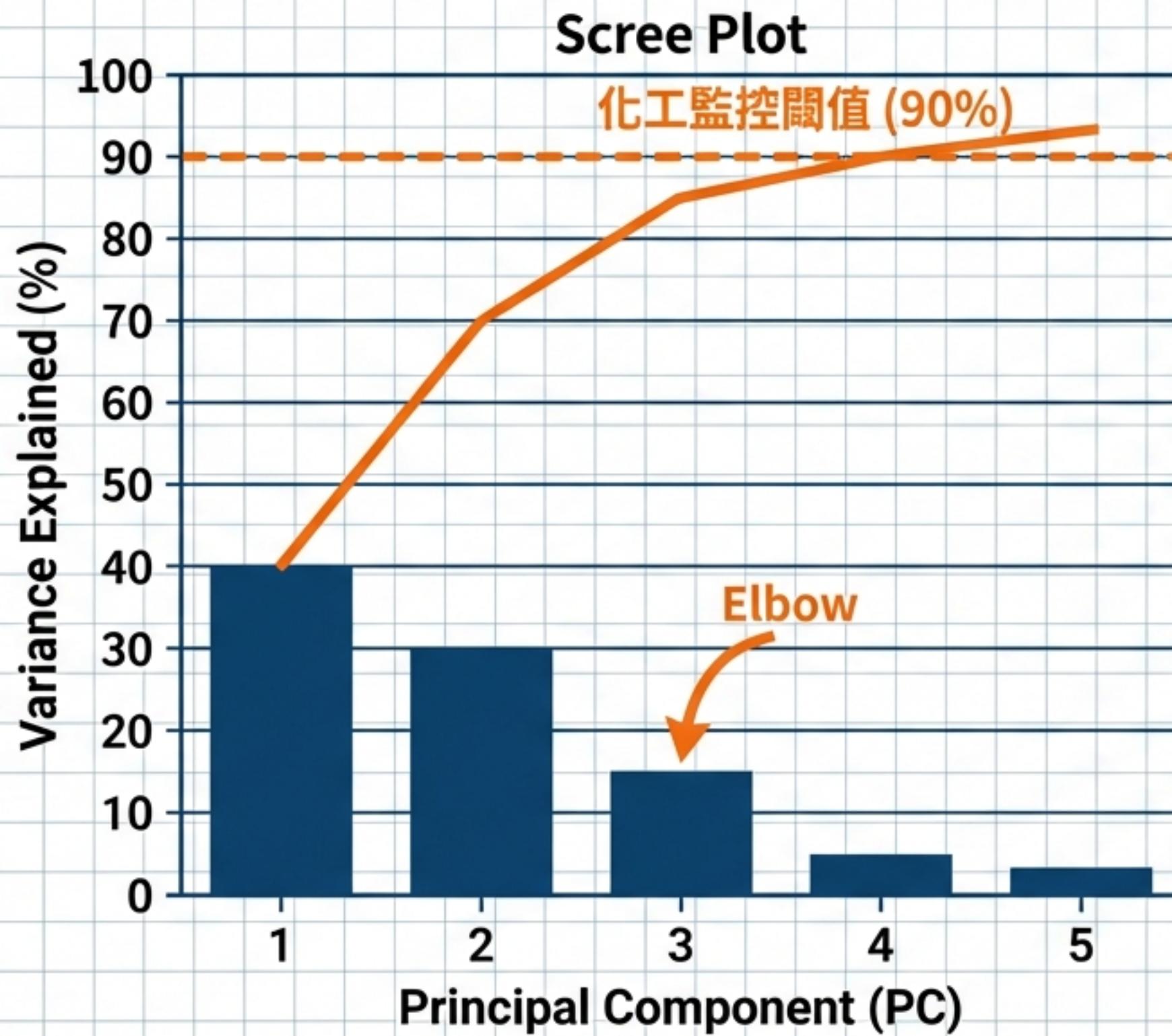
- 定義：原始樣本在主成分空間中的投影座標。
- 意義：降維後的新特徵，用於繪製圖表 (Score Plot)。

# PCA 演算法流程：從數據到模型



⚠ 極度重要!  
消除單位差異  
(°C vs bar)

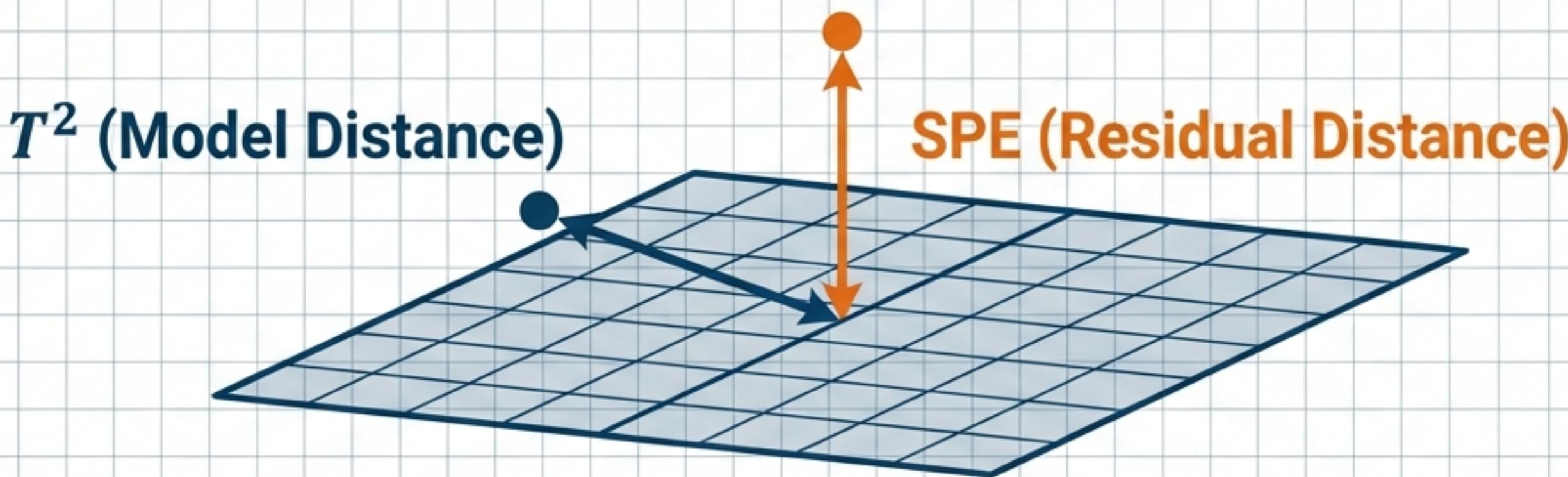
# 決策時刻：該保留多少主成分？



## 選擇準則 (Selection Criteria)

1. 累積解釋變異數 (Cumulative Explained Variance)：
  - 化工應用建議：90% - 95% (保留高信息量以偵測異常)。
2. 陡坡圖 (Scree Plot)：
  - 尋找「肘點」 (Elbow Point) – 斜率劇烈變化處，之後視為噪音。
3. Kaiser 準則：
  - 保留特徵值  $\lambda > 1$  的主成分。

# 製程監控雙指標： $T^2$ 與 SPE



## Deep Engineering Blue

### 1. Hotelling's $T^2$ (模型內監控)

- 定義：樣本在主成分空間中心點的距離。
- 偵測：「系統性偏移」(Systematic Shift)。
- 口語：我們還在正常的操作範圍內嗎？

## Safety Orange

### 2. SPE / Q-Statistic (模型外監控)

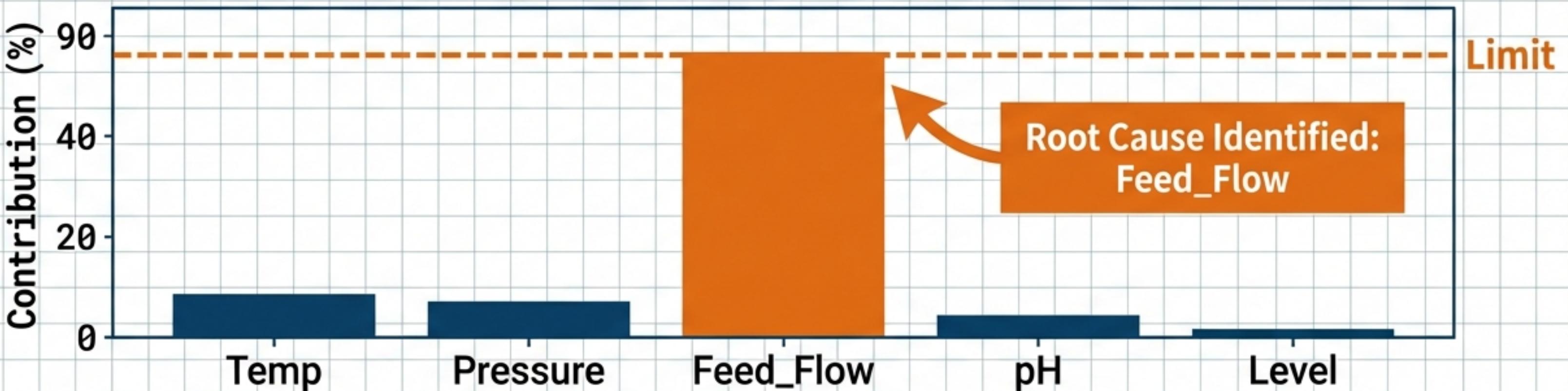
- 定義：樣本投影到模型平面後的殘差距離。
- 偵測：「新型態異常」(New Anomaly)。
- 口語：變數間的相關性結構是否被破壞？

# 異常偵測與診斷：貢獻圖 (Contribution Plot)

## 工作流程 (Workflow)

- 1. 偵測 (Detection): 警報響起 ( $T^2$  或  $SPE > Limit$ )
- 2. 診斷 (Diagnosis): 貢獻圖分析 (Contribution Plot)

Contribution to SPE Alert



## 化工情境應用：

- 快速定位故障源 (Root Cause Analysis)
- 縮短停機時間 (Downtime Reduction)
- 區分感測器故障 vs. 真實製程變動

# 實戰演練：反應器監控 (1/2) — 模型建立

## Case Context

- 場景: 連續式攪拌槽反應器 (CSTR)
- 數據: 10個變數 (Temp, Pressure, Flow, Conc, pH)
- 相關性結構: 溫度群組高度相關 ( $r = 0.82$ )，流量與濃度耦合。

## Results

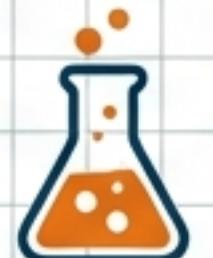
### PCA Loadings 分析 (物理意義):



PC1 (35.5% 變異數)  
熱力學狀態軸 (Reactor\_Temp, Jacket\_Temp, Pressure)



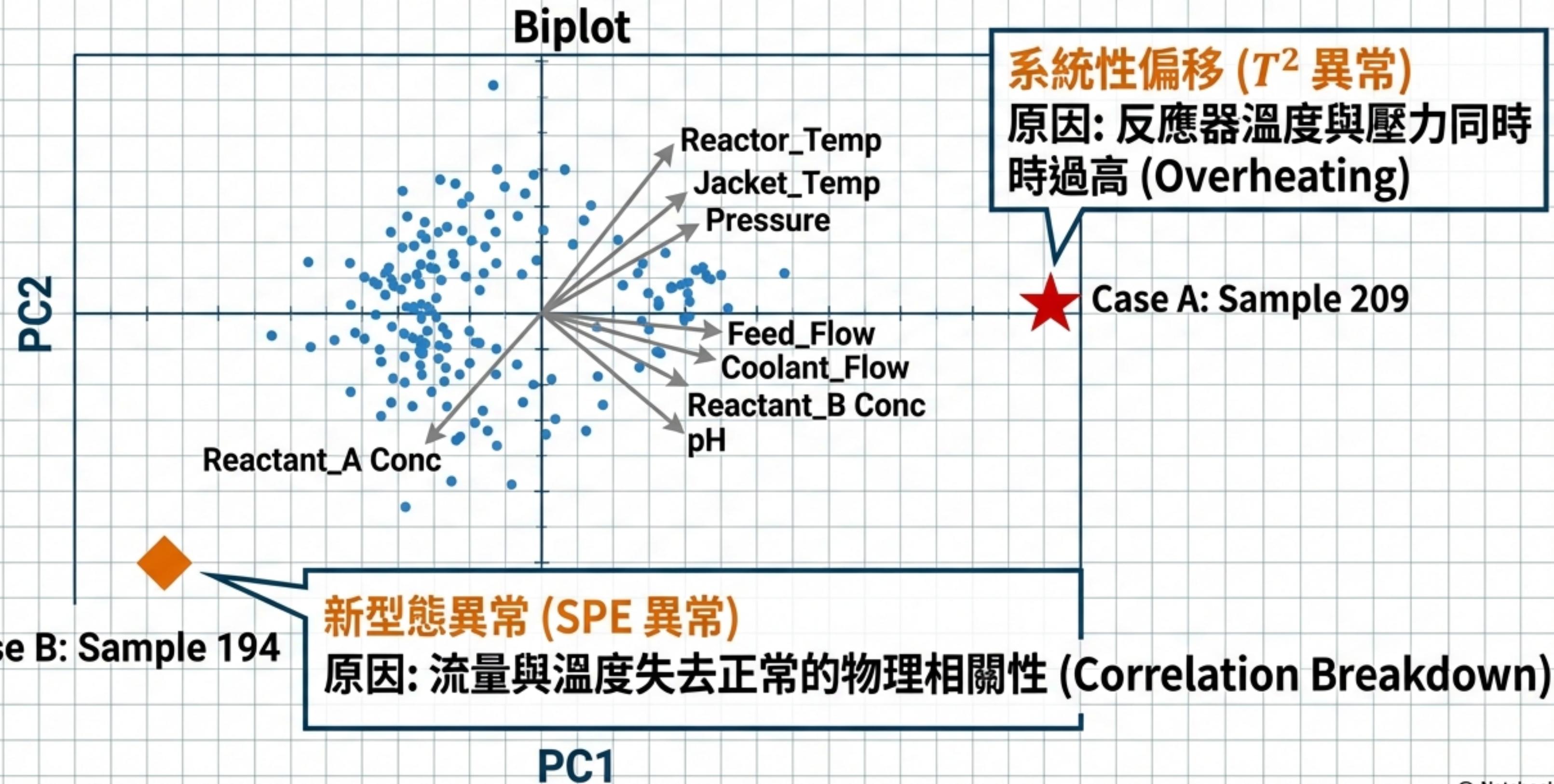
PC2 (31.1% 變異數)  
流量控制軸 (Feed\_Flow, Coolant\_Flow)



PC3 (18.7% 變異數)  
濃度平衡軸 (Reactant\_A/B Conc)

結論: 3 個主成分解釋了 85.26% 的總變異數，有效降維。

# 實戰演練：捕捉異常訊號 (2/2) – 診斷分析



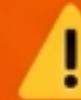
# PCA 的優勢與現實挑戰

## ✓ 優點 (Pros)

- ✓ 降維：10維 → 3維，效率高。
- ✓ 去相關：消除共線性，利於後續建模 (PCR)。
- ✓ 可解釋：Loadings 提供物理意義。
- ✓ 運算快：基於 SVD 矩陣分解。

## ✗ 限制 (Cons)

- ✗ 線性限制：僅捕捉線性關係 (Linearity Assumption)。
- ✗ 動態遺失：標準 PCA 不考慮時間序列 (需用 Dynamic PCA)。
- ✗ 離群值敏感：異常值會扭曲主成分方向。



### 關鍵檢查清單 (Checklist) :

1. 訓練數據必須是「正常操作數據」。
2. \*\*Garbage In, Garbage Out\*\*：必須進行標準化 (Standardization) !

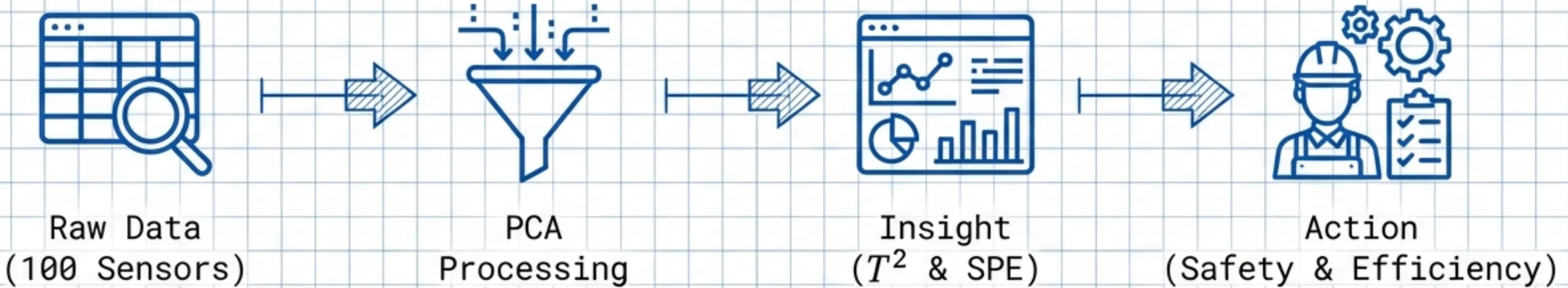
# 實作指南：Python Scikit-Learn

```
from sklearn.preprocessing import StandardScaler  
from sklearn.decomposition import PCA  
  
# 1. 前處理（標準化是關鍵！）  
scaler = StandardScaler()  
X_scaled = scaler.fit_transform(X_raw)  
  
# 2. 建模  
# n_components=0.90 表示保留 90% 變異數  
pca = PCA(n_components=0.90)  
X_pca = pca.fit_transform(X_scaled)  
  
# 3. 分析結果  
print(pca.explained_variance_ratio_) # 解釋比例  
print(pca.components_) # Loadings (載荷)
```



**Important:**  
**fit** 只能用在訓練集  
(Training Set) !

# 結語：您是數據的定義者



## 下一步 (Next Step)：

1. 開啟 Unit06\_PCA.ipynb 進行實作。
2. 進階主題：非線性降維 (Kernel PCA) 與動態 PCA (DPCA)。

PCA 是化工數據分析的基石，將抽象的矩陣運算轉化為可操作的監控指標。