PCA and predictive maintenance case study - machine learning 101

Albert Hsu

At ADLINK, We CARE

ADLINK TECHNOLOGY INC.

What is machine learning?

- “Machine learning teaches computers to do what comes naturally to humans and animals: learn from experience.”
- “Machine learning algorithms use computational methods to learn information directly from data without relying on a predetermined equation as a model.”
- “The algorithms adaptively improve their performance as the number of samples available for learning increase.”

--- Quoted lines from Matlab seminar

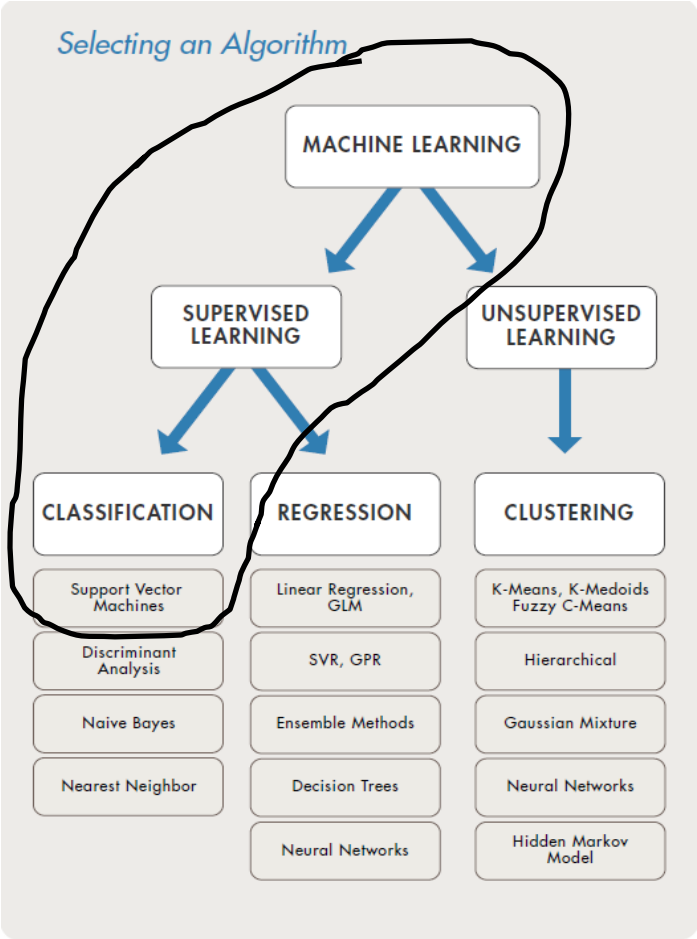
Selecting an algorithm

- **Supervised learning:**

  Learning from the know label data to create a model then predicting target class for the given input data.

- **Unsupervised learning:**

  Learning from the unlabeled data to differentiating the given input data.

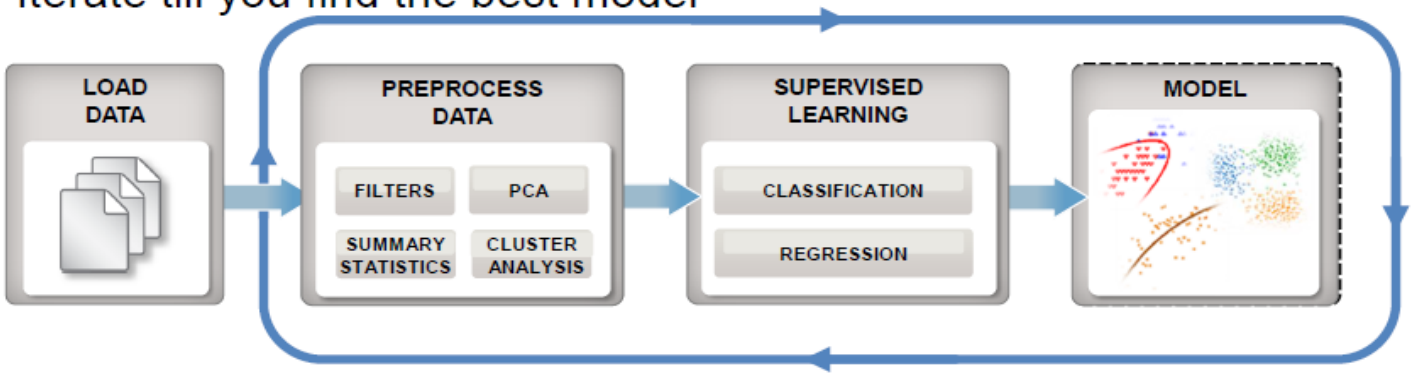


Work flow we use

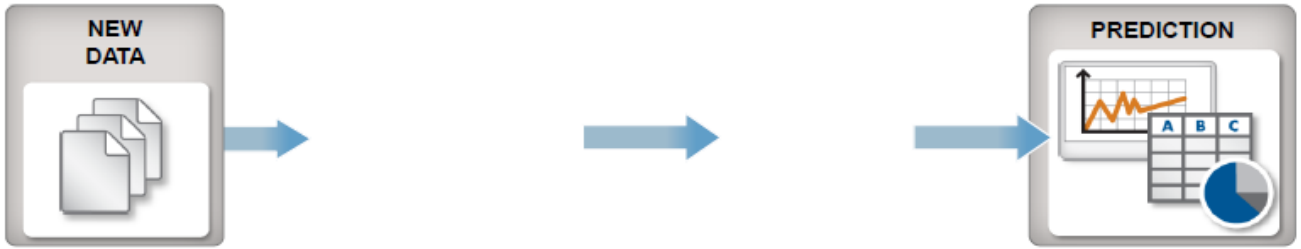
MathWorks

Machine Learning Workflow

Train: Iterate till you find the best model



Predict: Integrate trained models into applications



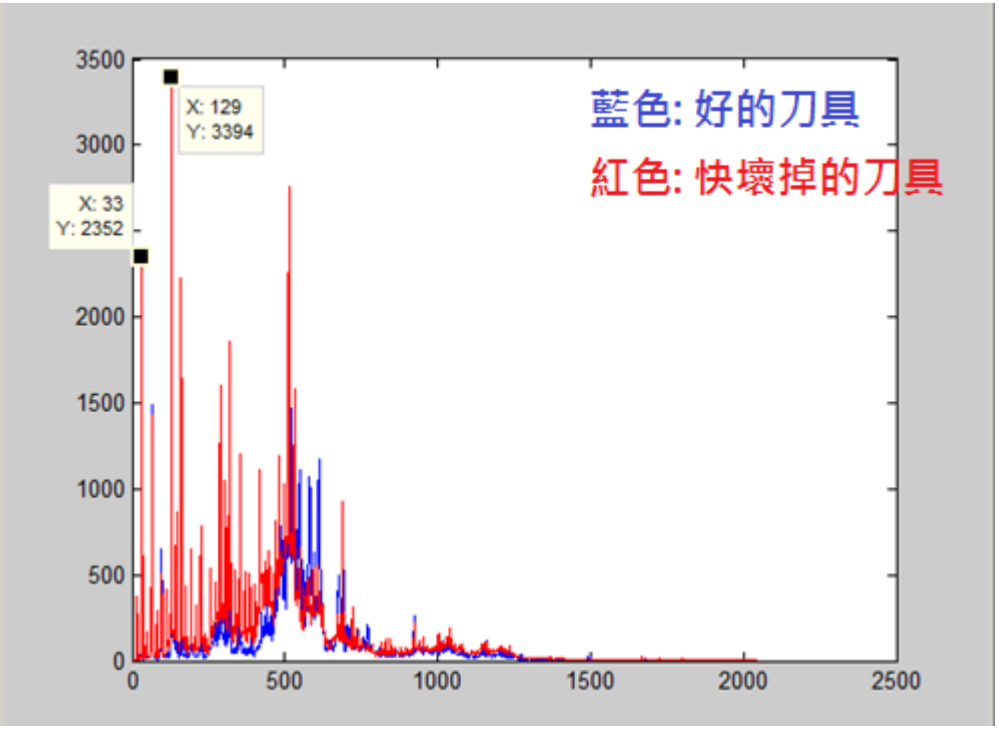
ADLINK TECHNOLOGY INC.

Case 1: 富士康刀具

- 鄭州富士康希望能用USB-2405 + 加速規，藉由振動分析監測車床上刀具的損耗情況。
- 大家第一時間想到的解法: FFT，從頻譜找出正常與即將損壞的差異 (try to find a equation)
- USB-2405 取樣率設為25.6k Hz，擷取4096點 (頻譜解析度 = 6.25Hz)。開始觀察>>>

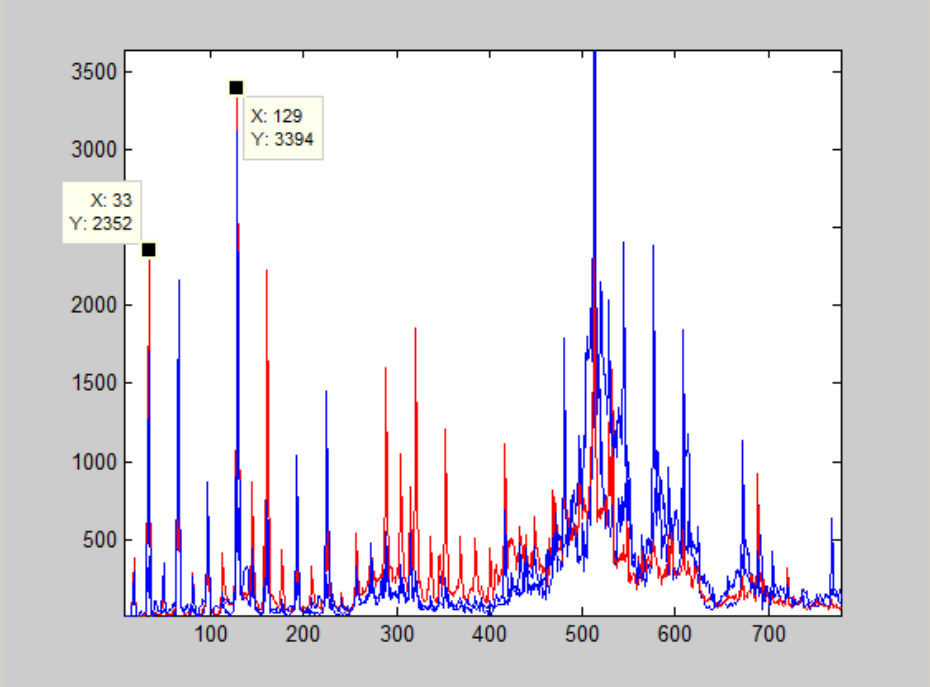
Try to find a predetermined equation

- 先疊加一個OK & almost broken的spectrum data，看起來200Hz & 800Hz是好壞的判定頻點(?)



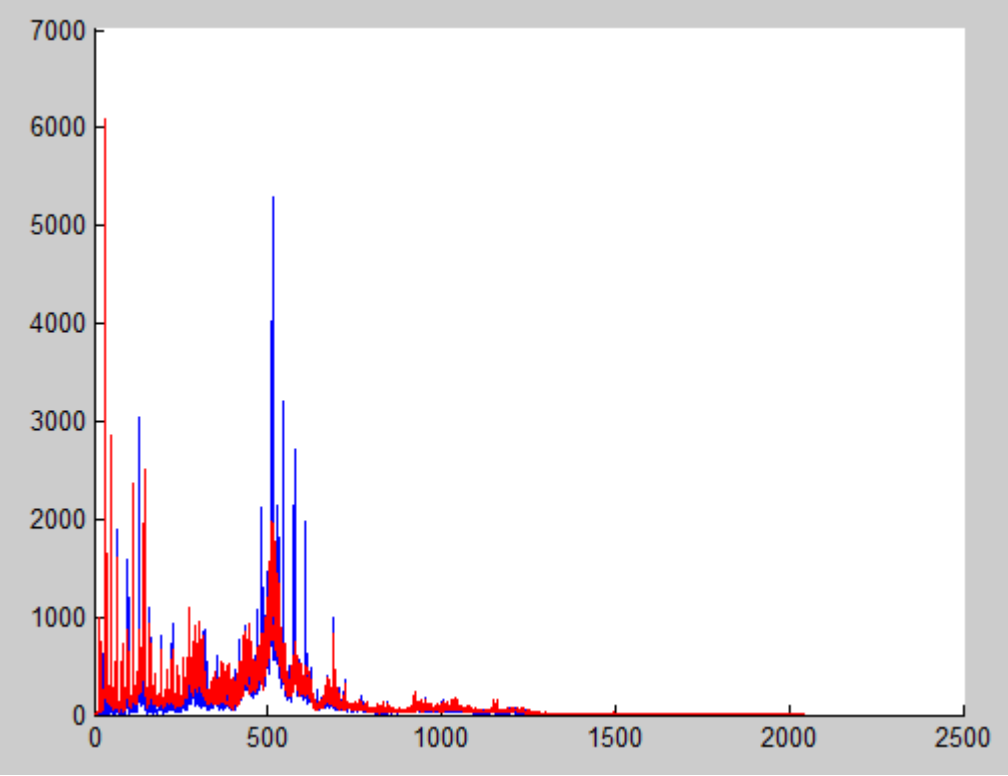
Try to find a predetermined equation

- 多疊加一組正常刀具的頻譜資料上去
- #馬上破功



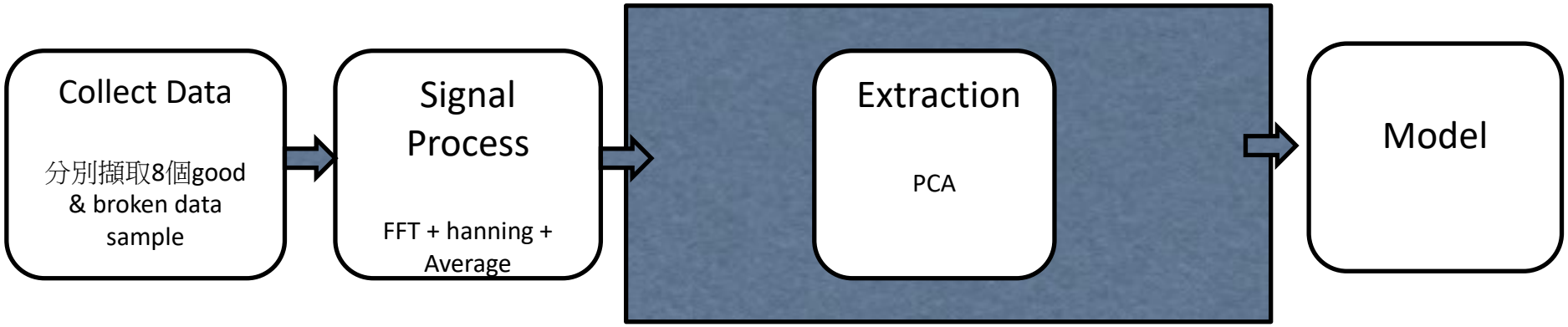
Try to find a predetermined equation

- Overlap all good and broken spectrum
- I am lost. 多達2048個頻譜點要觀察，超過人類能力所及。



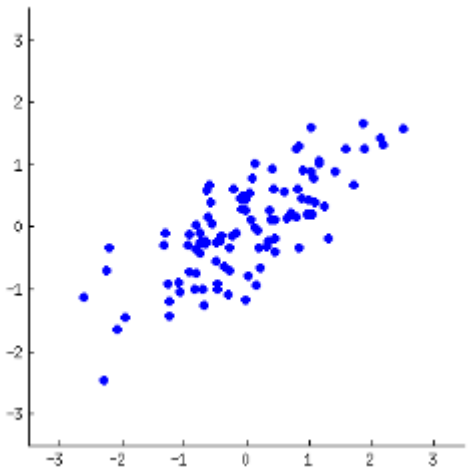
Basic machine learning can solve it

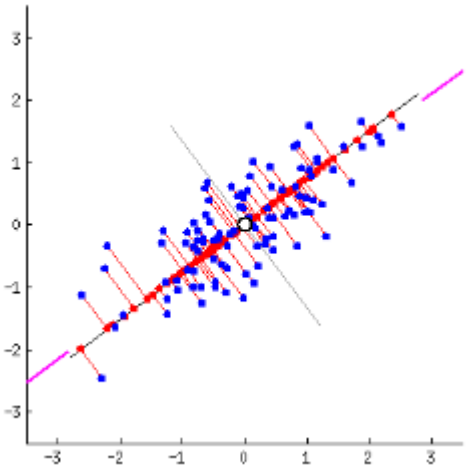
- Train(build model):



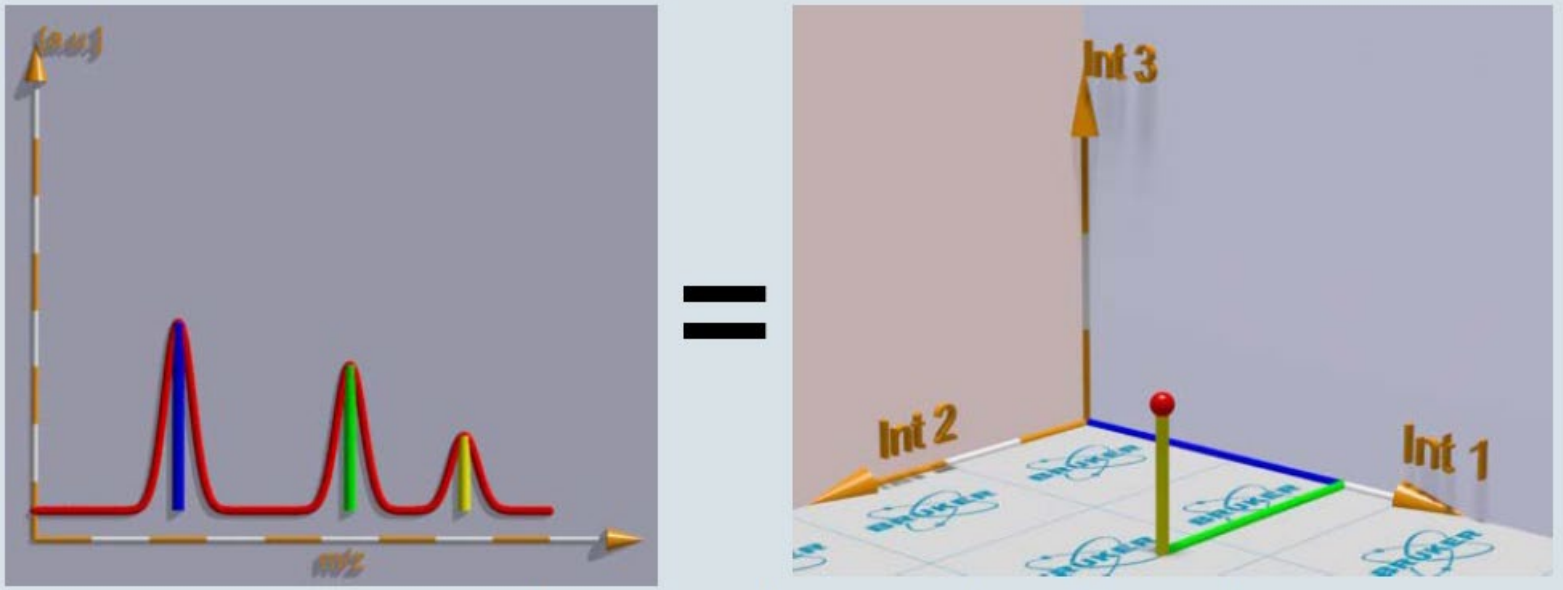
What is PCA

- PCA是一種減少維度(濃縮特徵)的數學計算，以利於做**資料分類**。
- EX: 假如有一群跑者，沿著45度方向的道路前進，可是觀察員2D感知能力極差，無法分辨跑者位置。是否能找到一個1D座標系統來替代?
- 跑者座標 => PCA演算法，計算出**(0.5, 0.5)**的投影算式。
- 2D跑者座標，**(x, y) * (0.5, 0.5)'** = 新的1d座標位置(最佳的座標軸)。





N個頻點 = N dimensions



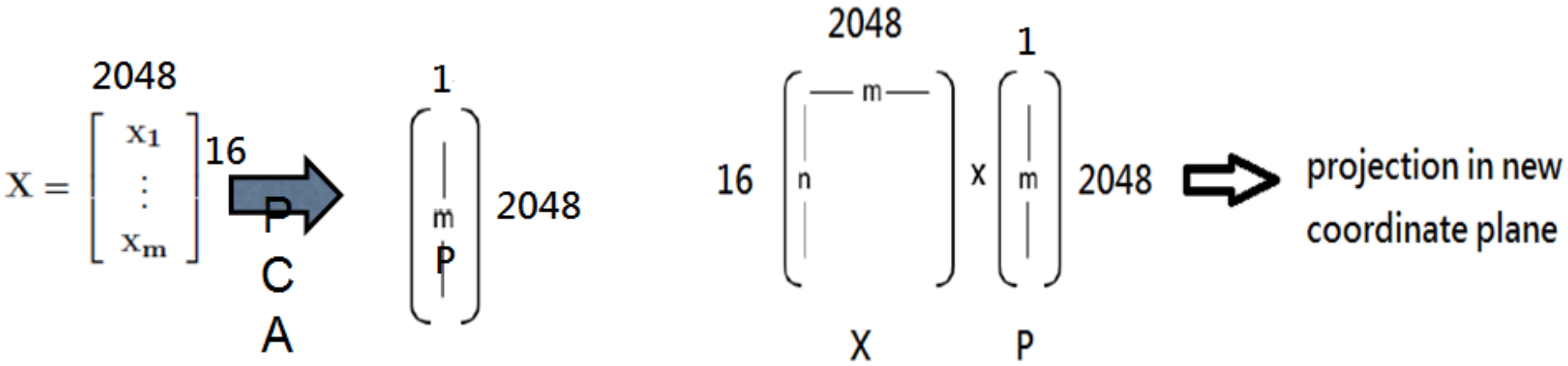
A spectrum with n peaks can be plotted in a n-dimensional space.
The two pictures are equivalent

Bruker Daltonics

再回到富士康刀具

- 2048個頻點，相當於有2048個維度要觀察。很難找到一組spectrum 頻段來當判斷點。所以用PCA減少維度。

- x1 = [x1(f0) x1(f1) ….. x1(2047)] , 把8筆OK以及8筆broken 頻譜資料放到一個 16 *2048的大矩陣X中。透過PCA計算，得到一個1*2048的投影矩陣P

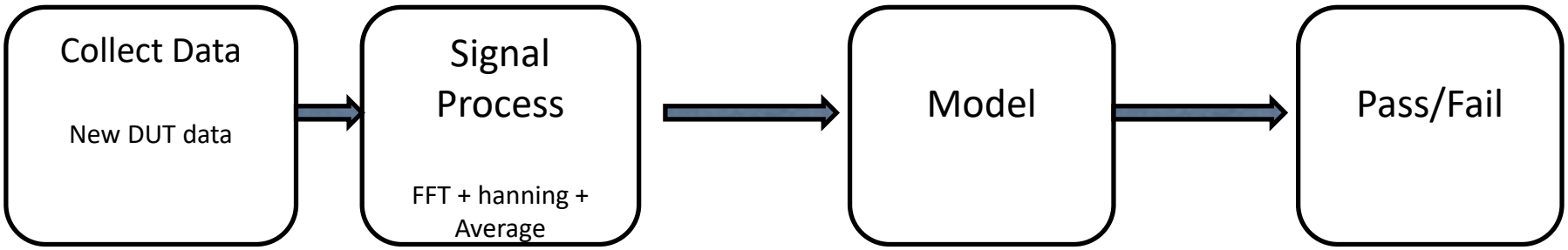


In new 1d plane

Good	Broken
-0.84	0.78
-0.77	0.86
-0.84	0.86
-0.35	0.79
-0.78	0.84
-0.79	0.89
-0.75	0.78

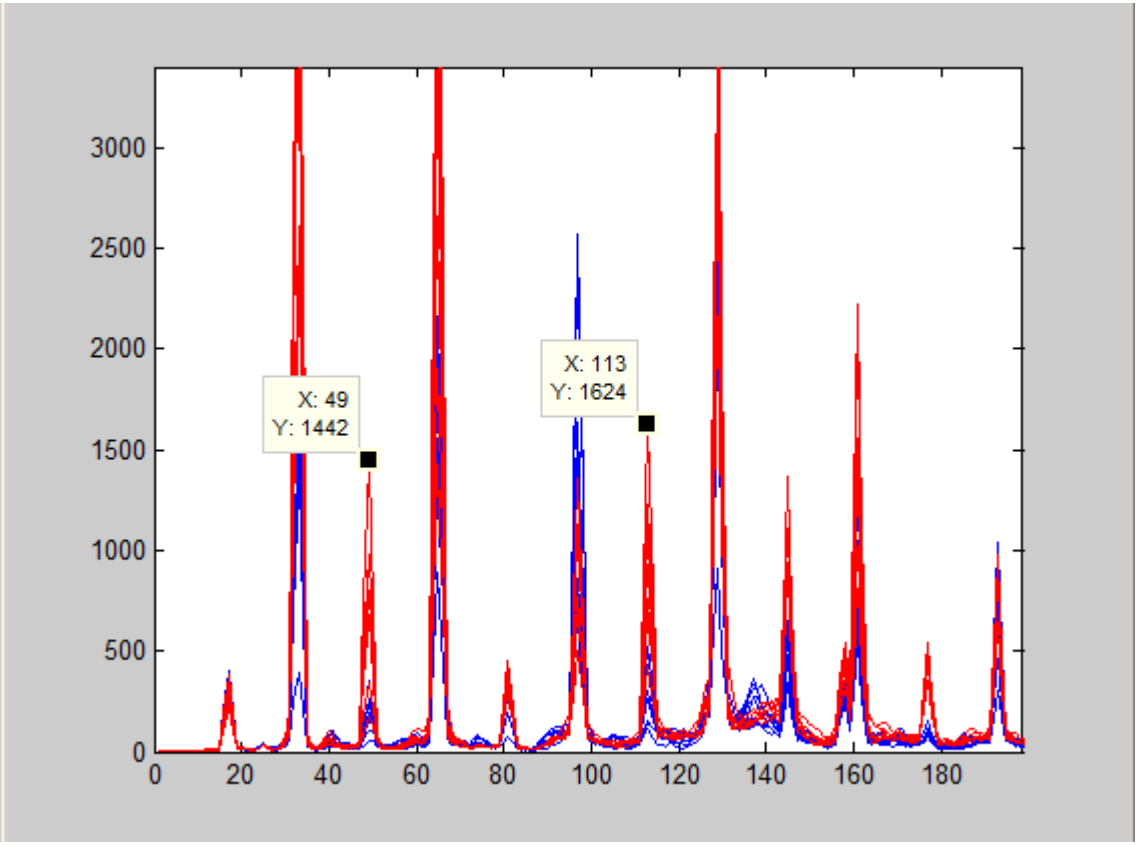
Test

- 以本案為例，train完的P矩陣，以後可重複使用在同種刀具上。
- 新的待測刀具來，一樣做FFT等處理，接著乘上矩陣P作投影。
- 根據之前的經驗，投影到接近-0.7就是OK；0.7左右就是broken。



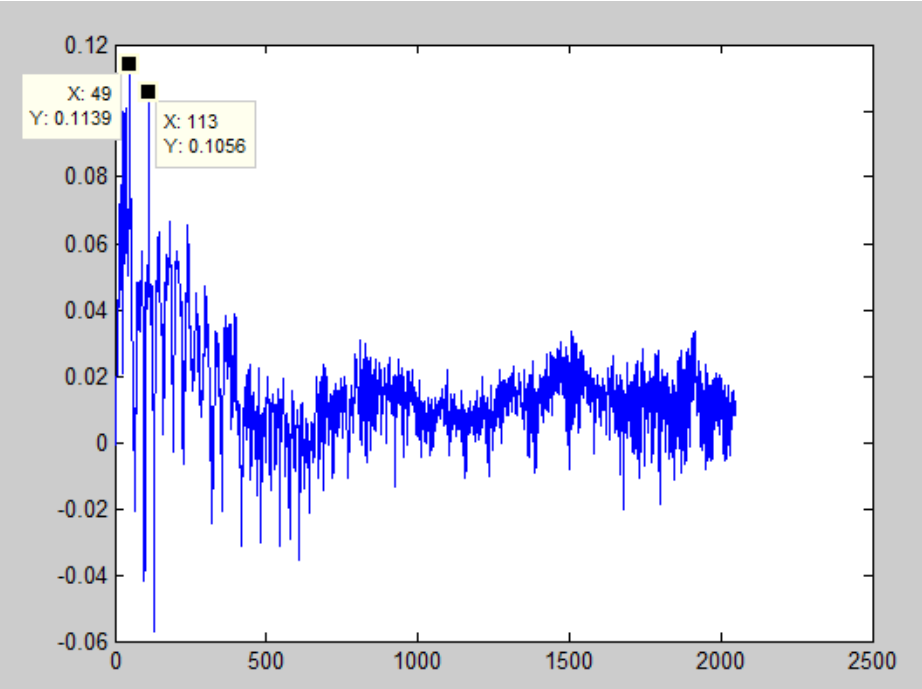
覺得PCA還是很虛無飄邈嗎

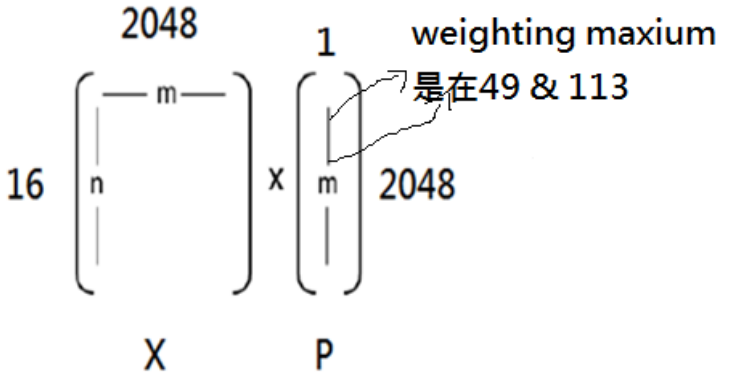
- 有經驗的SI告訴我們，觀察轉速3倍 5倍 or 7倍頻的頻點。



Plot out P matrix

- PCA training後的model之判斷，其實跟有經驗的SI一樣。



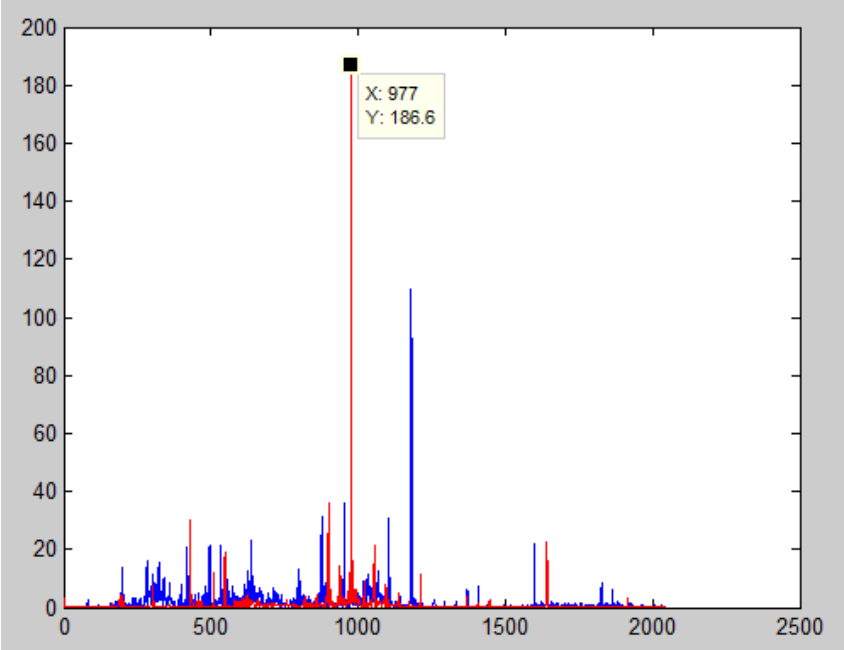


Case 2: DJI(大疆)

- 現況: 大疆產線，用人力監聽馬達震動的聲音，判斷馬達是否正常。
- 目標: 使用9527 or 2405錄取聲音&自動化判斷，取代人力並減少人耳失誤。
- 大家第一時間想到的解法，FFT，從頻譜找出正常與異常的差異。
- 聽個wave檔先。

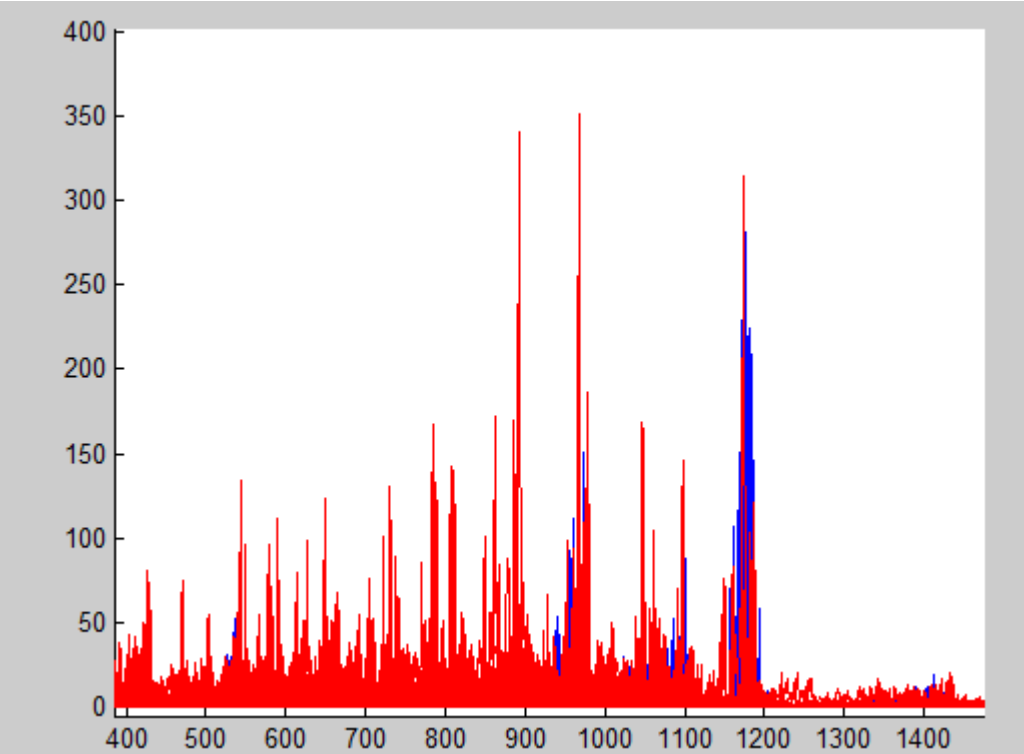
If rely on a predetermined equation

- Overlap one spectrum of good and fail respectively.
- It seems that the spur in x:977 is our target. Is it?



If rely on a predetermined equation

- When we overlap more spectrum data together ….
- Not the case!

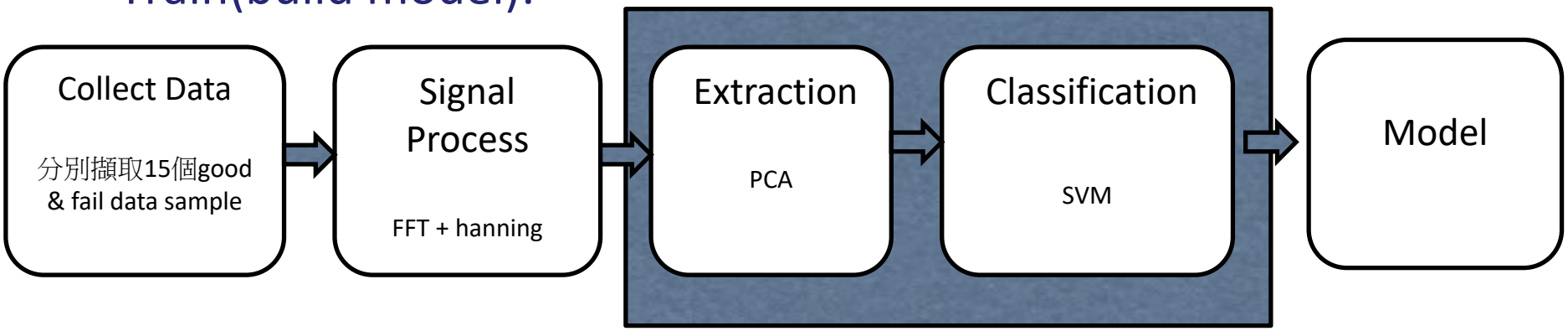


If rely on a predetermined equation

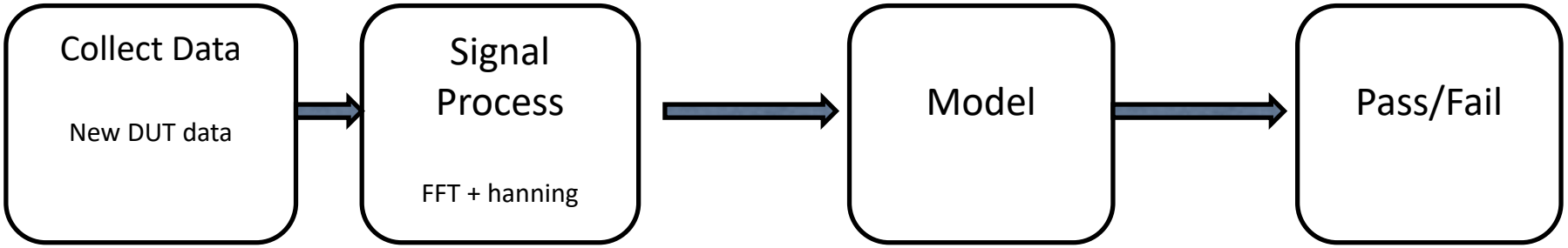
- 這個case更複雜。每個聲音聽起來都不大一樣。找不到專家對這種小馬達有經驗。
- SI商X奇嘗試在頻譜上找幾個點來判斷。這樣做的判斷成功率，只有70%。
- 即使真的找到了一組判斷頻點，以後若有新的產品，這個痛苦的觀察流程又要再來一次。

Use machine learning

- Train(build model):

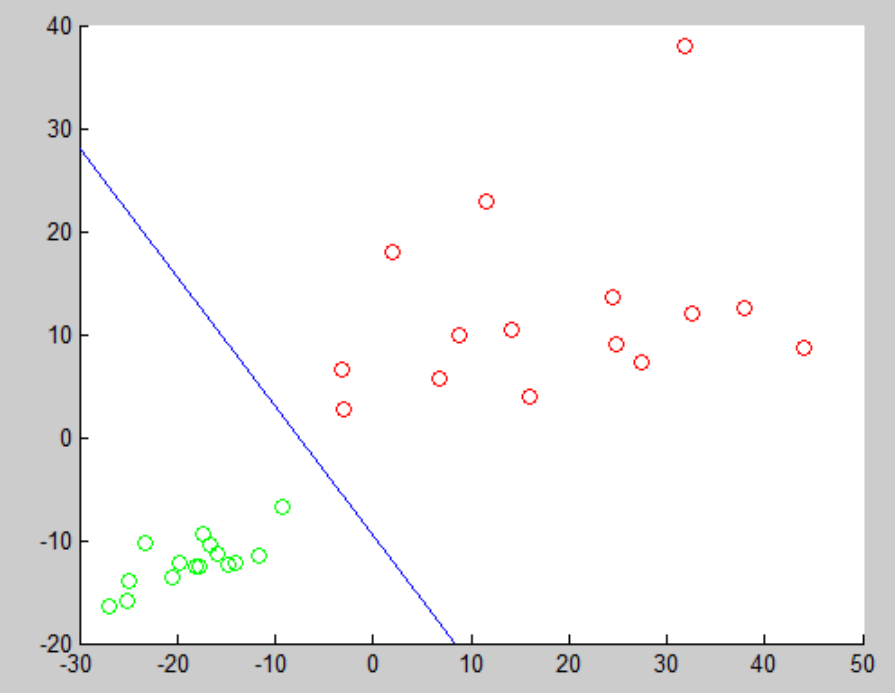


- Test



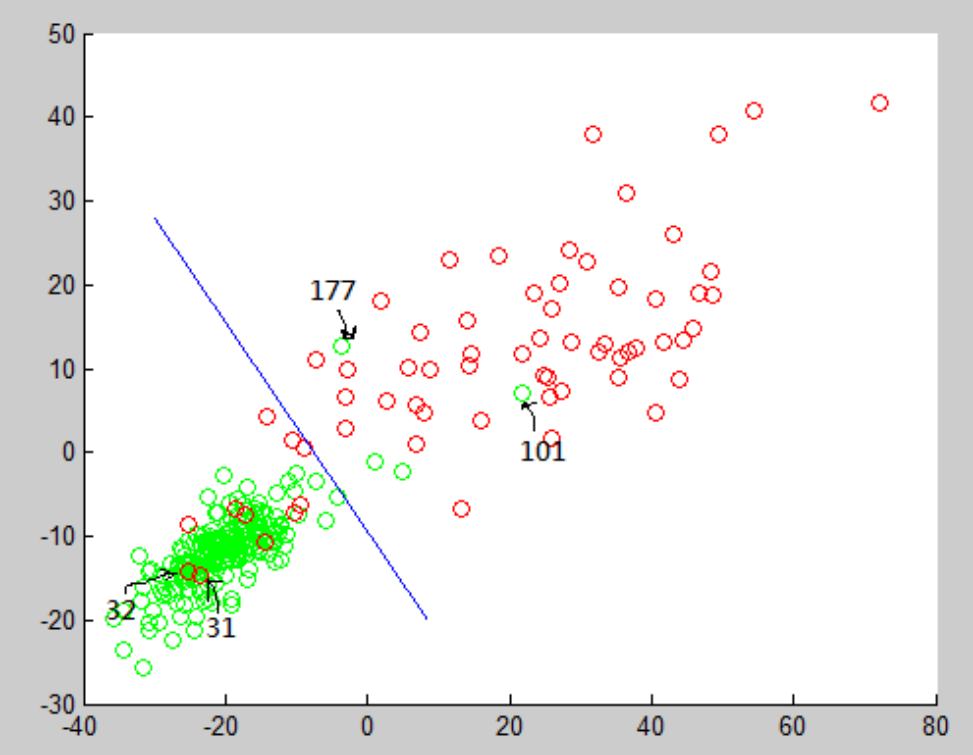
Modeling

- Pass data are plotted in green
- Fail data are plotted in red
- Determine the line for classification (SVM algorithm)



Test Result

- Next, project all DUT data to the new axis based on pre build model:
- Not too well. Let's hear those wave files.



Test Result

- 排除那些明顯人為誤判的wave files後的準確度:

- Accuracy of pass DUT: 99% (199/201)
- Accuracy of fail DUT: 85% (54/63)

- Not too shabby! 樂勝X奇。

How to improve

- “The algorithms adaptively improve their performance as the number of samples available for learning increase.”

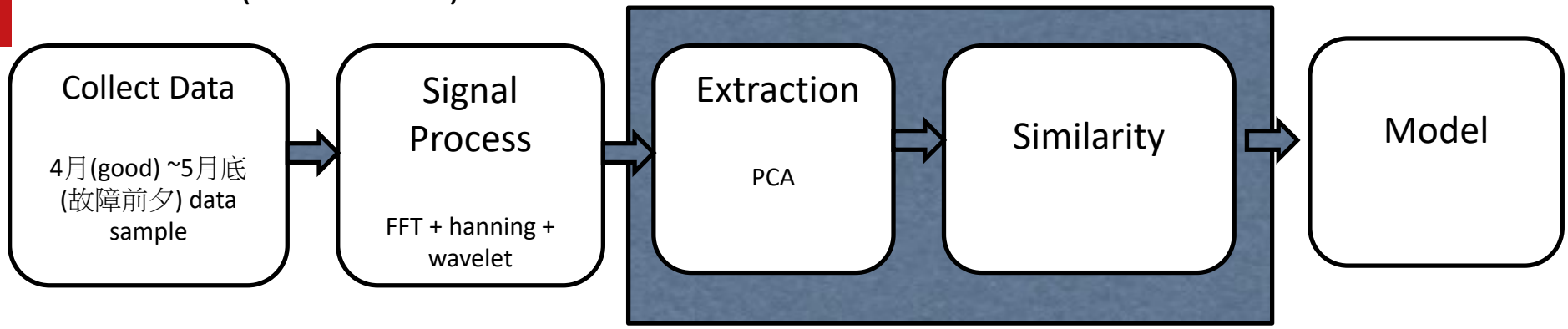
- 如果能請DJI的老師傅，再多提供一些broken motor的資料&分類，這個測試model會更強健。

Case 3: Adlink Compressor

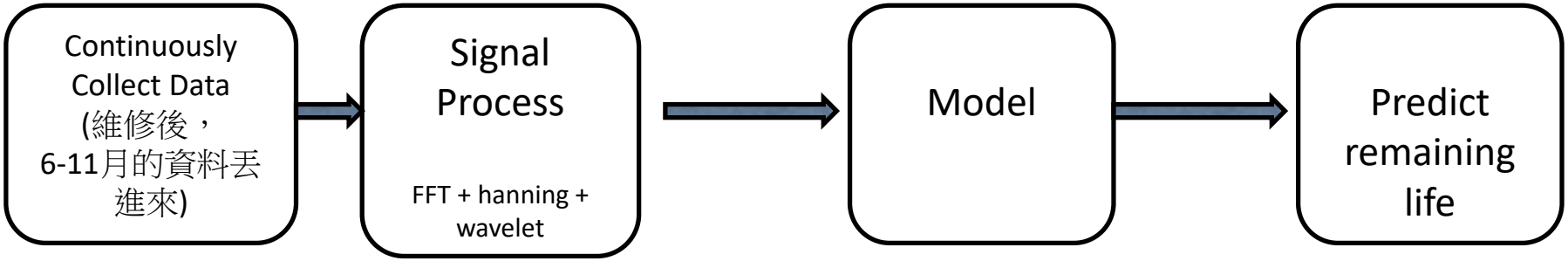
- 2017/04 佈置了一套USB-2405在2F產線的空壓機上，紀錄聲音與震動來判斷空壓機是否正常。
- 運氣很好，2017/05 月底空壓機就故障了。6月初第一次維修完畢。
- 運氣更好的是，維修過後到2017/11再次故障，而且兩次故障的原因都相同(這表示2017/06 ~ 11的資料，正好拿來驗證!)。
- 雖然故障後，震動Vrms值會暴增 => 軟體發警示；但是據產線人員所說，兩次故障發生前幾週，空壓機運作的聲音就不大對了。
- 有無可能分析聲音資料，提前預測故障的發生(prediction maintain)?

Prediction maintain

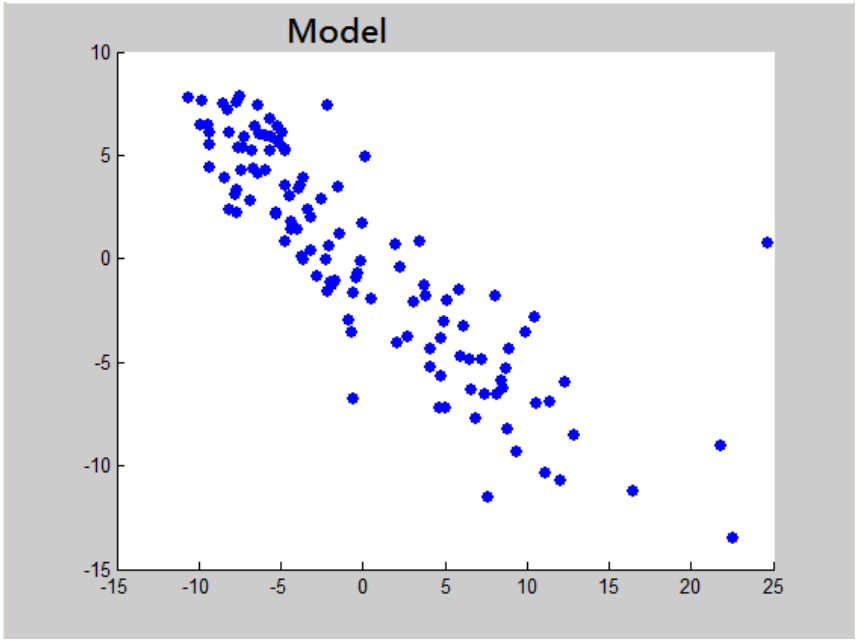
- Prediction maintain其實也是用類似的方法
- Train (Build Model)

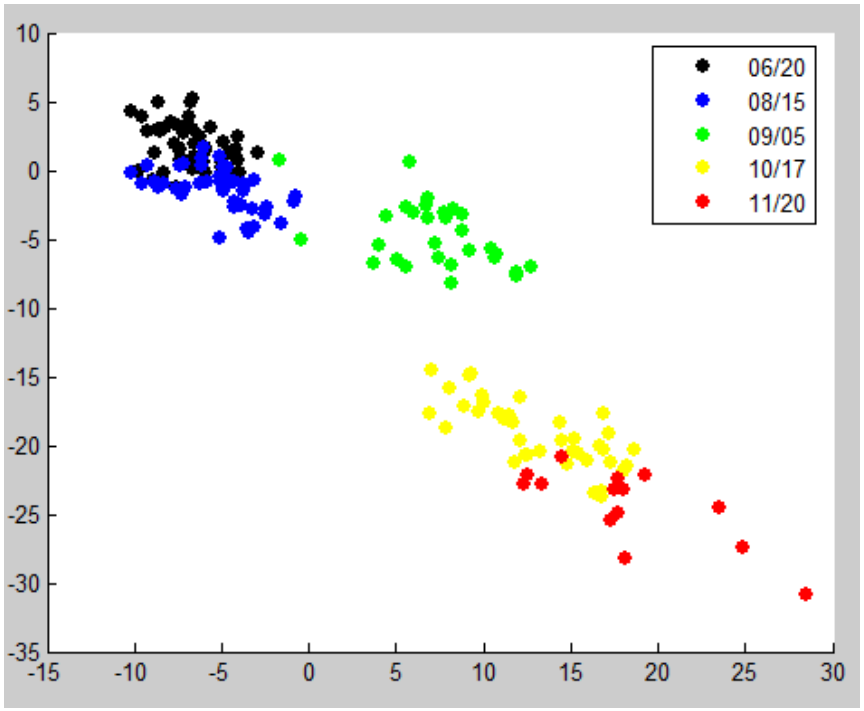


- Test



Result





Thanks