降維

『降維』用白話文說就是『減少特徵個數』，
那什麼時候需要減少特徵個數呢？

『保留原本特徵』 - 提高UI/UX體驗

身高	體重	年齡	~~性別~~
…	…	…	~~…~~

統計終究是要使用在生活中的，今天一個**『正確率99%，但需要使用30個問卷問題』**的模型可能是比**『正確率95%，但只使用5個特徵』**的模型差的，所以這時候推薦你可以使用有**『決定特徵重要性』**的模型(e.g. 決策樹)來有效

提高使用者UI/UX體驗

『組合新特徵』 - 解決維度災難

身高	體重
…	…

→

bmi
…

我們說過，如果特徵太多，有上千上萬個，一般的演算法就很難派上用場了，這個就叫做**『維度災難』**，但我們如果直接丟棄特徵，又會丟棄太多資訊，所以我們只好把**『數個特徵組合成一個特徵』**，其實這個**『組合』**的概念就跟我們人腦的神經系統很像！

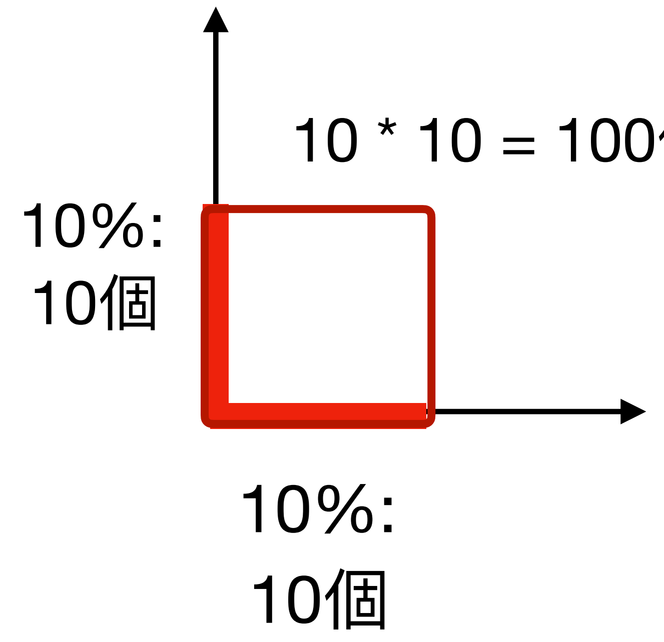
再探維度

維度太高第一個問題是，像『決策樹』這類型的演算法完全無法運作(我們在單純貝式提過這點)，那今天從**『訓練資料需求多寡』**的角度來觀察

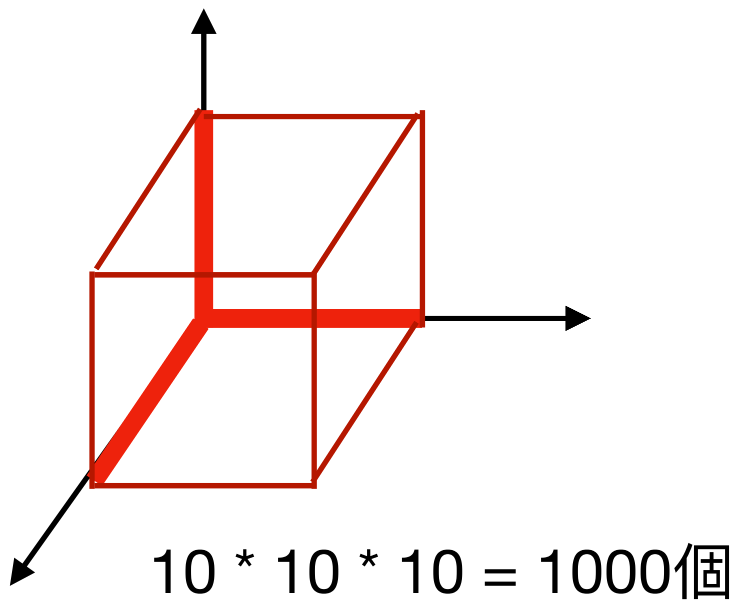
假設對於我們現在的『統計』，我們至少需要總體資料的10%才能有效統計

只有一個特徵

兩個特徵



三個特徵



你會發現，當你的特徵越多(問題越難)，你就需要更多的**『訓練資料』**才有辦法正確統計，而且這個增加的幅度是以『指數增加』，所以我們最後下了一個結論

健康的維度 -> 健康的『演算法』和健康的『訓練資料』

PCA

PCA我們又叫做『主成份分析』(Principal Component Analysis)，主要分做三步驟

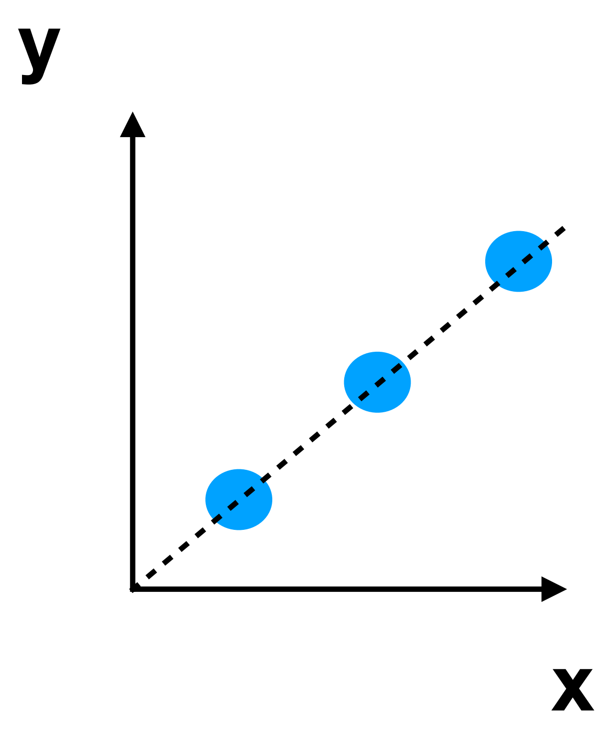
Step 1. 組合
- 把數個欄位組合成一個新的欄位

Step 2. 分解
- 讓新的座標軸互相垂直

Step 3. 篩選
- 篩選出比較重要的欄位

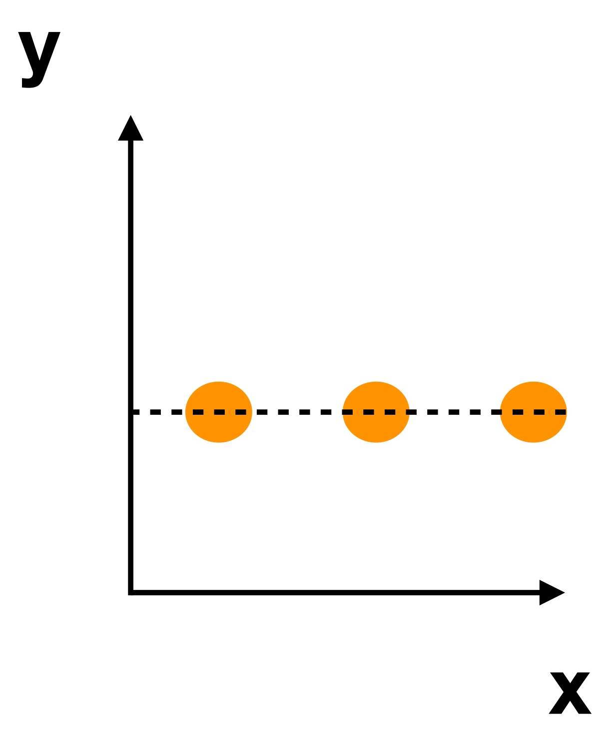
相關係數

相關係數：當一個變數變動的時候，另外一個變數變動的方向



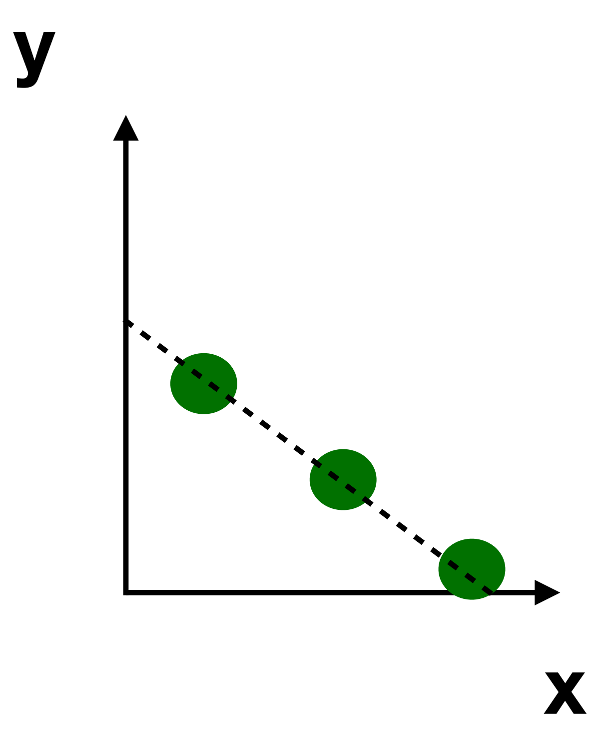
y和x變動同方向

正相關：正相關就像一個很愛你的女孩，你對他越好，他心情越好



x或y不為所動

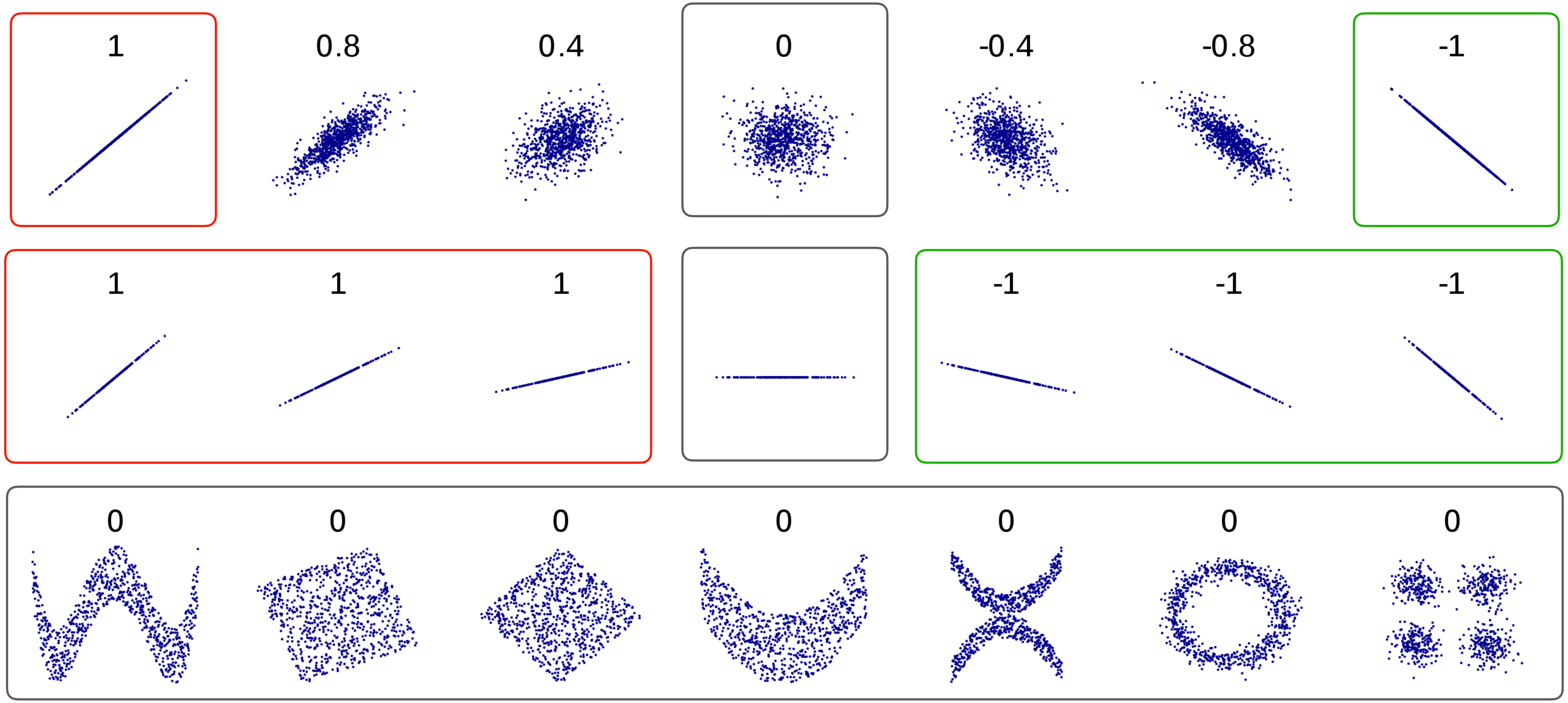
零相關：零相關是一個陌生人，你對他越好越壞，不影響



y和x變動反方向

負相關：負相關是一個不愛你的女孩，你對他越好，他心情越差

相關係數



所以！千萬別說『相關係數是0』就是兩個變數沒有關係，實際上應該說，兩個變數沒有『線性相關』的變動關係

共變異數

相關係數 (Correlation)

$$\rho = \frac{\sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\sum_{i=1}^{n}(x_i - \mu_x)^2}\sqrt{\sum_{i=1}^{n}(y_i - \mu_y)^2}}$$

共變異數 (Covariance)

x標準差 (x分佈多廣)

y標準差 (y分佈多廣)

相關係數跟共變異數的關係是

『相關係數』是Scaling過後的**『共變異數』**，將兩個變數的分佈範圍除掉，使**『相關係數』**只會落在-1~1之間，所以等下的課程中，只要你不懂**『共變異數』**，你都可以用**『相關係數』**來思考

組合

第一步：組合
我們要變少欄位，但是還是要盡量保持原有的資訊，那就只能把欄位的資訊想辦法組合成一個新欄位

這是你原本的一筆訓練資料

$$Data = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix}$$

你說：老師，我完全不懂！
合理！你懂就奇怪了，千萬別硬記這個式子，我們來解析一下

對完整訓練資料算出每一個欄位的『共變異數』

$$Covariance = \begin{bmatrix} Cov(col_1, col_1) & Cov(col_1, col_2) & Cov(col_1, col_3) \\ Cov(col_2, col_1) & Cov(col_2, col_2) & Cov(col_2, col_3) \\ Cov(col_3, col_1) & Cov(col_3, col_2) & Cov(col_3, col_3) \end{bmatrix}$$

你的新資料就是原本資料乘上『共變異數矩陣』

$$NewData = Covariance(Matrix) \times Data^{\top}$$

組合

其實你真的乘開就知道意思了

$$Covariance = \begin{bmatrix} \boxed{Cov(col_1, col_1) \quad Cov(col_1, col_2) \quad Cov(col_1, col_3)} \\ Cov(col_2, col_1) \quad Cov(col_2, col_2) \quad Cov(col_2, col_3) \\ Cov(col_3, col_1) \quad Cov(col_3, col_2) \quad Cov(col_3, col_3) \end{bmatrix}$$

$$Data^{\top} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$NewData = \begin{bmatrix} \boxed{x_1 Cov(1,1) + x_2 Cov(2,1) + x_3 Cov(3,1)} \\ x_1 Cov(1,2) + x_2 Cov(2,2) + x_3 Cov(3,3) \\ x_1 Cov(1,3) + x_2 Cov(2,3) + x_3 Cov(3,3) \end{bmatrix}$$

還不知道？我說過，『共變異數』想不通就用『相關係數』想

$$x_1 Cov(1,1) + x_2 Cov(2,1) + x_3 Cov(3,1)$$

沒錯，第一個新欄位就是把所有其他欄位對於x1的影響全部加進來變成一個新的欄位

同理，每一個欄位都是把『其他欄位的影響』組合進來而已

何謂矩陣

在講接下來的簡單數學前，我們要先聊聊**『矩陣乘法』**

首先，我們要先了解『向量』的『點積』

$$\begin{bmatrix}1 & 2\end{bmatrix} \cdot \begin{bmatrix}3 & 4\end{bmatrix} = 1 \times 3 + 2 \times 4 = 11$$

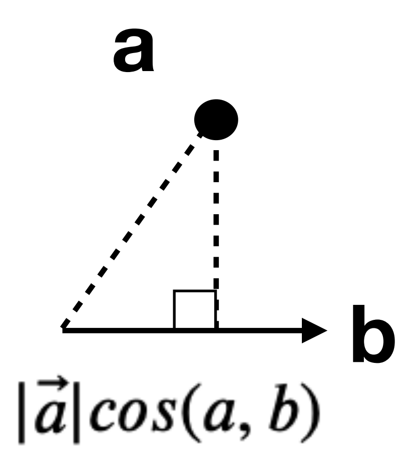
點積也可以寫成這樣的形式

$$\vec{a} \cdot \vec{b} = |\vec{a}||\vec{b}|cos(a, b)$$

簡化一下，假設b的長度是1

$$\vec{a} \cdot \vec{b} = |\vec{a}|cos(a, b)$$

如果你把a看成我們 (1,0) (0, 1)
座標系的一個點



這個式子就是把原本的點投影到另外一個座標軸的時候的長度，換句話說，就是把你原本的點放在另外一個『軸上』時候的座標啦!

$$\begin{bmatrix}1 & 3\\ 2 & 1\end{bmatrix} \times \begin{bmatrix}5\\ 6\end{bmatrix}$$

$$= \begin{bmatrix}(1,3)\cdot(5,6)\\ (2,1)\cdot(5,6)\end{bmatrix} = \begin{bmatrix}23\\ 16\end{bmatrix}$$

(5, 6)就是你的資料點，(1, 3),
(2,1)就是你投影的兩個座標軸

分解

$$x_1Cov(1,1) + x_2Cov(2,1) + x_3Cov(3,1)$$

但是你會發現，其實我們根本沒有減少欄位個數，只是做出一樣多的新欄位而已，那如何減少呢？

在講解決方法前，我們要先來聊聊『對稱矩陣』

如果有一個矩陣，他有以下的性質

$$A(matrix) = A(matrix)^T$$

轉置以後還是原本的矩陣，我們就叫做『對稱矩陣』

$$Covariance = \begin{bmatrix} Cov(col_1, col_1) & Cov(col_1, col_2) & Cov(col_1, col_3) \\ Cov(col_2, col_1) & Cov(col_2, col_2) & Cov(col_2, col_3) \\ Cov(col_3, col_1) & Cov(col_3, col_2) & Cov(col_3, col_3) \end{bmatrix}$$

這很明顯是一個對稱矩陣，『上三角』和『下三角』是一樣的值

分解

特徵值分解(別怕，就只是矩陣版本的因式分解的感覺)告訴我：
一個『對稱矩陣』A 一定可以被分成三個『矩陣』相乘

$$A = Q^T \times \Lambda \times Q$$

Q：正交(且單位)矩陣

正交的意思，就是不管把直的當『行向量』或者是把橫的當『列向量』都互相垂直，單位向量指的是距離原點1

垂直 垂直

$$\begin{bmatrix} vector_1 & vector_2 & vector_3 \end{bmatrix}$$

垂直

垂直

$$\begin{bmatrix} vector_1 \\ vector_2 \\ vector_3 \end{bmatrix}$$

垂直

垂直

Λ：對角矩陣

『上三角』和『下三角』都為0的矩陣

$$\begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}$$

對角矩陣有一個重要的性質

$$\begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix} \times \begin{bmatrix} vector_1 \\ vector_2 \\ vector_3 \end{bmatrix} = \begin{bmatrix} a \times vector_1 \\ b \times vector_2 \\ c \times vector_3 \end{bmatrix}$$

不信你試試看

$$\begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix} \times \begin{bmatrix} 1 & 2 \\ 5 & 6 \end{bmatrix} =$$

$$\begin{bmatrix} (3 \times 1 + 0 \times 5) & (3 \times 2 + 0 \times 6) \\ (0 \times 1 + 4 \times 5) & (0 \times 2 + 4 \times 6) \end{bmatrix} = \begin{bmatrix} 3 \times (1, 2) \\ 4 \times (5, 6) \end{bmatrix}$$

分解

於是我們的式子從，把你的資料放在 Covariance 座標軸上

$$NewData = Covariance(Matrix) \times Data^{\top}$$

因為Covariance矩陣是『對稱矩陣』，分解後

$$(Q_{cov}^{T} \times \Lambda_{cov} \times Q_{cov})$$

重點來了！我只拿出 Q_{cov} 來投影

為什麼呢？別忘記喔！要是『單位向量』(長度1)才是投影，如果長度不為1會有縮放效果，那要是『正交向量』才不會兩個軸具有『重疊的意義』(互相包含)，而我們的 Q 剛好都符合這兩個條件

分解最重要的就是把原本的組合座標軸，分解出『垂直單位座標軸』
於是我們本來的『式子』就變成

$$NewData = Q_{cov} \times Data^{\top}$$

篩選

目前我們的新特徵並沒有減少，但你仔細想想

$$\begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix} \times \begin{bmatrix} vector_1 \\ vector_2 \\ vector_3 \end{bmatrix} = \begin{bmatrix} a \times vector_1 \\ b \times vector_2 \\ c \times vector_3 \end{bmatrix}$$

如果 c ~> 0，那你整個 c * vector 基本就是 [0, 0, ……]

所以我們的最後一步！就是去掉影響相對小的軸

$$\begin{bmatrix} \boxed{\begin{matrix} a & 0 \\ 0 & b \end{matrix}} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 \end{matrix} & c \end{bmatrix} \times \begin{bmatrix} \boxed{\begin{matrix} vector_1 \\ vector_2 \end{matrix}} \\ vector_3 \end{bmatrix} = \begin{bmatrix} \boxed{\begin{matrix} a \times vector_1 \\ b \times vector_2 \end{matrix}} \\ c \times vector_3 \end{bmatrix}$$

那我們最後的新資料就會像是這樣

$$\begin{bmatrix} vector_1 \\ vector_2 \end{bmatrix} \times \begin{bmatrix} Col_1(old) \\ Col_2(old) \\ Col_3(old) \end{bmatrix} = \begin{bmatrix} Col_1(new) \\ Col_2(new) \end{bmatrix}$$

只剩兩個欄位！大功告成，確實減少了欄位，避免『欄位太多』(維度災難)

流程圖

1-2-3, PCA好簡單，跟我一起默念口訣

『組合』 - 『分解』 - 『篩選』