

讀範例資料

load hald

>> ans=ingredients

ans =

7	26	6	60
1	29	15	52
11	56	8	20
11	31	8	47
7	52	6	33
11	55	9	22
3	71	17	6
1	31	22	44
2	54	18	22
21	47	4	26
1	40	23	34
11	66	9	12
10	68	8	12

[coeff,score]=pca(ingredients);

>> coeff

coeff =

-0.0678	-0.6460	0.5673	0.5062
-0.6785	-0.0200	-0.5440	0.4933
0.0290	0.7553	0.4036	0.5156
0.7309	-0.1085	-0.4684	0.4844

若 svd 分解 $A=U*\Sigma*V$

coeff 便是cov(ingredients) 矩陣 做svd 的 U 矩陣

* 這裡 ingredients 有先減掉平均

奇異值矩陣 Σ =

517.7969	0	0	0
0	67.4964	0	0
0	0	12.4054	0
0	0	0	0.2372

最大的517則對應到U 第一列 column vector 也就是最大的 **Principal components** pc1

-0.0677

-0.6785

0.0290

0.7308

67.49 則對應 第二列 column vector 以此類推
其中U矩陣(coeff) 第一行可視為原是維度對應到第一行的相關係數

```
coeff = 4x4
```

-0.0678	-0.6460	0.5673	0.5062
-0.6785	-0.0200	-0.5440	0.4933
0.0290	0.7553	0.4036	0.5156
0.7309	-0.1085	-0.4684	0.4844

原本4X4 cov(ingredients) 矩陣 變為 4X4 U矩陣
用學科資料來說 便是原本4個科目 對應到第一 **components** 相關係數

前面的 score 便是 原始ingredients 透過相關係數project on pc1
score= (ingredients-mean(ingredients))*coeff

```
>> score
```

```
score =
```

36.8218	-6.8709	-4.5909	0.3967
29.6073	4.6109	-2.2476	-0.3958
-12.9818	-4.2049	0.9022	-1.1261
23.7147	-6.6341	1.8547	-0.3786
-0.5532	-4.4617	-6.0874	0.1424
-10.8125	-3.6466	0.9130	-0.1350
-32.5882	8.9798	-1.6063	0.0818
22.6064	10.7259	3.2365	0.3243
-9.2626	8.9854	-0.0169	-0.5437
-3.2840	-14.1573	7.0465	0.3405
9.2200	12.3861	3.4283	0.4352
-25.5849	-2.7817	-0.3867	0.4468
-26.9032	-2.9310	-2.4455	0.4116

36.8 29.6 那一行便是 ingredients 第一列 * pc1 所得到
可以比喻成是學生投影到pc1的分數

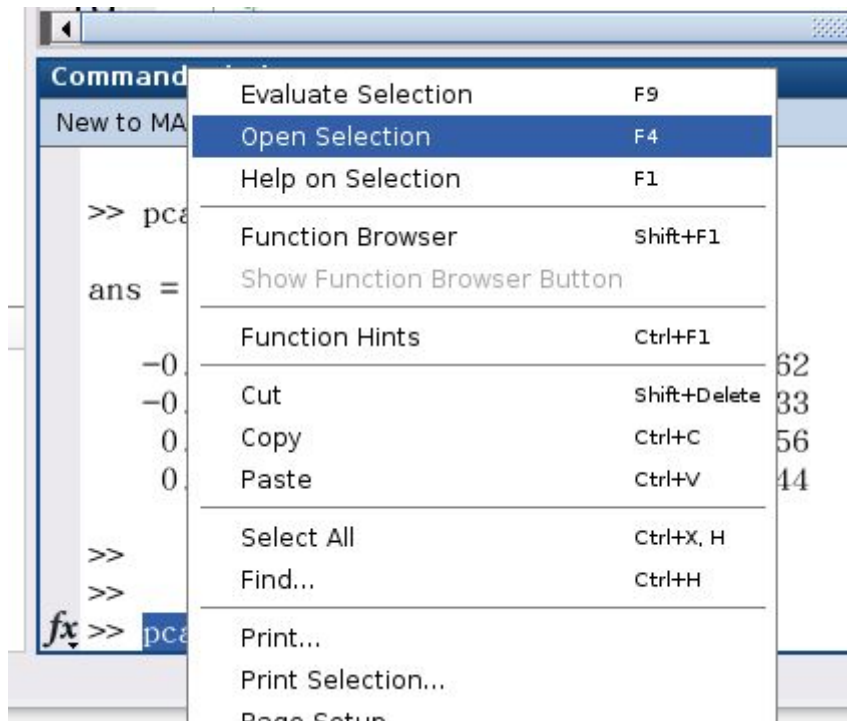
至於還原資料則可以透過

$$\text{ingredients} = \text{score} * \text{coeff}^T + \text{mean}(\text{ingredients})$$

視覺化可以透過 coeff 當軸, score 當點去呈現

可參考<https://www.mathworks.com/help/stats/biplot.html>

matlab 可以對pca函式滑鼠按右鍵open 看原始碼
或其函式也可以 有興趣可以直接參考



另外有找到一篇用octave實做的教學 可以了解看看

https://www.bytefish.de/blog/pca_lda_with_gnu_octave/

或是可以嘗試另一個 Factor analysis 因素分析方法

<https://www.mathworks.com/help/stats/factoran.html>

可以使dimension間不一定要獨立