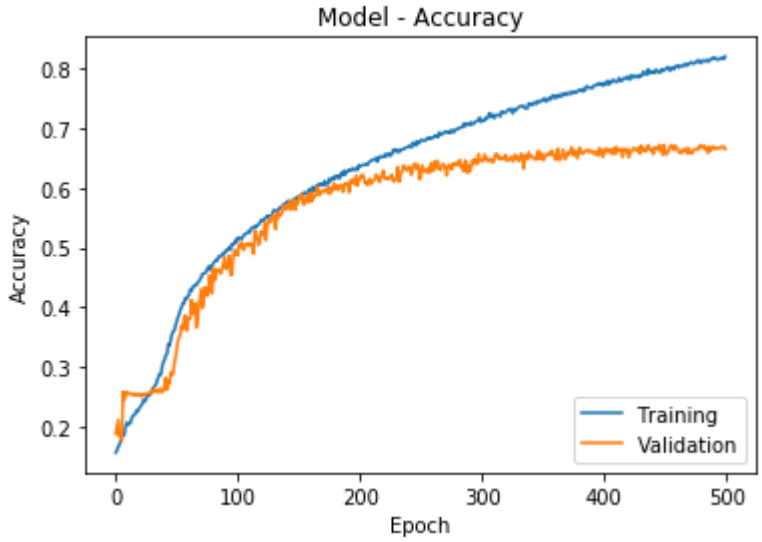
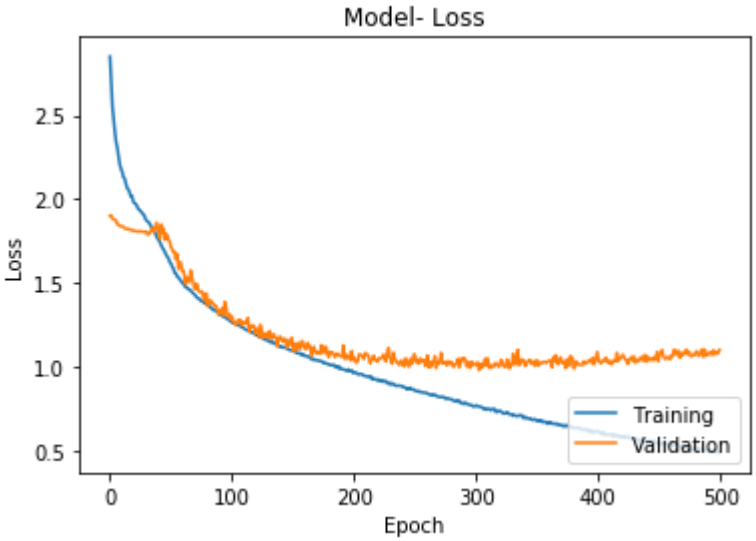
Homework3 Report

Professor Pei-Yuan Wu
EE5184 - Machine Learning

姓名：張廷維
學號：R07921078

1. (1%) 請說明你實作的 CNN model，其模型架構、訓練過程和準確率為何？
   A. 參考 AlexNet，5 層 convolution、3 層 fully connected
   B. 使用 BatchNormalization、Dropout = 0.5、ReLu
   C. Filter size 64/128/256/512/512，kernel size = (3, 3)
   D. 訓練過程:





由圖表可以發現 epoch 超過 200 後有明顯的 overfitting

   E. Batch size = 512、epochs = 500
   F. Accuracy/Loss (train): 0.4989/0.8186
   G. Accuracy/Loss (validation): 1.1217/0.6531

H. Accuracy (kaggle): 0.65143

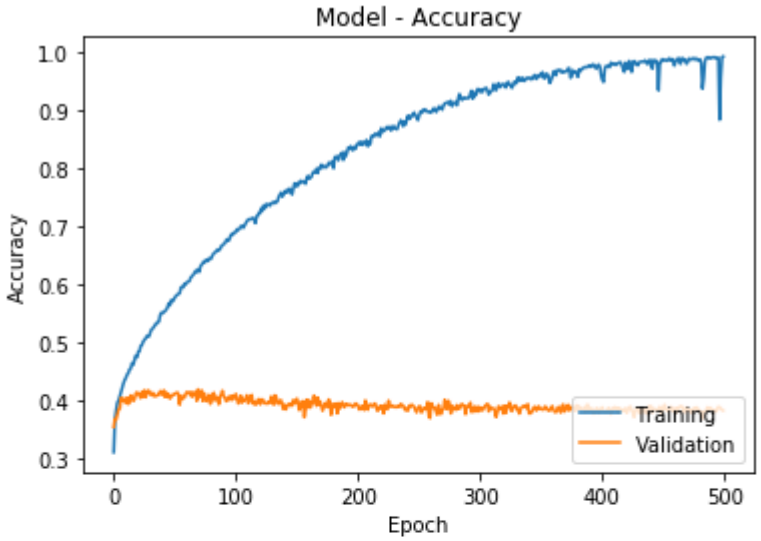
2. (1%) 承上題，請用與上述 CNN 接近的參數量，實做簡單的 DNN model，其模型架構、訓練過程和準確率為何？試與上題結果做比較，並說明你觀察到了什麼？

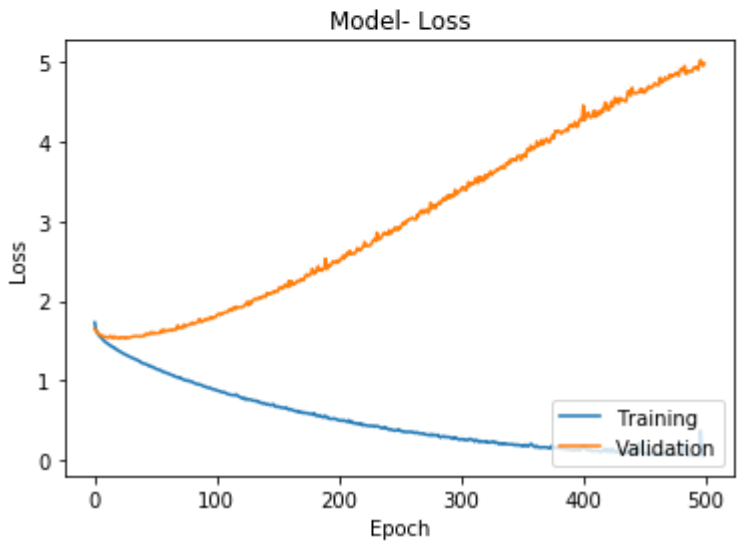
A. CNN 參數量: 4,448,519

B. DNN 參數量: 4,460,991

C. 模型架構:盡量將參數湊比較接近，input 48*48, 第一/二層 output 皆 44，最後一個 softmax 7

D. 訓練過程





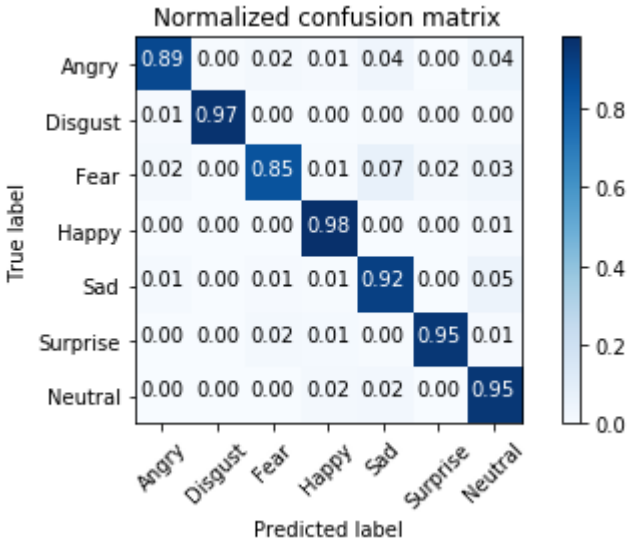
可以發現 overfitting 非常嚴重

E. Accuracy/Loss (train): 0.9925/0.0582

F. Accuracy/Loss (validation): 0.3824/4.9773

G. Accuracy (kaggle): 0.36806

3. (1%) 觀察答錯的圖片中，哪些 class 彼此間容易用混？ 並說明你觀察到了什麼？ [繪出 confusion matrix 分析]



A. 同第一題，因為有 overfitting 的關係，準確率在 train data 上都蠻高的

B. Angry 較容易與 Sad 及 Neutral 搞混

C. Fear 較容易與 Sad 搞混

----------Handwritten question----------

4. (1.5%,each 0.5%)CNN time/space complexity:

For a. b. Given a CNN model as

```
model = Sequential()
model.add(Conv2D(filters=6,
                 strides=(3, 3),
"""Layer A"""    padding ="valid",
                 kernel_size=(2,2),
                 input_shape=(8,8,5),
                 activation='relu'))
model.add(Conv2D(filters=4,
                 strides=(2, 2),
"""Layer B"""    padding ="valid",
                 kernel_size=(2,2),
                 activation='relu'))
```

And for the c. given the parameter as:

kernel size = (k,k);

channel size = c;

filter size = f;

input shape = (n,n);

padding = 1;

strides = (s,s);

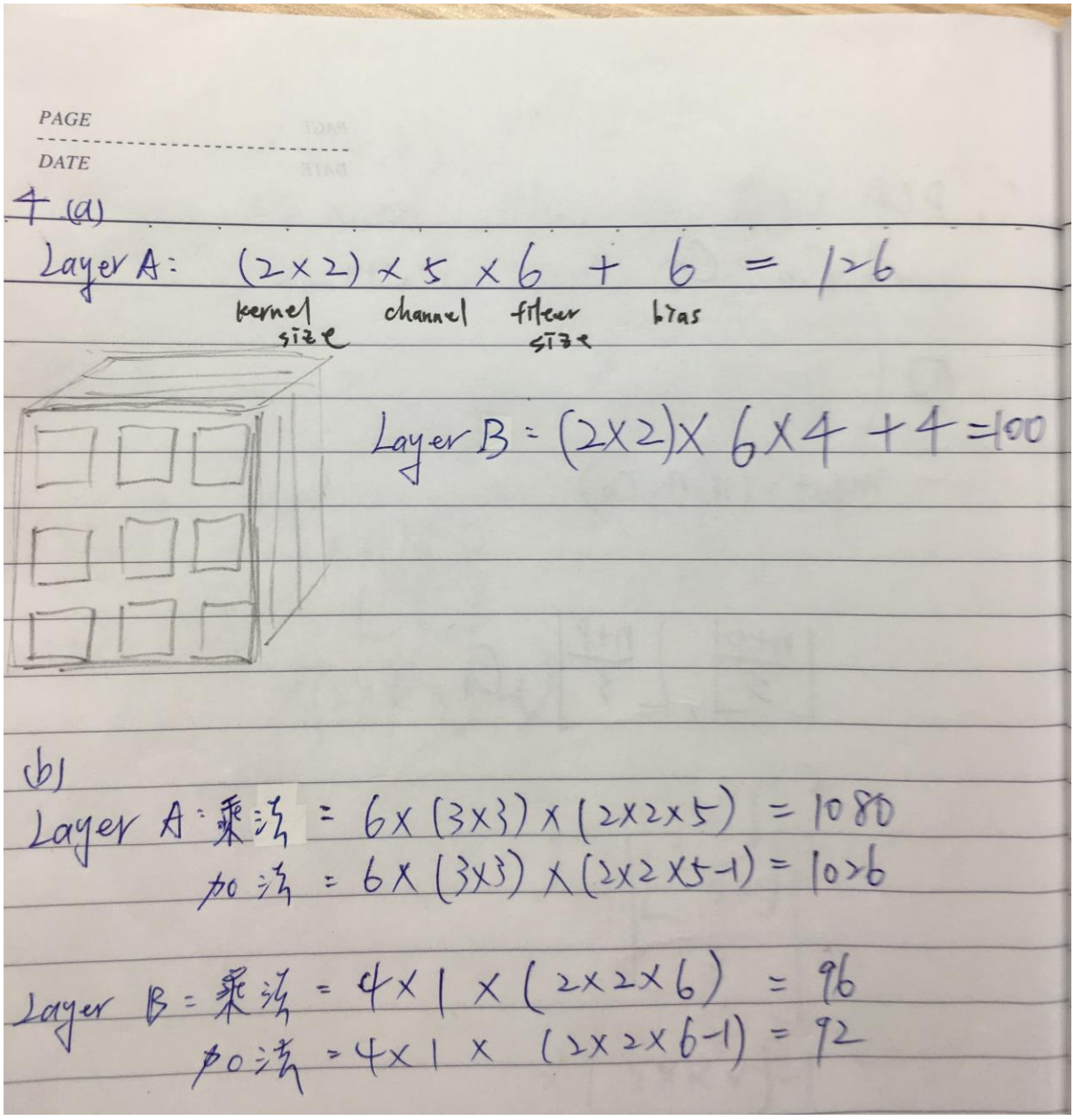
a. How many parameters are there in each layer(Hint: you may consider whether the number of parameter is related with)

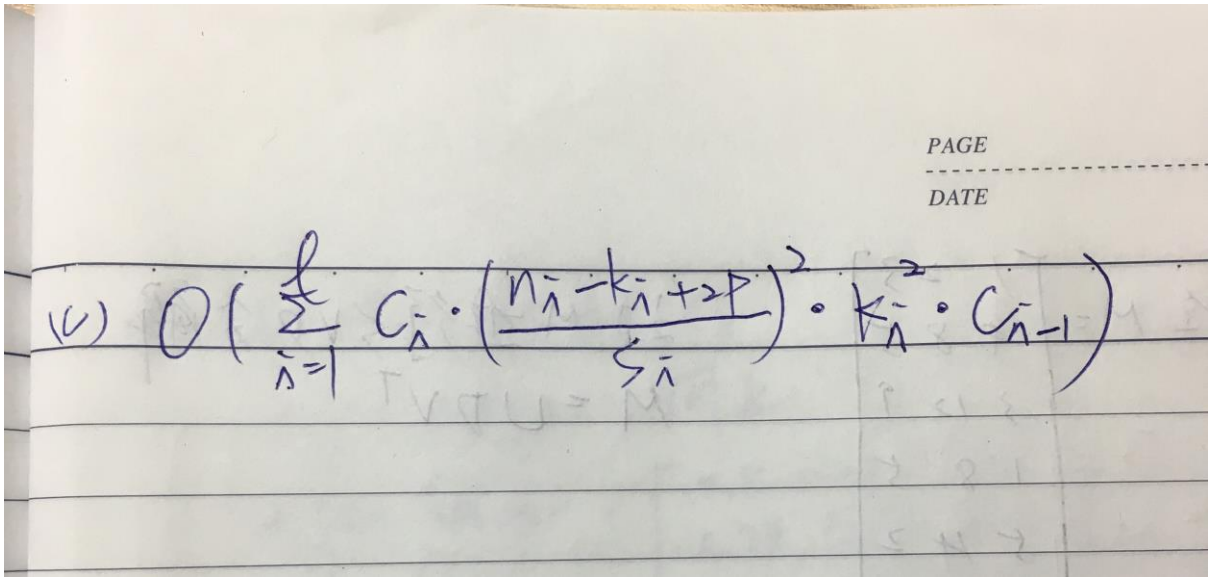
Layer A: 126

Layer B: 100

b. How many multiplications/additions are needed for a forward pass(each layer).
   Layer A: 1080/1026
   Layer B: 96/92

c. What is the time complexity of convolutional neural networks?(note: you must use big-O upper bound, and there are l layer, you can use $\square_l$, $\square_{l-1}$as lth and l-1th layer)





5. (1.5%,each 0.5%)PCA practice:Problem statement: Given 10 samples in 3D space.(1,2,3),(4,8,5),(3,12,9),(1,8,5),(5,14,2),(7,4,1),(9,8,9),(3,8,1),(11,5,6),(10,11,7)
   a. (1) What are the principal axes?
   b. (2) Compute the principal components for each sample.
   c. (3) Reconstruction error if reduced to 2D.(Calculate the L2-norm)

PAGE

DATE

5.

令 $M=\begin{bmatrix}1&2&3\\4&8&5\\3&12&9\\1&8&5\\5&4&2\\7&4&1\\9&8&9\\3&8&1\\11&5&6\\10&11&7\end{bmatrix}_{10\times3}$

對 M 進行 SVD 分解

$M=UDV^T$

(a) 對 M^TM 進行 eigen 分解 得到

$$\text{eigen value}=\begin{bmatrix}1313.18&0&0\\0&117.64&0\\0&0&55.17\end{bmatrix}$$

$$\text{eigen vector}=\begin{bmatrix}-0.5&0.78&-0.37\\-0.74&-0.61&-0.28\\-0.45&0.13&0.88\end{bmatrix}=V$$

(Principle axis)

PAGE
DATE

(b) 對 $M \cdot M^T$ 分解

eigen value = 同上

eigen vector = $\begin{bmatrix} -0.5 & 0.78 & -0.37 \\ -0.74 & -0.61 & -0.28 \\ -0.45 & 0.13 & 0.88 \\ 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix}_{10\times 3}$

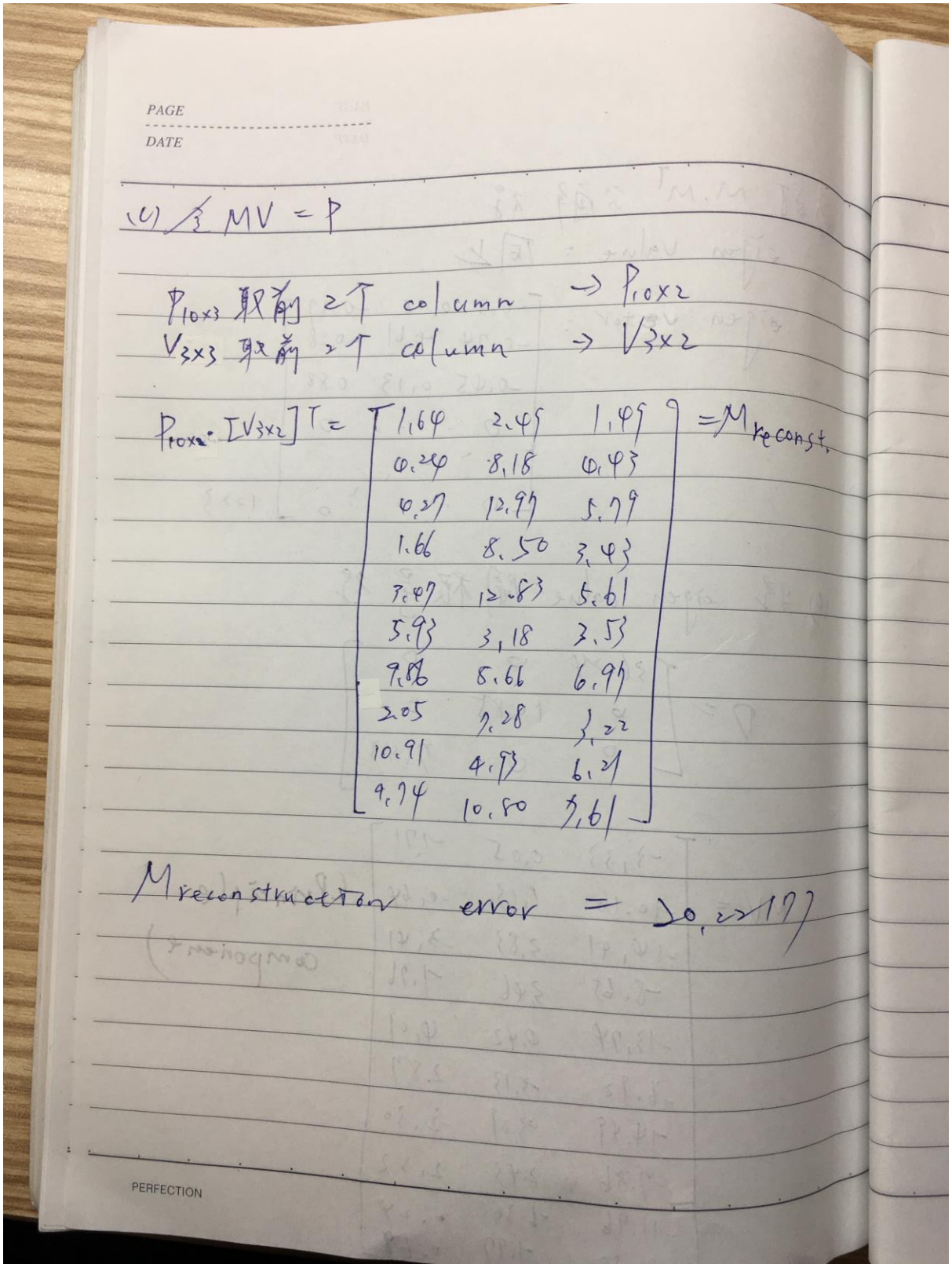
算 eigen value 開根号

$$D = \begin{bmatrix} 36.24 & 0 & 0 \\ 0 & 10.85 & 0 \\ 0 & 0 & 7.43 \end{bmatrix}$$

$$MV = \begin{bmatrix} -3.33 & 0.05 & -1.71 \\ -10.16 & 1.13 & -0.64 \\ -14.41 & 3.83 & -3.41 \\ -8.65 & 2.46 & -1.76 \\ -13.74 & 0.42 & 4.09 \\ -6.92 & -3.13 & 2.89 \\ -14.49 & -3.21 & -2.30 \\ -7.86 & 2.43 & 2.52 \\ -11.96 & -6.30 & 0.24 \\ -16.30 & -1.97 & 0.69 \end{bmatrix}$$

(Principle component)

PERFECTION



備註:與同學 R05621110 共同討論