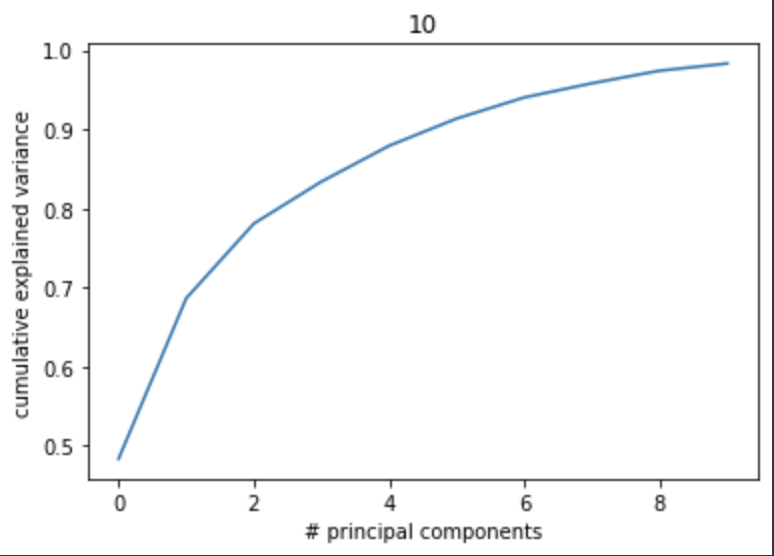
**Machine Learning**

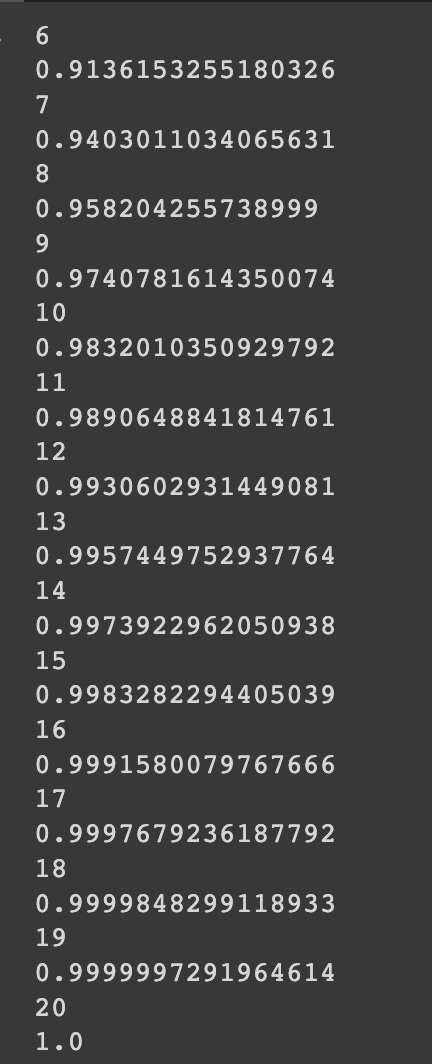
Homework 3 report b09508004 陳祈曄

（如果助教在跑code的時候出問題，再麻煩跟我說，因為我是在colab上跑的，有些路徑可能需要調整，謝謝！）



**用PCA的方式做Dimension reduction，透過以下程式碼去選擇能讓PCA投影後cumulative variance大於95%以上的feature數，決定要降到幾維，我最後是以10個features去做分類：因為他的cumulative variance到0.983，代表這10個eigenvectors的投影結果可以達到全部variance的98%，而且也降了一半的維數，達到降維的目的。**

一張含有 文字 的圖片

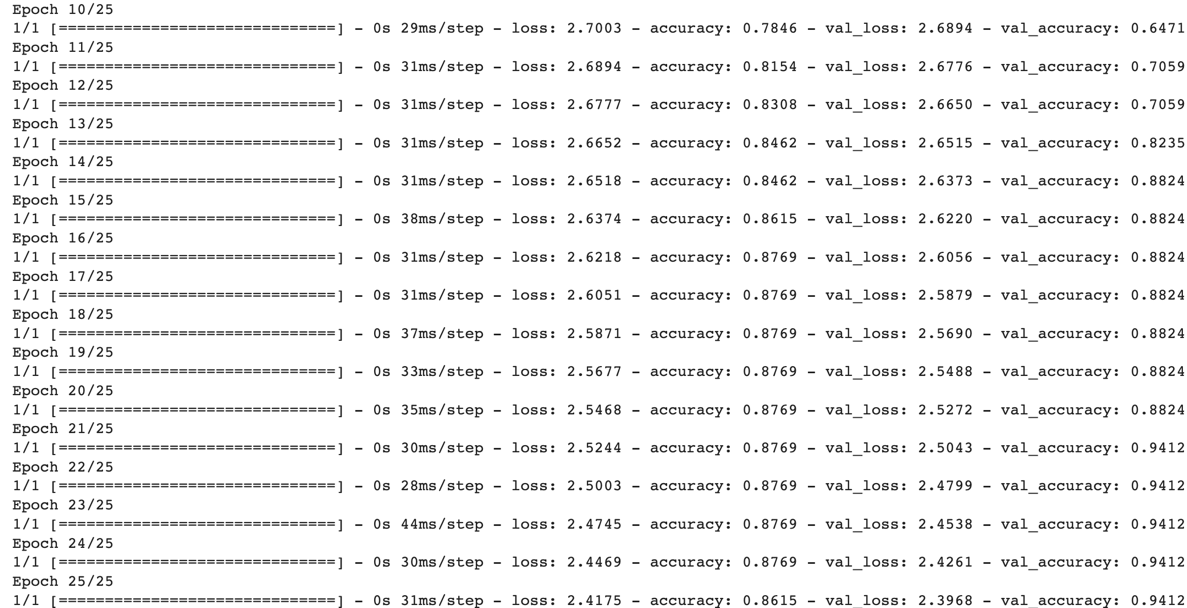
自動產生的描述 



**下面是我自己建的ANN model：我建了兩層hidden layer 和一層output layer，常見的hidden layer的activation function為relu，而output layer的activation function則為softmax，適合多分類使用。**

**至於層數和neuron數，我是嘗試幾個後找到一個訓練結果的accuracy和validation accuracy都表現的還不錯的。**

一張含有 文字 的圖片

自動產生的描述

**透過製造confusion matrix 來計算accuracy, sensitivity, specificity，並把五組的這三個值存到三個陣列，計算平均值和標準差。**

Logistic regression SVM

一張含有 文字 的圖片

自動產生的描述 一張含有 文字 的圖片

自動產生的描述

ANN model

一張含有 桌 的圖片

自動產生的描述

|  |  |  |  |
| --- | --- | --- | --- |
|  | Logistic | SVM | ANN |
| Accuracy |  |  |  |
| Sensitivity |  |  |  |
| Specificity |  |  |  |

**三者的accuracy表現都差不多在90%上下，sensitivity則是以我們的ANN最高，表示能靈敏判斷真正有病症的患者，但在Specification中，ANN的表現卻是最差的，表示沒病症的患者有可能被檢驗錯誤。**

**就結果而言，我認為三者之中效果最好的是SVM，accuracy和Specification都最高，sensitivity也足夠好，而我認為自己建立的ANN模型還不足夠穩定，數值變動較大，可能還有待改善！**