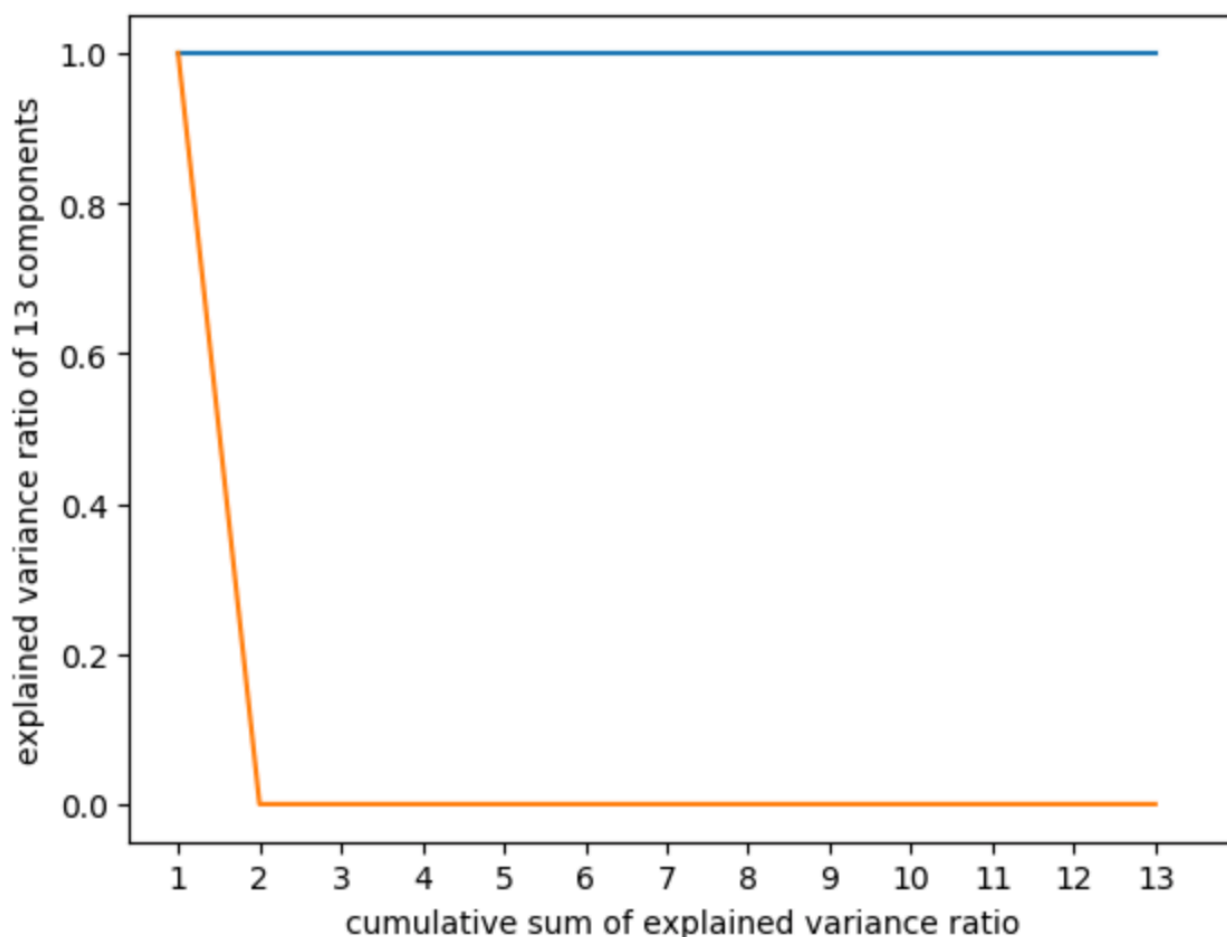


做好的dataset存成dataset.csv放在資料夾裡。

此次作業使用FinMind以及talib完成。

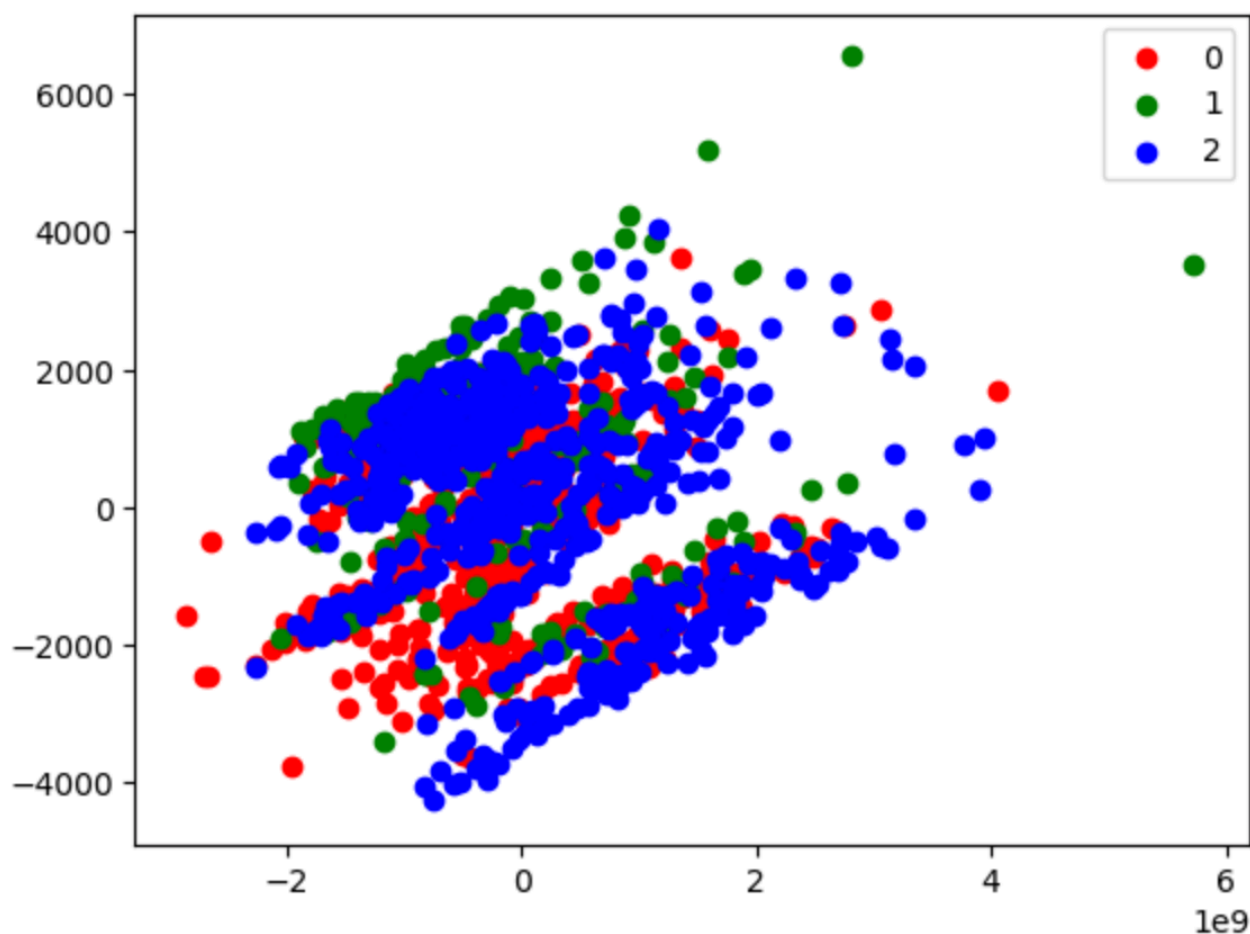
FinMind提供資料集，而talib提供好用的套件。

依據Spec所述，經過PCA後的13個component的explained variance ratio 跟其 cumulative sum如下圖所示：



我們可以觀察到，經過PCA後，演算法幾乎只用某一個component用來分label，所以在第一個component的cumulative sum就已經非常接近1了。

我們可以再看PCA壓成2維空間之後的圖：



可以發現雖然有一些群聚效應(大部分pattern是右上左下的斜線)，但是每個種類的data point混雜在一起，這代表PCA後的component在低維度空間是沒有辦法很好的將label區分開來的，可能要PCA到更高維度或是取得更多的feature才能更好的預測label。

PCA的好處是能夠避免因為distance的用法(例如歐式距離)，而造成某些重要的feature權重被稀釋，不重要的feature權重被放大。

但其壞處是，在低維度空間可能會因為表示空間的限制，造成沒辦法區分個類別的data point。

