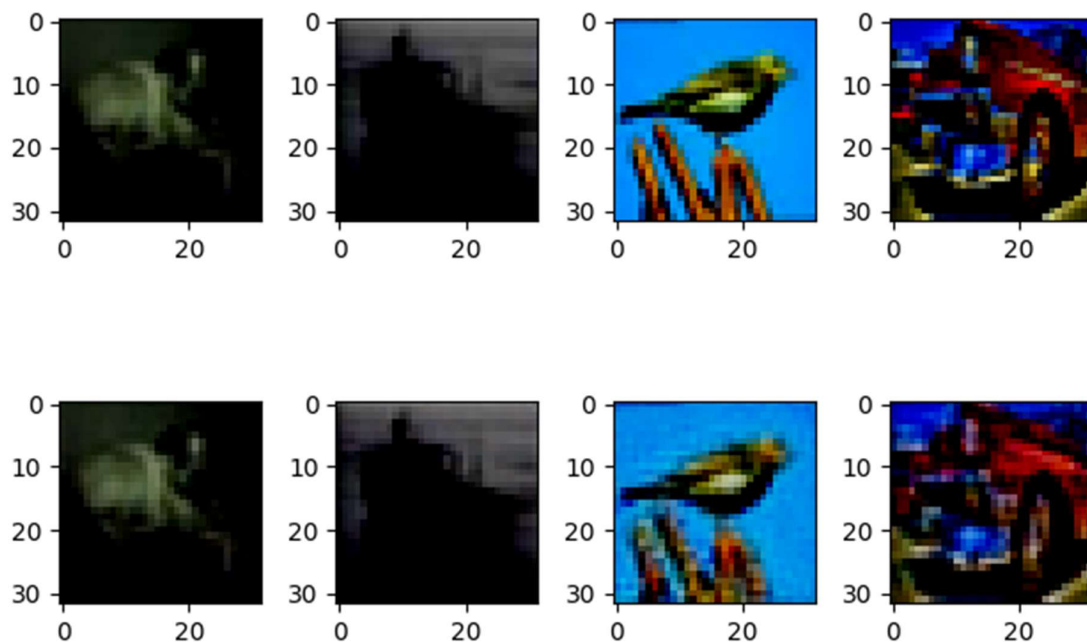
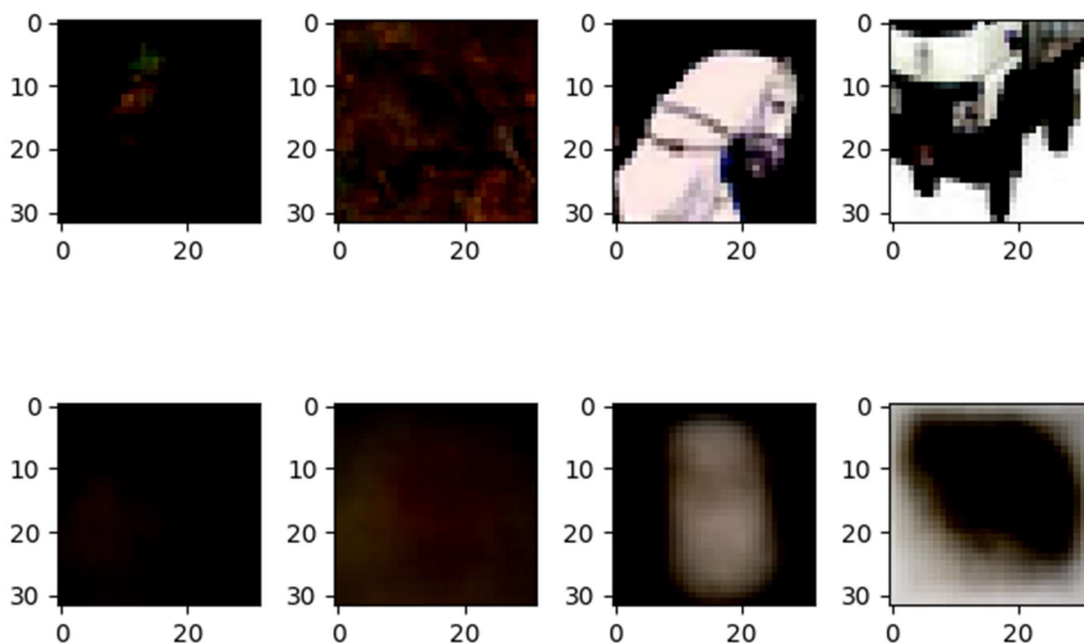


1. (2%) 任取一個 baseline model (sample code 裡定義的 fcn · cnn · vae) 與你在 kaggle leaderboard 上表現最好的單純 autoencoder 架構的 model (如果表現最好的 model 就是 sample code 裡定義的 model 的話就再任選一個 , e.g. 如果 cnn 最好那就再選 fcn) , 對各自重建的 testing data 的 image 中選出與原圖 mse 最大的兩張加上最小的兩張並畫出來。(假設有五張圖 , 每張圖經由 autoencoder A 重建的圖片與原圖的 MSE 分別為 [25.4, 33.6, 15, 39, 54.8] , 則 MSE 最大的兩張是圖 4、5 而最小的是圖 1、3) 。須同時附上原圖與經 autoencoder 重建的圖片。(圖片總數 : (原圖+重建)*(兩顆 model)*(mse 最大兩張+mse 最小兩張) = 16 張)

baseline model:



best model:

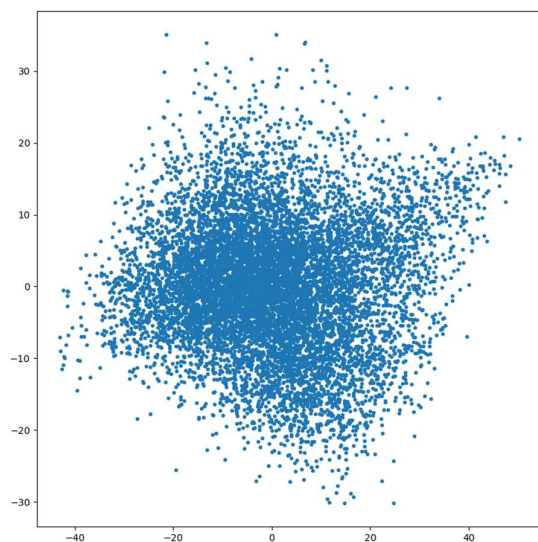


2. (1%) 嘗試把 sample code 中的 KNN 與 PCA 分別做在 autoencoder 的 encoder output 上，並回報兩者的 auc score 以及本來 model 的 auc。autoencoder 不限。不論分數與本來的 model 相比有上升還是下降，請同學簡述原因。

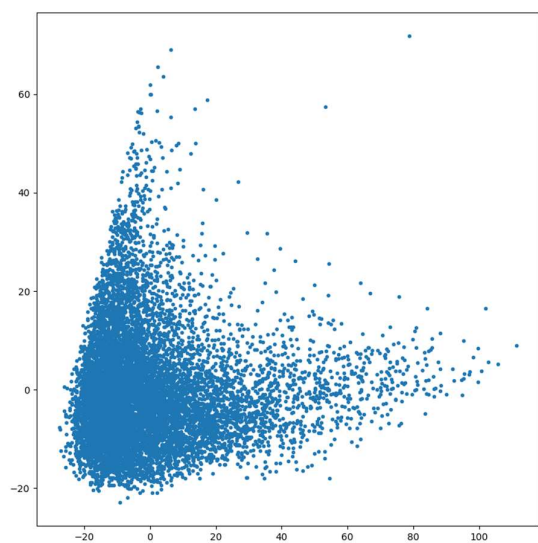
我嘗試把 sample code 中的 KNN 和 PCA 分別做在 fcn 的 encoder output 上，而原本的 fcn 在 kaggle 上的分數為 0.59，加入 KNN 後在 kaggle 上的分數提升到 0.62，但加入 PCA 後在 kaggle 上的分數卻降低到 0.44，其原因可能是 KNN 演算法比較適用於樣本容量比較大的類域的自動分類，而 PCA 在計算時丟棄的信息可能恰巧是是對數據分類有用的信息，因此才會產生這種結果

3. (1%) 如 hw9，使用 PCA 或 T-sne 將 testing data 投影在 2 維平面上，並將 testing data 經第 1 題的兩顆 model 的 encoder 降維後的 output 投影在 2 維平面上，觀察經 encoder 降維後是否分成兩群的情況更明顯。（因未給定 testing label，所以點不須著色）

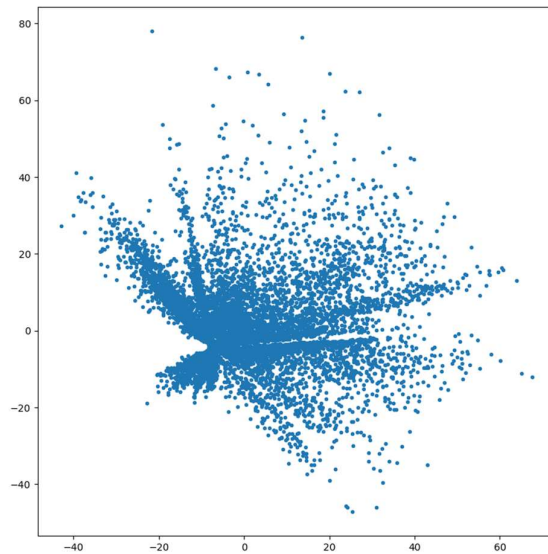
testing data:



baseline model:



best model:



我使用的是 PCA 降維，而透過視覺化分析可以觀察出原本的 testing data 是完全沒有分群的情形，而經過 baseline model 後，感覺有較為好一點，但仍然沒有明顯分為兩群的情況，而經過 best model 後，從圖片上可以發現大致上分為上下兩群，因此分群的情形 best model 明顯比 baseline model 來的好

4. (2%) 說明為何使用 auc score 來衡量而非 binary classification 常用的 f1 score。如果使用 f1 score 會有什麼不便之處？

auc scores 能夠反映了 classifier 的分類能力，並且結合考慮了分類器輸出概率的準確性，而 f1 score 是一個綜合考慮 precision 和 recall 的 metric，但是如果我們的兩個模型，分別一個 precision 比較高 recall 比較低，而另一個 recall 比較高 precision 比較低的話，f1 score 可能是差不多的，所以無法基於此來作出判斷