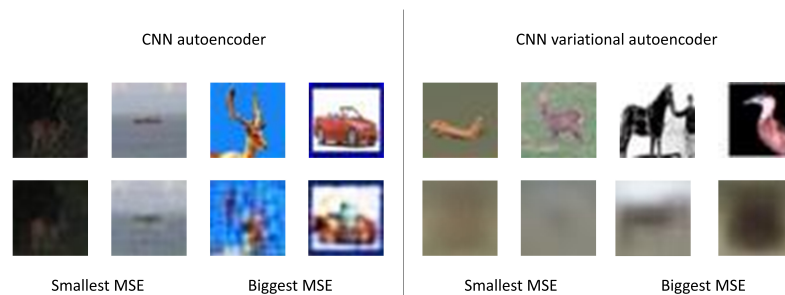


學號：B06902125 系級：資工三 姓名：黃柏瑋

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)(兩類model)mse**最大兩張+**mse**最小兩張）= 16張）



2. (1%)

嘗試把 **sample code**中的**Kmeans** 與 **PCA** 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。

在將**Kmeans**和**PCA**做在**CNN variational autoencoder**的**latent vector**上之前，**CNN variational autoencoder**的**auc score**為0.5974。在沒有取得**train data**的資料下，先將資料過**autoencoder**後取出**error**最小的前279筆資料作為**pseudo normal data**，**fit**完**Kmeans**或**PCA**後再將其他的資料丟進去**transform**。

首先討論**Kmeans**，嘗試了以下不同的**cluster**數：

N_CLUSTER	2	3	5	8	10
AUC_SCORE	0.6468	0.6224	0.6272	0.6112	0.6118

可以發現**Kmeans**讓**anomaly detection**的結果更進步，可能是因為我們用**data**與**centroid**的距離作為**anomaly score**，且**outlier**在通過**encoder**後的**feature**會大致上會散落在**train data cluster**的邊緣。

接著討論**PCA**，嘗試了以下不同的**component**數：

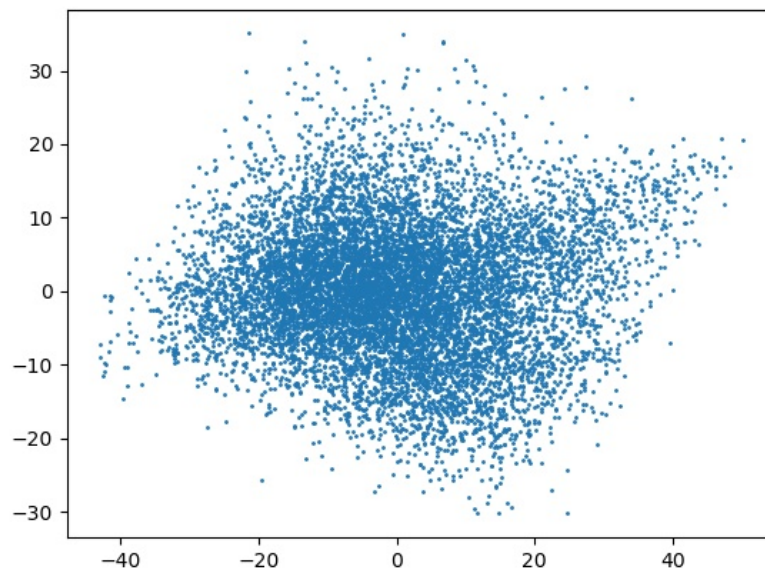
N_COMPONENT	8	16	32	64	128
AUC_SCORE	0.6299	0.6296	0.6294	0.6293	0.6293

可以發現PCA讓anomaly detection的結果更進步，可能是因為當被transform的資料離被fit的資料太遠時，就很難inverse transform回原本的資料。

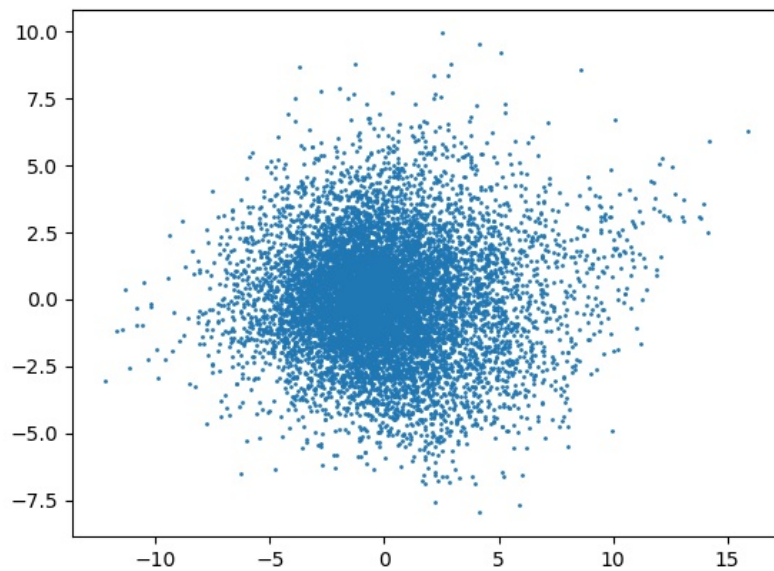
3. (1%)

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）

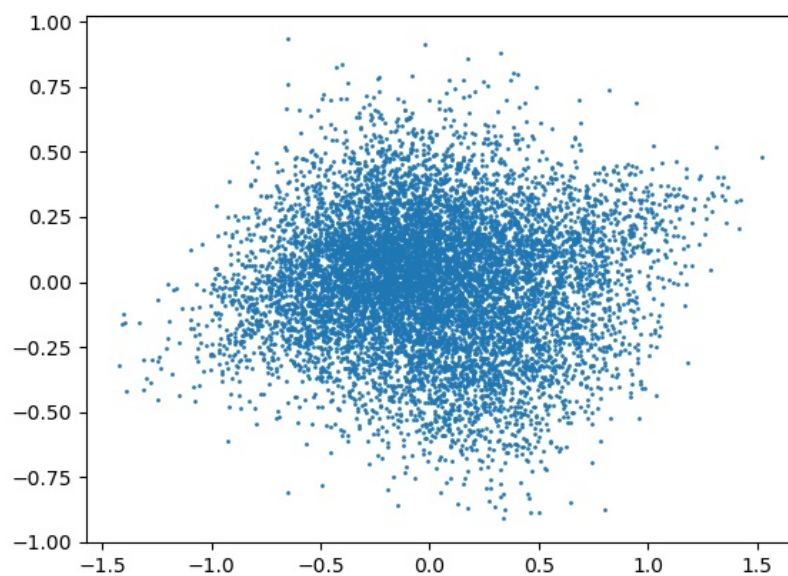
data:



CNN:



CVAE:



由上面可見，經過encoder降維後的結果並沒有將資料分成更明顯的兩群，但集中的部分有更緊密一些。

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？

F1-score和AUC-score間，最大的差別就是F1-score需要挑定一個threshold才能計算，而AUC-score是嘗試所有threshold後的結果；相比之下，後者能看出模型比較完整、全面的表現。此外，AUC-score在排名上也相對敏感，假設我們只看前20名的資料，當正確解答的排名越前面，AUC-score通常會越高，而F1-score則辨別不出這種情況下排名導致的差異。

