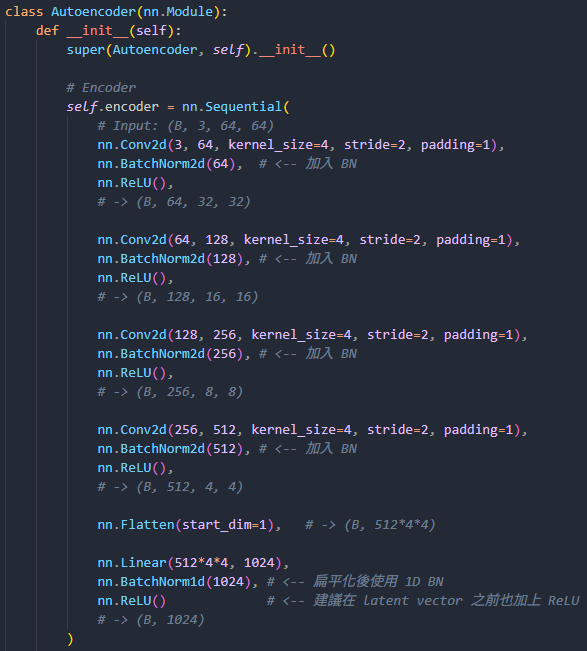
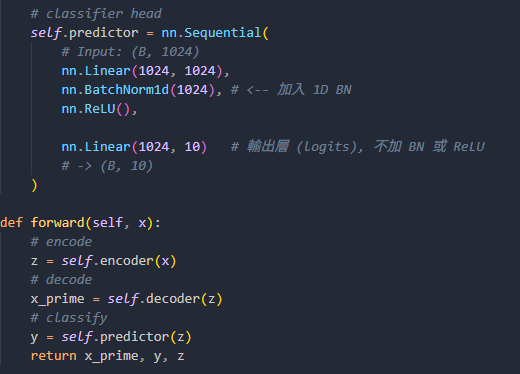
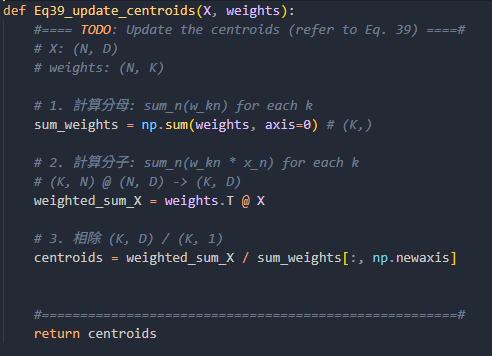
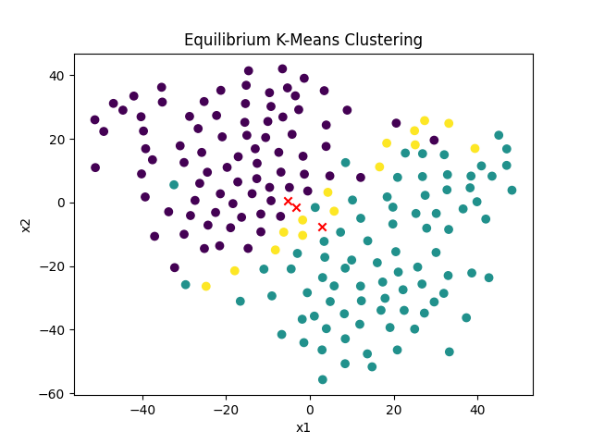
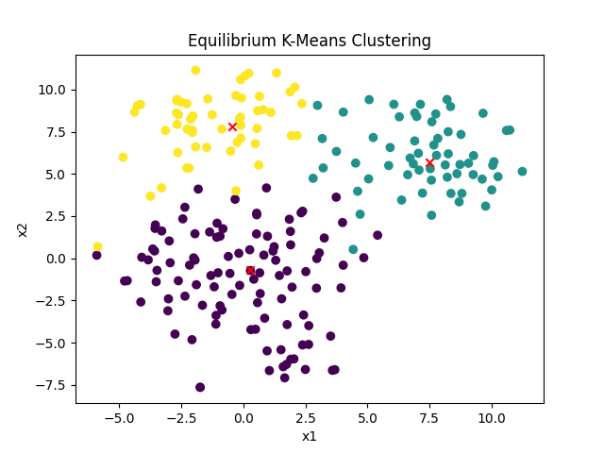
說明：請各位使用此 template 進行 report 撰寫，如果想要用其他排版模式也請註明題號以及題目內容(請勿擅自更改題號)，最後上傳前，請務必轉成PDF檔，並且命名為 report.pdf，否則將不予計分。

----------------------------------------------------------------------------------------------------------------

學號：r14921A13 系級：電機所資安組碩一 姓名：鄭皓中

1. (1.5%) **AutoEncoder**
   1. (0.5%) Paste the **complete code** of the **AutoEncoder** used in your private submission.  
        
      
   2. (1.0%) Choose **one** optimization (loss function, data augmentation, etc.) you applied during the entire training process (including both pre-training and fine-tuning). Paste the **public scores** obtained **with** and **without** this optimization, compare the two results, and try to explain the reason for the difference.  
        
         
      上面是with data augmentation 的public score，下圖則是只使用Resize及ToTensor的public score，透過觀察調整model時的Train跟Valid acc可以發現Train acc是一路上升到0.98，但Valid acc則是卡在0.5左右，這應該是overfitting的部分，因為我們沒有對data進行任何的augmentation，導致模型直接背圖片特徵而沒有關注整體，所以Train學得非常好但是Valid卻很糟糕。經過了Data augmentation調整圖片的狀態，包含翻轉、色彩變換、隨機仿射，讓model沒有辦法直接背圖片特徵，而是需要關注整體後，Train跟Valid的acc就一起緩慢上升，而非Overfitting下Train跟Valid差距懸殊的狀況。
2. (1.5%) **Equilibrium K-means Algorithm** (ref: <https://arxiv.org/pdf/2402.14490>)
   1. (0.5%) Paste the relevant code sections (**Eq38\_compute\_weights**, **Eq39\_update\_centroids**).  
        
      
   2. (1.0%) Adjust the value of **α (alpha)** until the centroids are separated and the sample size among the three clusters is approximately **2:1:1**. Then, using **10×** and **0.1×** of that α value, paste the corresponding three images and compare them.

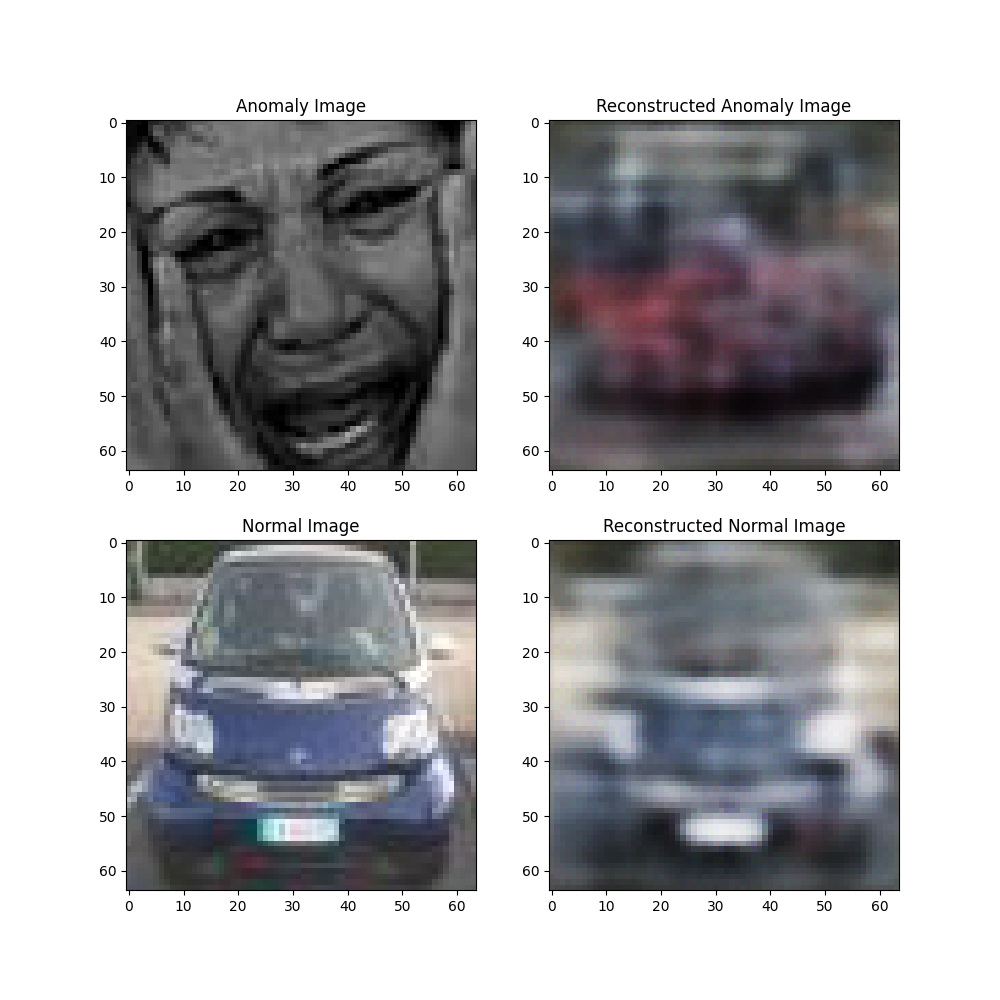
  
Left : alpha = 0.1, the three clusters are 100 : 55 : 45  
Right : alpha = 0.01  
在使用alpha = 1 的過程中出現了error message: TSNE does not accept missing values encoded as NaN natively。由這個錯誤訊息可以發現此時Equilibrium-K-means Algorithm產生了NaN的中心點，導致分群失敗。

而由上面的圖也可以發現當alpha=0.01的時候三個群的中心點非常接近，可以看出當alpha過小的時候會導致分群上面混在一起，無法清楚地進行clustering。

1. (1%) **Anomaly Detection**

Paste the **loss values** and the **corresponding images** from the results. (Choose **one** of the following options to answer.)

* If the loss and reconstruction quality differ significantly between **normal** and **anomalous** images, try to explain the reason.
* If the loss and reconstruction quality are **similar** (i.e., the model fails to distinguish anomalies), try to explain the reason.

Finally, use your **pre-trained model** or **fine-tuned model** to run the **last cell**  in the given .ipynb, observe the reconstruction results, and explain your observations.  


Anomaly loss: 0.055664725601673126

Normal loss : 0.01421459328146681

由於Anomaly loss跟Normal loss的差距很大，這可能是因為Autoencoder對於汽車這一個class已經非常熟悉，所以在判讀一般汽車圖片時可以很好的去Decode並重新生成，但當遇到非熟悉的class(Ex：人像)的時候，他則是會試圖去將其以汽車的class進行還原，產生極大的loss。透過這個方式我們可以判讀產生極大的loss的部分可能是Anomaly的圖片。