學號：R08921A01 系級： 電機碩一　姓名：陳允中

1. (1%) 請使用不同的 Autoencoder model，以及不同的降維方式(降到不同維度)，討論其 reconstruction loss & public / private accuracy。(因此模型需要兩種，降維方法也需要兩種，但 clustrering 不用兩種。)

   Model 1: 單純使用 linear 來做 Autoencoder(3*32*32->2048->1024->512->256-> 512 ->...-> 3*32*32)

   Model 2: 在 Model 1 encoder 的 linear 層之前加 LeakyReLU 和 BatchNorm1d

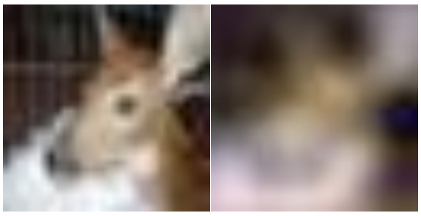
	reconstruction loss	public accuracy
Model 1	0.0233	0.52
Model 2	0.0278	0.79

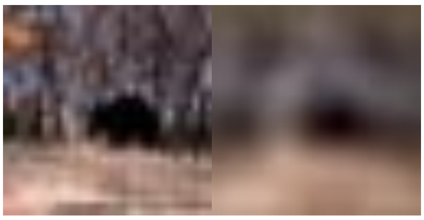
LeakyReLU 和 BatchNorm1d 可以讓產生的 latents 比較有意義

	reconstruction loss	public accuracy
Model 2 降到 128 維(latents)	0.0389	0.76
Model 2 降到 256 維(latents)	0.0278	0.79
Model 2 降到 512 維(latents)	0.0159	0.78

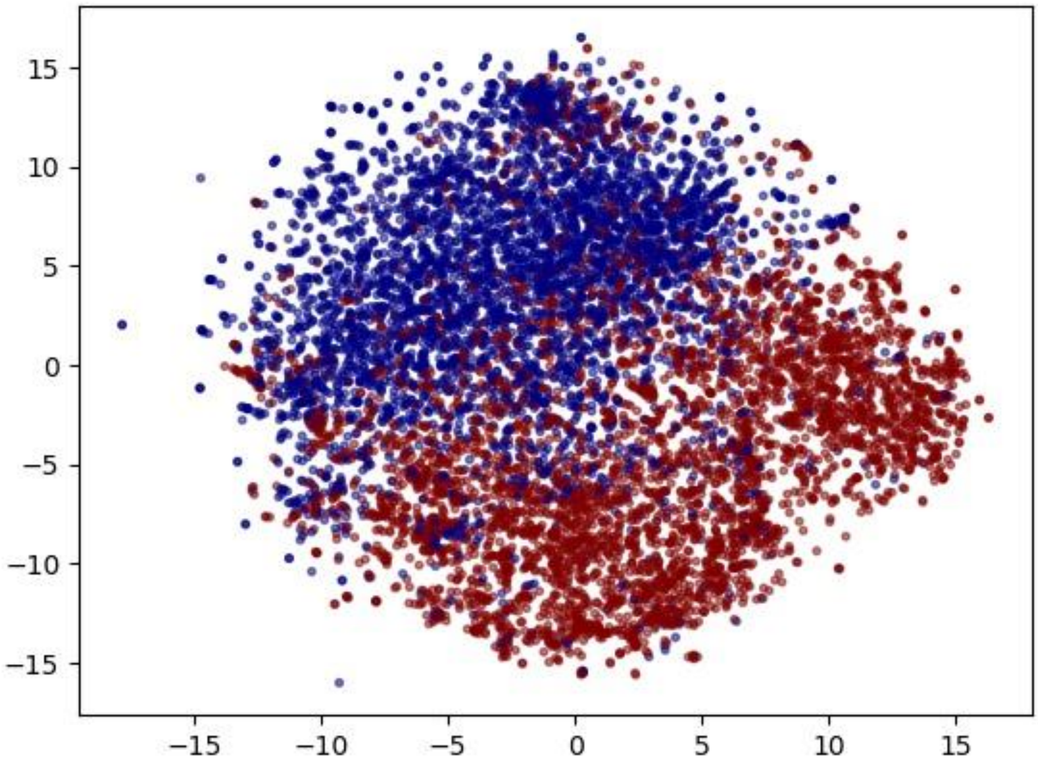
降到 256 維會使 accuracy 較高

2. (1%) 從 dataset 選出 2 張圖，並貼上原圖以及經過 autoencoder 後 reconstruct 的圖片。





3. (1%) 在之後我們會給你 dataset 的 label。請在二維平面上視覺化 label 的分佈。



math problem

1. X=(1, 2, 3), (4, 8, 5), (3, 12, 9), (1, 8, 5), (5, 14, 2), (7, 4, 1), (9, 8, 9), (3, 8, 1), (11, 5, 6), (10, 11, 7)

   X 減去平均值 = X_0 =

   [-4.4 -6. -1.8]
   [-1.4 0. 0.2]
   [-2.4 4. 4.2]
   [-4.4 0. 0.2]
   [-0.4 6. -2.8]
   [1.6 -4. -3.8]
   [3.6 0. 4.2]
   [-2.4 0. -3.8]
   [5.6 -3. 1.2]
   [4.6 3. 2.2]

   X_0 的共變異矩陣=

   [13.37777778 0.55555556 3.64444444]

[0.55555556 13.55555556   3.22222222]
[3.64444444   3.22222222   9.06666667]
共變異矩陣的特徵值及特徵向量:
6.08003657   [0.39985541 -0.67817891 -0.6165947]
12.9228041   [0.33758926   0.73439013 -0.58881629]
16.99715933   [-0.85214385 -0.02728563 -0.52259579]
(a) principal axes 為[0.39985541 -0.67817891 -0.6165947], [0.33758926   0.73439013 -0.58881629], [-0.85214385 -0.02728563 -0.52259579]
(b) principal components for each sample 等於特徵向量矩陣乘以 X_0=
[-2.25104047 -1.37323947   7.18658682]
[-0.73022635   0.94399334   0.75871342]
[-3.1883001   4.45059025 -3.07034019]
[-1.92979259   2.97853006   2.60849751]
[4.25159619   4.75401212 -1.82299166]
[2.52755823 -3.91896138   3.35457763]
[-2.13952468 -2.55604371 -4.41464321]
[2.27849363   1.73131477   3.46569126]
[0.2038499   -6.03371503 -2.31359638]
[0.97738622 -0.97648096 -5.75249521]
(c)取特徵值前 2 大的特徵向量為投影矩陣，乘上 X_0 後再乘上自己的轉置還原=
[1.90009072   2.75992709   1.08178971]
[4.29198496   8.24651657   4.37774211]
[4.27485905 13.07633588   6.28310968]
[1.77163801   8.65147726   3.35553912]
[3.29997625 12.56470677   5.62297154]
[5.98934216   3.14672348   3.1538432]
[9.85550052   8.72228056   7.17681721]
[2.08893199   7.23080501   2.94160433]
[10.91848951   4.93118246   6.17370944]
[9.60918683 10.6700449   7.83287366]
最後計算 squared loss =   54.72032912651863

2. (a)令 $Y = A^T A$，$A = [a1, a2, \ldots an]$，ai 為行向量
Y_{ij} =ai*aj
Y_{ji} =aj*ai
$Y_{ij} = Y_{ji}$
同理可得知$A^T A$和AA^T皆為 symmetric
$\exists x$，$x^T A^T Ax = (Ax)^T Ax = ||Ax||^2 \geq 0$(向量長度平方大於等於 0)

同理可得知A^TA和AA^T皆為 positive semi-definite

假設A^TA的一組非零特徵值和特徵向量為 x 和 λ ，即$A^TAx = \lambda x$
兩邊同乘以 A，得到$AA^TAx = \lambda Ax$，即$AA^T(Ax) = \lambda(Ax)$
所以A^TA和AA^T具有相同的非零特徵值。
(b)
令$X = [x_1, x_2, \dots, x_n]$
$X' = [x_1 - \mu, x_2 - \mu, \dots, x_n - \mu]$
將X'表示為其奇異值分解USV^T
$\Sigma = X'X'^T = USV^TVS^TU^T = US^2U^T$
S^2為對角矩陣，因此將Σ對角化可得 U 和 S
V為正交且為了讓$X' * [1,1,\dots,1]^T$為 0，$V^T * [1,1,\dots,1]^T$(n維)必須等於 0。
得到 U、S、V 後可得到X'，再加上μ後可得到X
(c)
$\mathrm{Trace}(\Phi^T \Sigma\Phi)$

$= \frac{1}{m}\mathrm{Trace}(\Phi^T XX^T\Phi)$

$=\frac{1}{m}\sum_{i=1}^{m}\|\Phi^T x_i\|^2$

$=\frac{1}{m}\sum_{i=1}^{m}\left\|\hat{x}_i^{(s)}\right\|^2$

Eckart-Young-Mirsky Theorem:
Variance after projection:

$$\sum_{i=1}^{m}\left\|\hat{x}_i^{(pca)}\right\|^2 \geq \sum_{i=1}^{m}\left\|\hat{x}_i^{(s)}\right\|^2$$

因此若要使 variance 最小，取Φ為 PCA 後 k 個 axes