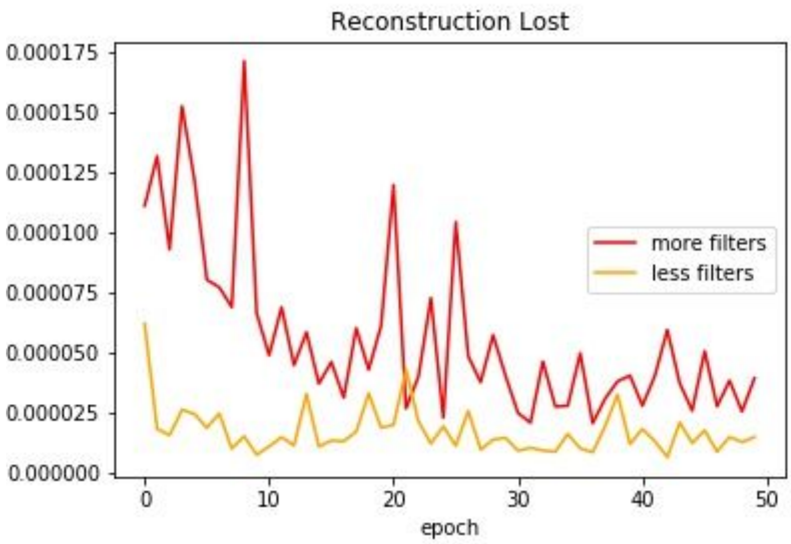
學號：B04303006 系級： 經濟四 姓名：劉傳筠

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。）

(a) Reconstruction Loss:



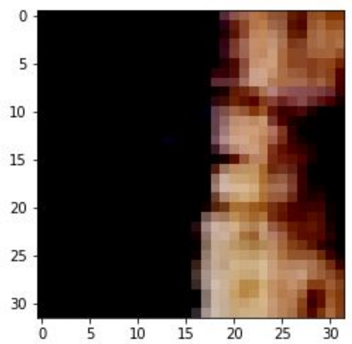
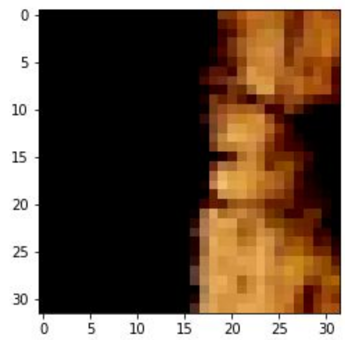
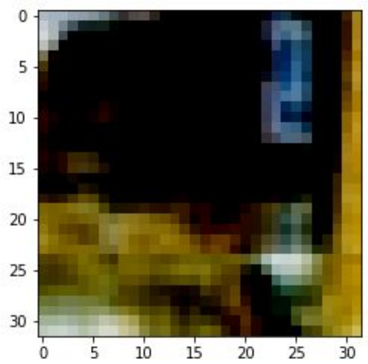
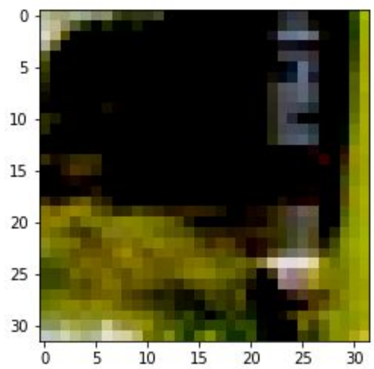
使用兩種不同的autoencoder: 第一種autoencoder每層的的filters較多(64,128,256)第二種autoencoder的filters較少(8,16,32)。觀察可以發現fulters較多會導致初期的reconstruction loss較高，應該是因為參數較多的原因。隨著epoch增加reconstruction loss皆逐漸下降，不過較多filters的仍然loss較高，應該是因為這裡epoch設較小的緣故。

(b) Private / Public Accuracy:

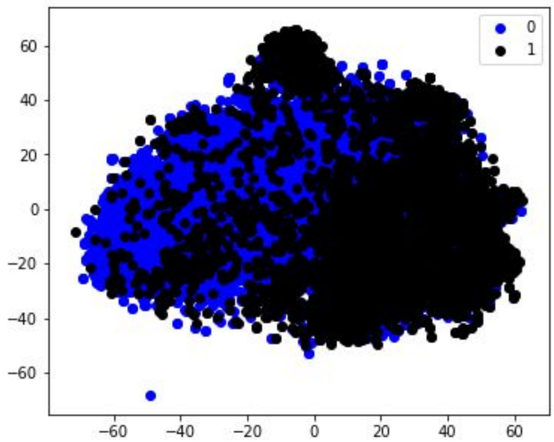
	較多channels	較少channels
w/ TSNE	0.82238 / 0.82259	0.83301 / 0.83629
wo/ TSNE	0.58333 / 0.58740	0.53539 / 0.54740

降維使用PCA(200 components)->FactorAnalysis(50 componenets)->(TSNE(2 components))->MinibatchKmeans(2 clusters)，其中一種有用tsne一種沒有。根據kaggle的private & public分數，有用tsne明顯好很多，不過多和少的channels就沒有像有沒有tsne的差距這麼巨大。

2. (1%) 從dataset選出2張圖，並貼上原圖以及經過autoencoder後reconstruct的圖片。
使用較多filters的model(64,128,256)，效果還不錯，雖然顏色會改變但是形狀都有抓到。

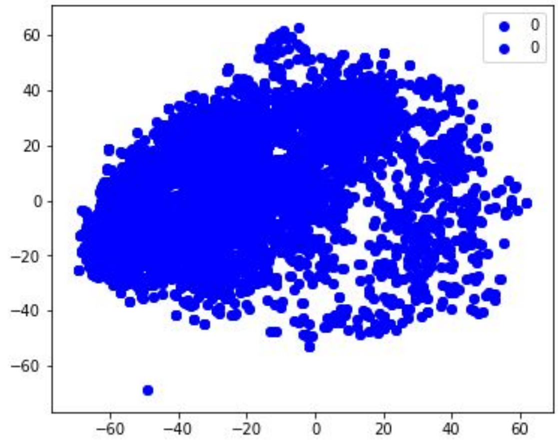
Reconstructed	Original
	
	

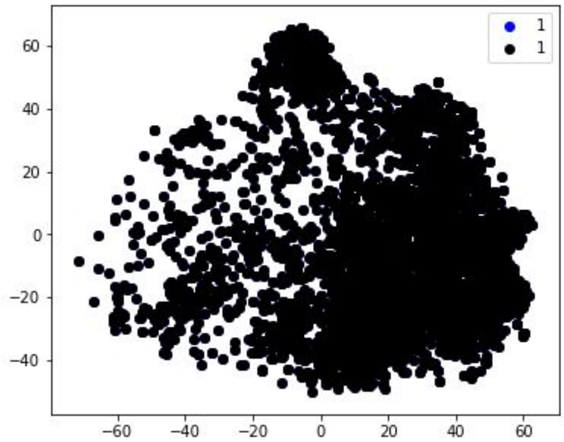
3. (1%) 在之後我們會給你dataset的label。請在二維平面上視覺化label的分佈。
True label:



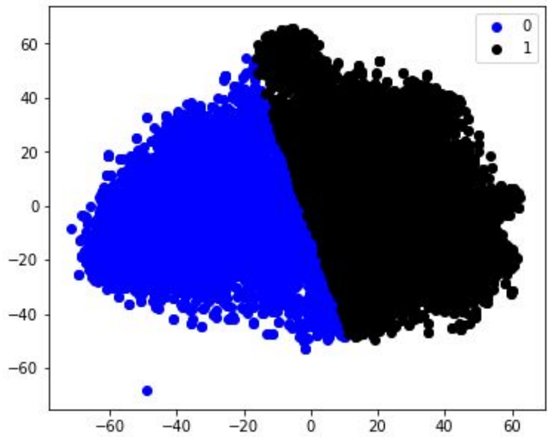
看起來好像沒有分得很開...不過彼此覆蓋的部份較少。用minibatch kmeans分群之後正確率約為83.3(both private & public, 無條件捨去至小數點第二位)

個別：





用MiniBatch Kmeans分群：



4. (3%)Refer to math problem
https://drive.google.com/file/d/1e_IDAV2yv0YEhIuVWpDdaH4Pzz5s1p2P/view?fbclid=IwAR0tO9NRxK9JZeUDNdawNuSbGTvqI7niuMX3Kkk9arauC8O6p6iJc7oMz84

MLHW4 Handwritten. ◆ date/ 11/29 ◆ page/ 1

1\. (a) (無條件捨去至小數第三位)

[-0.616, -0.588, -0.522]

[-0.678, 0.734, -0.027]

[0.399, 0.337, -0.852]

(b) (無條件捨去至小數第二位)

(-3.36, 0.70, -1.48)

(-9.78, 3.02, 0.03)

(-13.61, 6.53, -2.41)

(-7.94, 5.06, -1.16)

(-12.37, 6.83, 5.02)

(-7.19, -1.83, 3.29)

(-14.96, -0.47, -1.36)

(-11.08, 3.81, 3.04),

(-12.86, -3.95, 0.97)

(-16.30, 1.10, 1.74)

(c) 6.064

2\. (a) • $x^T A^T A x = (Ax)^T (Ax) = \|Ax\|^2 \geq 0, \forall x \in R^n$ and

$y^T A A^T y = (A^T y)^T (A^T y) = \|A^T y\|^2 \geq 0, \forall y \in R^m$

Also $(A^T A)^T = A^T A$ & $(AA^T)^T = AA^T$

∴ They're both symmetric positive semi-definite

∴ Their eigen values are non-negative.

• Next, show they share the same eigen values

Let λ be an eigen value of $A^T A$,

$$A^T A x = \lambda \cdot x$$

both multiply by A from left:

$$A \cdot A^T (Ax) = \lambda (Ax)$$

∴ For each eigen value of $A^T A$ (corresponding to eigen vector x), $A A^T$ has the same eigen value (corresponding to Ax) ∴ They share the same non-zero eigne values

(b) $\because \Sigma$ is positive semi-definite,

It can be diagonalized to $C \cdot \lambda^{-1} C^T$

$\because C \cdot \lambda^{-1} \cdot C^T = C \cdot \lambda^{-\frac{1}{2}} \cdot (C \cdot \lambda^{-\frac{1}{2}})^T$

$\therefore C \cdot \lambda^{-\frac{1}{2}}$ is the $(X-\mu)$ in our case.

$\therefore$ given Σ & μ, we can derive X.

(c) $\Phi \in R^{m \times k}$

$\min Tr[\Phi^T \Sigma \Phi]$

s.t. $\Phi^T \Phi = I$

$Tr[\Phi^T \Sigma \Phi]$

$= Tr[\Phi^T \underbrace{(I-W)(I-W)^T}_{\Sigma} \Phi]$

$= Tr[\Phi^T (I-W)[\Phi^T (I-W)]^T]$

$= \sum_i \| y_i - \sum_j W_{ij} y_j \|^2$ in LLE

The solution of the problem is obtained from the set of eigen vectors associated with the k smallest non-zero eigen values of Σ (or $(I-W)(I-W^T)$).

<pf> $\min \sum_j \phi_j^T \Sigma \phi_j$

s.t. $\sum \phi_j^T \phi_j = k$

$(\Phi^T \Phi = I_k)$

Lagrange: $L = \sum_j \phi_j^T \Sigma \phi_j - \lambda(\sum_j \phi_j^T \phi_j - k)$

$\left.\begin{array}{l} \frac{\partial L}{\partial \phi_j} = 2\Sigma\phi_j - 2\lambda\phi_j = 0 \\ \frac{\partial L}{\partial \lambda} = \sum_j \phi_j^T \phi_j - k = 0 \end{array}\right\} \Rightarrow \Sigma \Phi = \lambda \Phi$

eigen vectors

to minimize $Tr(\Phi^T \Sigma \Phi)$

$\Rightarrow$ choose smallest k eigen vectors.