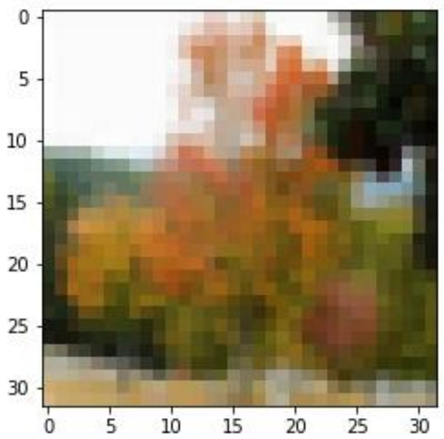
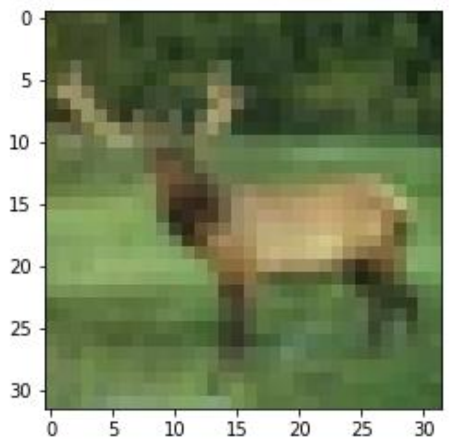
學號：R06522709　系級： 機械碩三　姓名：鄭呈毅

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。）
   (Autoencoder 將一張圖片壓到 64*4*4)+(tsne 降到 2 維) = 0.77
   剛好過 simple baseline。
   (Autoencoder 將一張圖片壓到 64*4*4)+(pca 降到 16 維+tsne 降到 2 維) = 0.68
   整組壞掉。
   (Autoencoder 將一張圖片壓到 128*4*4+dense layer with 32 nodes)+(pca 降到 64 維+tsne 降到 2 維) = 0.75
   不怎麼樣。
   (Autoencoder 將一張圖片壓到 128*4*4+dense layer with 12 nodes)+(tsne 降到 2 維) = 0.81
   剛好過了 strong baseline。
   至於 reconstruction loss 都十分接近，我就不一一列出來了。
   這次因為幾乎沒有指標可以判斷自己 train 出來的結果，本來想要透過工人智慧來幫小部分的資料上標籤，以衡量準確率，結果工人智慧的真實準確度大概也只有 0.7 吧，我看到那個原圖的解析度之後馬上就感受到助教們杜絕工人智慧的決心了。也因為這樣，這次作業真的好難，還好我後來發現 tsne 的 random seed 只要有包含 87，分數都會比較高一些。

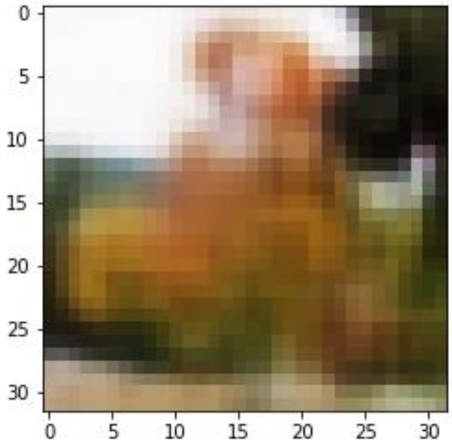
2. (1%) 從 dataset 選出 2 張圖，並貼上原圖以及經過 autoencoder 後 reconstruct 的圖片。

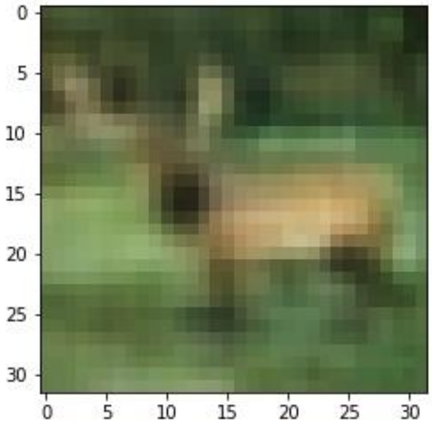
   原圖:





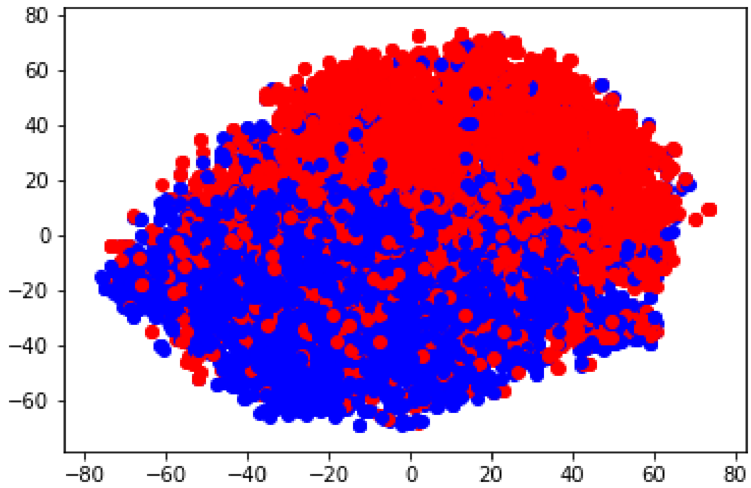
   Reconstruct:



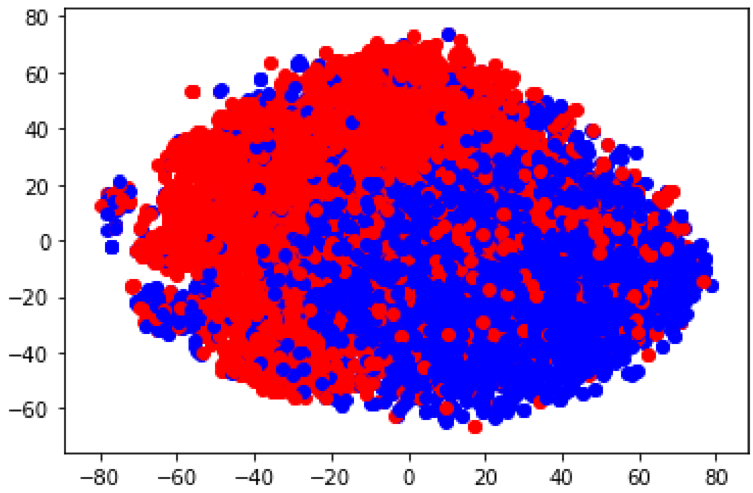


從兩組圖可以看出 reconstruct 過後的圖與原圖十分相近，只是相較之下輪廓又更加模糊。

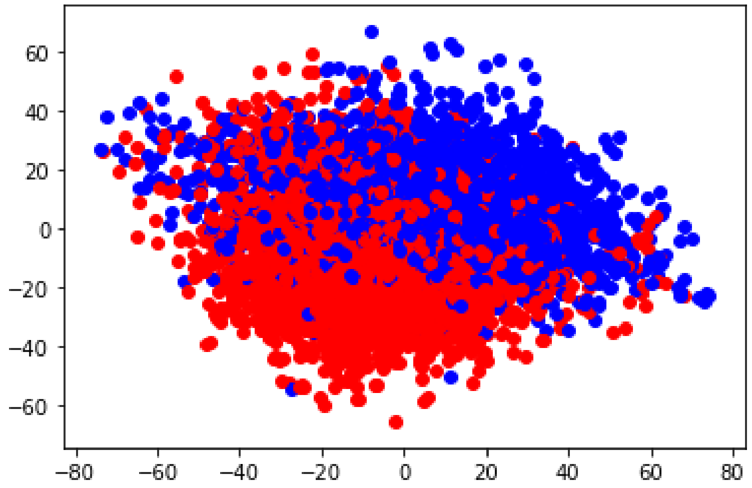
3. (1%) 在之後我們會給你 `dataset` 的 `label`。請在二維平面上視覺化 `label` 的分佈。



(pca+tsne)



(only tsne)



(pca , conponents=16, 前兩維)

從這三張圖中可以發現，TSNE+PCA 的效果與單獨使用 TSNE 的效果很接近，而比起單獨使用 PCA，前兩者的兩個標籤在特徵空間中的距離較大，因此更容易正確的判別兩種標籤 (雖然只取 PCA 的前兩維特徵來觀察對他而言有些不公平，但從結果來看它也的確較差沒錯)。

4. (3%)Refer to math problem

ML HW4 R06522709 鄭呈毅

1\. (a)

mean = (5.4, 8, 4.8)

$$conv = \begin{bmatrix} 120.4 & 5 & 32.8 \\ 5 & 122 & 29 \\ 32.8 & 29 & 81.6 \end{bmatrix}.$$

singular value = [152.9, 116.3, 54.7].

eigen vector = [-0.62, -0.59, -0.52].

[0.68, -0.73, 0.03].

[-0.40, -0.34, 0.85]

(b)

[7.19, 1.37, 2.25].

[0.76, -0.94, 0.73].

[-3.07, -4.45, 3.19].

[2.61, -2.98, 1.93]

[-1.82, -4.75, -4.25].

[3.35, 3.92, -2.53].

[-4.41, 2.56, 2.14].

[3.47, -1.73, -2.28].

[-2.31, 6.03, -0.20].

[-5.75, -0.98, -0.98]

average loss = 54.72

(c)

$$\mathrm{Trace}(\Phi^T \Sigma \Phi)$$

$$= \frac{1}{N}\mathrm{Trace}(\Phi^T X X^T \Phi)$$, From (b), $\Sigma = \frac{1}{N} X X^T$.

$$= \frac{1}{N}\|\Phi^T X\|_F^2$$

$$= \frac{1}{N}\sum_{i=1}^{N}\|\Phi^T x_i\|^2$$ → 投影到 Φ_1, Φ_2.

$$= \frac{1}{N}\sum_{i=1}^{N}\|\hat{x}_i^{(c)}\|^2$$

3.

$$L(g_T^1, \dots g_T^k) = \sum_{i=1}^{n} \exp\left(\frac{1}{k-1}\sum_{k\neq \hat{y}_i} g_T^k(x_i) - g_T^{\hat{y}_i}(x_i)\right)$$

$$= \sum_{i=1}^{n} \exp\left(\frac{1}{k-1}\sum_{k\neq y_i}\sum_{t=1}^{T}\alpha_t^k f_t(x) - \sum_{t=1}^{T}\alpha_t^{\hat{y}_i} f_t(x)\right)$$

$$= \sum_{\substack{i=1\\(\hat{y}_i\neq k)}}^{n} \exp\left[\frac{1}{k-1}\sum_{k\neq \hat{y}_i} g_{T-1}^k(x_i) + \frac{\alpha_t^k f_t(x)}{k-1} - g_t^{\hat{y}_i}(x_i)\right]$$

$$+ \sum_{\substack{i=1\\(\hat{y}_i=k)}}^{m} \exp\left[\frac{1}{k-1}\sum_{k\neq \hat{y}_i} g_{T-1}^k(x_i) - g_T^{\hat{y}_i}(x_i) - \alpha_t^k f_t(x)\right].$$

$$L(g_T^k) - L(g_{T-1}^k) = \sum_{i=1}^{n} \exp\left[\frac{1}{k-1}\sum_{k\neq \hat{y}_i} g_T^k(x_i) - g_T^{\hat{y}_i}(x_i)\right] \exp\left[\frac{\alpha_t^k f_t(x)}{k-1} - 1\right]$$

$$+ \sum_{i=1}^{m} \exp\left[\frac{1}{k-1}\sum_{k\neq \hat{y}_i} g_T^k(x_i) - g_T^{\hat{y}_i}(x_i)\right] \exp\left[-\alpha_t^k f_t(x_i) - 1\right]$$

differentiate with respect to α_t^k.

$$\approx \frac{f_t(x)}{k-1}\sum_{i=1}^{n} \exp\left[\frac{1}{k-1}\sum_{k\neq \hat{y}_i} g_T^k(x_i) - g_T^{\hat{y}_i}(x_i)\right] -$$

$$f_t(x)\sum_{i=1}^{m} \exp\left[\frac{1}{k-1}\sum_{k\neq \hat{y}_i} g_T^k(x_i) - g_T^{\hat{y}_i}(x_i)\right]$$

2. (a)

symmetric:

$(AA^T)^T = (A^T)^T A^T = AA^T$.

$\Rightarrow$ symmetric

$(A^TA)^T = A^T(A^T)^T = A^TA$

positive definite:

$x^T AA^T x = (A^Tx)^T A^Tx$ = two-norm of $A^Tx \geq 0$.

$\Rightarrow AA^T$ is positive definitive

$x^TA^TAx = (Ax)^T Ax =$ two-norm of $Ax \geq 0$

$\Rightarrow A^TA$ is positive definitive

$\because$ A^TA and AA^T are symmetric

$\Rightarrow$ they share the same non-zero eigenvalues

(b) $\frac{1}{n}\sum_{i=1}^{n} x_i x_i^T = \Sigma$ $\frac{1}{n}\sum_{i=1}^{n} x_i = u$

$\Sigma \in R^{m\times m}$ be a symmetric positive semi-definitive matrix,

since Σ is symmetric, Σ can be written as $V\Lambda V^T$

where $V = [u_1 \ldots u_m]$ is orthogonal matrix of eigenvalue of Σ, $\Lambda = \operatorname{diag}(\lambda_1 \ldots \lambda_m)$ is associated eigenvalues.

$\because$ Σ is semi positive definitive

$\Rightarrow \lambda_1 \ldots \lambda_m \geq 0$.

$\Rightarrow \Sigma = V\Lambda V^T = V\Lambda^{\frac{1}{2}}\Lambda^{\frac{1}{2}}$

$= V\Lambda^{\frac{1}{2}} (V\Lambda^{\frac{1}{2}})^T$

$= \frac{1}{n}\sum_{i=1}^{n} (x_i - u)(x_i - u)^T$.