HW4 Report

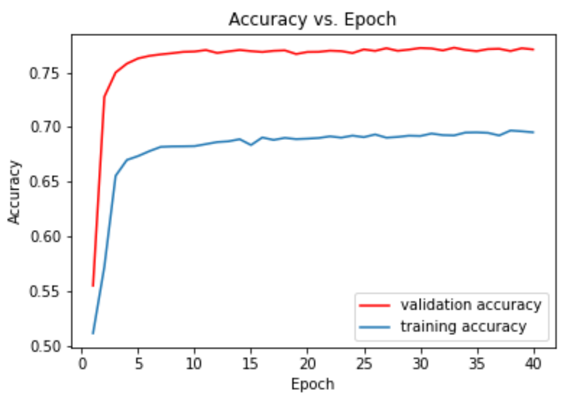
學號：b06705024　系級：資管四 姓名：郭宇軒

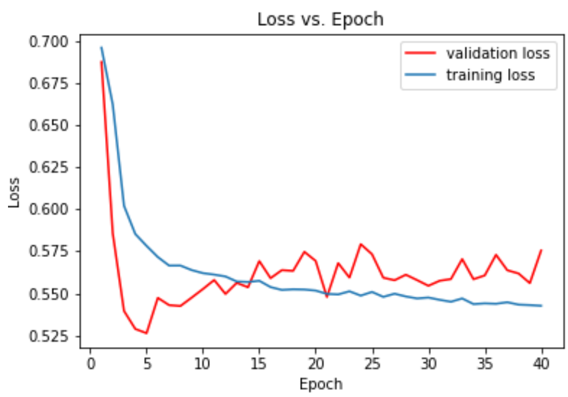
1. (0.5%) 請說明你實作之 RNN 模型架構及使用的 word embedding 方法，回報模型的正確率並繪出訓練曲線*

模型結構：

Embedding	Embedding(embedding.size(0),embedding.size(1))
RNN	LSTM(embedding_dim=256, hidden_dim=5, num_layers=1, dropout=0.5, fix_embedding= True, bidirectional=True)
Linear	Dropout(0.8)
	Linear(hidden_dim*2, 32)
	LeakyReLU()
	BatchNorm1d(32)
	Linear(32,2)

我的RNN 模型使用雙向的 LSTM， 利用最後一個 outputs 的值連結 linear classifier 和 LeakyReLU function 和 BatchNorm1d 後得到決定分類的分數，epochs 設為 40 後，最終訓練曲線如下圖：



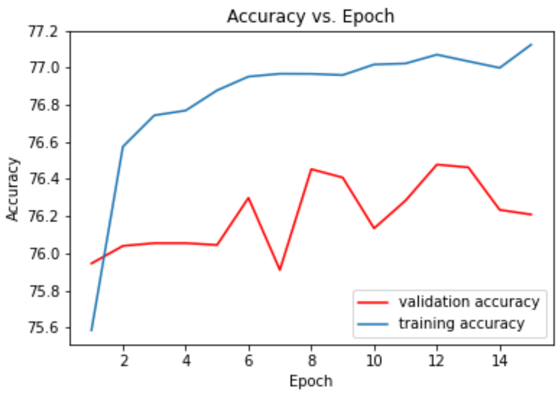


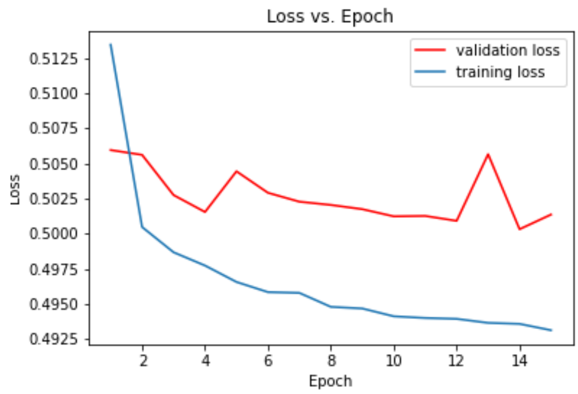
2. (0.5%) 請實作 BOW+DNN 模型，敘述你的模型架構，回報模型的正確率並繪出訓練曲線*。

DNN 模型結構：

Linear	Linear(embedding_dim, 512)
	nn.Linear(512, 128)
	nn.Linear(128, 1)
	nn.Sigmoid()

最後得到 kaggle Public score: 0.76960，訓練曲線如下圖：





3. (0.5%) 請敘述你如何 improve performance（preprocess, embedding, 架構等），並解釋為何這些做法可以使模型進步。

word_embedding:

利用 gensim.models 中的 word2vec 模型，並將 iter 調整為 20，雖然訓練時間變久，但實際上可以看到後來的準確率有從kaggle public score從 0.75 提升的 0.77左右。

模型架構:

將LSTM改成雙向的LSTM模型，並增加linear 層數，實際訓練後結果也有更進一步，kaggle public score 提升到 0.78左右。

4. (0.5%) 請比較 RNN 與 BOW 兩種不同 model 對於 "Today is hot, but I am happy" 與 "I am happy, but today is hot" 這兩句話的分數（model output），並討論造成差異的原因。

BOW 沒有學習字詞先後順序的能力，只考慮了每個詞的出現次數而已，無法

	"Today is hot, but I am happy"	"I am happy, but today is hot"
RNN	[0.76534625,0.23465375]	[0.71983659,0.28016341]
BOW	[0.90035638,0.09964362]	[0.90035638,0.09964362]

分辨出這次兩個不同的句子，因此可以發現兩者得到的分數相同； RNN 則可以判斷字詞的不同順序會產生不同的語意，並分辨出這是兩個不同的句子，讓兩者輸出的值有些微的不同。

5. (3%)Math problem:https://drive.google.com/file/d/1fEu87banB4s6Yjku1dA5sMcnwCugEPBF/view?usp=sharing

1. (a).

$$\mu = \frac{1}{10}\sum_{n=1}^{10} x_n = \begin{bmatrix} 5.4 \\ 8 \\ 4.8 \end{bmatrix}$$

$$\Sigma = \frac{1}{10}\sum (x_i - \bar{x})(x_i - \bar{x})^T = \begin{bmatrix} 12.04 & 0.5 & 3.28 \\ 0.5 & 12.2 & 2.9 \\ 3.28 & 2.9 & 8.16 \end{bmatrix}$$

$\Rightarrow$ Eigenvalues, Eigenvector of S:

$$\lambda_1 = 15.29, \quad u_1 = \begin{bmatrix} -0.62 \\ -0.59 \\ -0.52 \end{bmatrix}$$

$$\lambda_2 = 11.63, \quad u_2 = \begin{bmatrix} -0.68 \\ 0.73 \\ -0.03 \end{bmatrix}$$

$$\lambda_3 = 5.47 \quad u_3 = \begin{bmatrix} 0.40 \\ 0.34 \\ -0.85 \end{bmatrix}$$

$\Rightarrow u_1, u_2, u_3$ are principal axes. ※

(b.) $$W = \begin{bmatrix} -0.62 & -0.59 & -0.52 \\ -0.68 & 0.73 & -0.03 \\ 0.4 & 0.34 & -0.85 \end{bmatrix}$$

$$z_1 = \begin{bmatrix} -3.36 \\ 0.71 \\ 1.48 \end{bmatrix} \quad z_2 = \begin{bmatrix} -9.78 \\ 3.03 \\ -0.04 \end{bmatrix} \quad z_3 = \begin{bmatrix} -13.61 \\ 6.53 \\ 2.42 \end{bmatrix}$$

$$z_4 = \begin{bmatrix} -7.94 \\ 5.06 \\ 1.16 \end{bmatrix} \quad z_5 = \begin{bmatrix} -12.37 \\ 6.84 \\ -5.02 \end{bmatrix} \quad z_6 = \begin{bmatrix} -7.19 \\ -1.84 \\ -3.3 \end{bmatrix}$$

$$z_7 = \begin{bmatrix} -14.96 \\ -0.47 \\ 1.37 \end{bmatrix} \quad z_8 = \begin{bmatrix} -7.08 \\ 3.81 \\ -3.05 \end{bmatrix} \quad z_9 = \begin{bmatrix} -12.86 \\ -3.95 \\ -0.97 \end{bmatrix}$$

$$z_{10} = \begin{bmatrix} -16.30 \\ 1.11 \\ -1.75 \end{bmatrix}$$

(c.) $\frac{1}{10}\sum_{i=1}^{10}(x_i-\hat{y}_i)^2 = 6.064$ # 3D→2D

2. (a.)

$\because (AA^T)^T = (A^T)^T A^T = AA^T$, $(A^TA)^T = A(A^T)^T = A^TA$

$\therefore$ AA^T and A^TA are both symmetric,

$\forall\ x\in R^m, x\neq 0$, $\forall\ y\in R^n, y\neq 0$,

$x^T(AA^T)x = (A^Tx)^T A^Tx = \|A^Tx\|^2 \geq 0$

$x^T(A^TA)x = (Ax)^T Ax = \|Ax\|^2 \geq 0$

$\Rightarrow$ A^TA and AA^T are both positive semi-definite

再令 λ 為 AA^T 的其中一 eigenvalue, $\lambda\neq 0$

$\forall\ \underline{v}\in R^m$, $(AA^T)\underline{v} = \lambda\underline{v}$

$\Rightarrow (A^TA)(A^T\underline{v}) = A^T((AA^T)\underline{v}) = A^T(\lambda\underline{v}) = \lambda(A^T\underline{v})$

$\Rightarrow$ λ 也是 (A^TA) 的 eigenvalue, $A^T\underline{v}$ 為其 eigenvector.

令 Δ 為 AA^T 的其中一 eigenvalue, $\Delta\neq 0$

同理 $\forall\ \underline{u}\in R^n$, $(A^TA)\underline{u} = \Delta\underline{u}$,

$\Rightarrow (AA^T)(A\underline{u}) = A((A^TA)\underline{u}) = A(\Delta\underline{u}) = \Delta(A\underline{u})$

$\Rightarrow$ Δ 也是 (AA^T) 的 eigenvalue, $A\underline{u}$ 為其 eigenvector

$\Rightarrow$ AA^T and A^TA 有相同 eigenvalue #

(b.) 設 $z_1, z_2 \cdots z_m, z_{2m} \in R^m$ 為

$$\begin{bmatrix}\sqrt{m}\\0\\\vdots\\0\end{bmatrix}, \begin{bmatrix}-\sqrt{m}\\0\\\vdots\\0\end{bmatrix} \cdots\cdots \begin{bmatrix}0\\0\\0\\\sqrt{m}\\\vdots\\0\end{bmatrix}\begin{bmatrix}0\\0\\0\\-\sqrt{m}\\\vdots\\0\end{bmatrix} \cdots\cdots \begin{bmatrix}0\\0\\\vdots\\\sqrt{m}\end{bmatrix}\begin{bmatrix}0\\\vdots\\0\\0\\-\sqrt{m}\end{bmatrix}$$

$\Rightarrow \frac{1}{2m}\sum_{k=1}^{2m} z_k = 0$, $\mathrm{Cov}(z_k) = \frac{1}{2m}\sum_{1}^{2m}(z_k)(z_k)^T = \frac{I_m}{2m}$

Let $M = (x_i - \mu)(x_i - \mu)^T$, M is symmetric and positive semidefinite matrix

$\therefore z^T M z \geq 0$, $\forall z \in \mathbb{R}^m$.

$\Rightarrow E(z^T M z) \geq 0 \Rightarrow z^T E(M) z \geq 0$

$\alpha\ E(M) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)(x_i - \mu)^T = \Sigma \Rightarrow z \Sigma z^T \geq 0$

$\Rightarrow \Sigma$ is positive semi-definite matrix. — ①

$\exists A \in \mathbb{R}^{m \times m}$ let $\Sigma = AA^T$

when $x_i = Az_k + \mu$

$$\frac{1}{2m} \sum_{k=1}^{2m} (Az_k + \mu) = \frac{A}{2m} \sum_{k=1}^{2m} z_k + \mu = A \cdot 0 + \mu = \mu.$$

此时 $Cov(x_i) = Cov(Az_k + \mu)$

$$= \frac{1}{2m} \sum_{k=1}^{2m} ((Az_k + \mu) - \mu)((Az_k + \mu) - \mu)^T$$

$$= \frac{1}{2m} \sum_{1}^{2m} (Az_k)(Az_k)^T$$

$$= \frac{1}{2m} \sum_{1}^{2m} (A z_k^T A^T) = \frac{A}{2m} \sum_{1}^{2m} (z_k z_k^T) A^T = A \cdot I_n \cdot A^T$$

$$= A \cdot A^T = \Sigma$$

(C.) Since Σ and $\phi\phi^T$ are symmetric, exist orthogonal matrix,

$Trace(\phi \Sigma \phi^T) = \frac{1}{N} Trace(\phi^T X X^T \phi) = \frac{1}{N} \| \phi X \|_F^2$

$= \frac{1}{N} \sum_{1}^{N} \| \phi^T x_i \|^2 = \frac{1}{N} \sum_{1}^{N} \| \hat{x}_i^{(s)} \|^2$ — ?

let $\phi = v_1, v_2, \dots, v_k$, is a orthogonal vector set in $\mathbb{R}^m$.

$(\phi\phi^T) v_i = \phi(\phi^T v_i) = \phi \cdot e_i = v_i = 1 \cdot v_i$, v_i is eigenvector of $\phi\phi^T$

eigenvalue $= 0$.

3. we want to find f_t s.t. $L(g_t^1 \dots g_t^k) < L(g_{t+1}^1 \dots g_{t+1}^k)$

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

$= \sum_{\substack{i=1 \\ \text{if } k\neq y_i}}^n \exp\left(\frac{1}{k-1}\sum_{k\neq \hat{y}_i} g_{t-1}^k(x_i) + \frac{a_t^k f_t(x)}{k-1} - g_t^{\hat{y}_i}(x_i)\right) +$

$\sum_{\substack{i=1 \\ \text{if } k=y_i}}^n \exp\left(\frac{1}{k-1}\sum_{k\neq \hat{y}} g_{t-1}^k(x_i) - a_t^k f_t(x) - g_t^{\hat{y}_i}(x_i)\right)$

$\Delta L = L(g_T^k) - L(g_{t-1})^k$

$= \sum_{i=1}^n \exp\left(\frac{1}{k-1}\sum_{k\neq \hat{y}_i} g_{t-1}^k(x_i) - g_{t-1}^{\hat{y}_i}(x_i)\right)\left(\exp\left(\frac{a_t^k f_t(x_i)}{k-1}\right)-1\right)+$

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

$\Delta L = \sum_{i=1}^n \exp\left(\frac{1}{k-1}\sum_{k\neq \hat{y}_i} g_{t-1}^k(x_i) - g_{t-1}^{\hat{y}_i}(x_i)\right)\cdot\left(\frac{a_t^k f_t(k)}{k-1} - a_t^k f_t(x_i)\right)$

$\Rightarrow \frac{\partial \Delta L}{\partial a_t} = \frac{f_t(k)}{k-1}\sum \underline{\exp\left(\frac{1}{k-1}\sum_{k\neq \hat{y}_i} g_{t-1}^k(x_i) - g_{t-1}^{\hat{y}_i}(x_i)\right)} -$

$f_t(x_i)\sum_{k=\hat{y}_i} \underline{\exp\left(\frac{1}{k-1}\sum_{k\neq \hat{y}_i} g_{t-1}^k(x_i) - g_{t-1}^{\hat{y}_i}(x_i)\right)}$

$\Rightarrow$ we can update f_t by $\left[f_t \longleftarrow f_{t-1} - \eta\frac{\partial L}{\partial a_t}\right]$..

we want to find a_t that minimize $L(g_t)$.

let $\frac{\partial L}{\partial a_k^t} = 0$,

$$\frac{\partial L}{\partial \alpha_k^t} = \sum_{k=\theta_i} u_t^n e^{a_t} \cdot \frac{1}{k-1} - \sum_{k\neq\theta_i} u_t^n e^{-a_t}$$

$$\Rightarrow Z_t \varepsilon_t e^{a_t} \frac{1}{k-1} - Z_t (1-\varepsilon_t) e^{-a_t} = 0$$

$$\Rightarrow e^{a_t} = (k-1) \cdot \left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)$$

$$\Rightarrow a_t = \ln\left(\sqrt{(k-1)\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)}\right) \ \#.$$