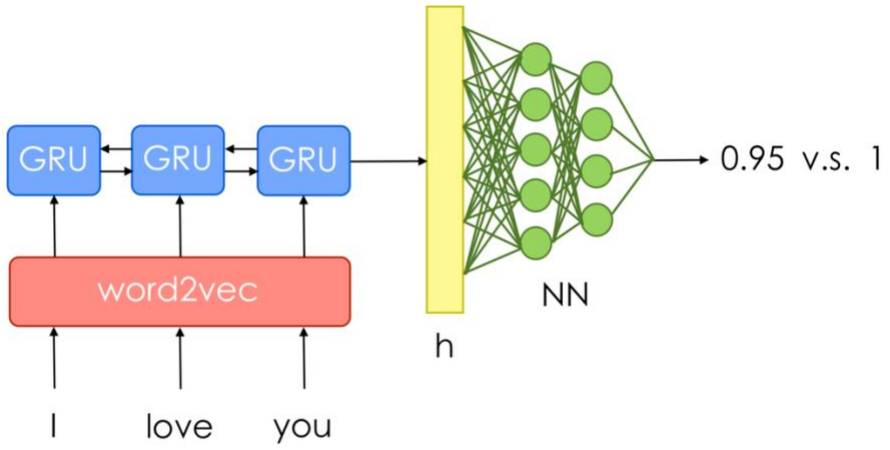
學號：R08725021 系級：資管碩一 姓名：王鼎元

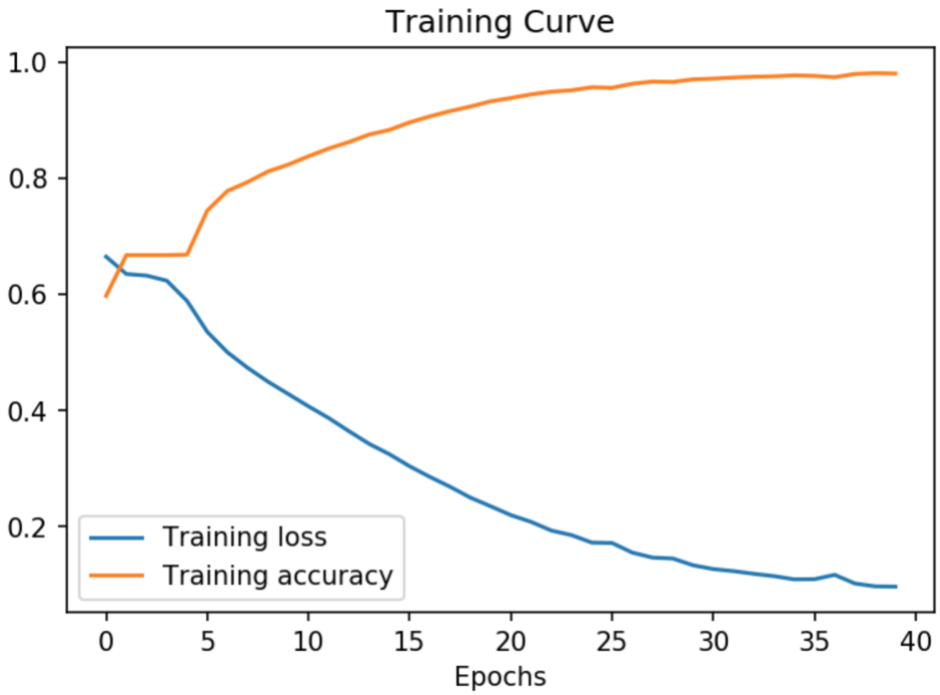
1. (1%) 請說明你實作之 RNN 模型架構及使用的 word embedding 方法，回報模型的正確率並繪出訓練曲線*

我的模型架構如下：



先訓練 250 維的 word2vec 模型來產生 embedding，接著過一個 3 層的雙向 GRU，hidden layer 的維度是 128，dropout rate 是 0.8，把最後一層的 hidden layer 拿出來當作 representation，過 (256, 128)、(128, 64) 的全連接層，最後產出預測值。以下為我的分數與訓練曲線：

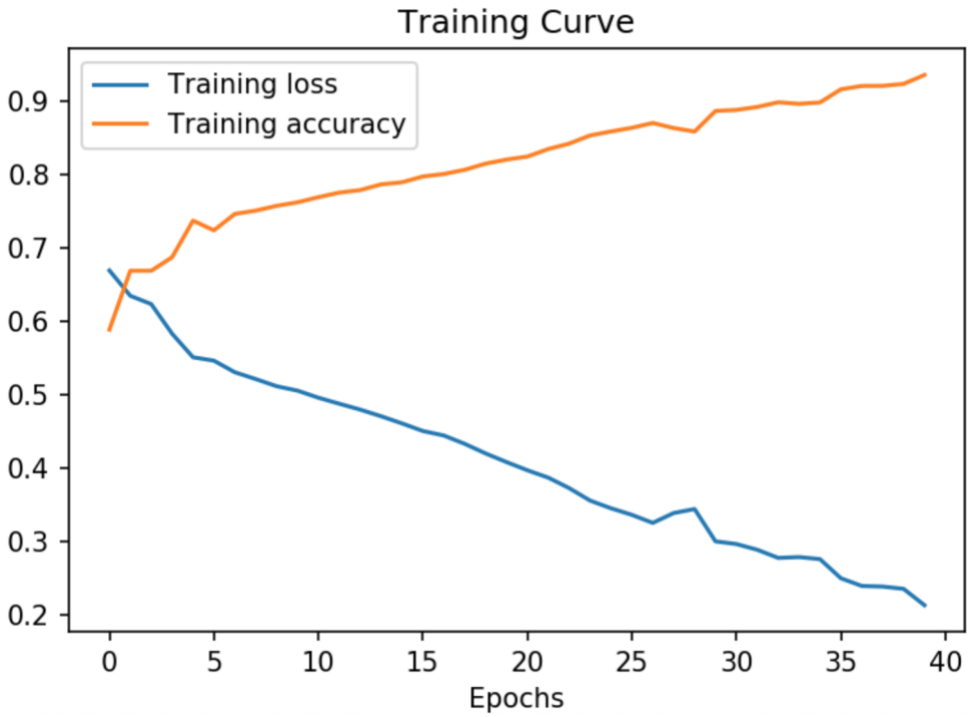
	Private score	Public score
GRU_Model	0.80000	0.79767



2. (1%) 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*。

我將處理過後的句子通過 word2vec 模型後產生出 BOW 向量，在接上 (1024, 512)、(512, 256) 的全連接層，最後通過 softmax 產出預測值。以下為我的分數與訓練曲線：

	Private score	Public score
BOW_Model	0.74883	0.74186



3. (1%) 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。

   a. 將 word2vec 模型的維度從 100 維調整到 250，因為我覺得維度太少的時候會無法完整表現出詞義。

   b. 將 GRU 的 dropout rate 從 0.2 調高至 0.8，因為剛開始訓練時 validation accuracy 總是無法提升，我認為是因為資料集不大容易 overfitting，故調整 dropout rate。

   c. 前處理的時候將無意義的詞（如：@user）、非字（emoji）刪掉，並把有數字出現的地方都換成 1 以減少 token 量。

   d. 稍微對字做 stemming 以減少 token 量。

4. (1%) 請比較不做斷詞 (e.g.,用空白分開) 與有做斷詞，兩種方法實作出來的效果差異，並解釋為何有此差別。

有沒有做斷詞所造成的影響如下：

	Private score	Public score
有斷詞	0.80000	0.79767
沒斷詞	0.76279	0.74816

可以發現有斷詞的模型在預測的成績明顯比較好，我想是因為標點符號在檢測惡意語言的前提下有著一定的幫助，純粹用空白隔開詞可能會造成一些訊息的損失以及沒辦法精確的抽出 token。

5. (1%) 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。

兩者的差異如下：

	Sentence 1	Sentence 2
GRU_Model	0.78424	0.63841
BOW_Model	0.83135	0.83135

由上可知，BOW 模型因為兩者的組成相同而得到一樣的成績，而 GRU 模型則會考慮到前後文、考慮句子結構，故兩個句子的成績不同。

數學題

Q1

- $t = 1$
  - 先算出 z, z_i, z_f, z_o

$$z = w \cdot x^1 + b = [0,0,0,1][0,1,0,3]^T + 0 = 3$$
$$z_i = w_i \cdot x^1 + b_i = [100,100,0,0][0,1,0,3]^T - 10 = 90$$
$$z_f = w_f \cdot x^1 + b_f = [-100,-100,0,0][0,1,0,3]^T + 110 = 10$$
$$z_o = w_o \cdot x^1 + b_o = [0,0,100,0][0,1,0,3]^T - 10 = -10$$

  - 再算出 c'

$$c' = f(z_i)g(z) + cf(z_f) = 1 \times 3 + 0 \times 1 = 3$$

  - 最後算出 y_1

$$y_1 = f(z_o)h(c') = 0 \times 3 = 0$$

- $t = 2$
  - 先算出 z, z_i, z_f, z_o

$$z = w \cdot x^2 + b = [0,0,0,1][1,0,1,-2]^T + 0 = -2$$
$$z_i = w_i \cdot x^2 + b_i = [100,100,0,0][1,0,1,-2]^T - 10 = 90$$
$$z_f = w_f \cdot x^2 + b_f = [-100,-100,0,0][1,0,1,-2]^T + 110 = 10$$
$$z_o = w_o \cdot x^2 + b_o = [0,0,100,0][1,0,1,-2]^T - 10 = 90$$

  - 再算出 c' , 用上一個 cell 裡的 c' 當作這邊的 c

$$c' = f(z_i)g(z) + cf(z_f) = 1 \times -2 + 1 \times 3 = 1$$

  - 最後算出 y_2

$$y_2 = f(z_o)h(c') = 1 \times 1 = 1$$

- 以此類推, $t = 3$
  - 先算出 z, z_i, z_f, z_o

$$z = 4$$
$$z_i = 190$$
$$z_f = -90$$
$$z_o = 90$$

  - 再算出 c' , 用上一個 cell 裡的 c' 當作這邊的 c

$$c' = 4$$

  - 最後算出 y_3

$$y_3 = 4$$

- $t = 4$
  - 先算出 z, z_i, z_f, z_o

$$z = 0$$
$$z_i = 90$$
$$z_f = 10$$
$$z_o = 90$$

  - 再算出 c' , 用上一個 cell 裡的 c' 當作這邊的 c

$$c' = 4$$

  - 最後算出 y_4

$$y_4 = 4$$

- $t = 5$
  - 先算出 z, z_i, z_f, z_o

$$z = 2$$
$$z_i = 90$$
$$z_f = 10$$
$$z_o = -10$$

  - 再算出 c' , 用上一個 cell 裡的 c' 當作這邊的 c

$$c' = 6$$

  - 最後算出 y_5

$$y_5 = 0$$

- $t = 6$
  - 先算出 z, z_i, z_f, z_o

$$z = -4$$
$$z_i = -10$$
$$z_f = 110$$
$$z_o = 90$$

  - 再算出 c' , 用上一個 cell 裡的 c' 當作這邊的 c

$$c' = 6$$

  - 最後算出 y_6

$$y_6 = 6$$

- $t = 7$
  - 先算出 z, z_i, z_f, z_o

$$z = 1$$
$$z_i = 190$$
$$z_f = -90$$
$$z_o = 90$$

  - 再算出 c' , 用上一個 cell 裡的 c' 當作這邊的 c

$$c' = 1$$

  - 最後算出 y_7

$$y_7 = 1$$

- $t = 8$
  - 先算出 z, z_i, z_f, z_o

$$z = 2$$
$$z_i = 90$$
$$z_f = 10$$
$$z_o = 90$$

  - 再算出 c' , 用上一個 cell 裡的 c' 當作這邊的 c

$$c' = 3$$

  - 最後算出 y_8

$$y_8 = 3$$

Q2

- 已知下式, 其中 j^* 為向量中為 1 的地方

$$L = -logP(w_t|w_c) = -log\frac{exp\,\mu_{j^*}}{\Sigma_i exp\,\mu_i} = -\mu_{j^*} + log\Sigma_i exp(\mu_i)$$

$$\mu_k = \Sigma_{m=1}^{N}\Sigma_{l=1}^{V} W'^{T}_{mk} W^{T}_{lm} x_l$$

- $\frac{\partial L}{\partial W'^{T}_{ij}}$
  - 因連鎖率

$$\frac{\partial L}{\partial W'^{T}_{ij}} = \Sigma_{k=1}^{V}\frac{\partial L}{\partial \mu_k}\frac{\partial \mu_k}{\partial W'^{T}_{ij}}$$

  - 因只有 $k = j$ 時不為 0, 故

$$\frac{\partial L}{\partial W'^{T}_{ij}} = \frac{\partial L}{\partial \mu_j}\frac{\partial \mu_j}{\partial W'^{T}_{ij}}$$

  - 先算 $\frac{\partial L}{\partial \mu_j}$, 其中 δ_{jj^*} 只有在 $j = j^*$ 為 1, 其餘狀況為 0

$$\frac{\partial L}{\partial \mu_j} = -\delta_{jj^*} + \frac{exp\,\mu_j}{\Sigma_i exp(\mu_i)} = -\delta_{jj^*} + y_j$$

  - 再算 $\frac{\partial \mu_j}{\partial W'^{T}_{ij}}$

$$\frac{\partial \mu_j}{\partial W'^{T}_{ij}} = \Sigma_{k=1}^{V} W^{T}_{ik} x_k$$

  - 故

$$\frac{\partial L}{\partial W'^{T}_{ij}} = (-\delta_{jj^*} + y_j)(\Sigma_{k=1}^{V} W^{T}_{ik} x_k)$$

- $\frac{\partial L}{\partial W^{T}_{ij}}$
  - 因連鎖率

$$\frac{\partial L}{\partial W^{T}_{ij}} = \Sigma_{k=1}^{V}\frac{\partial L}{\partial \mu_k}\frac{\partial \mu_k}{\partial W^{T}_{ij}}$$

  - $\frac{\partial L}{\partial \mu_j}$ 承上
  - 再算 $\frac{\partial \mu_j}{\partial W^{T}_{ij}}$

$$\frac{\partial \mu_j}{\partial W^{T}_{ij}} = W'^{T}_{jk} x_i$$

  - 故

$$\frac{\partial L}{\partial W^{T}_{ij}} = \Sigma_{k=1}^{V}(-\delta_{kk^*} + y_k) W'^{T}_{jk} x_i$$