學號：r08246009　系級：應數碩二　　姓名：許哲維

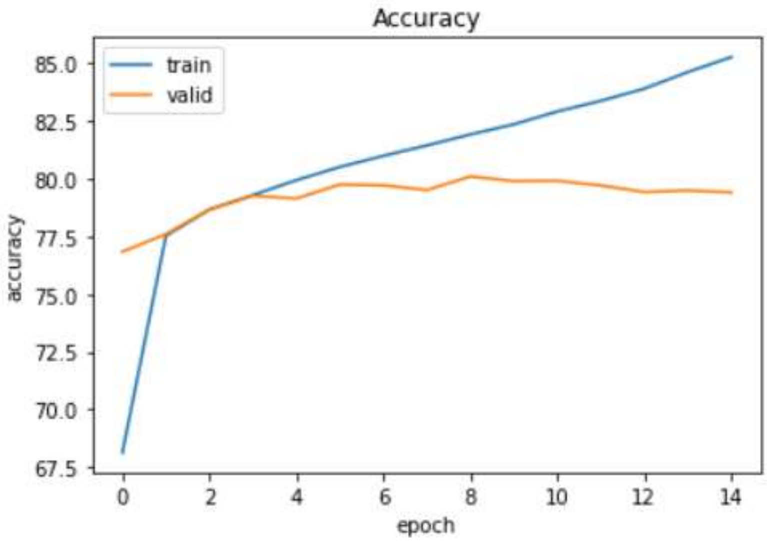
1. **(0.5%) 請說明你實作之 RNN 模型架構及使用的 word embedding 方法，回報模型的正確率並繪出訓練曲線**

   embedding_dim=250, hidden_dim=120

   句子長度設定成sen_len = 37, epoch = 15

   使用Adam with lr = 0.001

   最後的到kaggle正確率0.79460



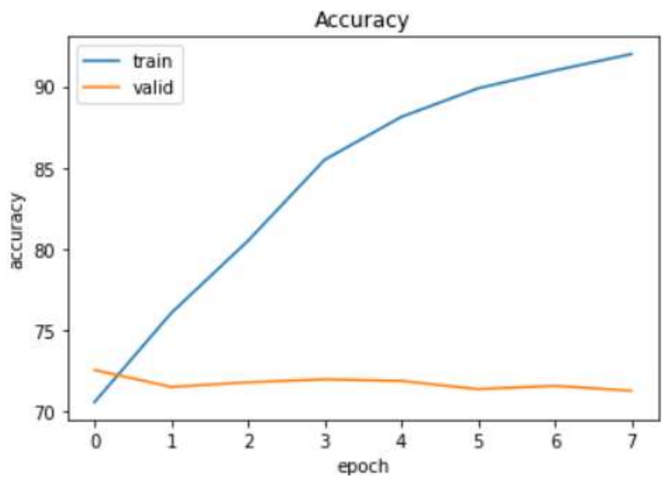
2. **(0.5%) 請實作 BOW+DNN 模型，敘述你的模型架構，回報模型的正確率並繪出訓練曲線。**

   用兩層的layer, sen_length 設定為20

   使用adam with lr=0.001

   epoch=8

   最後得到kaggel上面的正確率是0.7374



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。**

   我用了一些方式增加準度

   1. sen_len 選擇37，我算了一下training set的每句詞與個數，如果設定為30幾乎可以包含所有句子，因此我從30開始往後一個一個試，得到37時的cv score最佳。

   2. embedding_dim=250, hidden_dim=120利用更高維度以及更多層的解述複雜的語句。

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。**

RNN

sent1 = " Today is hot, but I am happy " → 0.7116

sent2 = " I am happy, but today is hot " → 0.5915

BOW

sent1 = " Today is hot, but I am happy " → 0.9571

sent2 = " I am happy, but today is hot " → 0.8649

第一句話是強調” 今天很熱” ，第二句話則是” 我很開心” ；實際上是完全不同的意思，但是如果使用bow+dnn的話，對於這兩句話來說詞語出現頻率是相同的，因此比較沒有辦法區分出來兩句話的差異。

5. **(3%)Math problem**

r08246009
許哲維

HW4.

1.

(a).

The principal axes are eigenvectors :

$$v_1 = [0.399, -0.678, -0.616]$$

$$v_2 = [0.337, 0.734, -0.588]$$

$$v_3 = [-0.85, -0.027, -0.522]$$

(b).

Let $T = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$, the principal components of each samples are $y_i = T(x_i - \mu)$ $\forall i = 1, \dots, 10$.

$$y_1 = (3.41, -4.83, 4.85)$$

$$y_2 = (-0.68, -0.59, 1.08)$$

$$y_3 = (-6.26, -0.34, -0.25)$$

$$y_4 = (-1.88, -1.60, 3.64)$$

$$y_5 = (-2.05, 0.91, 1.64)$$

$$y_6 = (5.69, -0.15, 0.73)$$

$$y_7 = (-1.15, -1.25, -5.26)$$

$$y_8 = (1.38, 1.42, 4.03)$$

$$y_9 = (3.53, -1.019, -5.31)$$

$$y_{10} = (-1.55, 2.46, -5.15)$$ #.

(c)

The MSE $= tr(\Sigma) - \lambda_1 \lambda_2$

(Σ : the covariance of the data).

MSE $= 16.99$ #.

2.

(a).

Symmetric:

$$(AA^T)^T = AA^T$$

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

positive semi definite:

$$x^TAA^Tx = \|A^Tx\|^2 \geq 0 \quad \forall x \in \mathbb{R}^m \ \#$$

$$x^TA^TAx = \|Ax\|^2 \geq 0 \quad \forall x \in \mathbb{R}^n$$

Hence, A^TA and AA^T are both p.s.d #.

The same non-zero eigenvalues:

Since A^TA is symmetric, we have.

$$A^TA = P\Lambda P^T$$

$$\Rightarrow A^TAx = \lambda x \quad \forall x \in \mathbb{R}^n$$

$$\Rightarrow AA^TAx = \lambda Ax.$$

(let $y = Ax$) $\Rightarrow AA^Ty = \lambda y$, hence AA^T and A^TA share the same non-zero eigenvalue #.

(c).

Since Σ is symmetric and p.s.d, we have

$\Sigma = T \Lambda T^T$, Λ: diagonal matrix associate with the eigenvalues of Σ

T: orthonormal matrix, s.t. $T^T T = I$

Consider

$$tr(T^T \Lambda T) = tr(\Lambda) = \sum_{i=1}^{k} \lambda_i \quad (\lambda_i : \text{eigenvalues}).$$

$\forall k = 1, \dots, m.$ #.

(c).

$$tr(\Phi^T \Sigma \Phi) = tr(\Sigma \Phi \Phi^T).$$

$\Phi \Phi^T$ is also symmetric, by the decomposition, we have eigenvalues of $\Phi \Phi^T \Rightarrow \mu_1 \geq \dots \geq \mu_m$

the eigenvalues of Σ are $\lambda_1 \geq \dots \geq \lambda_m$.

Consider $tr(\Sigma(\Phi\Phi^T)) \leq \sum_{i=1}^{m} \mu_i \lambda_i$

$\quad\quad\quad \geq \quad\quad\quad\quad \downarrow$

$\sum_{i=1}^{m} \lambda_i \mu_{m-i+1}$ \quad Von Neumann's inequality.

$\quad \overset{0}{\|} \quad\quad\quad \overset{0}{\|} \quad\quad\quad\quad \overset{1}{\|} \quad\quad\quad \overset{1}{\|}$

$= \lambda_1 \mu_m + \lambda_2 \mu_{m-1} + \dots + \lambda_{m-k+1} \mu_k + \dots + \lambda_m \mu_1$

$= \lambda_{m-k+1} + \dots + \lambda_m$

Hence, we get the lower bound of $tr(\Phi^T \Sigma \Phi)$ which is

$\lambda_{m-k+1} + \dots + \lambda_m$

No.
Date

We denote the eigenvector of Σ corresponding to the eigenvalues $\lambda_1, \dots, \lambda_m$ be $v_1, \dots, v_m$, $v_i \in \mathbb{R}^m$

Let $\Phi = \begin{pmatrix} v_{m-k+1} & \cdots & v_m \end{pmatrix}_{m\times k}$.

then $tr(\Phi^T \Sigma \Phi) = \lambda_{m-k+1} + \cdots + \lambda_m$, which reaches the minimum of $tr(\Phi^T \Sigma \Phi)$ #.

(b).

Consider points to be.

$$x_1 = \begin{pmatrix} \sqrt{m} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad x_2 = \begin{pmatrix} -\sqrt{m} \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \dots, x_{2m-1} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \sqrt{m} \end{pmatrix}, x_{2m} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ -\sqrt{m} \end{pmatrix}$$

Let $n = 2m$.

$$u = \frac{1}{n} \sum_{i=1}^{n} x_i = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} = 0$$

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

$$= \frac{1}{n} \sum_{i=1}^{n} x_i x_i^T = \frac{1}{n} \begin{pmatrix} 2m & & 0 \\ & 2m & \\ 0 & & \ddots \; 2m \end{pmatrix}_{m\times m}$$

$= I_m$ which is obviously symmetric and p.s.d.