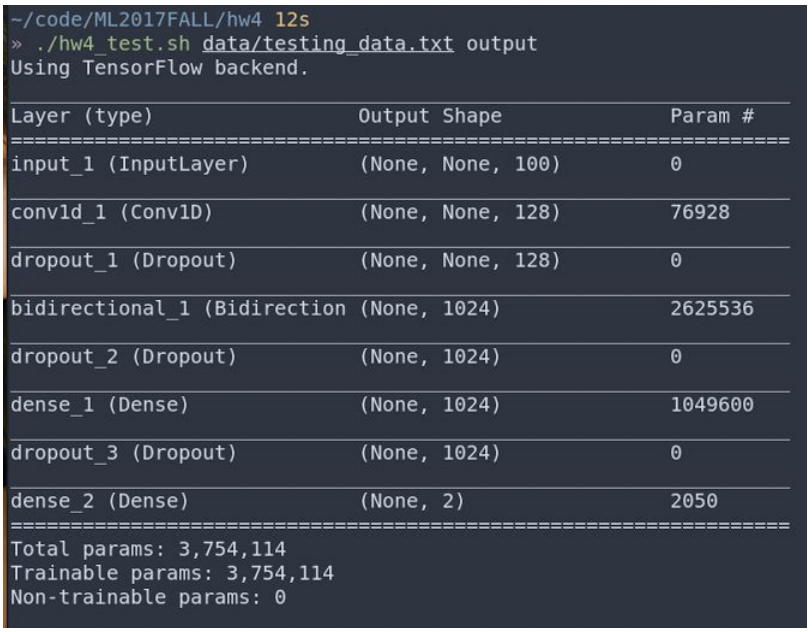
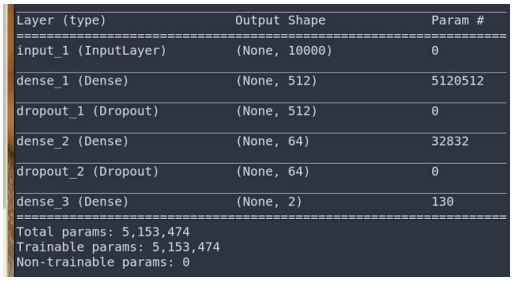
學號：B03705049 系級： 資管四 姓名：廖寬璿

1. (1%) 請說明你實作的 RNN model，其模型架構、訓練過程和準確率為何？
   (Collaborators:)



用gensim Word2Vec (iter=25) 再進以上的model
training 用 K-fold (K=4) train 4次
準確率： private: 0.80713, public:0.80859

2. (1%) 請說明你實作的 BOW model，其模型架構、訓練過程和準確率為何？
   (Collaborators:)



用keras tokenizer斷詞和轉成binary incidence matrix再進以上model
training 4 個epoch
準確率： private: 0.79862, public: 0.80105

3. (1%) 請比較bag of word與RNN兩種不同model對於"today is a good day, but it is hot"與"today is hot, but it is a good day"這兩句的情緒分數，並討論造成差異的原因。
   (Collaborators:)

BOW: [0.22046296, 0.77953702] [0.22046296, 0.77953702]
RNN: [0.35092029, 0.64907968] [0.13977341, 0.86022657]

這兩句用的字都一樣對bag of words來說是一樣的句字, 分數也一樣
但因為RNN會看字的順序,評分的結果既會不一樣

4. (1%) 請比較"有無"包含標點符號兩種不同tokenize的方式，並討論兩者對準確率的影響。

有標點符號：private: 0.80713, public: 0.80859
無標點符號：private: 0.80611, public: 0.80735
標點符號在文字裡也是用來表達情緒的, 加了對model有些許的幫助

5. (1%) 請描述在你的semi-supervised方法是如何標記label，並比較有無semi-surpervised training對準確率的影響。
(Collaborators:)

threshold > 0.97 或 threshold > 0.95 的加label
但這樣好像只會讓model更容易overfit
準確率大概是：private: 0.79266, public: 0.79451