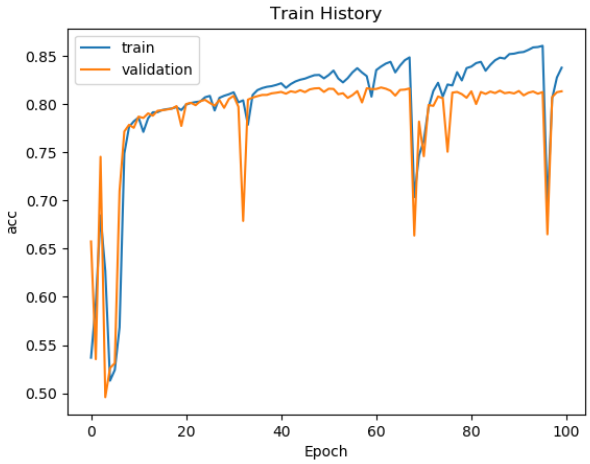
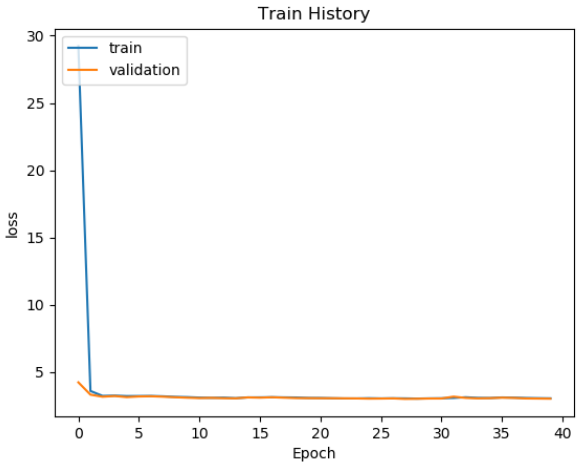
1. (1%) 請說明你實作的 RNN model，其模型架構、訓練過程和準確率為何？
   (Collaborators:)
   答：

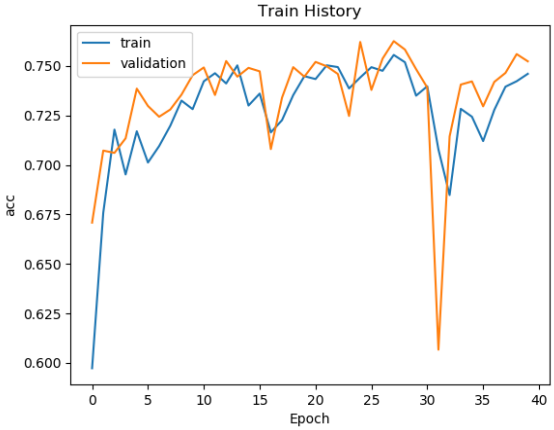
```
Layer (type)                 Output Shape              Param #
=================================================================
embedding_1(Embedding)       (None, None, 260)         14376440
_________________________________________________________________
lstm_1(LSTM)                 (None, 256)               529408
_________________________________________________________________
dense_1(Dense)               (None, 1)                 257
=================================================================
Total params: 14,906,105
Trainable params: 529,665
Non-trainable params: 14,376,440
```

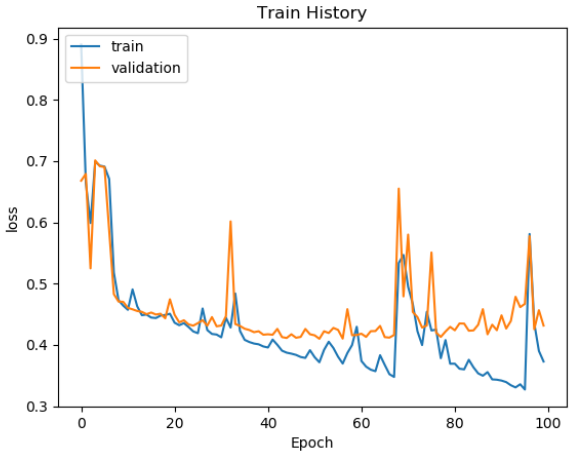




Accruacy: 0.82

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:)
   答：





```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense_3 (Dense)              (None, 1024)              20481024
_________________________________________________________________
dense_4 (Dense)              (None, 256)               262400
_________________________________________________________________
dropout_1 (Dropout)          (None, 256)               0
_________________________________________________________________
dense_5 (Dense)              (None, 1)                 257
=================================================================
Total params: 20,743,681
Trainable params: 20,743,681
Non-trainable params: 0
```

Accruacy: 0.74

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:)
   答：

   BoW 只有使用到句子中的各個單字出現頻率，沒有順序性；因此最終的 score(情緒分數)會是一樣的。

Model	today is a good day, but it is hot	today is hot, but it is a good day
RNN(LSTM)	score:0.67884731	score:0.94832146
BoW+DNN	score:0.67285758	score:0.67285758

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。
   (Collaborators:)
   答：

   無標點符號之 **Accruacy: 0.82235**　　有標點符號之 **Accruacy: 0.81932**

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。
   答：

   accruacy = 0.8066,which is worse than pure supervised learning.

   標記方法: (建字庫)->利用 gensim.model.Word2Vec 建立 embedding matrix 及 Word2Vec model ->利用有 label 之 training Keras.LSTM 訓練模型 ->預測

training_no_label ->通過 Labeling Criterion 的再取出來合併為 training 重新訓練。

Labeling criterion: filter out samples which its prediction of class 1 probability is greater than 0.95 or less than 0.45