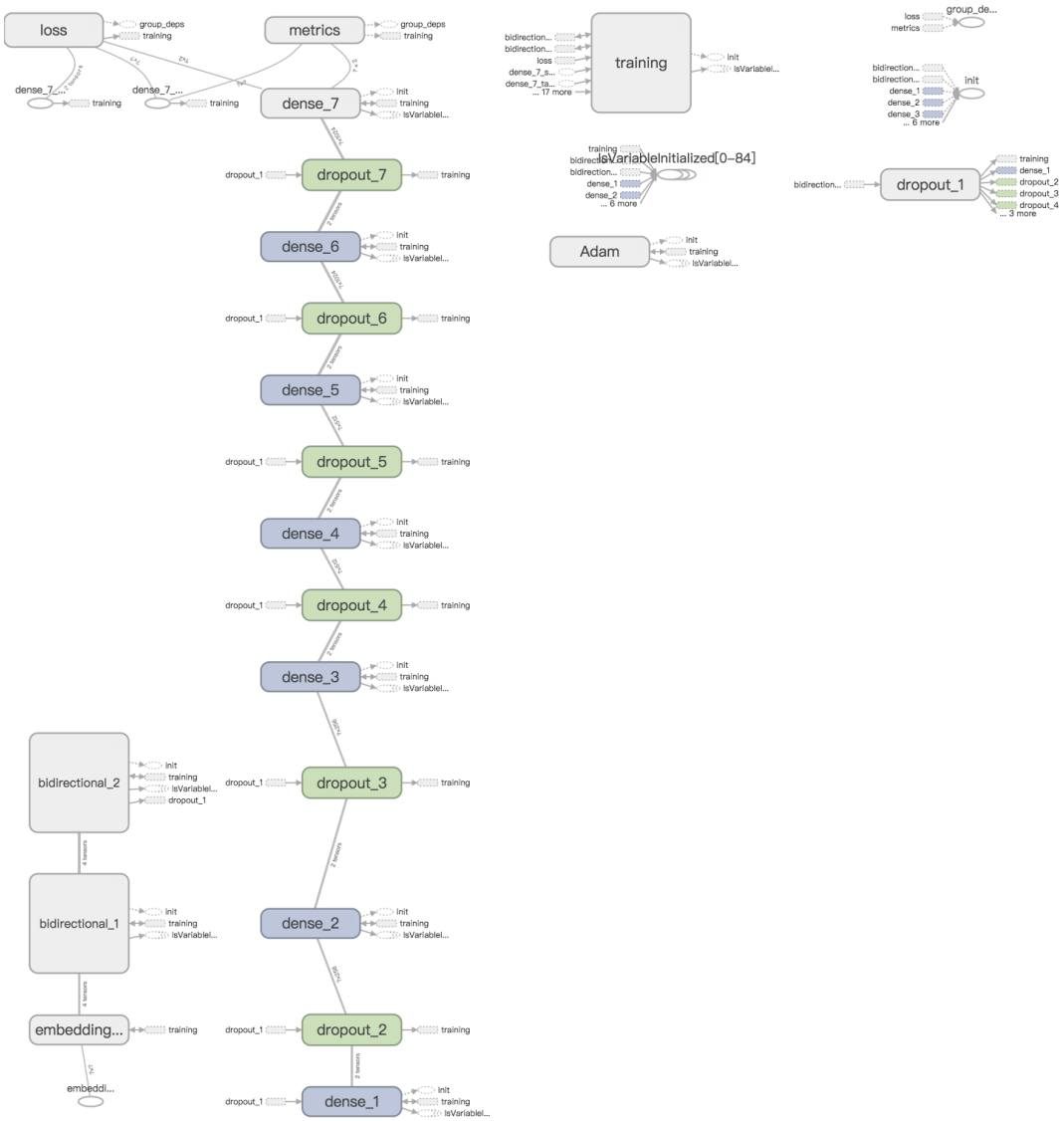
學號：R06922129 系級： 資工碩一 姓名：丁縉楷

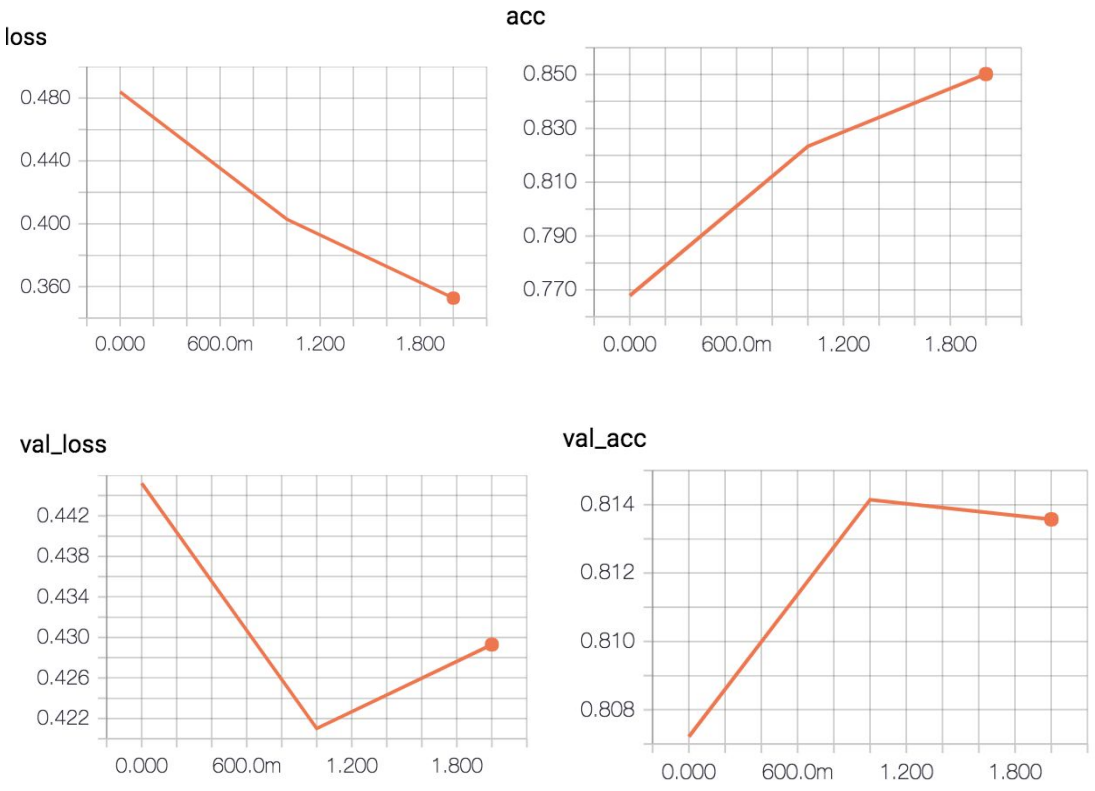
1. (1%) 請說明你實作的 RNN model，其模型架構、訓練過程和準確率為何？(Collaborators: 葉韋辰、黃禹程)

答：

model架構：



word2vec的dim設為100，通常train到第二個epoch, validation accuracy就會最高

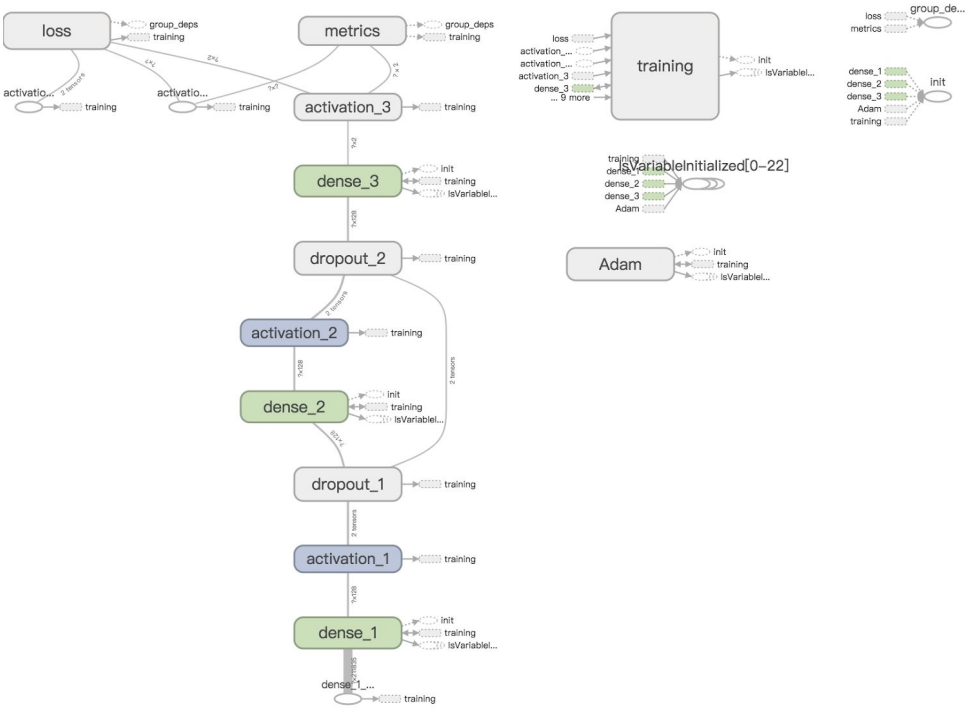


public score	private score
0.81890	0.81731

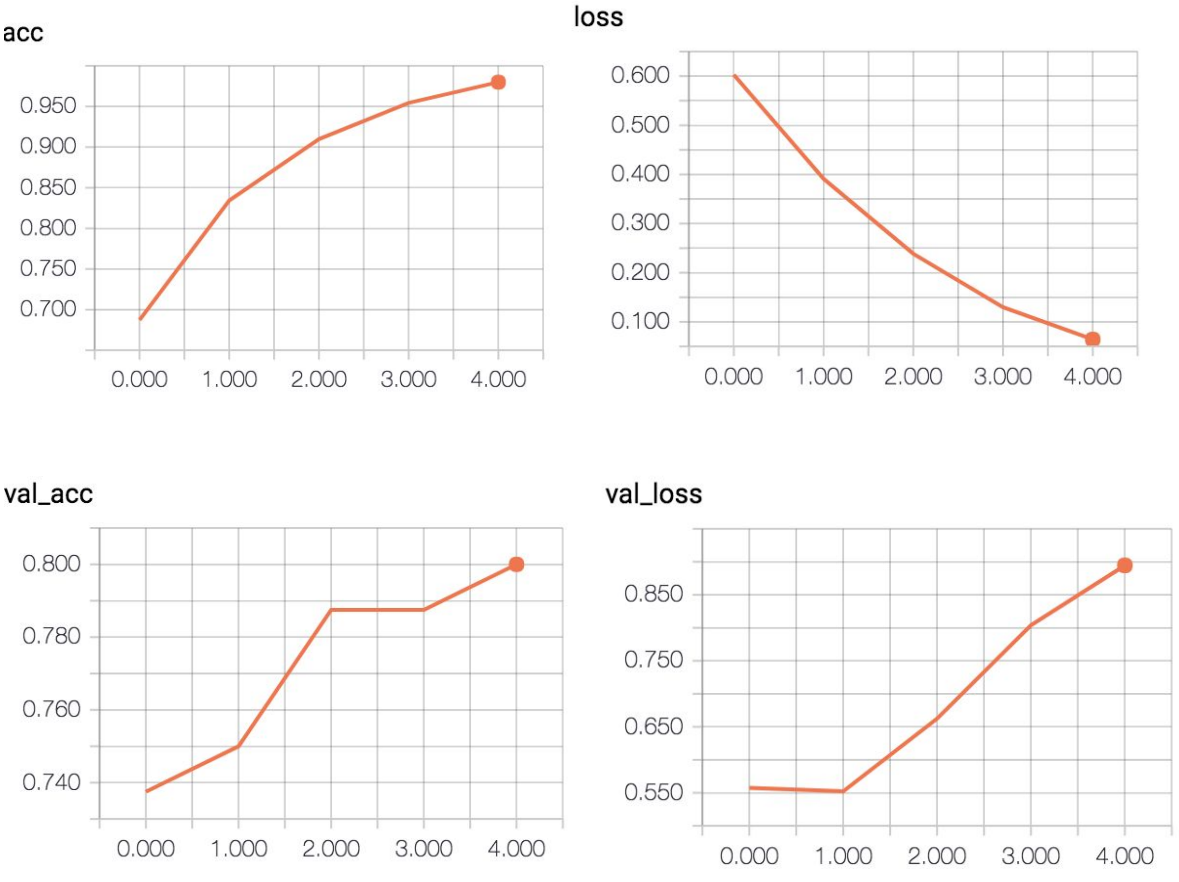
2. (1%) 請說明你實作的 BOW model，其模型架構、訓練過程和準確率為何？ (Collaborators:葉韋辰、黃禹程)

答：

model架構：



訓練過程：因為是bag of words，所以先對data做stemming



可以看到用bow模型很容易會overfitting，雖然val acc可以到0.8，但loss非常大

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:葉韋辰、黃禹程)

   答：

RNN:

today is a good day, ...	0.50820035	0.49179965
today is hot, ...	0.10042092	0.89957905

BOW:

today is a good day, ...	0.10367276	0.8963272
today is hot, ...	0.10367276	0.8963272

bag of words 模型不考慮詞之間的順序，因此這兩句的input是相同的，反之RNN模型有考慮順序，所以可以比較精準的判斷出語意的差別

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:葉韋辰、黃禹程)

   答：

	public score	private score
無標點符號	0.81890	0.81731
有標點符號	0.81416	0.81394

實作結果是filter掉標點符號效果較好，推測是加入標點符號可能使建字典的時候變得雜亂，或是逗號句號等標點符號沒有和語意有太直接的關聯。

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:葉韋辰、黃禹程)

答：

標記方法為不設定閾值，直接label全部的data再下去train，結果如下

	public score	private score
沒有semi-supervised	0.81890	0.81731
semi-supervised	0.81370	0.81414

semi-supervised對準確率沒有幫助，可能是這個方法只適用於model很強的時候，準確率本來就不高時反而有反效果。