sample code大概在10個epoch之後就會開始overfitting

學號： 系級： 姓名：

1. (1%) 請以block diagram或是文字的方式說明這次表現最好的 model 使用哪些layer module(如 Conv/RNN/Linear 和各類 normalization layer) 及連接方式(如一般forward 或是使用 skip/residual connection)，並概念性逐項說明選用該 layer module 的理由。
2. (1%) 請比較 word2vec embedding layer 初始設為 non-trainable/trainable 的差別，列上兩者在 validation/public private testing 的結果，並嘗試在訓練過程中設置一策略改變 non-trainable/trainable 設定，描述自己判斷改變設定的機制以及該結果。

單純設為  
由於發現將word2vec embedding layer設為trainable很容易有overfitting的問題，因此選擇以先設為non-trainable，validation data的準確率沒有繼續上升後，再改設為trainable，提高準確率。測試後發現，通常在改設為trainable後的第一個epoch跑完後，validation data的準確率會最高。

1. (1%) 請敘述你如何對文字資料進行前處理，並概念性的描述你在資料中觀察到什麼因此你決定採用這些處理，並描述使用這些處理時作細節，以及比較其實際結果，該結果可以不用具備真正改進。如果你沒有作任何處理，請給出一段具體描述來說服我們為什麼不做處理可以得到好的結果，這個理由不能是因為表現比較好。  
   這次我沒有做任何文字資料的前處理，原本有想過把標點符號拿掉，但拿掉後準確率是下降的，推測標點符號也是有它的情緒上的意義(特別是驚嘆號、刪節號等等)，同時像是逗號雖然對情緒可能影響不大，但對於句子的分段也是有重要的意義，因此拿掉標點符號的前處理認為不是個好選擇。我認為以助教原本的code以空白分割，就已經滿適合的了，也沒有想到其他更好的前處理方式了。
2. (1%) 請「自行設計」兩句具有相同單字但擺放位置不同的語句，使得你表現最好的模型產生出不同的預測結果，例如 "Today is hot, but I am happy" 與 "I am happy, but today is hot"，並討論造成差異的原因。  
   用我最好的模型，"Today is hot, but I am happy" 與 "I am happy, but today is hot"這兩句是可以準確分辨的，前者的label為1，後者為0。因為這次使用的模型是LSTM，是會將前面看過的字都記錄下來的，因此文字的順序造成的語意差異是能夠被準確分辨的。
3. (4%) Math problem:

https://hackmd.io/@hAe95tLdTVqEePbZsJyqrw/BkWSTuqPF