

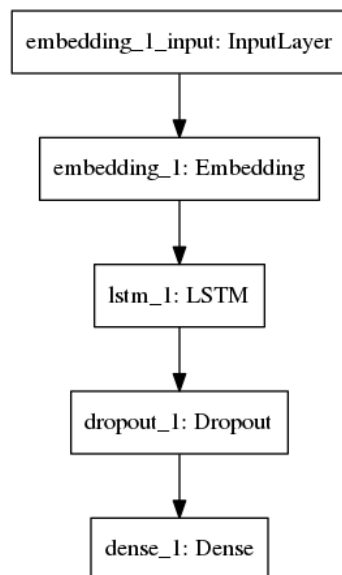
學號：R06922093 系級：資工碩一 姓名：陳禹齊

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？

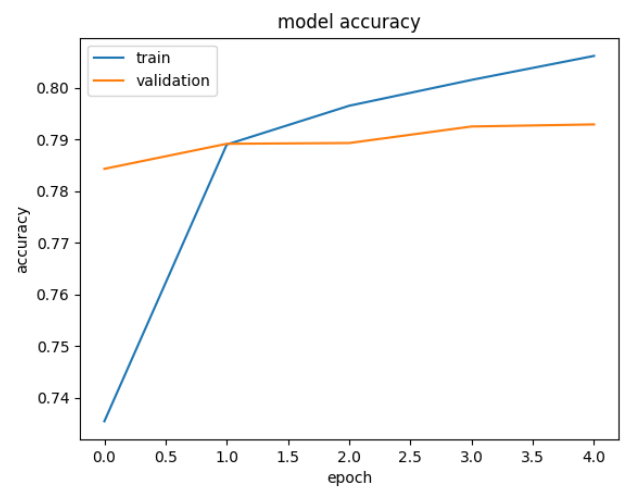
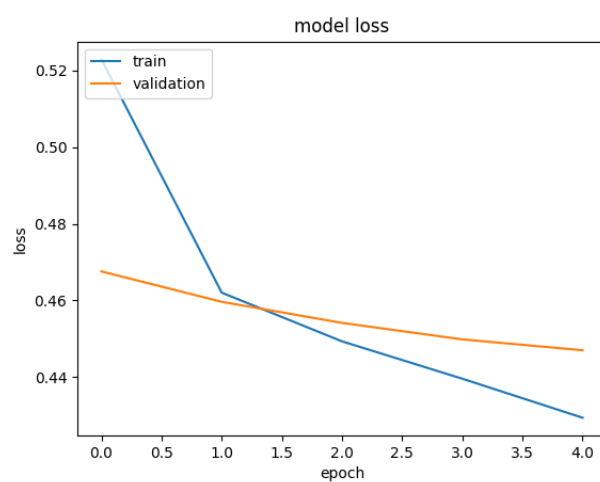
(Collaborators: R0692096 洪子翔)

答：kaggle public score:0.79963

模型架構:



訓練過程:

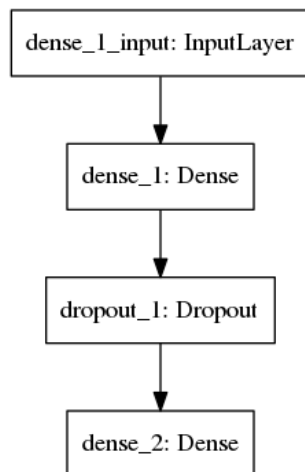


2. (1%) 請說明你實作的 BOW model，其模型架構、訓練過程和準確率為何？

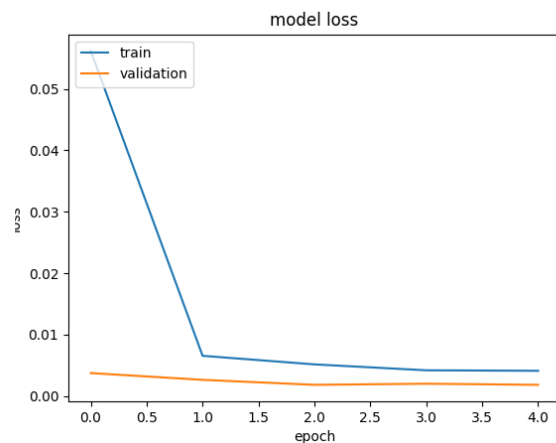
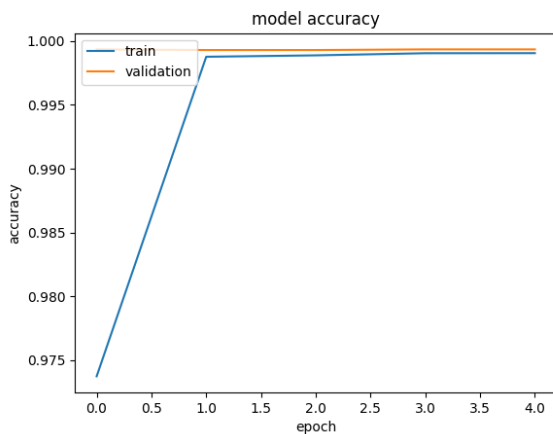
(Collaborators: R0692096 洪子翔)

答：kaggle public score:0.57496

模型架構:



訓練過程:



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: R0692096 洪子翔)

答：

用 bag of word predict 出來的結果,兩句話都一樣是 0.77192,結果都是正面,會造成這種現象的原因,是因為組成這兩句話的詞都一樣,所以等於送進去 predict 的 input 是一樣的,才會有一樣的結果。

用 RNN 出來的結果就會不一樣,因為 RNN 會考慮到詞與詞之間的關係,兩個句子 predict 的結果,第一句話是 0.54678,結果是正面,不過我想應該是我的 Model 還不夠好,差一點點就可以判斷成負面,第二句話是的結果 0.89092,就是正面情緒。

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: R0692096 洪子翔)

答：透過 `keras` 的 `tokenizer` 的 `filter`, 可以去掉標點符號, 不過實作之後對於準確率的提升也沒有很大的提升, 只有 0.000 幾的影響。

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

(Collaborators: R0692096 洪子翔)

答：先 `pretrain` 一個 `model`, 然後把沒有 `label` 的 `data`, 用 `pretrain` 的 `model` 去做 `predict`, 然後直接把 `predict` 的結果當作那些 `data` 的 `label`, `predict` 結果大於 0.5 `label` 成 1 小於 0.5 的 `label` 成 0, 然後再把那些 `data` 繼續放進去 `train`, 有做 `semi-supervised` 的準確率有略微上升一點, 不過還是沒有顯著的提升。