ML hw4 Report

學號：B05611033 系級： 生機二 姓名：杜杰翰

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？

我將句子中一些“can ‘ t”, “wouldn ‘ t”之類的字結合，並將label及unlabel data都丟進gensim的Word2Vec並設window=5, min_count=5，將每個字轉換成一個200維的vector，並針對長度不足或沒在dictionary裡面的字補上0向量。

之後我使用三個model做ensemble且在三個model中的GRU層都有使用0.3的dropout，optimization都是使用Adam，learning rate=0.001並使用L2 regularization lambda=0.001。

Model 1

Layer Name	Input_shape	Output_shape	Comments
GRU_1	(40, 200)	(40, 200*2)	Bidirectional
GRU_2	(40, 200*2)	(40, 16*2)	Bidirectional
Linear_1	(40*16*2)	(40*16)	Swish
Linear_2	(40*16)	(40*8)	Swish
Linear_3	(40*8)	(40*4)	Swish
Output	(40*4)	(2)	Softmax

Model 2

Layer Name	Input_shape	Output_shape	Comments
GRU_1	(40, 200)	(40, 512*2)	Bidirectional
GRU_2	(40, 512*2)	(40, 8*2)	Bidirectional
Linear_1	(40*8*2)	(40*8)	Swish
Linear_2	(40*8)	(40*4)	Swish
Linear_3	(40*4)	(40*2)	Swish
Output	(40*2)	(2)	Softmax

Model 3

Layer Name	Input_shape	Output_shape	Comments
GRU_1	(40, 200)	(40, 128*2)	Bidirectional
GRU_2	(40, 128*2)	(40, 64*2)	Bidirectional
Linear_1	(40*64*2)	(40*64)	Swish
Linear_2	(40*64)	(40*32)	Swish
Linear_3	(40*32)	(40*16)	Swish
Output	(40*16)	(2)	Softmax

之後使用CrossEntropy來計算loss。經過15個epoch之後，在kaggle上的public成績為0.82337。

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？

我將每個句子表達成BOW並使用全部出現的字詞，共82945維，並在最後用softmax輸出成2維做CrossEntropy，其中沒有使用任何dropout。

我的optimization使用Adam，learning rate = 0.001並使用L2 regularization lambda=0.001。

Layer Name	Input_shape	Output_shape	Activation
Linear_1	(82945)	(768)	Swish
Linear_2	(768)	(384)	Swish
Linear_3	(384)	(128)	Swish
Output	(128)	(2)	Softmax

經過8個epoch之後在kaggle上public的結果是0.79355。

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。

	BOW	RNN
today is a good day, but it is hot	(0.4241, 0.5759)	(0.4776, 0.5224)
today is hot, but it is a good day	(0.4241, 0.5759)	(0.0620, 0.9380)

從結果可以看出因兩句中各單字出現的次數都一樣，所以bag of words分不出其差異，而RNN因考慮到字詞出現的先後順序，因而產生了不同的結果，而在此RNN對兩句話可能還是會歸在同一類，但是可以明顯地看出兩句的分數有著明顯的差距。

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。

我使用第一題的model2 來做本題，除了有無標點符號之外都一樣。

	public score
有標點符號	0.81866
無標點符號	0.81148

從結果來看，有標點符號的結果比較好，我想是因為一些如”!, ?”之類的標點符號可以表現出句子的情緒，將其移除反而會使判斷label的標的減少，因而產生較差的結果。

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。

我先用第一題的model2 train 6個epoch之後對unlabel data直接predict他的label，再將所有unlabel data連同原本的data一起丟回去model裡繼續train 3個epoch，最後對testing data predict的結果在kaggle上的成績為0.81325。

然而在沒加入unlabel data前的準確率為0.81866，因此我的方法在加入unlabel data後會使準確率下降一些，其中有可能是因為我標記的unlabel data中有很多label是錯誤的，使得用這些錯誤的data去做training會使結果變差。