Assignment2 Report
(COMP-5900V 2021 Winter)

Ao Peng

1 Part 1

This part includes 4 architecture which consists of a two-level bi-LSTM model and 3 RNN models containing 2-level bi-RNN, 1-level bi-RNN and normal RNN. The task is to analyse the sentiment based on IMDB movie reviews. For each model, the training epoch is set to 5. The number of parameters in RNN/LSTM layers and test accuracy show in Table. 1

Here are the information of the structure of each model.

1) 2-level bi-LSTM

```
MyRNN(
  (embedding): Embedding(25002, 100, padding_idx=1)
  (lstm): LSTM(100, 200, num_layers=2, dropout=0.5, bidirectional=True)
  (fc): Linear(in_features=400, out_features=1, bias=True)
  (dropout): Dropout(p=0.5, inplace=False)
)
```

2) 2-level bi-RNN

```
biRNN_2(
  (embedding): Embedding(25002, 100, padding_idx=1)
  (rnn): RNN(100, 200, num_layers=2, dropout=0.5, bidirectional=True)
  (fc): Linear(in_features=400, out_features=1, bias=True)
  (dropout): Dropout(p=0.5, inplace=False)
)
```

3) bi-RNN

```
biRNN_1(
    (embedding): Embedding(25002, 100, padding_idx=1)
    (rnn): RNN(100, 200, dropout=0.5, bidirectional=True)
    (fc): Linear(in_features=400, out_features=1, bias=True)
    (dropout): Dropout(p=0.5, inplace=False)
)
```

4) RNN

```
RNN_1(
  (embedding): Embedding(25002, 100, padding_idx=1)
  (rnn): RNN(100, 200, dropout=0.5)
  (fc): Linear(in_features=200, out_features=1, bias=True)
  (dropout): Dropout(p=0.5, inplace=False)
)
```

	Total # of parameters in RNN/LSTM layer(s)	Test accuracy %
2-level bi-LSTM	1,446,400	85.81
2-level bi-RNN	361,600	50.73
bi-RNN	120,800	49.51
RNN	60,400	69.49

Table 1: Results

2 Part 2

2.1 Q2

Set

$$F = \frac{1}{2}\|f_a - f_p\|^2 + \alpha - \frac{1}{2}\|f_a - f_n\|^2$$

Therefore

$$L^T = \max(0, F)$$
$$\frac{\partial L^T}{\partial f_a} = \frac{\partial L^T}{\partial F}\frac{\partial F}{\partial f_a}$$
$$\frac{\partial L^T}{\partial f_p} = \frac{\partial L^T}{\partial F}\frac{\partial F}{\partial f_p}$$
$$\frac{\partial L^T}{\partial f_n} = \frac{\partial L^T}{\partial F}\frac{\partial F}{\partial f_n}$$

1) $\mathbf{F < 0}$

$$L^t \equiv 0 \Longrightarrow \frac{\partial L^T}{\partial f_a} = \frac{\partial L^T}{\partial f_p} = \frac{\partial L^T}{\partial f_n} = 0$$

2) $\mathbf{F = 0}$

$$L^T = \max(0, F) = 0$$
$$\lim_{\Delta h \to 0^-} \frac{f(F + \Delta h) - f(F)}{\Delta h} = 0 \tag{1}$$
$$\lim_{\Delta h \to 0^+} \frac{f(F + \Delta h) - f(F)}{\Delta h} = 1 \tag{2}$$

(1)≠(2)
Therefore, F cannot be differentiated when $F = 0$

2.2 Q3

F > 0

$$L^T = F \Longrightarrow \frac{\partial L^T}{\partial F} = 1$$
$$\frac{\partial L^T}{\partial f_a} = f_n - f_p$$
$$\frac{\partial L^T}{\partial f_p} = f_p - f_a$$
$$\frac{\partial L^T}{\partial f_n} = f_a - f_n$$

2.3 Q4

There are 8 unique triplets. As in this dataset, there are only 2 different values. The combination shows below.

$$\begin{array}{ll} (2,2,2) & (1,1,1) \\ (2,2,1) & (1,1,2) \\ (2,1,1) & (1,2,1) \\ (2,1,2) & (1,2,2) \end{array}$$

2.4 Q5

2.4.1 Q5.1

$$\frac{\partial L^1}{\partial f_1} = f_5 - f_2$$
$$\frac{\partial L^1}{\partial f_2} = f_2 - f_1$$
$$\frac{\partial L^1}{\partial f_3} = 0$$
$$\frac{\partial L^1}{\partial f_4} = 0$$
$$\frac{\partial L^1}{\partial f_5} = f_1 - f_5$$
$$\frac{\partial L^1}{\partial f_6} = 0$$

2.4.2 Q5.2

$$\frac{\partial L^2}{\partial f_1} = f_4 - f_3$$
$$\frac{\partial L^2}{\partial f_2} = 0$$
$$\frac{\partial L^2}{\partial f_3} = f_3 - f_1$$
$$\frac{\partial L^2}{\partial f_4} = f_1 - f_4$$
$$\frac{\partial L^2}{\partial f_5} = 0$$
$$\frac{\partial L^2}{\partial f_6} = 0$$

2.4.3 Q5.3

$$\frac{\partial L^3}{\partial f_1} = f_1 - f_2$$
$$\frac{\partial L^3}{\partial f_2} = f_6 - f_1$$
$$\frac{\partial L^3}{\partial f_3} = 0$$
$$\frac{\partial L^3}{\partial f_4} = 0$$
$$\frac{\partial L^3}{\partial f_5} = 0$$
$$\frac{\partial L^3}{\partial f_6} = f_2 - f_6$$

2.4.4 Q5.4

$$\frac{\partial L^4}{\partial f_1} = 0$$
$$\frac{\partial L^4}{\partial f_2} = f_4 - f_3$$
$$\frac{\partial L^4}{\partial f_3} = f_3 - f_2$$
$$\frac{\partial L^4}{\partial f_4} = f_2 - f_4$$

$$\frac{\partial L^4}{\partial f_5} = 0$$
$$\frac{\partial L^4}{\partial f_6} = 0$$

2.4.5 Q5.5

$$\frac{\partial L^5}{\partial f_1} = f_1 - f_3$$
$$\frac{\partial L^5}{\partial f_2} = 0$$
$$\frac{\partial L^5}{\partial f_3} = f_5 - f_1$$
$$\frac{\partial L^5}{\partial f_4} = 0$$
$$\frac{\partial L^5}{\partial f_5} = f_3 - f_5$$
$$\frac{\partial L^5}{\partial f_6} = 0$$

2.4.6 Q5.6

$$\frac{\partial L^6}{\partial f_1} = 0$$
$$\frac{\partial L^6}{\partial f_2} = f_5 - f_2$$
$$\frac{\partial L^6}{\partial f_3} = 0$$
$$\frac{\partial L^6}{\partial f_4} = 0$$
$$\frac{\partial L^6}{\partial f_5} = f_2 - f_6$$
$$\frac{\partial L^6}{\partial f_6} = f_6 - f_5$$

2.4.7 Q5.7

$$\frac{\partial L^7}{\partial f_1} = 0$$
$$\frac{\partial L^7}{\partial f_2} = 0$$
$$\frac{\partial L^7}{\partial f_3} = f_6 - f_3$$
$$\frac{\partial L^7}{\partial f_4} = f_4 - f_6$$
$$\frac{\partial L^7}{\partial f_5} = 0$$
$$\frac{\partial L^7}{\partial f_6} = f_3 - f_4$$

2.4.8 Q5.8

$$\frac{\partial L^8}{\partial f_1} = 0$$
$$\frac{\partial L^8}{\partial f_2} = f_6 - f_2$$
$$\frac{\partial L^8}{\partial f_3} = 0$$
$$\frac{\partial L^8}{\partial f_4} = f_4 - f_6$$
$$\frac{\partial L^8}{\partial f_5} = 0$$
$$\frac{\partial L^8}{\partial f_6} = f_2 - f_4$$

2.5 Q6

$$\frac{\partial L}{\partial f_1} = 2(f_1 - f_2 - f_3) + f_4 + f_5$$
$$\frac{\partial L}{\partial f_2} = 2(f_6 - f_1) - f_2 - f_3 + f_4 + f_5$$
$$\frac{\partial L}{\partial f_3} = f_3 + f_5 + f_6 - 2f_1 - f_2$$
$$\frac{\partial L}{\partial f_4} = f_1 + f_2 - 2f_6$$

$$\frac{\partial L}{\partial f_5} = f_1 + f_2 + f_3 - 2f_5 - f_6$$

$$\frac{\partial L}{\partial f_6} = 2(f_2 - f_4) + f_3 - f_5$$