Task 1

1.1 Model – LSTM

LSTM, or Long-short term memory network is used to predict the daily confirmed new cases. It is widely used to address long-term dependencies. The key idea is to use memory cells and gates:

$$c_{(t)} = f_t c_{(t-1)} + i_t a^{(t)}$$

$c_{(t)}$ is memory state at t; $a^{(t)}$ is new input as t.

If the forget gate f_t is open (i.e, 1) and the input gate it is closed (i.e., 0), the current memory is kept. If the forget gate f_t is closed (i.e, 0) and the input gate it is open (i.e., 1), the current memory is erased and replaced by new input. If we can learn ft and it from data, then we can automatically determine how much history to remember/forget.

f_t is determined based on current input x(t) and previous hidden unit h(t−1):

$$\mathbf{f}_t = \sigma\left(\mathbf{W}_f \mathbf{h}^{(t-1)} + \mathbf{U}_f \mathbf{x}^{(t)} + \mathbf{b}_f\right)$$

i_t is also determined based on current input x(t) and previous hidden unit h(t −1) :

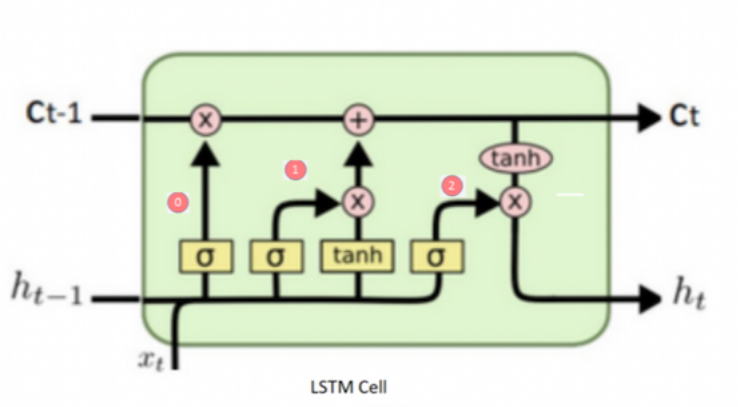
$$\mathbf{i}_t = \sigma\left(\mathbf{W}_i \mathbf{h}^{(t-1)} + \mathbf{U}_i \mathbf{x}^{(t)} + \mathbf{b}_i\right)$$

And the output gate is defined as :

$$\mathbf{o}_t = \sigma\left(\mathbf{W}_o \mathbf{h}^{(t-1)} + \mathbf{U}_o \mathbf{x}^{(t)} + \mathbf{b}_o\right)$$

Following figure demonstrates the structure of LSTM cell.

Above materials are referred from MSBD5012 Machine Learning Slides.



LSTM Cell

1.2 Implement details:

First, extract data column which belongs to US. Since the scale of data varies, I perform the Z-score normalization to process the data. The mean and std are recorded.

Since the task is to predict the daily new confirmed in next 7 days, the window size of ground truth data is set to 7. And the window size of training data (e.g., x) is set to 21. In other words, three weeks of data is used to predict the data of next week. Consider the time is very limited, a simple LSTM model is used:

```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
lstm_1 (LSTM)                (None, 256)               264192
_________________________________________________________________
dense_3 (Dense)              (None, 64)                16448
_________________________________________________________________
dense_4 (Dense)              (None, 32)                2080
_________________________________________________________________
dense_5 (Dense)              (None, 7)                 231
=================================================================
Total params: 282,951
Trainable params: 282,951
Non-trainable params: 0
_________________________________________________________________
```

I train the model on the confirmed data and death data using epochs = 100, batch_size = 32, optimizer = Adam with learning rate 0.0001, respectively.

1.3 Prediction

For the confirmed data, the daily new confirmed is [65365. 45872. -52760. 183020. 122548. 89584. 24904.], compared to the ground truth provided by Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE): [117447.0,138505.0, 138433.0, 153696.0, 61187.0, 47573.0, 182337.0]. It seems that the model tends to give a lower prediction number than the real one.

For the death data, the daily new confirmed is [-26072.875 , -7449, 4584.25, 4344.75, -618.5, -4412.25,1553.625], compared to the real one is : [1445.0, 1961.0, 3801.0, 1532.0, 507.0, 161.0, 1384.0]. The model even give the negative number , which means the situation was getting worse and model can not give a real prediction.

Task 2