Q2 COVID-19 Prediction

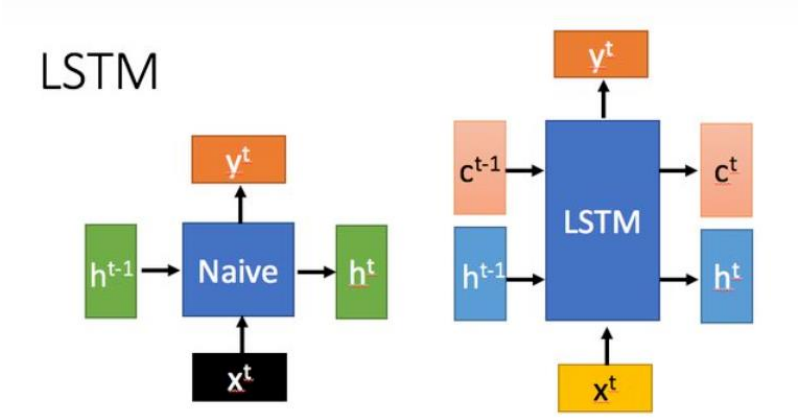
Author: Jiyao Wang 20797324

Date: 2021.12.13

In dataset, for each record, its features can be divided into 2 classes: structed and time series. So for the prediction task, I just use time series.

	Province/State	Country/Region	Lat	Long	1/22/20	1/23/20	1/24/20	1/25/20	1/26/20	1/27/20	1/28/20	1/29/20	1/3
0	NaN	Afghanistan	33.939110	67.709953	0	0	0	0	0	0	0	0	
1	NaN	Albania	41.153300	20.168300	0	0	0	0	0	0	0	0	
2	NaN	Algeria	28.033900	1.659600	0	0	0	0	0	0	0	0	
3	NaN	Andorra	42.506300	1.521800	0	0	0	0	0	0	0	0	
4	NaN	Angola	-11.202700	17.873900	0	0	0	0	0	0	0	0	

As we all know, COVID19 is always shown after 14 days. So I set the time window as 14. Meanwhile, I choose LSTM (Long Short Term Memory) as the prediction model. Long short-term memory (Long short-term memory, LSTM) is a special RNN, mainly to solve the problem of gradient disappearance and gradient explosion in the training process of long sequences. Simply put, compared to ordinary RNNs, LSTM can perform better in longer sequences.



I use LSTM to learn the distribution of one time window and predict the next day. Firstly, I need to rescale each time series into [-1,1] to better LSTM's performance and optimization. I use Adam to optimize a MSELoss. Following is the structure of my LSTM model.

```
LSTM(
  (lstm): LSTM(1, 16, batch_first=True)
  (fc): Linear(in_features=16, out_features=1, bias=True)
)
```

After training, I output results and compared with true data from JHU CSSE as following:

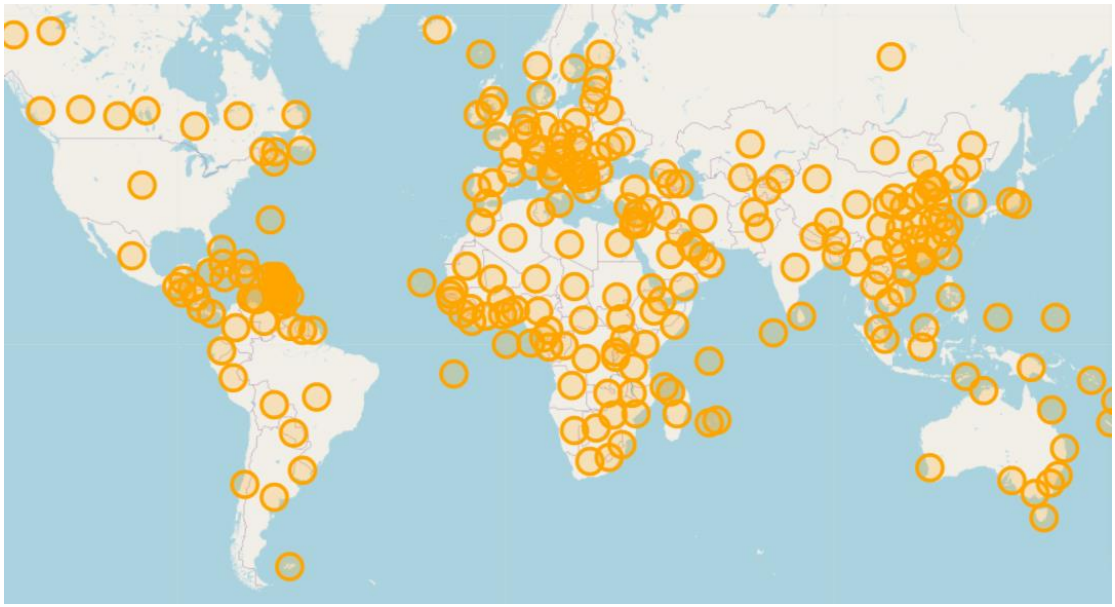
	1	2	3	4	5	6	7
Prediction	91609	91240	90498	89846	88685	92045	94998
True	117447	138505	138433	153696	61187	47573	182337

Cofirm data in US

	1	2	3	4	5	6	7
Prediction	844	817	782	759	728	730	736
True	1445	1961	3801	1532	507	161	1384

Death data in US

For task2, I draw a map according to the structed data. I utilize folium.CircleMarker which is a public library. The radius size is decided by the proportion of each location‘s data of the sum. Part of the map is following:



Reference:
[1] http://python-visualization.github.io/folium/
[2] https://zhuanlan.zhihu.com/p/32085405