1 Team

Team Members	*Jinzhao Kang c6n1b 32356362, Xiaoyu Zou x8f1b 27344167*
Kaggle Team Name	*still fantasy*

2 Solution Summary

Answer:
The step that we approach this problem can be divided into: data reformation, feature selection, training and testing
In the data reformation step, we reformatted the original data for the our convenience for next steps. We transformed the data into the form that each row contains the data among the all countries within one day
In order to find the connection between the daily deaths of Canada with all the additional data, we made a feature selection to pick out the most relevant features about Canadian deaths.
In the training phase, we used aggregated Auto Linear Auto Regressive Models to training our model. The "aggregated" here means that the Linear Auto Regressive Model has updated feature matrix that contains not only the Canadian death data, but also the prediction of features that we picked in the feature selection step by apply independent Linear Auto Regressive Model on each of them.
Finally, we used our trained model to fit the test data and got the testing error. By comparing with the testing error, we can adjust our hyper-parameters and adjust our model to get a better fit.

3 Experiments

Answer:
We used the method of forward selection to do the feature selection. To determine the grading criteria for the "score" to pick the most relevant features, we tried and compared the features picked by AIC,BIC,correlation and cosine similarity. By the result we found that the features picked by AIC/BIC has the least training error. And it is also noteworthy that some features are bad for linear regression. Therefore we made a filter like "do not consider the features with training error bigger than r" inside our feature selection method.
The hyper-parameters that we need to decide are the number of previous datas("K") for each features and the number of intersting features that we want to take consider("f") selected by our forward selection method. Each pair of "K" and "f" lead to a specific feature matrix. In order to select these hyper-parameters, we created a for loop to iterate through each possible pair and obtained the pair of "K" and "f" that has the smallest testing error of phase 1.
In the training phase, we use the Linear Autoregressive Models with the daily data of Canada death and the extra features that we found by our model selection and our selected hyperparameter "K" and "f". In order to consider the extra features and make the daily prediction of the Canada death not only about the previous data of Canada death,but also the previous data of the selected features, we updated the X in the Linear Autogressive Model into our feature matrix. The N^{th} row of our feature matrix can be represented as:

$$[1,\ d_{N-K+1,Canadadeath},\ d_{N-K+1,feature1},\ \dots\ d_{N-K+1,featuref}\ \dots\ d_{N,Canadadeath},\ d_{N,feature1},\ \dots,\ d_{N,featuref}]$$

Note that for each feature in our feature matrix, we apply Linear Autoregressive Model to get a predict data from d_K to d_T. In other words, we aggregate a series of Linear Autogregressive Model(for each intersting features) into a updated Linear matrix model(for canada death, as decribed above). In this way, we can adjust the Linear Autoregressive Models to predict based on all the features that we need to consider.

4 Results

Team Name	Kaggle Phase 1 Score	Kaggle Phase 2 Score
still fantasy	*11.82062*	*your Phase 2 Kaggle score*

5 Conclusion

Answer:
In reality, the models can be in different forms. We can only use them after filtering and reformatting them. From this Kaggle contest, We learnt that this step is challenging and time-consuming than we expected, even if we are just using the most simple model.
Also we learnt the difficulties is very tough when we try to predicting the cases in reality. The relationship among different features in reality is always very complicated and obscure. When try to choose the model, we should always consider complex model like the aggregated model to make prediction.
If given more time, one thing that we can try is that we can use cross validation to pick hyper-parameters. What is more, we can try and compare other kinds of linear regression models such as Polynomial Basis, Weighed Least Squares or Robust Regression under log-sum-exp approximation to see whether they can provide better result.

6 Codes

Team Name	date	our prediction
still fantasy	*2020-10-26*	*9948.313044*
still fantasy	*2020-10-27*	*9977.194834*
still fantasy	*2020-10-28*	*10004.36783*
still fantasy	*2020-10-29*	*10030.18055*
still fantasy	*2020-10-30*	*10055.51062*