Group	Model#	Model des	regularization	epoch	performance	compare with previous model	train time
1-models without regularization	1	common-sense baseline	NA	NA	Test MAE: 2.62	NA	41.7s
	2	regular densely connected model, one flatten layer	NA	10	Test MAE: 2.70, val_mae: 2.5783@9epoch	worse	2m7.2s
	3	1D convnet	NA	10	Test MAE: 3.16, val_mae: 2.8983@5epoch	worse	1m32s
	4	model with a LSTM(16) layer	NA	10	Test MAE: 2.52, val_mae: 2.3776@2epoch	better	1m54s
	5	model with a GRU(16) layer	NA	10	Test MAE: 2.47, val_mae: 2.3990@epoch5	better	1m54s
2-models with dropout	6	model with a LSTM(32) layer	dropout	40	val_mae: 2.3710@epoch3	better	246m37.0s
	7	model with a GRU(32) layer	dropout	40	val_mae: 2.3052@epoch6	better	220m18.6s
3- model with stacked	8	model with stacked GRU 32,64	dropout	30 based on my previous models, I think 30 epochs are enough	val_mae: 2.2870@epoch5	better	241m30.7s
4-model with CNN+RNN	9	model with 1dcnn and rnn. Three 1d and two maxpooling and one GRU	dropout only applied to GRU	30	val_mae: 2.9414@epoch3	worse	17m54s

In this assignment, I run 9 models. Basically, there are four groups of different models.
The first group is noted as models without regularization, which contains the commonsense baseline model, densely connected model, 1dCNN, and models of RNN with LTSM and GRU. For this forecasting task, CNN doesnot perform well. But RNN models which can capture memory performed better among these models. The achieved val_mae is round 2.5 which is pretty good.
Then for my second group, which is noted as RNN with regularization, the LSTM and GRU both performed better compared to the RNN without regularization. However, the training time substantially increased from 1 minute to over 200 minutes. Unlike the textbook which documents that with regularization, the overfitting comes later after 30 epochs, my models did not indicate this. It mitigates overfitting but that is really limited. The interesting part is that my GRU even performed better compared to LSTM which requires way more computational power.
For my third group which is noted as GRU with stacking, with more regularization, the GRU model provided the best model overall, with val_mae: 2.2870@epoch5. But it also required more time to train the model(241m30s).
At my last part, I tried the 1dCNN plus RNN, however, it performed surprisingly bad. This model provided the worst validation results among all models. It could be due to my tuning is not good enough, but it does required much less time to train(17m54s).
Conclusion: for forecasting task, RNN performs better because of its memory. And my trials turn out that most of the time GRU with drop and stacking is good enough and there is not much need to spend more time to train LSTM. RNN with regularization can mitigate overfitting problem but not too significantly in my trials.