TIME-SERIES DATA

Assignment 3

Gupta, Shefali

KSU ID - 811276732

OBJECTIVE OF THIS ASSIGNMENT

In this assignment, our objective is to forecast weather conditions 24 hours ahead. To achieve this, the data set was recorded at the Weather Station in Jena, Germany. This dataset comprises 14 distinct variables, encompassing factors such as temperature, atmospheric pressure, humidity, and more. Notably, these variables have been meticulously documented at 10-minute intervals since the year 2003. It's important to note that our current focus for the purpose of this assignment, is limited to data spanning the years 2009 to 2016.

RESULTS

Group	Model No	Model description	Regularization	Epoch	Performance		Comparison with previous model
					Test_MAE	VAL_MAE	
1- models without regularization	1	non-machine-learning baseline	NA	NA	5.24	4.73	NA
	2	regular densely connected model, one flatten layer	NA	10	2.61	2.81	worse
	3	1D convnet	NA	10	3.09	2.99	worse
	4	model with a LSTM(16) layer	NA	10	2.55	2.52	better
	5	model with a GRU(16) layer	NA	10	2.56	2.39	better
2-models with dropout	6	model with a LSTM(32) layer	dropout	10	-	2.393@epoch3 2.47@epoch10	better
	7	model with a GRU(32) layer	dropout	10	-	2.3744@epoch3 2.46@epoch10	better
3- model with stacked	8	model with stacked GRU 32,64	dropout	10	-	2.304@epoch5 2.34@epoch 10	better

4-model with CNN+RN N	9	model with 1dcnn and rnn. Three 1d and two maxpooling and one GRU	dropout only applied to GRU	10	-	2.99@epoch3 3.11@epoch 10	worse
5 - Bidirectional LSTM	10	Model with LSTM(16) Layer	without dropout	10	-	2.55@epoch 10	worse

RESULTS FROM THE ABOVE TABLE

We examined the performance of 10 models grouped into five categories. Here's what we have discovered:
In this assignment, I executed nine different models organized into four distinct groups.

The 1st group, labeled as "models without regularization," included models like the commonsense baseline, a densely connected model, a 1D CNN, and RNN models utilizing LSTM and GRU architectures. Surprisingly, the CNN model did not perform well in the forecasting task, while RNN models, known for their memory-capturing capabilities, outperformed the rest. Notably, they achieved a commendable validation MAE of around 2.5.
The 2nd group, focusing on "RNN models with regularization," saw both LSTM and GRU models outperforming their non-regularized counterparts. Unlike common expectations, my models didn't exhibit delayed overfitting with regularization. While regularization did help mitigate overfitting to some extent, its effectiveness remained relatively limited. Interestingly, the GRU model surpassed the LSTM in performance, despite the latter's greater computational demands.
In 3rd group, where the emphasis was on "GRU models with stacking and increased regularization," the GRU architecture emerged as the top-performing model. Achieving a validation MAE of 2.304 at epoch 5, this GRU model demonstrated remarkable forecasting capabilities. However, this superior performance came at the cost of significantly longer training time as compared to the others.
In 4^{th} group, I explored the combination of "1D CNN and RNN models." Surprisingly, this hybrid model performed poorly, yielding the least favorable validation results among all the models. My insufficient tuning might be the reason for the performance shortfall. Nonetheless, the model required significantly less training time.

Summary:
In conclusion, among the models presented, Model 8 stands out as the top-performing model with the lowest VAL_MAE at epoch 10. For forecasting tasks, RNN models excel due to their innate memory capabilities. It's evident that, in most cases, GRU models with dropout and stacking provide a robust solution, negating the necessity for prolonged training of LSTM models. While

regularization proves beneficial in mitigating overfitting, its impact was relatively modest in my trials.

TOP 3 MODELS

Based on the VAL_MAE (Validation Mean Absolute Error), the top 3 models are:

TEST_Model	VAL_MAE	TEST_MAE	VAL_LOSS
Model 8 - Model with Stacked GRU 32,64 (With Dropout)	2.34	2.47	9.49
Model 7 - Model with a GRU(32) Layer (With Dropout)	2.39	2.54	10.788
Model 5 - Model with a GRU(16) Layer (No Regularization)	2.46	2.51	10.07

the models exhibited relatively small differences between Validation MAE and Test MAE, suggesting reasonable generalization capabilities" means that the models, when tested on unseen data (Test MAE), performed quite similarly to how they performed on the validation data (Validation MAE).

This observation suggests that the models were able to generalize well from the patterns they learned during training on the validation data to make accurate predictions on the test data. Small differences between Validation MAE and Test MAE are generally a positive sign, as it indicate that the model is not overfitting (performing well on the validation set but poorly on the test set) or underfitting (performing poorly on both sets). Instead, it suggests that the model's performance on the validation data is a good indicator of its performance on new, unseen data, which is a desirable characteristic for predictive models.