IE 8990

Spring 2022

Homework #3

Due Date: 04/07/2022 5PM CST

Submission: Please put your answer and code in a PDF file and upload on Canvas

Q1. In LSTM model, why sigmoid or tanh are used? Can we use ReLU to replace tanh?

* The sigmoid and hyperbolic tangent functions are used to have binary values for decision making. The sigmoid sections are used to retain somewhere between 0 and 1 of the previous state. The tanh function is used to allow for negative values between -1 and 1.
* The ReLU function would not work since the activation function needs to be bounded. If the forget gate can be weighted past 1 then it is technically remembering more than 100% of the information. The negative values of the hyperbolic tangent are also important when highlighting the main issues of the functions.

Q2. For the HW2 CIFAR-10 Base Model (provided in attached Jupyter Notebook file), let’s fix the epochs = 10. Please modify the model structure to improve the model performance based on the tips we discussed in DL 10 (target: test error lower than 0.25 in 10 epochs training). Discuss your approach.

Q3. Based on the jena climate 2009 01.csv data. Please develop a LSTM model to predict the next 24 hours’ temperature (Celsius) based on the previous 24 hours’ information. (Note: you can use all 14 climate features or part of them). Hint: here is a link might be useful: link). Please plot your predicted value and the true value in one plot.

Q4. For the following models AlexNet, ZFNet, VGG, GooLeNet, ResNet, MobileNet,DenseNet, EfficientNet

* Discuss the advantages and disadvantages of each model
* Discuss the motivations of how each model was developed