HW2

Git: https://github.com/Fallt0earth/Real_Time_ML/tree/main/HW2

1a:

Training Time: 0.16614627838134766

Training loss: 0.0233

Evaluation Accuracy: 0.9565734267234802

Gmacs: 0.0

Params: 57

Discussion: The model accuracy was surprisingly high. I have a suspicion that the accuracy model may be flawed. The training loss converged rather quickly, and training was incredibly fast for such a light network. I am highly concerned that there is overfitting present as well.

1b:

Training Time: 0.2971830368041992

Training loss: 0.1608

Evaluation Accuracy: 56.73%

Gmacs: 0.02

Params: 42.75k

Discussion: This model accuracy was more than I expected although there are still concerns over its actual accuracy. The training time was around 3 times longer which makes sense with two additional layers. The model complexity also exploded from basically nothing and not even registering GMacs to real values. This indicates the exponential nature of adding additional layers. The number of parameters exploded several orders of magnitude as well greater. The loss values were higher with the more complex network and as such clearly the simple problem does poorly against more complex models.

2a:

Training Time: 1838.7329592704773 sec

Training loss: 0.015110

Evaluation Accuracy: 46.74%

Gmacs: 1.59

Params: 1.58M

Discussion: The low accuracy of the model makes sense as it is a more complex problem, there is likely not enough non-linearity in the function. The training time was massive, and I seemed unable to offload the computation to my GPU even though I followed all of the steps. It seems that the tensors are actually executing on the GPU but the model size is not large enough to leverage the SIMT architecture. This means that the CPU and GPU times are highly similar at this small scale.

2b:

Training Time: 1859.5388488769531

Training loss: 0.000380

Evaluation Accuracy: 45.92%

Gmacs: 1.59

Params: 1.58M

Discussion: I was very surprised by the low accuracy of the model considering the training loss was incredibly low. This would indicate that the model is overfitting to the training data and is unable to generalize well. The other surprise was that the execution time was almost identical.