Report: Neural Collaborative Filtering Attention Blocks

Goal: To minimize the loss under the attention block integration.

Accuracy Function:

I used a simple accuracy function to measure the accuracy, with threshold of 0.5, as the ratings are on the scale of 1-5

$$\text{accuracy} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left(\left|y_{\text{true},i} - y_{\text{pred},i}\right| \leq \tau\right)$$

Where:

- $\left|y_{\text{true},i} - y_{\text{pred},i}\right|$ is the absolute difference between the true and predicted ratings for the i^{th} instance.
- $\mathbf{1}(\cdot)$ is the indicator function that returns 1 if the condition is true and 0 otherwise.
- τ is the tolerance value.
- n is the total number of instances.

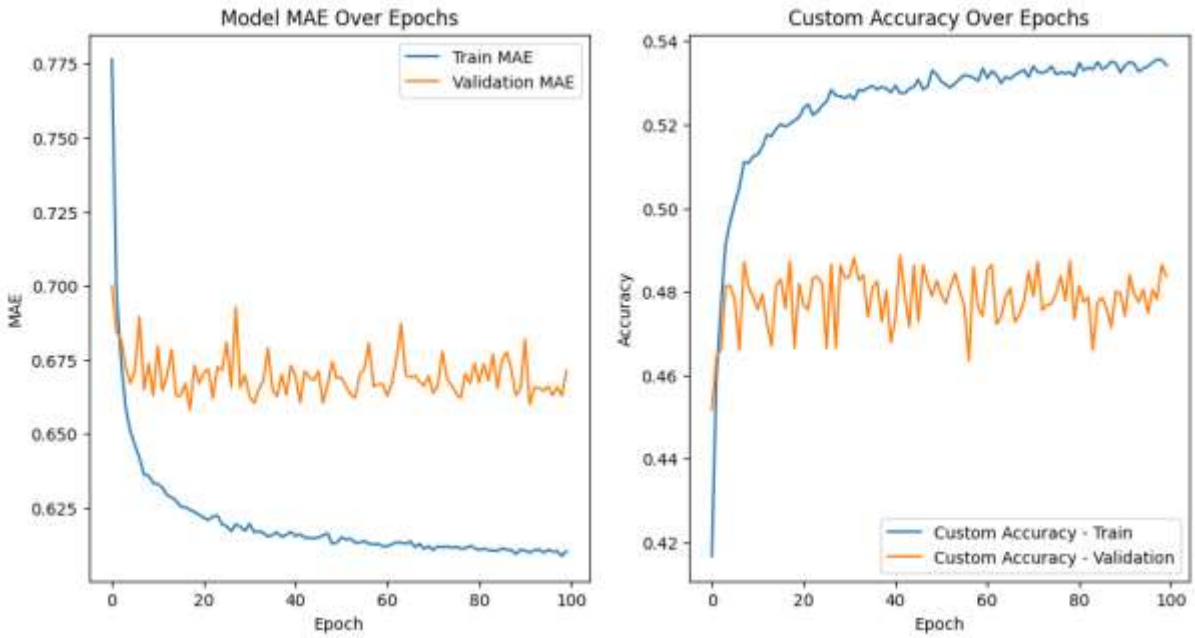
We tested the attention block with multiple possibilities of Matrix Factorization and MLP. Few examples include:

Attention Block

- After MLP and MF layers
- After MLP layer and MF layer
- With MLP layers and MF layers
- Tested with MF
- Tested with MLP

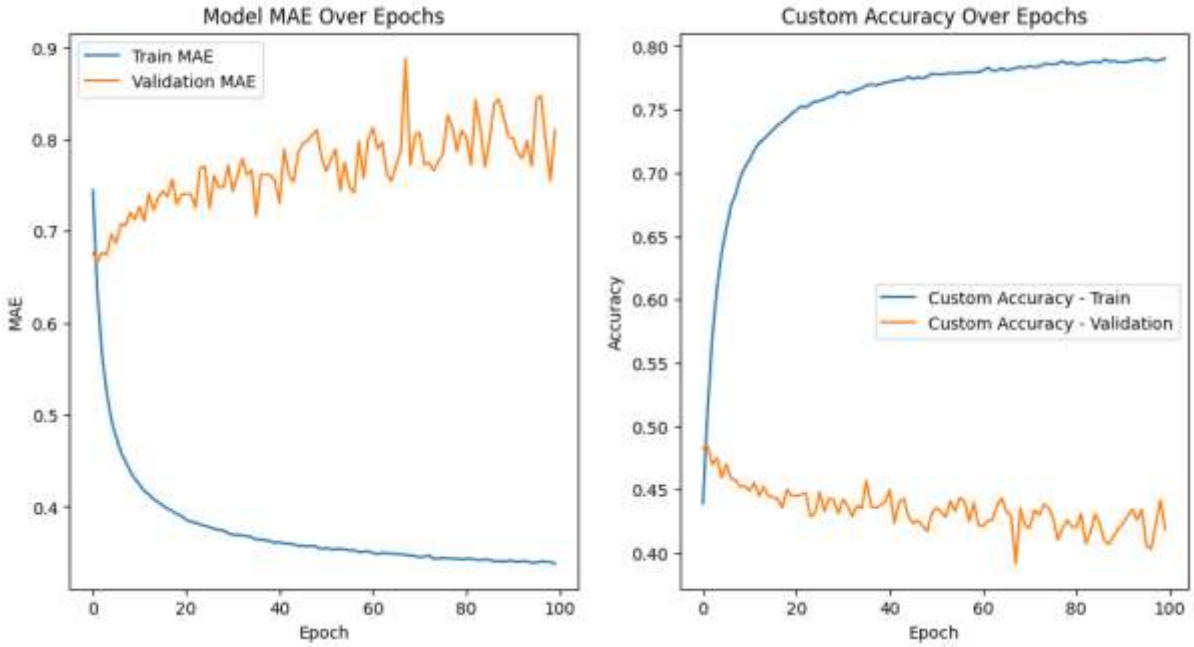
Under the above all conditions, we found two conditions in the architecture to be promising:

1. **Attention Block After MLP only, with no MF:**
   The model seems to be promising in terms of reducing overfitting and underfitting. It maintains the Validation loss throughout the epochs.



Mean Absolute Error on Test Set: 0.6651288739618731
Custom Accuracy on Test Set: 0.4817532725109084

2. **Attention Block concatenated with MF only, with no MLP:**
   The model is overfit too much and does not perform well on validation loss, but we could use regularization techniques to improve it's generalizability:

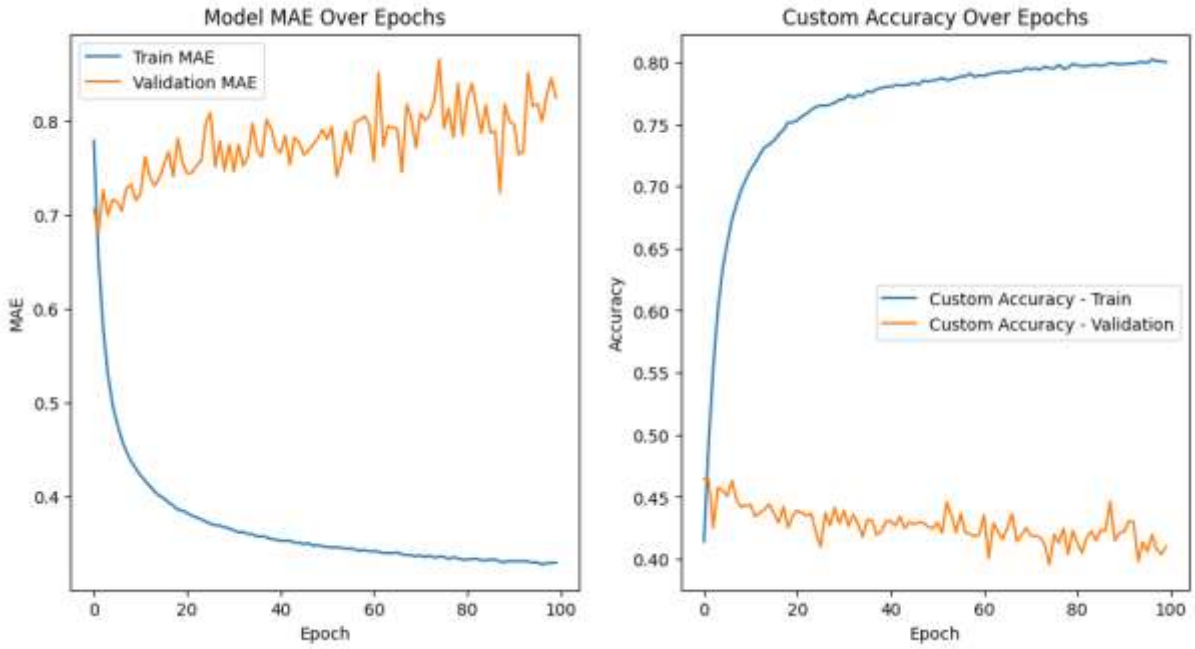


Mean Absolute Error on Test Set: 0.8106766776615832

Custom Accuracy on Test Set: 0.4187326457754859

Paper Architecture Results:

Kindly note the following architecture results from the paper, this can be used as reference for improvement. The model utilizes MF and MLP both:



Mean Absolute Error on Test Set: 0.75038956555522

Custom Accuracy on Test Set: 0.43856604522015075

Conclusion:

Based on the above, the **Attention Block After MLP only, with no MF** does better in terms of bias variance trade off. It is worth noting that we are having true validation loss, with **num_heads = 3** and drop out in attention block = 0.1. Kindly note that the drop out values for the other layers such as Concatenation, MLP layers are kept same as 0.2.