Training and Loss curves:

A graph with blue lines and orange dots

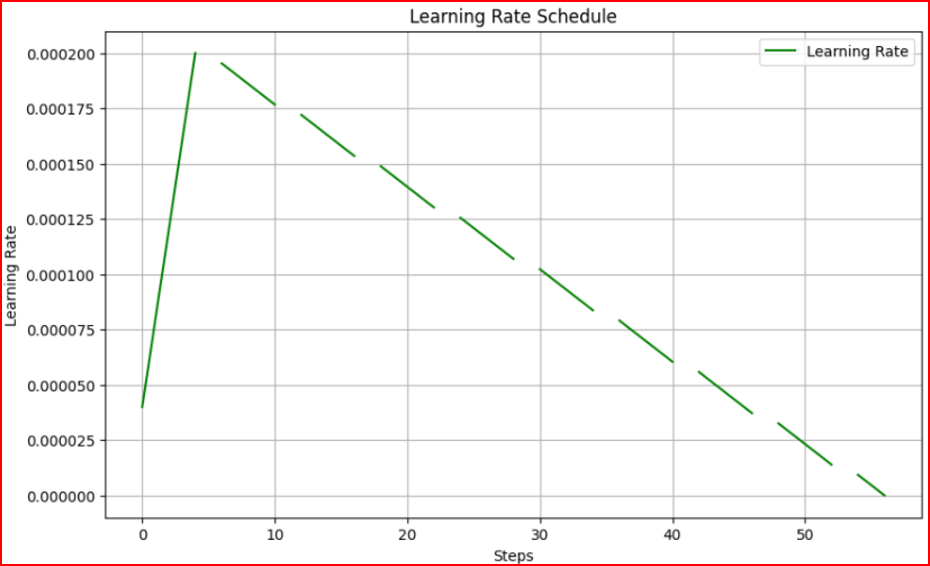
Description automatically generated

During the model training process, the training and validation loss are depicted in the plot that is supplied. The y-axis displays the associated loss values, and the x-axis depicts the training stages. The validation loss is represented by orange crosses, while the training loss is shown as a blue line with round markers. Reducing both losses—which represent the model's performance on training and validation datasets—is the aim.

The training loss is significant at first, around 2.0, and gradually drops as the model gains knowledge. However, especially after the 20th training step, there are discernible swings and spikes. A high learning rate, noisy data, or inadequate regularization are some possible causes of the training process instability shown by these swings. However, at the tenth step, the validation loss plateaus at about 0.8, although being quite stable. This could mean that the model is beginning to overfit the training set or that more hyperparameter tweaking is required.

Overall, the training and validation losses level off after about 20 to 30 steps, indicating that further training yields decreasing results. The model's performance could be enhanced by using techniques like regularization (e.g., dropout or L2 regularization), scheduler-assisted learning rate adjustment, and early halting to avoid overfitting. To further stabilize the training process and improve performance, more hyperparameters may need to be optimized or alternative model designs investigated.

Learning Curve:



The learning rate schedule plot highlights distinct points of interest where the rate goes high and low during training. Initially, the learning rate starts at a very low value and gradually increases during the warmup phase, peaking at around the 10th step. This upward trend signifies the model's gradual adaptation to the optimization process. The peak around step 10 ensures that the model can explore the parameter space effectively without drastic updates, which might destabilize training early on.

After step 10, the learning rate begins a linear decline, consistently reducing as the training progresses. This downward trajectory ensures smaller updates to the model’s parameters, allowing for precise adjustments as the optimization approaches convergence. The lowest points of the learning rate, observed towards the final steps, reflect the model's fine-tuning stage, where smaller learning rates help avoid overshooting and ensure stability. The balance between high learning rates during the warmup and low rates during decay supports both effective exploration and controlled convergence in training.