Learning and Fine-Tuning steps

Contents

[All parameters in model are FROZEN. 2](#_Toc194324077)

[All parameters in model are FROZEN. 4](#_Toc194324078)

[Adjusted controls 4](#_Toc194324079)

[Perform a PEFT application – params unfrozen 5](#_Toc194324080)

# All parameters in model are FROZEN.

|  |  |  |  |
| --- | --- | --- | --- |
| Phases | Settings | Output | Notes and recommendation |
| Phase 1 – PC | SPLIT\_TRAINING = "1%" SPLIT\_EVAL = "1%" LEARN\_RATE = 2e-3 TRAIN\_BATCH\_SIZE = 4 EVAL\_BATCH\_SIZE = 4 TRAIN\_EPOCHS = 1 WGT\_DECAY = 0.01 MY\_EVAL\_STRAT = "epoch" S\_STRAT = "epoch" BEST\_MODEL = True | Evaluation Metrics:  Loss Accuracy Runtime (s) Samples/sec Steps/sec Epoch  1.09216 0.614565 126.997 4.433 1.11 1.0  Training Metrics:  Runtime (s) Samples/sec Steps/sec Loss Epoch  1394.274 2.956 0.739 1.252511 1.0 | **Data Splits:** Increase from 1% to a larger subset (e.g., 10% or more). **Epochs:** Increase TRAIN\_EPOCHS from 1 to at least 3–5.  **Learning Rate:** Lower LEARN\_RATE from 2e-3 to a value in the range of 2e-5 to 5e-5.  **Batch Size:** Increase TRAIN\_BATCH\_SIZE/EVAL\_BATCH\_SIZE |
| Phase 2 - Jupyter | SPLIT\_TRAINING = "25%" SPLIT\_EVAL = "25%" LEARN\_RATE = 5e-5 TRAIN\_BATCH\_SIZE = 16 EVAL\_BATCH\_SIZE = 16 TRAIN\_EPOCHS = 4 WGT\_DECAY = 0.01 MY\_EVAL\_STRAT = "epoch" S\_STRAT = "epoch" BEST\_MODEL = True | Tabular Evaluation Metrics:  Loss Accuracy Runtime (s) Samples/sec Steps/sec Epoch  1.088676 0.599943 226.0095 62.289 3.894 4.0 | These setting proved to be too resource intensive for PC. Reducing them to see what can be achieved.  Moved these to Jupyter and course platform. |

|  |  |  |  |
| --- | --- | --- | --- |
| Phase 2 - PC | SPLIT\_TRAINING = "10%" SPLIT\_EVAL = "10%" LEARN\_RATE = 2e-3 TRAIN\_BATCH\_SIZE = 8 EVAL\_BATCH\_SIZE = 8 TRAIN\_EPOCHS = 2 WGT\_DECAY = 0.01 MY\_EVAL\_STRAT = "epoch" S\_STRAT = "epoch" BEST\_MODEL = True | No results possible | Running adjusted setting on PC platform.  This took too long on PC resource |
| Phase 2.1 - PC | SPLIT\_TRAINING = "2%" SPLIT\_EVAL = "2%" LEARN\_RATE = 2e-3 TRAIN\_BATCH\_SIZE = 4 EVAL\_BATCH\_SIZE = 4 TRAIN\_EPOCHS = 1 WGT\_DECAY = 0.01 MY\_EVAL\_STRAT = "epoch" S\_STRAT = "epoch" BEST\_MODEL = True | Tabular Evaluation Metrics  Loss Accuracy Runtime (s) Samples/sec Steps/sec Epoch  1.128501 0.619893 264.9978 4.249 1.064 1.0 | Adjusted everything back to original setting except for % size of the test and eval samples |

# All parameters in model are FROZEN.

## Adjusted controls

|  |  |  |  |
| --- | --- | --- | --- |
| Phases | Settings | Output | Notes and recommendation |
| Test 1- Jupyter | SPLIT\_TRAINING = "25%" SPLIT\_EVAL = "25%" LEARN\_RATE = 5e-5 TRAIN\_BATCH\_SIZE = 16 EVAL\_BATCH\_SIZE = 16 TRAIN\_EPOCHS = 4 WGT\_DECAY = 0.01 MY\_EVAL\_STRAT = "epoch" S\_STRAT = "epoch" BEST\_MODEL = True | Stopped Process  [ 2200/25756 26:16 < 4:41:35, 1.39 it/s, Epoch 0.34/4] | In both these instances of un-frozen layers, resource usage and time lead to a stop on processing |
| Test 1 - PC | SPLIT\_TRAINING = "2%" SPLIT\_EVAL = "2%" LEARN\_RATE = 2e-3 TRAIN\_BATCH\_SIZE = 4 EVAL\_BATCH\_SIZE = 4 TRAIN\_EPOCHS = 1 WGT\_DECAY = 0.01 MY\_EVAL\_STRAT = "epoch" S\_STRAT = "epoch" BEST\_MODEL = True |  |  |

# Perform a PEFT application – params unfrozen

|  |  |  |  |
| --- | --- | --- | --- |
| Test 1 - Jupyter | SPLIT\_TRAINING = "2%" SPLIT\_EVAL = "2%" LEARN\_RATE = 2e-3 TRAIN\_BATCH\_SIZE = 4 EVAL\_BATCH\_SIZE = 4 TRAIN\_EPOCHS = 1 WGT\_DECAY = 0.01 MY\_EVAL\_STRAT = "epoch" S\_STRAT = "epoch" BEST\_MODEL = True | Loss Accuracy Runtime (s) Samples/sec Steps/sec Epoch  1.474885 0.483126 20.2109 55.713 13.953 1.0 | Not very good attainment compared to previous models |
| Test 2 -Jupyter | SPLIT\_TRAINING = "2%" SPLIT\_EVAL = "2%" LEARN\_RATE = 5e-5 TRAIN\_BATCH\_SIZE = 16 EVAL\_BATCH\_SIZE = 16 TRAIN\_EPOCHS = 4 WGT\_DECAY = 0.01 MY\_EVAL\_STRAT = "epoch" S\_STRAT = "epoch" BEST\_MODEL = True | Loss Accuracy Runtime (s) Samples/sec Steps/sec Epoch  1.019231 0.645648 20.4111 55.166 3.478 4.0 | Much better outcome with each epoch showing an accuracy improvement.  Speed on both Test 1 and 2 far outperformed earlier non tuned comparisons |