COMPUTE ENVIRONMENT
NVIDIA L4 GPU, g2-standard-4 (4 vCPUs, 16 GB Memory) machine for everything

C2:

This screenshot contains all the answers for C2.1, C2.2, and C2.3.

```
RUNNING EXPERIMENT: C2 - Default experiment


python3 main.py


================================================
Epoch 1 Summary:
- Loss: 1.9055
- Accuracy (Top-1): 30.77%
- Total Data Loading Time: 4.5966s
- Total Training Time: 7.9523s
- Average Training Time per Batch: 0.0203s
- Total Epoch Time: 13.3623s
================================================


================================================
Epoch 2 Summary:
- Loss: 1.4962
- Accuracy (Top-1): 44.70%
- Total Data Loading Time: 4.2482s
- Total Training Time: 7.4377s
- Average Training Time per Batch: 0.0190s
- Total Epoch Time: 12.5102s
================================================


================================================
Epoch 3 Summary:
- Loss: 1.3023
- Accuracy (Top-1): 52.94%
- Total Data Loading Time: 4.3642s
- Total Training Time: 7.3371s
- Average Training Time per Batch: 0.0188s
- Total Epoch Time: 12.5342s
================================================


================================================
Epoch 4 Summary:
- Loss: 1.1300
- Accuracy (Top-1): 59.67%
- Total Data Loading Time: 4.6559s
- Total Training Time: 7.2438s
- Average Training Time per Batch: 0.0185s
- Total Epoch Time: 12.7545s
================================================


================================================
Epoch 5 Summary:
- Loss: 1.0045
- Accuracy (Top-1): 64.66%
- Total Data Loading Time: 6.9076s
- Total Training Time: 7.7193s
- Average Training Time per Batch: 0.0197s
- Total Epoch Time: 15.4743s
================================================
```

C3:

C3.1

```
Running command: python3 main.py --num_workers 0


================================================
Epoch 1 Summary:
- Loss: 1.8943
- Accuracy (Top-1): 30.84%
- Total Data Loading Time: 19.1252s
- Total Training Time: 4.5867s
- Average Training Time per Batch: 0.0117s
- Total Epoch Time: 25.0753s
================================================


================================================
Epoch 2 Summary:
- Loss: 1.4719
- Accuracy (Top-1): 45.86%
- Total Data Loading Time: 19.1013s
- Total Training Time: 4.1523s
- Average Training Time per Batch: 0.0106s
- Total Epoch Time: 24.6509s
================================================


================================================
Epoch 3 Summary:
- Loss: 1.2586
- Accuracy (Top-1): 54.60%
- Total Data Loading Time: 18.9617s
- Total Training Time: 4.1299s
- Average Training Time per Batch: 0.0106s
- Total Epoch Time: 24.4902s
================================================


================================================
Epoch 4 Summary:
- Loss: 1.1007
- Accuracy (Top-1): 60.76%
- Total Data Loading Time: 19.0576s
- Total Training Time: 4.1610s
- Average Training Time per Batch: 0.0106s
- Total Epoch Time: 24.6022s
================================================


================================================
Epoch 5 Summary:
- Loss: 0.9793
- Accuracy (Top-1): 65.38%
- Total Data Loading Time: 19.1425s
- Total Training Time: 4.1132s
- Average Training Time per Batch: 0.0105s
- Total Epoch Time: 24.6739s
================================================


num_workers: 0 -> Total Runtime: 123.4925 s


num_workers: 0 -> Total Data Loading Time: 95.3883 s
```

```
Running command: python3 main.py --num_workers 4


================================================
Epoch 1 Summary:
- Loss: 1.9395
- Accuracy (Top-1): 30.22%
- Total Data Loading Time: 2.1701s
- Total Training Time: 8.6026s
- Average Training Time per Batch: 0.0220s
- Total Epoch Time: 11.6665s
================================================


================================================
Epoch 2 Summary:
- Loss: 1.4877
- Accuracy (Top-1): 45.40%
- Total Data Loading Time: 2.0581s
- Total Training Time: 7.6949s
- Average Training Time per Batch: 0.0197s
- Total Epoch Time: 10.6166s
================================================


================================================
Epoch 3 Summary:
- Loss: 1.2601
- Accuracy (Top-1): 54.60%
- Total Data Loading Time: 2.3053s
- Total Training Time: 8.5668s
- Average Training Time per Batch: 0.0219s
- Total Epoch Time: 11.7611s
================================================


================================================
Epoch 4 Summary:
- Loss: 1.0715
- Accuracy (Top-1): 61.96%
- Total Data Loading Time: 1.9551s
- Total Training Time: 7.8265s
- Average Training Time per Batch: 0.0200s
- Total Epoch Time: 10.6802s
================================================


================================================
Epoch 5 Summary:
- Loss: 0.9259
- Accuracy (Top-1): 67.45%
- Total Data Loading Time: 1.9827s
- Total Training Time: 7.8284s
- Average Training Time per Batch: 0.0200s
- Total Epoch Time: 10.7267s
================================================


num_workers: 4 -> Total Runtime: 55.4511 s


num_workers: 4 -> Total Data Loading Time: 10.4713 s
```

```
Running command: python3 main.py --num_workers 8

/opt/conda/lib/python3.10/site-packages/torch/utils/data/dataloader.py:624: UserWarning: This DataLoader will c
4, which is smaller than what this DataLoader is going to create. Please be aware that excessive worker creatio
ial slowness/freeze if necessary.
  warnings.warn(

================================================
Epoch 1 Summary:
- Loss: 1.9343
- Accuracy (Top-1): 30.77%
- Total Data Loading Time: 2.0790s
- Total Training Time: 8.6778s
- Average Training Time per Batch: 0.0222s
- Total Epoch Time: 11.7589s
================================================


================================================
Epoch 2 Summary:
- Loss: 1.5060
- Accuracy (Top-1): 44.51%
- Total Data Loading Time: 2.1725s
- Total Training Time: 8.3618s
- Average Training Time per Batch: 0.0214s
- Total Epoch Time: 11.5405s
================================================


================================================
Epoch 3 Summary:
- Loss: 1.3215
- Accuracy (Top-1): 52.36%
- Total Data Loading Time: 2.1369s
- Total Training Time: 8.0145s
- Average Training Time per Batch: 0.0205s
- Total Epoch Time: 11.2213s
================================================


================================================
Epoch 4 Summary:
- Loss: 1.1590
- Accuracy (Top-1): 58.56%
- Total Data Loading Time: 1.8530s
- Total Training Time: 8.0138s
- Average Training Time per Batch: 0.0205s
- Total Epoch Time: 10.8932s
================================================


================================================
Epoch 5 Summary:
- Loss: 1.0295
- Accuracy (Top-1): 63.44%
- Total Data Loading Time: 1.8624s
- Total Training Time: 7.9952s
- Average Training Time per Batch: 0.0204s
- Total Epoch Time: 10.9297s
================================================


num_workers: 8 -> Total Runtime: 56.3436 s


num_workers: 8 -> Total Data Loading Time: 10.1038 s
```

```
Running command: python3 main.py --num_workers 12

/opt/conda/lib/python3.10/site-packages/torch/utils/data/dataloader.py:624: UserWarning: This DataLoader will crea
 4, which is smaller than what this DataLoader is going to create. Please be aware that excessive worker creation
tial slowness/freeze if necessary.
  warnings.warn(

================================================
Epoch 1 Summary:
- Loss: 2.0263
- Accuracy (Top-1): 27.57%
- Total Data Loading Time: 1.9407s
- Total Training Time: 8.8459s
- Average Training Time per Batch: 0.0226s
- Total Epoch Time: 11.9316s
================================================


================================================
Epoch 2 Summary:
- Loss: 1.5225
- Accuracy (Top-1): 43.88%
- Total Data Loading Time: 2.2939s
- Total Training Time: 8.5641s
- Average Training Time per Batch: 0.0219s
- Total Epoch Time: 12.0233s
================================================


================================================
Epoch 3 Summary:
- Loss: 1.3287
- Accuracy (Top-1): 51.93%
- Total Data Loading Time: 1.9682s
- Total Training Time: 7.8730s
- Average Training Time per Batch: 0.0201s
- Total Epoch Time: 11.0255s
================================================


================================================
Epoch 4 Summary:
- Loss: 1.1533
- Accuracy (Top-1): 58.89%
- Total Data Loading Time: 2.0792s
- Total Training Time: 7.8233s
- Average Training Time per Batch: 0.0200s
- Total Epoch Time: 11.0501s
================================================


================================================
Epoch 5 Summary:
- Loss: 1.0328
- Accuracy (Top-1): 63.43%
- Total Data Loading Time: 1.9481s
- Total Training Time: 7.9660s
- Average Training Time per Batch: 0.0204s
- Total Epoch Time: 11.1119s
================================================


num_workers: 12 -> Total Runtime: 57.1424 s


num_workers: 12 -> Total Data Loading Time: 10.2301 s
```

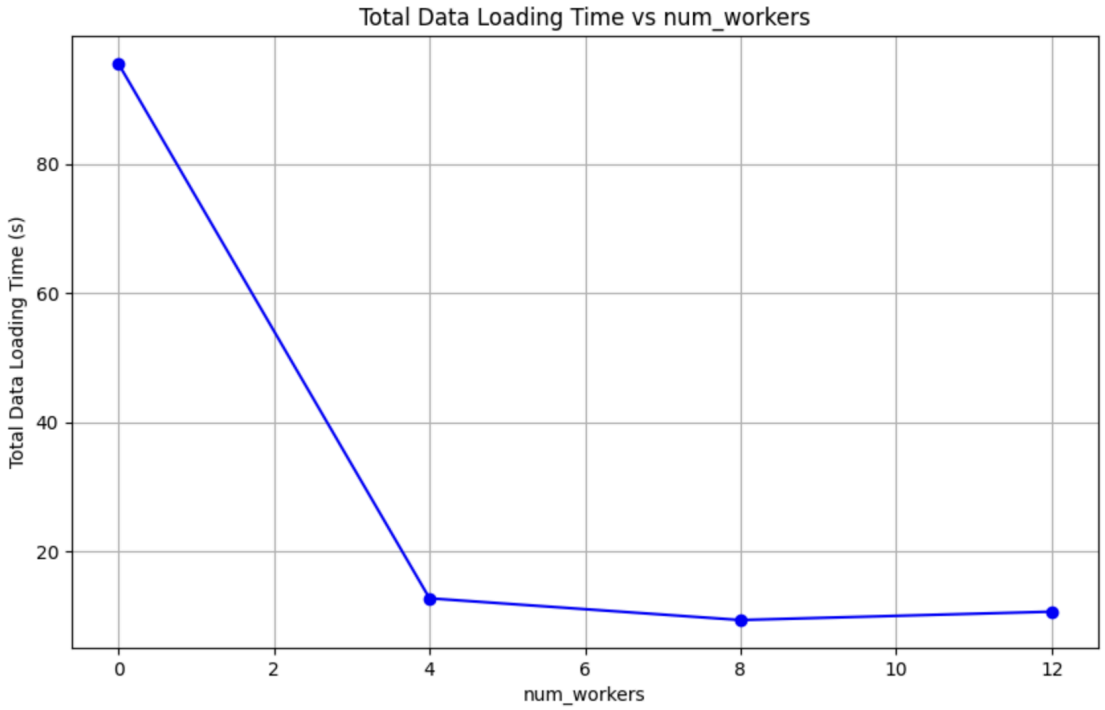
The total data loading times and total epoch times are as follows:

```
num_workers: 0 -> Total Runtime: 123.4925 s

num_workers: 0 -> Total Data Loading Time: 95.3883 s
```

```
num_workers: 4 -> Total Runtime: 55.4511 s

num_workers: 4 -> Total Data Loading Time: 10.4713 s
```

```
num_workers: 8 -> Total Runtime: 56.3436 s

num_workers: 8 -> Total Data Loading Time: 10.1038 s
```

```
num_workers: 12 -> Total Runtime: 57.1424 s

num_workers: 12 -> Total Data Loading Time: 10.2301 s
```



C3.2
From the data, it looks like the optimal number of workers is 4 for the runtime performance. The optimal number of workers for data loading time only is 8, however.

C3.3
After 4 workers, increasing the number of workers does not lead to further speedup. This is because there is a tradeoff between data loading time and overall runtime performance. When we go from 4 to 8 workers, the data loading time decreases slightly, which might lead us to expect a slight decrease in overall runtime. But we see an increase in overall runtime. This is probably because when we create more workers for the data loader, we improve just the data loading time. However, this causes more processes to be spawned and affects how much contention there is for the processes on the CPU. Thus the Python code to run the model will probably be slower and result in a slowdown in the rest of the training loop.

C4:

```
(base) si2468@instance-20250306-074853:~/hpml_assignments/HW2$ python3 driver.py


RUNNING EXPERIMENT: C4 - 1 worker vs optimal num workers



python3 main.py --num_workers 1


================================================
Epoch 1 Summary:
- Loss: 1.9375
- Accuracy (Top-1): 30.31%
- Total Data Loading Time: 15.5939s
- Total Training Time: 6.1555s
- Average Training Time per Batch: 0.0157s
- Total Epoch Time: 22.8543s
================================================


================================================
Epoch 2 Summary:
- Loss: 1.5135
- Accuracy (Top-1): 44.52%
- Total Data Loading Time: 15.2216s
- Total Training Time: 5.5800s
- Average Training Time per Batch: 0.0143s
- Total Epoch Time: 21.9051s
================================================


================================================
Epoch 3 Summary:
- Loss: 1.3057
- Accuracy (Top-1): 53.01%
- Total Data Loading Time: 15.1844s
- Total Training Time: 6.3151s
- Average Training Time per Batch: 0.0162s
- Total Epoch Time: 22.4378s
================================================


================================================
Epoch 4 Summary:
- Loss: 1.1596
- Accuracy (Top-1): 58.64%
- Total Data Loading Time: 14.8880s
- Total Training Time: 5.2975s
- Average Training Time per Batch: 0.0135s
- Total Epoch Time: 21.3497s
================================================


================================================
Epoch 5 Summary:
- Loss: 1.0436
- Accuracy (Top-1): 63.06%
- Total Data Loading Time: 14.6652s
- Total Training Time: 5.4441s
- Average Training Time per Batch: 0.0139s
- Total Epoch Time: 21.2728s
================================================
```

```
python3 main.py --num_workers 4


================================================
Epoch 1 Summary:
- Loss: 2.0970
- Accuracy (Top-1): 24.63%
- Total Data Loading Time: 2.1550s
- Total Training Time: 8.7491s
- Average Training Time per Batch: 0.0224s
- Total Epoch Time: 11.7747s
================================================


================================================
Epoch 2 Summary:
- Loss: 1.6233
- Accuracy (Top-1): 39.31%
- Total Data Loading Time: 1.9568s
- Total Training Time: 7.8536s
- Average Training Time per Batch: 0.0201s
- Total Epoch Time: 10.6927s
================================================


================================================
Epoch 3 Summary:
- Loss: 1.4386
- Accuracy (Top-1): 47.16%
- Total Data Loading Time: 2.2175s
- Total Training Time: 8.2014s
- Average Training Time per Batch: 0.0210s
- Total Epoch Time: 11.3012s
================================================


================================================
Epoch 4 Summary:
- Loss: 1.2888
- Accuracy (Top-1): 53.66%
- Total Data Loading Time: 2.2338s
- Total Training Time: 8.1044s
- Average Training Time per Batch: 0.0207s
- Total Epoch Time: 11.2318s
================================================


================================================
Epoch 5 Summary:
- Loss: 1.1416
- Accuracy (Top-1): 59.26%
- Total Data Loading Time: 1.9041s
- Total Training Time: 7.8724s
- Average Training Time per Batch: 0.0201s
- Total Epoch Time: 10.6769s
================================================
```

From the results above, we can see that the Total Data Loading Time for the optimal number of workers (4) is about 2 seconds per epoch. However, the Total Data Loading Time for 1 worker is about 15 seconds per epoch. Thus the Total Data Loading Time is much slower for 1 worker than for 4.

For the total compute time, we can see that the training times per epoch for 1 worker are about 6 seconds per epoch. The training times per epoch for 4 workers is about 8 seconds per epoch. Thus, the optimal number of workers (4) has a larger training time than that of the single worker.

The total epoch time is about 11 seconds for the optimal number of workers and about 22 seconds for 1 worker, so the optimal number of workers pretty much halves the runtime.

C5:

```
(base) si2468@instance-20250306-074853:~/hpml_assignments/HW2$ python3 driver.py


RUNNING EXPERIMENT: C5 - GPU vs CPU



python3 main.py --num_workers 4


================================================
Epoch 1 Summary:
- Loss: 1.9875
- Accuracy (Top-1): 28.75%
- Total Data Loading Time: 2.2396s
- Total Training Time: 8.6573s
- Average Training Time per Batch: 0.0221s
- Total Epoch Time: 11.7855s
================================================


================================================
Epoch 2 Summary:
- Loss: 1.5292
- Accuracy (Top-1): 43.59%
- Total Data Loading Time: 2.0342s
- Total Training Time: 7.9109s
- Average Training Time per Batch: 0.0202s
- Total Epoch Time: 10.8215s
================================================


================================================
Epoch 3 Summary:
- Loss: 1.3221
- Accuracy (Top-1): 51.90%
- Total Data Loading Time: 1.9350s
- Total Training Time: 7.9129s
- Average Training Time per Batch: 0.0202s
- Total Epoch Time: 10.7484s
================================================


================================================
Epoch 4 Summary:
- Loss: 1.1354
- Accuracy (Top-1): 59.42%
- Total Data Loading Time: 1.9887s
- Total Training Time: 7.8493s
- Average Training Time per Batch: 0.0201s
- Total Epoch Time: 10.7235s
================================================


================================================
Epoch 5 Summary:
- Loss: 1.0079
- Accuracy (Top-1): 64.57%
- Total Data Loading Time: 2.4626s
- Total Training Time: 8.5211s
- Average Training Time per Batch: 0.0218s
- Total Epoch Time: 11.8545s
================================================
```

```
python3 main.py --num_workers 4 --disable_cuda


================================================
Epoch 1 Summary:
- Loss: 2.0947
- Accuracy (Top-1): 24.09%
- Total Data Loading Time: 0.9056s
- Total Training Time: 333.4509s
- Average Training Time per Batch: 0.8528s
- Total Epoch Time: 334.5173s
================================================


================================================
Epoch 2 Summary:
- Loss: 1.6169
- Accuracy (Top-1): 39.65%
- Total Data Loading Time: 0.8690s
- Total Training Time: 332.8691s
- Average Training Time per Batch: 0.8513s
- Total Epoch Time: 333.9233s
================================================


================================================
Epoch 3 Summary:
- Loss: 1.4308
- Accuracy (Top-1): 47.52%
- Total Data Loading Time: 0.8981s
- Total Training Time: 335.5258s
- Average Training Time per Batch: 0.8581s
- Total Epoch Time: 336.6421s
================================================


================================================
Epoch 4 Summary:
- Loss: 1.2713
- Accuracy (Top-1): 53.88%
- Total Data Loading Time: 0.8917s
- Total Training Time: 335.2924s
- Average Training Time per Batch: 0.8575s
- Total Epoch Time: 336.3769s
================================================


================================================
Epoch 5 Summary:
- Loss: 1.1114
- Accuracy (Top-1): 60.33%
- Total Data Loading Time: 0.8881s
- Total Training Time: 332.8631s
- Average Training Time per Batch: 0.8513s
- Total Epoch Time: 333.9518s
================================================
```

Average Runtime per epoch on GPU: 11.19 seconds
Average Runtime per epoch on CPU: 335.08 seconds

C6:

```
python3 main.py --num_workers 4 --optimizer adam


================================================
Epoch 1 Summary:
- Loss: 2.2213
- Accuracy (Top-1): 20.92%
- Total Data Loading Time: 2.1536s
- Total Training Time: 8.6105s
- Average Training Time per Batch: 0.0220s
- Total Epoch Time: 11.9035s
================================================


================================================
Epoch 2 Summary:
- Loss: 1.8972
- Accuracy (Top-1): 26.73%
- Total Data Loading Time: 2.5293s
- Total Training Time: 8.8396s
- Average Training Time per Batch: 0.0226s
- Total Epoch Time: 12.5200s
================================================


================================================
Epoch 3 Summary:
- Loss: 1.8753
- Accuracy (Top-1): 27.41%
- Total Data Loading Time: 3.0686s
- Total Training Time: 9.1143s
- Average Training Time per Batch: 0.0233s
- Total Epoch Time: 13.3960s
================================================


================================================
Epoch 4 Summary:
- Loss: 1.8723
- Accuracy (Top-1): 26.94%
- Total Data Loading Time: 1.7760s
- Total Training Time: 7.8862s
- Average Training Time per Batch: 0.0202s
- Total Epoch Time: 10.8152s
================================================


================================================
Epoch 5 Summary:
- Loss: 1.8657
- Accuracy (Top-1): 27.89%
- Total Data Loading Time: 2.3062s
- Total Training Time: 8.4605s
- Average Training Time per Batch: 0.0216s
- Total Epoch Time: 11.9271s
================================================
```

```
python3 main.py --num_workers 4 --optimizer adagrad


================================================
Epoch 1 Summary:
- Loss: 2.1167
- Accuracy (Top-1): 26.01%
- Total Data Loading Time: 2.1472s
- Total Training Time: 8.7174s
- Average Training Time per Batch: 0.0223s
- Total Epoch Time: 11.8546s
================================================


================================================
Epoch 2 Summary:
- Loss: 1.6632
- Accuracy (Top-1): 38.02%
- Total Data Loading Time: 1.8833s
- Total Training Time: 8.0383s
- Average Training Time per Batch: 0.0206s
- Total Epoch Time: 10.8863s
================================================


================================================
Epoch 3 Summary:
- Loss: 1.5054
- Accuracy (Top-1): 44.26%
- Total Data Loading Time: 1.8835s
- Total Training Time: 8.0606s
- Average Training Time per Batch: 0.0206s
- Total Epoch Time: 10.9843s
================================================


================================================
Epoch 4 Summary:
- Loss: 1.3550
- Accuracy (Top-1): 50.23%
- Total Data Loading Time: 2.3044s
- Total Training Time: 8.6920s
- Average Training Time per Batch: 0.0222s
- Total Epoch Time: 11.9703s
================================================


================================================
Epoch 5 Summary:
- Loss: 1.2033
- Accuracy (Top-1): 56.62%
- Total Data Loading Time: 1.9279s
- Total Training Time: 7.9309s
- Average Training Time per Batch: 0.0203s
- Total Epoch Time: 10.8396s
================================================
```

```
python3 main.py --num_workers 4 --optimizer sgd


================================================
Epoch 1 Summary:
- Loss: 1.9070
- Accuracy (Top-1): 33.27%
- Total Data Loading Time: 2.1025s
- Total Training Time: 8.7295s
- Average Training Time per Batch: 0.0223s
- Total Epoch Time: 11.7390s
================================================


================================================
Epoch 2 Summary:
- Loss: 1.4366
- Accuracy (Top-1): 47.08%
- Total Data Loading Time: 1.8429s
- Total Training Time: 7.8803s
- Average Training Time per Batch: 0.0202s
- Total Epoch Time: 10.6128s
================================================


================================================
Epoch 3 Summary:
- Loss: 1.2259
- Accuracy (Top-1): 55.82%
- Total Data Loading Time: 1.8346s
- Total Training Time: 7.9170s
- Average Training Time per Batch: 0.0202s
- Total Epoch Time: 10.6271s
================================================


================================================
Epoch 4 Summary:
- Loss: 1.0679
- Accuracy (Top-1): 62.11%
- Total Data Loading Time: 2.3191s
- Total Training Time: 8.4341s
- Average Training Time per Batch: 0.0216s
- Total Epoch Time: 11.6515s
================================================


================================================
Epoch 5 Summary:
- Loss: 0.9364
- Accuracy (Top-1): 66.98%
- Total Data Loading Time: 1.9303s
- Total Training Time: 7.8062s
- Average Training Time per Batch: 0.0200s
- Total Epoch Time: 10.6519s
================================================
```

```
python3 main.py --num_workers 4 --optimizer sgdnesterov


================================================
Epoch 1 Summary:
- Loss: 1.9186
- Accuracy (Top-1): 31.27%
- Total Data Loading Time: 2.2553s
- Total Training Time: 8.4447s
- Average Training Time per Batch: 0.0216s
- Total Epoch Time: 11.8179s
================================================


================================================
Epoch 2 Summary:
- Loss: 1.4598
- Accuracy (Top-1): 46.29%
- Total Data Loading Time: 1.9213s
- Total Training Time: 7.6906s
- Average Training Time per Batch: 0.0197s
- Total Epoch Time: 10.6950s
================================================


================================================
Epoch 3 Summary:
- Loss: 1.2274
- Accuracy (Top-1): 55.81%
- Total Data Loading Time: 1.8367s
- Total Training Time: 7.7924s
- Average Training Time per Batch: 0.0199s
- Total Epoch Time: 10.7307s
================================================


================================================
Epoch 4 Summary:
- Loss: 1.0621
- Accuracy (Top-1): 62.23%
- Total Data Loading Time: 2.2843s
- Total Training Time: 8.2671s
- Average Training Time per Batch: 0.0211s
- Total Epoch Time: 11.7084s
================================================


================================================
Epoch 5 Summary:
- Loss: 0.9602
- Accuracy (Top-1): 66.14%
- Total Data Loading Time: 1.9022s
- Total Training Time: 7.7081s
- Average Training Time per Batch: 0.0197s
- Total Epoch Time: 10.6985s
================================================
```

```
python3 main.py --num_workers 4 --optimizer adadelta


================================================
Epoch 1 Summary:
- Loss: 1.5511
- Accuracy (Top-1): 42.98%
- Total Data Loading Time: 1.7267s
- Total Training Time: 8.6770s
- Average Training Time per Batch: 0.0222s
- Total Epoch Time: 12.1853s
================================================


================================================
Epoch 2 Summary:
- Loss: 1.1635
- Accuracy (Top-1): 58.03%
- Total Data Loading Time: 1.4208s
- Total Training Time: 7.8783s
- Average Training Time per Batch: 0.0201s
- Total Epoch Time: 11.0670s
================================================


================================================
Epoch 3 Summary:
- Loss: 0.9635
- Accuracy (Top-1): 65.82%
- Total Data Loading Time: 1.5166s
- Total Training Time: 8.1016s
- Average Training Time per Batch: 0.0207s
- Total Epoch Time: 11.3884s
================================================


================================================
Epoch 4 Summary:
- Loss: 0.8321
- Accuracy (Top-1): 70.47%
- Total Data Loading Time: 1.7382s
- Total Training Time: 8.2533s
- Average Training Time per Batch: 0.0211s
- Total Epoch Time: 11.8342s
================================================


================================================
Epoch 5 Summary:
- Loss: 0.7412
- Accuracy (Top-1): 73.90%
- Total Data Loading Time: 1.3943s
- Total Training Time: 7.9371s
- Average Training Time per Batch: 0.0203s
- Total Epoch Time: 11.0661s
================================================
```

Recap:

Adam:

Average Training Time: 8.5822 seconds
Average Loss: 1.9464
Average Accuracy: 25.978%

Adagrad:

Average Training Time: 8.2878 seconds
Average Loss: 1.5687
Average Accuracy: 42.828%

SGD:

Average Training Time: 8.1534 seconds
Average Loss: 1.3148
Average Accuracy: 53.052%

Nesterov:

Average Training Time: 7.9806 seconds
Average Loss: 1.3256
Average Accuracy: 52.388%

Adadelta:

Average Training Time: 8.1695 seconds
Average Loss: 1.0503
Average Accuracy: 62.24%

Discussion:

1. Training Time:

1. **Nesterov** – 7.9806 seconds
2. **SGD** – 8.1534 seconds
3. **Adadelta** – 8.1695 seconds
4. **Adagrad** – 8.2878 seconds
5. **Adam** – 8.5822 seconds

2. Loss:

1. **Adadelta** – 1.0503
2. **SGD** – 1.3148
3. **Nesterov** – 1.3256
4. **Adagrad** – 1.5687
5. **Adam** – 1.9464

3. Accuracy):

1. **Adadelta** – 62.24%
2. **SGD** – 53.05%
3. **Nesterov** – 52.39%
4. **Adagrad** – 42.83%
5. **Adam** – 25.98%

Adam places last in Training Time, Loss, and Accuracy. Thus this optimizer is probably the worst for this dataset.

Adadelta has the best loss and accuracy, which means it's probably the best for this dataset. It was not significantly slower than SGD or Nesterov, so it seems like this performs the best by far.

SGD and SGD w Nesterov perform about the same and are the fastest optimizers for this run.

Adagrad performs relatively poor on accuracy as well as timing.

Reasoning:

Adam performed the worst in terms of both accuracy and loss, while also being the slowest optimizer. This result might come from Adam’s reliance on adaptive learning rates, which dynamically change based on gradient estimates. This might be overly complex for this dataset in particular. Additionally, Adam’s has an increased computational overhead - it needs to track both first and second-order moments. This likely contributed to its slower training time.

Adadelta performed the best, achieving the lowest loss and highest accuracy. Its success can be attributed to its dynamic learning rate adjustment, which scales updates based on the magnitude of recent gradients. This adaptability allows Adadelta to make larger updates early in training while gradually refining its learning rate as convergence nears. It is less computationally expensive than Adam, allowing it to maintain competitive training speeds.

Despite the differences in optimizer behavior, the total training times across all optimizers were relatively close. This outcome reflects the fact that PyTorch efficiently handles gradient computations, and the fundamental backpropagation process remains identical across optimizers. While Adam and Adadelta have slightly higher computational overhead due to their

adaptive learning rate mechanisms, these operations are lightweight enough that they don't really impact overall runtime that much.

C7:

```
python3 main.py --num_workers 4 --disable_batch_normalization


================================================
Epoch 1 Summary:
- Loss: 1.9436
- Accuracy (Top-1): 26.75%
- Total Data Loading Time: 4.0468s
- Total Training Time: 5.9585s
- Average Training Time per Batch: 0.0152s
- Total Epoch Time: 11.0368s
================================================


================================================
Epoch 2 Summary:
- Loss: 1.5765
- Accuracy (Top-1): 42.33%
- Total Data Loading Time: 3.7461s
- Total Training Time: 5.4392s
- Average Training Time per Batch: 0.0139s
- Total Epoch Time: 10.2070s
================================================


================================================
Epoch 3 Summary:
- Loss: 1.4097
- Accuracy (Top-1): 49.03%
- Total Data Loading Time: 3.7556s
- Total Training Time: 5.3925s
- Average Training Time per Batch: 0.0138s
- Total Epoch Time: 10.1907s
================================================


================================================
Epoch 4 Summary:
- Loss: 1.2687
- Accuracy (Top-1): 54.68%
- Total Data Loading Time: 3.6221s
- Total Training Time: 5.4917s
- Average Training Time per Batch: 0.0140s
- Total Epoch Time: 10.2013s
================================================


================================================
Epoch 5 Summary:
- Loss: 1.1555
- Accuracy (Top-1): 59.02%
- Total Data Loading Time: 4.7336s
- Total Training Time: 5.3910s
- Average Training Time per Batch: 0.0138s
- Total Epoch Time: 11.1837s
================================================
```

```
python3 main.py --num_workers 4


================================================
Epoch 1 Summary:
- Loss: 2.0807
- Accuracy (Top-1): 25.86%
- Total Data Loading Time: 2.3712s
- Total Training Time: 8.6696s
- Average Training Time per Batch: 0.0222s
- Total Epoch Time: 12.0196s
================================================


================================================
Epoch 2 Summary:
- Loss: 1.5860
- Accuracy (Top-1): 41.08%
- Total Data Loading Time: 3.3952s
- Total Training Time: 8.9312s
- Average Training Time per Batch: 0.0228s
- Total Epoch Time: 13.3404s
================================================


================================================
Epoch 3 Summary:
- Loss: 1.4001
- Accuracy (Top-1): 48.78%
- Total Data Loading Time: 2.3163s
- Total Training Time: 8.0280s
- Average Training Time per Batch: 0.0205s
- Total Epoch Time: 11.3759s
================================================


================================================
Epoch 4 Summary:
- Loss: 1.2107
- Accuracy (Top-1): 56.68%
- Total Data Loading Time: 2.3411s
- Total Training Time: 8.3623s
- Average Training Time per Batch: 0.0214s
- Total Epoch Time: 11.6552s
================================================


================================================
Epoch 5 Summary:
- Loss: 1.0397
- Accuracy (Top-1): 63.47%
- Total Data Loading Time: 1.8646s
- Total Training Time: 7.7198s
- Average Training Time per Batch: 0.0197s
- Total Epoch Time: 10.5443s
================================================
```

With Batch Normalization:

Average Training Time: 8.3422 seconds
Average Loss: 1.4634
Average Accuracy: 47.974%

Without Batch Normalization:

Average Training Time: 5.5346 seconds
Average Loss: 1.4708
Average Accuracy: 46.762%

Discussion:

The main difference between the two runs is that without batch normalization, the average training per epoch was about 3 seconds slower than the run with batch normalization. This is because there are mean and variance calculations for each batch, and extra operations like scaling and shiting the normalized values.

There is a slight performance boost with batch normalization, for both loss and accuracy. This is most likely because batch normalization prevents covariance shifts across batches and helps the model learn better. However, in this case, it did not have a huge impact on the performance.

The batch normalization run converged a bit better than the run without batch normalization. This makes sense because batch normalization helps convergence by reducing covariance shifts. The accuracy of the 5th epoch was around 63 percent with batch normalization, and without batch normalization, it was about 59 %.

C8:

The following 4 screenshots are for compile modes: default, reduce_overhead, and max_autotune, followed by the eager mode.

```
python3 main.py --num_workers 4 --torch_compile default --num_epochs 10

default

================================================
Epoch 1 Summary:
- Loss: 2.0101
- Accuracy (Top-1): 28.48%
- Total Data Loading Time: 1.7721s
- Total Training Time: 22.9558s
- Average Training Time per Batch: 0.0587s
- Total Epoch Time: 25.6416s
================================================


================================================
Epoch 2 Summary:
- Loss: 1.5347
- Accuracy (Top-1): 43.28%
- Total Data Loading Time: 2.1600s
- Total Training Time: 7.9539s
- Average Training Time per Batch: 0.0203s
- Total Epoch Time: 11.0455s
================================================


================================================
Epoch 3 Summary:
- Loss: 1.3454
- Accuracy (Top-1): 51.37%
- Total Data Loading Time: 2.9576s
- Total Training Time: 8.2128s
- Average Training Time per Batch: 0.0210s
- Total Epoch Time: 12.1298s
================================================


================================================
Epoch 4 Summary:
- Loss: 1.1608
- Accuracy (Top-1): 58.75%
- Total Data Loading Time: 2.1173s
- Total Training Time: 8.0198s
- Average Training Time per Batch: 0.0205s
- Total Epoch Time: 11.0977s
================================================


================================================
Epoch 5 Summary:
- Loss: 1.0312
- Accuracy (Top-1): 63.52%
- Total Data Loading Time: 2.1510s
- Total Training Time: 8.0023s
- Average Training Time per Batch: 0.0205s
- Total Epoch Time: 11.1124s
================================================


================================================
Epoch 6 Summary:
- Loss: 0.9340
- Accuracy (Top-1): 67.31%
- Total Data Loading Time: 2.1851s
- Total Training Time: 7.8384s
- Average Training Time per Batch: 0.0200s
- Total Epoch Time: 10.9930s
================================================


================================================
Epoch 7 Summary:
- Loss: 0.8668
- Accuracy (Top-1): 69.56%
- Total Data Loading Time: 2.0977s
- Total Training Time: 7.9328s
- Average Training Time per Batch: 0.0203s
- Total Epoch Time: 10.9830s
================================================


================================================
Epoch 8 Summary:
- Loss: 0.7998
- Accuracy (Top-1): 72.26%
- Total Data Loading Time: 2.7286s
- Total Training Time: 8.3166s
- Average Training Time per Batch: 0.0213s
- Total Epoch Time: 11.9920s
================================================


================================================
Epoch 9 Summary:
- Loss: 0.7425
- Accuracy (Top-1): 74.24%
- Total Data Loading Time: 2.1977s
- Total Training Time: 7.8493s
- Average Training Time per Batch: 0.0201s
- Total Epoch Time: 10.9833s
================================================


================================================
Epoch 10 Summary:
- Loss: 0.7057
- Accuracy (Top-1): 75.49%
- Total Data Loading Time: 2.1217s
- Total Training Time: 7.9434s
- Average Training Time per Batch: 0.0203s
- Total Epoch Time: 11.0066s
================================================
```

```
python3 main.py --num_workers 4 --torch_compile reduce-overhead --num_epochs 10

reduce-overhead

================================================
Epoch 1 Summary:
- Loss: 1.9531
- Accuracy (Top-1): 30.68%
- Total Data Loading Time: 3.8912s
- Total Training Time: 18.1106s
- Average Training Time per Batch: 0.0463s
- Total Epoch Time: 24.4874s
================================================


================================================
Epoch 2 Summary:
- Loss: 1.5101
- Accuracy (Top-1): 44.33%
- Total Data Loading Time: 4.6669s
- Total Training Time: 4.7339s
- Average Training Time per Batch: 0.0121s
- Total Epoch Time: 11.7527s
================================================


================================================
Epoch 3 Summary:
- Loss: 1.3074
- Accuracy (Top-1): 52.57%
- Total Data Loading Time: 4.4339s
- Total Training Time: 3.4539s
- Average Training Time per Batch: 0.0088s
- Total Epoch Time: 10.1334s
================================================


================================================
Epoch 4 Summary:
- Loss: 1.1505
- Accuracy (Top-1): 58.76%
- Total Data Loading Time: 4.4692s
- Total Training Time: 3.3402s
- Average Training Time per Batch: 0.0085s
- Total Epoch Time: 10.1091s
================================================


================================================
Epoch 5 Summary:
- Loss: 1.0124
- Accuracy (Top-1): 64.18%
- Total Data Loading Time: 4.5370s
- Total Training Time: 3.3158s
- Average Training Time per Batch: 0.0085s
- Total Epoch Time: 10.1253s
================================================


================================================
Epoch 6 Summary:
- Loss: 0.9252
- Accuracy (Top-1): 67.36%
- Total Data Loading Time: 4.4633s
- Total Training Time: 3.2617s
- Average Training Time per Batch: 0.0083s
- Total Epoch Time: 10.0814s
================================================


================================================
Epoch 7 Summary:
- Loss: 0.8327
- Accuracy (Top-1): 71.15%
- Total Data Loading Time: 5.2144s
- Total Training Time: 3.5445s
- Average Training Time per Batch: 0.0091s
- Total Epoch Time: 11.0741s
================================================


================================================
Epoch 8 Summary:
- Loss: 0.7751
- Accuracy (Top-1): 73.12%
- Total Data Loading Time: 4.5995s
- Total Training Time: 3.2048s
- Average Training Time per Batch: 0.0082s
- Total Epoch Time: 10.1409s
================================================


================================================
Epoch 9 Summary:
- Loss: 0.7207
- Accuracy (Top-1): 75.00%
- Total Data Loading Time: 4.4626s
- Total Training Time: 3.3860s
- Average Training Time per Batch: 0.0087s
- Total Epoch Time: 10.1199s
================================================


================================================
Epoch 10 Summary:
- Loss: 0.6880
- Accuracy (Top-1): 76.24%
- Total Data Loading Time: 4.4175s
- Total Training Time: 3.4414s
- Average Training Time per Batch: 0.0088s
- Total Epoch Time: 10.1005s
================================================
```

```
python3 main.py --num_workers 4 --torch_compile max-autotune --num_epochs 10

max-autotune

================================================
Epoch 1 Summary:
- Loss: 1.9679
- Accuracy (Top-1): 29.06%
- Total Data Loading Time: 3.8943s
- Total Training Time: 18.4626s
- Average Training Time per Batch: 0.0472s
- Total Epoch Time: 24.7163s
================================================


================================================
Epoch 2 Summary:
- Loss: 1.5390
- Accuracy (Top-1): 43.19%
- Total Data Loading Time: 4.1191s
- Total Training Time: 4.7061s
- Average Training Time per Batch: 0.0120s
- Total Epoch Time: 11.0558s
================================================


================================================
Epoch 3 Summary:
- Loss: 1.3625
- Accuracy (Top-1): 50.46%
- Total Data Loading Time: 4.1942s
- Total Training Time: 3.4452s
- Average Training Time per Batch: 0.0088s
- Total Epoch Time: 9.9979s
================================================


================================================
Epoch 4 Summary:
- Loss: 1.1867
- Accuracy (Top-1): 57.50%
- Total Data Loading Time: 4.3170s
- Total Training Time: 3.4254s
- Average Training Time per Batch: 0.0088s
- Total Epoch Time: 10.0639s
================================================


================================================
Epoch 5 Summary:
- Loss: 1.0612
- Accuracy (Top-1): 62.46%
- Total Data Loading Time: 4.2824s
- Total Training Time: 3.2697s
- Average Training Time per Batch: 0.0084s
- Total Epoch Time: 9.9311s
================================================


================================================
Epoch 6 Summary:
- Loss: 0.9725
- Accuracy (Top-1): 65.81%
- Total Data Loading Time: 4.2245s
- Total Training Time: 3.3482s
- Average Training Time per Batch: 0.0086s
- Total Epoch Time: 9.9056s
================================================


================================================
Epoch 7 Summary:
- Loss: 0.8960
- Accuracy (Top-1): 68.53%
- Total Data Loading Time: 5.3657s
- Total Training Time: 3.3746s
- Average Training Time per Batch: 0.0086s
- Total Epoch Time: 11.0857s
================================================


================================================
Epoch 8 Summary:
- Loss: 0.8220
- Accuracy (Top-1): 71.28%
- Total Data Loading Time: 4.2072s
- Total Training Time: 3.4221s
- Average Training Time per Batch: 0.0088s
- Total Epoch Time: 9.9491s
================================================


================================================
Epoch 9 Summary:
- Loss: 0.7788
- Accuracy (Top-1): 72.96%
- Total Data Loading Time: 4.0886s
- Total Training Time: 3.4999s
- Average Training Time per Batch: 0.0090s
- Total Epoch Time: 9.9479s
================================================


================================================
Epoch 10 Summary:
- Loss: 0.7222
- Accuracy (Top-1): 75.06%
- Total Data Loading Time: 4.2599s
- Total Training Time: 3.4081s
- Average Training Time per Batch: 0.0087s
- Total Epoch Time: 9.9610s
================================================


python3 main.py --num_workers 4 --torch_compile none --num_epochs 10
```

```
python3 main.py --num_workers 4 --torch_compile none --num_epochs 10

none

================================================
Epoch 1 Summary:
- Loss: 1.9892
- Accuracy (Top-1): 27.99%
- Total Data Loading Time: 2.1486s
- Total Training Time: 8.2746s
- Average Training Time per Batch: 0.0212s
- Total Epoch Time: 11.3758s
================================================


================================================
Epoch 2 Summary:
- Loss: 1.5685
- Accuracy (Top-1): 41.65%
- Total Data Loading Time: 2.3674s
- Total Training Time: 8.5978s
- Average Training Time per Batch: 0.0220s
- Total Epoch Time: 11.8905s
================================================


================================================
Epoch 3 Summary:
- Loss: 1.3814
- Accuracy (Top-1): 49.83%
- Total Data Loading Time: 1.9288s
- Total Training Time: 7.9735s
- Average Training Time per Batch: 0.0204s
- Total Epoch Time: 10.8794s
================================================


================================================
Epoch 4 Summary:
- Loss: 1.1971
- Accuracy (Top-1): 56.96%
- Total Data Loading Time: 2.1471s
- Total Training Time: 7.8217s
- Average Training Time per Batch: 0.0200s
- Total Epoch Time: 10.9306s
================================================


================================================
Epoch 5 Summary:
- Loss: 1.0621
- Accuracy (Top-1): 62.53%
- Total Data Loading Time: 3.3060s
- Total Training Time: 9.1939s
- Average Training Time per Batch: 0.0235s
- Total Epoch Time: 13.5098s
================================================


================================================
Epoch 6 Summary:
- Loss: 0.9718
- Accuracy (Top-1): 65.67%
- Total Data Loading Time: 2.2547s
- Total Training Time: 8.5129s
- Average Training Time per Batch: 0.0218s
- Total Epoch Time: 11.7946s
================================================


================================================
Epoch 7 Summary:
- Loss: 0.8898
- Accuracy (Top-1): 68.92%
- Total Data Loading Time: 2.3436s
- Total Training Time: 8.5642s
- Average Training Time per Batch: 0.0219s
- Total Epoch Time: 11.8883s
================================================


================================================
Epoch 8 Summary:
- Loss: 0.8295
- Accuracy (Top-1): 71.13%
- Total Data Loading Time: 2.0134s
- Total Training Time: 7.8882s
- Average Training Time per Batch: 0.0202s
- Total Epoch Time: 10.8736s
================================================


================================================
Epoch 9 Summary:
- Loss: 0.7786
- Accuracy (Top-1): 73.04%
- Total Data Loading Time: 1.9373s
- Total Training Time: 8.0376s
- Average Training Time per Batch: 0.0206s
- Total Epoch Time: 10.9296s
================================================


================================================
Epoch 10 Summary:
- Loss: 0.7230
- Accuracy (Top-1): 74.93%
- Total Data Loading Time: 1.9232s
- Total Training Time: 8.0336s
- Average Training Time per Batch: 0.0205s
- Total Epoch Time: 10.9338s
================================================
```

	No Torch.compile	Torch.compile		
	Eager Mode	Default	reduce-overhead	max-autotune
Time for first epoch (s)	8.27	22.96	18.11	18.46
Avg times for epochs 6 - 10 (s)	8.21	7.98	3.37	3.41

1. Eager Mode (No torch.compile)
    - First epoch time is low (8.27s) because there is no compilation overhead.
    - Subsequent epochs remain similar (8.21s), as no optimizations are applied.
2. Default torch.compile
    - First epoch time is significantly higher (22.96s) due to the compilation overhead.
    - Subsequent epochs improve slightly (7.98s), but not drastically.
3. reduce-overhead mode
    - First epoch time (18.11s) is lower than the default torch.compile, suggesting reduced compilation time.
    - Avg time for later epochs (3.37s) is much lower, showing that the optimization is effective.
4. max-autotune mode
    - First epoch time (18.46s) is similar to reduce-overhead, implying a comparable initial compilation cost.
    - Subsequent epochs (3.41s) are also well-optimized.
    - Seems like max-autotune mode does not provide that much benefit here

Q1

How many convolutional layers are in the ResNet-18 model?

There are a total of 20 convolutional layers in the ResNet-18 model. Running print(model) shows that there are 2 convolutions per basic block. There are 2 basic blocks per layer, and 4 layers in the model. This is a total of 4 x 2 x 2 = 16 convolutional layers in the main portion of the model. There is also 1 initial convolution before any of the original layers. Additionally, there are 3 more 1 x 1 convolutions in the 2nd, 3rd, and 4th layers of the model for the skip connection to change the number of channels. This yields 20 convolutional layers in the entire model.

Q2

What is the input dimension of the last linear layer?
The input dimension of the last linear layer is 512. This can be seen by performing print(model.fc), which shows that the last linear layer takes 512 inputs and has 10 outputs corresponding to the 10 classes possible for prediction.

Q3

How many trainable parameters and how many gradients in the ResNet-18 model that you build (please show both the answer and the code that you use to count them), when using SGD optimizer?

To find the number of trainable parameters: I used this code snippet:

```python
import torch.nn as nn
from Resnet18 import ResNet18

model = ResNet18(num_classes=10)

def count_conv_layers(model):
    return sum(1 for layer in model.modules() if isinstance(layer, nn.Conv2d))

# Count convolutional layers
num_conv_layers = sum(1 for layer in model.modules() if isinstance(layer, nn.Conv2d))
print(num_conv_layers)

for name, param in model.named_parameters():
    print(f"{name}: {param.size()}")

#summary(model, input_size=(3, 32, 32))  # 3 channels, 32x32 image size

num_trainable_params = sum(p.numel() for p in model.parameters() if p.requires_grad)

print(f"Number of trainable parameters: {num_trainable_params}")

# Count the number of parameters that require gradients
num_gradients = sum(p.numel() for p in model.parameters() if p.requires_grad)

print(f"Number of gradients: {num_gradients}")
```

It looks like the total number of trainable parameters is: 11,173, 962 for this ResNet-18.

```
Number of trainable parameters: 11173962
Number of gradients: 11173962
```

This shows that the number of gradients was the same as the number of trainable parameters.

Q4

Same question as Q3, except now using Adam (only the answer is required, not the code).

There are still 11,173, 962 trainable parameters and 11,173, 962 gradients with Adam.

Q5

1) What effect does torch.compile have, 2) how does it make implementations faster and

3) why is the first epoch significantly slower in torch.compile compared to eager mode?

1. This allows for faster execution during training and inference. It makes the first epoch slower due to compilation overhead, but faster subsequent epochs due to optimized execution.
2. torch.compile() optimizes Pytorch models by converting them into a more efficient, lower-level representation. It combines multiple tensor operations to reduce memory access and improve efficiency. It converts eager-mode execution into an optimized computation graph. It uses some backend specific accelerations, like TorchInductor to optimize execution. It also reduces Pythonic overhead by minimizing Python involvement in code.
3. The first epoch is much slower in torch.compile() compared to eager mode because of compilation overhead. This is because torch.compile converts the PyTorch model into an optimized computation graph, and performs the optimizations mentioned above. This happens during the first epoch and causes a large slowdown. Eager mode executes operations immediately with no compilation, so the first epoch does not incur this costly slowdown. However, it is slower than the compiled version in subsequent epochs.

Extra Credit