

total\_frozen\_params=210236928  
trainable parameters: 28366848  
/tmp/ipython-input-32-347286690.py:16: FutureWarning: `torch.cuda.amp.GradScaler(args...)` is deprecated. Please use `torch.amp.GradScaler('cuda', args...)` instead.  
self.scaler = GradScaler()  
/tmp/ipython-input-32-347286690.py:38: UserWarning: Argument(s) 'always\_apply' are not valid for transform Normalize  
A.Normalize(mean=[0.5,0.5,0.5],std=[0.5,0.5,0.5],always\_apply=True),  
validating: 100% 10/10 [1:14:27<00:00, 440.77s/it]  
  
train loss: 6.121: 100% 782/782 [06:35<00:00, 1.76it/s]  
  
/tmp/ipython-input-32-347286690.py:58: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.  
with autocast():  
valid loss: 6.059: 100% 79/79 [00:30<00:00, 4.09it/s]  
  
/tmp/ipython-input-32-347286690.py:88: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.  
with autocast():  
train\_loss train\_perplexity val\_loss val\_perplexity  
0 6.891201 983.581961 5.829692 340.254015  
saving best model...  
train loss: 4.320: 100% 782/782 [06:40<00:00, 1.70it/s]  
  
/tmp/ipython-input-32-347286690.py:58: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.  
with autocast():  
valid loss: 4.628: 100% 79/79 [00:30<00:00, 3.82it/s]  
  
/tmp/ipython-input-32-347286690.py:88: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.  
with autocast():  
train\_loss train\_perplexity val\_loss val\_perplexity  
1 5.261354 192.742303 4.076043 58.911878  
saving best model...  
unfreezing GPT2 entirely...  
total\_frozen\_params=0  
train loss: 2.593: 100% 782/782 [06:59<00:00, 2.39it/s]  
  
/tmp/ipython-input-32-347286690.py:58: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.  
with autocast():  
valid loss: 2.970: 100% 79/79 [00:29<00:00, 4.01it/s]  
  
/tmp/ipython-input-32-347286690.py:88: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.  
with autocast():  
train\_loss train\_perplexity val\_loss val\_perplexity  
2 3.047644 21.065648 2.54525 12.746413  
saving best model...  
train loss: 2.261: 100% 782/782 [07:02<00:00, 2.52it/s]  
  
/tmp/ipython-input-32-347286690.py:58: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.  
with autocast():  
valid loss: 2.804: 100% 79/79 [00:29<00:00, 3.76it/s]  
  
/tmp/ipython-input-32-347286690.py:88: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.  
with autocast():  
train\_loss train\_perplexity val\_loss val\_perplexity  
3 2.432715 11.389758 2.394751 10.965465  
saving best model...  
train loss: 2.171: 100% 782/782 [07:03<00:00, 2.48it/s]  
  
/tmp/ipython-input-32-347286690.py:58: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.  
with autocast():  
valid loss: 2.780: 100% 79/79 [00:29<00:00, 3.96it/s]  
  
/tmp/ipython-input-32-347286690.py:88: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.  
with autocast():  
train\_loss train\_perplexity val\_loss val\_perplexity  
4 2.232986 9.327677 2.337731 10.357712  
saving best model...  
train loss: 1.921: 100% 782/782 [07:02<00:00, 2.20it/s]  
  
/tmp/ipython-input-32-347286690.py:58: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.  
with autocast():  
valid loss: 2.710: 100% 79/79 [00:29<00:00, 3.90it/s]  
  
/tmp/ipython-input-32-347286690.py:88: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.  
with autocast():  
train\_loss train\_perplexity val\_loss val\_perplexity  
5 2.088745 8.074779 2.301299 9.987152  
saving best model...  
train loss: 2.016: 100% 782/782 [07:02<00:00, 3.61it/s]  
  
/tmp/ipython-input-32-347286690.py:58: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.  
with autocast():  
valid loss: 2.637: 100% 79/79 [00:29<00:00, 4.03it/s]  
  
/tmp/ipython-input-32-347286690.py:88: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.  
with autocast():  
train\_loss train\_perplexity val\_loss val\_perplexity  
6 1.970605 7.175016 2.280509 9.78166  
saving best model...  
train loss: 1.966: 100% 782/782 [07:01<00:00, 2.43it/s]  
  
/tmp/ipython-input-32-347286690.py:58: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.  
with autocast():  
valid loss: 2.674: 100% 79/79 [00:29<00:00, 3.86it/s]  
  
/tmp/ipython-input-32-347286690.py:88: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.  
with autocast():  
train\_loss train\_perplexity val\_loss val\_perplexity  
7 1.870606 6.492229 2.283502 9.810977  
train loss: 1.827: 100% 782/782 [06:38<00:00, 2.40it/s]  
  
/tmp/ipython-input-32-347286690.py:58: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.  
with autocast():  
valid loss: 2.708: 100% 79/79 [00:29<00:00, 3.88it/s]  
  
/tmp/ipython-input-32-347286690.py:88: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.  
with autocast():  
train\_loss train\_perplexity val\_loss val\_perplexity  
8 1.796747 6.03 2.293451 9.909073

```
train loss: 1.808: 100%
782/782 [06:39<00:00, 2.37it/s]
/tmp/ipython-input-32-347286690.py:58: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
  with autocast():
valid loss: 2.722: 100%
79/79 [00:29<00:00, 3.96it/s]
/tmp/ipython-input-32-347286690.py:88: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
  with autocast():
  train_loss train_perplexity val_loss val_perplexity
9    1.760392         5.814714  2.299583         9.97002
```