01_Homework_6_Language_Modeling_Release

March 28, 2022

0.1 Homework 6

0.1.1 Language Modeling

Welcome to Homework 6!

The homework contains several tasks. You can find the amount of points that you get for the correct solution in the task header. Maximum amount of points for each homework is *four*.

The **grading** for each task is the following: - correct answer - **full points** - insufficient solution or solution resulting in the incorrect output - **half points** - no answer or completely wrong solution - **no points**

Even if you don't know how to solve the task, we encourage you to write down your thoughts and progress and try to address the issues that stop you from completing the task.

When working on the written tasks, try to make your answers short and accurate. Most of the times, it is possible to answer the question in 1-3 sentences.

When writing code, make it readable. Choose appropriate names for your variables (`a = 'cat'` - not good, `word = 'cat'` - good). Avoid constructing lines of code longer than 100 characters (79 characters is ideal). If needed, provide the commentaries for your code, however, a good code should be easily readable without them :)

Finally, all your answers should be written only by yourself. If you copy them from other sources it will be considered as an academic fraud. You can discuss the tasks with your classmates but each solution must be individual.

Important!: before sending your solution, do the `Kernel -> Restart & Run All` to ensure that all your code works.

[1]:
```
!nvcc --version
```

```
/bin/bash: nvcc: command not found
```

[2]:
```
!pip3 install --quiet torchtext==0.11.0 datasets torchinfo
```

```
ERROR: Could not find a version that satisfies the requirement
torchtext==0.11.0 (from versions: 0.1.1, 0.2.0, 0.2.1, 0.2.3, 0.3.1, 0.4.0,
0.5.0, 0.6.0, 0.12.0)
ERROR: No matching distribution found for torchtext==0.11.0
```

```
WARNING: You are using pip version 21.3.1; however, version 22.0.4 is
available.
You should consider upgrading via the
'/gpfs/space/software/jupyterhub/python/jupyter/bin/python -m pip install
--upgrade pip' command.
```

[4]:
```
from datasets import load_dataset
from torchtext.data.utils import get_tokenizer
from torchtext.vocab import build_vocab_from_iterator
from torchinfo import summary
```

Load the Penn Treebank dataset. Structurally, it is the same as the Wikitext-2 dataset used in the Lab 6. Please, refer to the Lab materials for more details on data structure and loading.

[5]:
```
train_dataset = load_dataset("ptb_text_only", split="train")
```

```
Downloading builder script: 6.50kB [00:00, 1.94MB/s]
Downloading metadata: 2.15kB [00:00, 639kB/s]

Downloading and preparing dataset ptb_text_only/penn_treebank (download: 5.68
MiB, generated: 5.72 MiB, post-processed: Unknown size, total: 11.40 MiB) to /gp
fs/space/home/chenghan/.cache/huggingface/datasets/ptb_text_only/penn_treebank/1
.1.0/8d1b97746fb9765d140e569ec5ddd35e20af4d37761f5e1bf357ea0b081f2c1f…

Downloading data files:   0%|          | 0/3 [00:00<?, ?it/s]
Downloading data:   0%|          | 0.00/1.70M [00:00<?, ?B/s]
Downloading data: 5.10MB [00:00, 27.0MB/s]
Downloading data files:  33%|        | 1/3 [00:01<00:03,  1.56s/it]
Downloading data: 400kB [00:00, 6.30MB/s]
Downloading data files:  67%|       | 2/3 [00:02<00:00,  1.05it/s]
Downloading data: 450kB [00:00, 7.12MB/s]
Downloading data files: 100%|      | 3/3 [00:02<00:00,  1.14it/s]
Extracting data files: 100%|      | 3/3 [00:00<00:00, 260.29it/s]

Dataset ptb_text_only downloaded and prepared to /gpfs/space/home/chenghan/.cach
e/huggingface/datasets/ptb_text_only/penn_treebank/1.1.0/8d1b97746fb9765d140e569
ec5ddd35e20af4d37761f5e1bf357ea0b081f2c1f. Subsequent calls will reuse this
data.
```

[6]:
```
tokenizer = get_tokenizer('basic_english')
vocab = build_vocab_from_iterator(map(tokenizer, train_dataset['sentence']),␣
 ↪specials=['<pad>', '<unk>', '<bos>', '<eos>'])
vocab.set_default_index(vocab['<unk>'])
```

[7]:
```
import torch
from torch import nn, Tensor
from torch.utils.data import dataset
```

```
def data_process(raw_text_iter: dataset.IterableDataset, device: torch.device)␣
↪-> Tensor:
    """Converts raw text into a flat Tensor."""
    data = [torch.tensor(vocab(['<bos>']) + vocab(tokenizer(item["sentence"])),␣
↪dtype=torch.long, device=device) for item in raw_text_iter]
    return list(filter(lambda t: t.numel() > 2, data))
```

[8]:
```
device = torch.device('cuda' if torch.cuda.is_available() else 'cpu')

dataset = load_dataset("ptb_text_only")
train_data = data_process(dataset["train"], device)
val_data = data_process(dataset["validation"], device)
test_data = data_process(dataset["test"], device)
```

```
Reusing dataset ptb_text_only (/gpfs/space/home/chenghan/.cache/huggingface/data
sets/ptb_text_only/penn_treebank/1.1.0/8d1b97746fb9765d140e569ec5ddd35e20af4d377
61f5e1bf357ea0b081f2c1f)
100%|      | 3/3 [00:00<00:00, 250.35it/s]
```

[9]:
```
from torch.nn.utils.rnn import pad_sequence
def _collate_fn(batch):
    inp = pad_sequence(batch, batch_first=True)
    target = pad_sequence([torch.cat((item[1:], torch.tensor(vocab['<eos>'],␣
↪device=device).unsqueeze(0))) for item in batch], batch_first=True)
    return inp, target
```

[10]:
```
batch_size = 16

train_dataloader = torch.utils.data.DataLoader(train_data,␣
↪batch_size=batch_size, collate_fn=_collate_fn, shuffle=True)
val_dataloader = torch.utils.data.DataLoader(val_data, batch_size=batch_size,␣
↪collate_fn=_collate_fn, shuffle=False)
test_dataloader = torch.utils.data.DataLoader(test_data, batch_size=batch_size,␣
↪collate_fn=_collate_fn, shuffle=False)
```

0.2 Task 1. Highway Model (3.5 points)

In this task, you will have to modify the model from the Lab 6.

You will have to add: - Three more convolutional layers - Highway layer

To add extra convolutional layers, you can just copy `conv_block_1` and call it `conv_block_2`, for example.

Then, you will need to add a `gate_layer` which is a simple linear layer that outputs the same dimension as it takes.

In the forward pass, add a `transform_gate` which is a nonlinear transformation of the embedded inputs, i.e. the gate layer followed by a sigmoid. Then, add a `carry_gate` which is simply `1 - transform_gate`. Finally, the output of the highway layer is the element-wise multiplication of the

input by the carry gate plus the element-wise multiplication of the previous layer output and the transform gate. You can see more information about the highway layer here.

Finally, carry the output of the highway layer to the next convolutional block.

Hint: To perform an element-wise multiplication of tensor `a` and `b`, you can do `a * b` or `torch.mul(a, b)`.

[11]:
```
class CNNLM(nn.Module):
    def __init__(self, num_words, emb_dim, hid_dim, kernel_size,␣
↪tie_weights=False):
        super().__init__()
        self.emb = nn.Embedding(num_words, emb_dim, padding_idx=0)
        pad_size = kernel_size - 1
        self.conv_block_1 = nn.Sequential(nn.ConstantPad1d((pad_size, 0), 0),
                                  nn.Conv1d(emb_dim, hid_dim, kernel_size),
                                  nn.ConstantPad1d((pad_size, 0), 0),
                                  nn.Conv1d(hid_dim, hid_dim, kernel_size),
                                  nn.ConstantPad1d((pad_size, 0), 0),
                                  nn.Conv1d(hid_dim, hid_dim, kernel_size))

        # TODO: Define another three convolutional layers
        self.conv_block_2 = nn.Sequential(nn.ConstantPad1d((pad_size, 0), 0),
                                  nn.Conv1d(emb_dim, hid_dim, kernel_size),
                                  nn.ConstantPad1d((pad_size, 0), 0),
                                  nn.Conv1d(hid_dim, hid_dim, kernel_size),
                                  nn.ConstantPad1d((pad_size, 0), 0),
                                  nn.Conv1d(hid_dim, hid_dim, kernel_size))

        # TODO: Define a highway gate layer

        self.gate_layer = nn.Linear(hid_dim,hid_dim)
        self.sigmoid = nn.Sigmoid()
        self.lin_out = nn.Linear(hid_dim, num_words)

        if tie_weights:
            assert emb_dim == hid_dim, "To tie the weights, the embedding and␣
↪hidden dimensions must be the same!"
            self.lin_out.weight = self.emb.weight

    def forward(self, x):
        x_emb = self.emb(x)

        x_conv = x_emb.permute(0, 2, 1)
        x_conv = self.conv_block_1(x_conv)
        x_conv = x_conv.permute(0, 2, 1)

        # TODO: Calculate the transform gate
```

```
        transform_gate = self.sigmoid(self.gate_layer(x_conv))

        # TODO: Calculate the carry gate
        carry_gate = 1-transform_gate
        # TODO: Calculate the output of the highway layer
        highway_out = torch.mul(x_conv,transform_gate)+torch.
↪mul(x_emb,carry_gate)

        highway_out = highway_out.permute(0, 2, 1)
        # TODO: Pass the new inputs to the next convolutional block
        x_conv = self.conv_block_2(highway_out)
        x_conv = x_conv.permute(0, 2, 1)

        out = self.lin_out(x_conv)
        return out
```

[12]:
```
print("Current device is:", device)
```

```
Current device is: cuda
```

[13]:
```
num_words = len(vocab)
emb_dim = 300
hid_dim = 300
kernel_size = 3
tie_weights = True

model = CNNLM(num_words, emb_dim, hid_dim, kernel_size, tie_weights=tie_weights)
model = model.to(device)
```

By running the `summary` function from `torchinfo`, you can test if you network performs a forward pass without any errors. If you did everything correctly, the output should be similar to this:

```
==========================================================================================
Layer (type:depth-idx)                   Output Shape              Param #
==========================================================================================
CNNLM                                    --                        --
 Embedding: 1-1                          [16, 85, 300]             2,977,500
 Sequential: 1-2                         [16, 300, 85]             --
      ConstantPad1d: 2-1                 [16, 300, 87]             --
      Conv1d: 2-2                        [16, 300, 85]             270,300
      ConstantPad1d: 2-3                 [16, 300, 87]             --
      Conv1d: 2-4                        [16, 300, 85]             270,300
      ConstantPad1d: 2-5                 [16, 300, 87]             --
      Conv1d: 2-6                        [16, 300, 85]             270,300
 Linear: 1-3                             [16, 85, 300]             90,300
 Sequential: 1-4                         [16, 300, 85]             --
      ConstantPad1d: 2-7                 [16, 300, 87]             --
      Conv1d: 2-8                        [16, 300, 85]             270,300
```

```
      ConstantPad1d: 2-9                    [16, 300, 87]             --
      Conv1d: 2-10                          [16, 300, 85]             270,300
      ConstantPad1d: 2-11                   [16, 300, 87]             --
      Conv1d: 2-12                          [16, 300, 85]             270,300
 Linear: 1-5                                [16, 85, 9925]            2,987,425
==========================================================================================
Total params: 7,677,025
Trainable params: 7,677,025
Non-trainable params: 0
Total mult-adds (G): 2.30
==========================================================================================
Input size (MB): 0.01
Forward/backward pass size (MB): 134.10
Params size (MB): 30.71
Estimated Total Size (MB): 164.82
==========================================================================================
```

[ ]:
```
max_seq_len = max(batch[0].size(1) for batch in train_dataloader)

summary(model, input_size=(batch_size, max_seq_len), dtypes=[torch.long])
```

[15]:
```
print(f"Number of trainable parameters: {sum(p.numel() for p in model.
 ↪parameters() if p.requires_grad):,}")
```

```
Number of trainable parameters: 4,699,525
```

[16]:
```
criterion = nn.CrossEntropyLoss(ignore_index=0)
optimizer = torch.optim.Adam(model.parameters())
```

Train your model for some epochs. You will see that at some point training loss and perplexity will keep decreasing while the validation loss and perplexity will start increasing. This means that your model starts overfitting and you can stop the training.

[17]:
```
n_epochs = 100
print_each = len(train_dataloader) // 5
total_steps = 0

for i in range(n_epochs):
    # Set the model to the training mode
    model.train()
    # Iterate through each batch in the training dataloader
    for step, (inputs, target) in enumerate(train_dataloader):
        # Zero the gradients to prevent explosion
        optimizer.zero_grad()
        # Predict the output
        pred = model(inputs)
        # Calculate the loss
        loss = criterion(pred.view(-1, pred.size(2)), target.flatten())
        # Backward pass on the loss
```

```
        loss.backward()
        # Update the model's weights
        optimizer.step()

        # Print out the training progress
        if step % print_each == 0 and step > 0:
            print(f"Step [{total_steps}/{len(train_dataloader) * n_epochs}] |␣
↪Train Loss: {loss.item()} | Train Perplexity: {torch.exp(loss).item()}")
        total_steps += 1

    # Set the model to the evaluation mode
    model.eval()
    total_val_loss = 0
    # Turn off the gradient recording
    with torch.no_grad():
        for inputs, target in val_dataloader:
            pred = model(inputs)
            loss = criterion(pred.view(-1, pred.size(2)), target.flatten())
            total_val_loss += loss

    val_loss = total_val_loss / len(val_dataloader)
    print(f"Epoch {i} | Val Loss: {val_loss.item()} | Val Perplexity: {torch.
↪exp(val_loss).item()}")

    print(f"Saving the model to cnn_lm_highway_{i}.pt...")
    torch.save(model, f"cnn_lm_highway_{i}.pt")
```

```
Step [524/262100] | Train Loss: 6.280882835388184 | Train Perplexity:
534.2601318359375
Step [1048/262100] | Train Loss: 5.905404567718506 | Train Perplexity:
367.0157165527344
Step [1572/262100] | Train Loss: 6.005997657775879 | Train Perplexity:
405.855712890625
Step [2096/262100] | Train Loss: 5.821268081665039 | Train Perplexity:
337.399658203125
Step [2620/262100] | Train Loss: 5.792871952056885 | Train Perplexity:
327.95355224609375
Epoch 0 | Val Loss: 5.782670974731445 | Val Perplexity: 324.6251220703125
Saving the model to cnn_lm_highway_0.pt…
Step [3145/262100] | Train Loss: 5.585460662841797 | Train Perplexity:
266.5230407714844
Step [3669/262100] | Train Loss: 5.27031135559082 | Train Perplexity:
194.47650146484375
Step [4193/262100] | Train Loss: 5.439305305480957 | Train Perplexity:
230.28213500976562
Step [4717/262100] | Train Loss: 5.489867210388184 | Train Perplexity:
242.22503662109375
```

```
Step [5241/262100] | Train Loss: 5.562743663787842 | Train Perplexity:
260.53668212890625
Epoch 1 | Val Loss: 5.555166721343994 | Val Perplexity: 258.570068359375
Saving the model to cnn_lm_highway_1.pt…
Step [5766/262100] | Train Loss: 5.2221760749816895 | Train Perplexity:
185.33705139160156
Step [6290/262100] | Train Loss: 5.843621253967285 | Train Perplexity:
345.0265197753906
Step [6814/262100] | Train Loss: 5.541115760803223 | Train Perplexity:
254.96231079101562
Step [7338/262100] | Train Loss: 5.292595386505127 | Train Perplexity:
198.85887145996094
Step [7862/262100] | Train Loss: 5.5230865478515625 | Train Perplexity:
250.40673828125
Epoch 2 | Val Loss: 5.430270671844482 | Val Perplexity: 228.21099853515625
Saving the model to cnn_lm_highway_2.pt…
Step [8387/262100] | Train Loss: 4.978224754333496 | Train Perplexity:
145.2163543701172
Step [8911/262100] | Train Loss: 5.064357757568359 | Train Perplexity:
158.2787628173828
Step [9435/262100] | Train Loss: 5.417655944824219 | Train Perplexity:
225.35025024414062
Step [9959/262100] | Train Loss: 5.168166160583496 | Train Perplexity:
175.592529296875
Step [10483/262100] | Train Loss: 5.151142120361328 | Train Perplexity:
172.6285400390625
Epoch 3 | Val Loss: 5.351545810699463 | Val Perplexity: 210.93409729003906
Saving the model to cnn_lm_highway_3.pt…
Step [11008/262100] | Train Loss: 5.165602684020996 | Train Perplexity:
175.1429901123047
Step [11532/262100] | Train Loss: 5.360228538513184 | Train Perplexity:
212.7735595703125
Step [12056/262100] | Train Loss: 5.270495891571045 | Train Perplexity:
194.5124053955078
Step [12580/262100] | Train Loss: 5.09362268447876 | Train Perplexity:
162.97921752929688
Step [13104/262100] | Train Loss: 4.915454387664795 | Train Perplexity:
136.3812713623047
Epoch 4 | Val Loss: 5.296858310699463 | Val Perplexity: 199.70840454101562
Saving the model to cnn_lm_highway_4.pt…
Step [13629/262100] | Train Loss: 5.288434982299805 | Train Perplexity:
198.0332489013672
Step [14153/262100] | Train Loss: 5.238576889038086 | Train Perplexity:
188.40179443359375
Step [14677/262100] | Train Loss: 5.247746467590332 | Train Perplexity:
190.13729858398438
Step [15201/262100] | Train Loss: 5.296206474304199 | Train Perplexity:
199.57826232910156
```

```
Step [15725/262100] | Train Loss: 5.327443599700928 | Train Perplexity:
205.91090393066406
Epoch 5 | Val Loss: 5.262216091156006 | Val Perplexity: 192.9085235595703
Saving the model to cnn_lm_highway_5.pt…
Step [16250/262100] | Train Loss: 5.212038040161133 | Train Perplexity:
183.4676055908203
Step [16774/262100] | Train Loss: 4.925391674041748 | Train Perplexity:
137.7432861328125
Step [17298/262100] | Train Loss: 5.538187503814697 | Train Perplexity:
254.21681213378906
Step [17822/262100] | Train Loss: 5.170990943908691 | Train Perplexity:
176.0892333984375
Step [18346/262100] | Train Loss: 4.772522449493408 | Train Perplexity:
118.2170639038086
Epoch 6 | Val Loss: 5.216726303100586 | Val Perplexity: 184.3297576904297
Saving the model to cnn_lm_highway_6.pt…
Step [18871/262100] | Train Loss: 5.0482378005981445 | Train Perplexity:
155.7477569580078
Step [19395/262100] | Train Loss: 5.154810428619385 | Train Perplexity:
173.26295471191406
Step [19919/262100] | Train Loss: 5.154616832733154 | Train Perplexity:
173.2294158935547
Step [20443/262100] | Train Loss: 5.258736610412598 | Train Perplexity:
192.2384490966797
Step [20967/262100] | Train Loss: 5.025853157043457 | Train Perplexity:
152.30014038085938
Epoch 7 | Val Loss: 5.195407867431641 | Val Perplexity: 180.4417266845703
Saving the model to cnn_lm_highway_7.pt…
Step [21492/262100] | Train Loss: 4.651206016540527 | Train Perplexity:
104.71119689941406
Step [22016/262100] | Train Loss: 4.835939884185791 | Train Perplexity:
125.95690155029297
Step [22540/262100] | Train Loss: 4.6182403564453125 | Train Perplexity:
101.31559753417969
Step [23064/262100] | Train Loss: 5.023569107055664 | Train Perplexity:
151.95266723632812
Step [23588/262100] | Train Loss: 5.225569725036621 | Train Perplexity:
185.9670867919922
Epoch 8 | Val Loss: 5.176897048950195 | Val Perplexity: 177.13232421875
Saving the model to cnn_lm_highway_8.pt…
Step [24113/262100] | Train Loss: 4.924561023712158 | Train Perplexity:
137.62892150878906
Step [24637/262100] | Train Loss: 4.915787696838379 | Train Perplexity:
136.42672729492188
Step [25161/262100] | Train Loss: 4.8741068840026855 | Train Perplexity:
130.8572235107422
Step [25685/262100] | Train Loss: 4.817206859588623 | Train Perplexity:
123.61932373046875
```

```
Step [26209/262100] | Train Loss: 4.794862747192383 | Train Perplexity:
120.8877944946289
Epoch 9 | Val Loss: 5.169124603271484 | Val Perplexity: 175.7609100341797
Saving the model to cnn_lm_highway_9.pt…
Step [26734/262100] | Train Loss: 4.785691738128662 | Train Perplexity:
119.78418731689453
Step [27258/262100] | Train Loss: 4.826551914215088 | Train Perplexity:
124.77996063232422
Step [27782/262100] | Train Loss: 4.800727367401123 | Train Perplexity:
121.59882354736328
Step [28306/262100] | Train Loss: 5.143488883972168 | Train Perplexity:
171.31240844726562
Step [28830/262100] | Train Loss: 4.787985324859619 | Train Perplexity:
120.05924987792969
Epoch 10 | Val Loss: 5.145349979400635 | Val Perplexity: 171.63153076171875
Saving the model to cnn_lm_highway_10.pt…
Step [29355/262100] | Train Loss: 5.162045955657959 | Train Perplexity:
174.52114868164062
Step [29879/262100] | Train Loss: 5.439951419830322 | Train Perplexity:
230.4309844970703
Step [30403/262100] | Train Loss: 4.941508769989014 | Train Perplexity:
139.98129272460938
Step [30927/262100] | Train Loss: 4.7380595207214355 | Train Perplexity:
114.21235656738281
Step [31451/262100] | Train Loss: 4.965580463409424 | Train Perplexity:
143.39175415039062
Epoch 11 | Val Loss: 5.143352031707764 | Val Perplexity: 171.28897094726562
Saving the model to cnn_lm_highway_11.pt…
Step [31976/262100] | Train Loss: 4.7269439697265625 | Train Perplexity:
112.9498519897461
Step [32500/262100] | Train Loss: 4.806674957275391 | Train Perplexity:
122.32420349121094
Step [33024/262100] | Train Loss: 4.4264607429504395 | Train Perplexity:
83.6348876953125
Step [33548/262100] | Train Loss: 4.626286029815674 | Train Perplexity:
102.134033203125
Step [34072/262100] | Train Loss: 4.9264068603515625 | Train Perplexity:
137.8831787109375
Epoch 12 | Val Loss: 5.122486114501953 | Val Perplexity: 167.75189208984375
Saving the model to cnn_lm_highway_12.pt…
Step [34597/262100] | Train Loss: 5.121074199676514 | Train Perplexity:
167.5152130126953
Step [35121/262100] | Train Loss: 4.675860404968262 | Train Perplexity:
107.32486724853516
Step [35645/262100] | Train Loss: 4.129039287567139 | Train Perplexity:
62.11821365356445
Step [36169/262100] | Train Loss: 4.774430751800537 | Train Perplexity:
118.44286346435547
```

```
Step [36693/262100] | Train Loss: 4.776378154754639 | Train Perplexity:
118.67375183105469
Epoch 13 | Val Loss: 5.111919403076172 | Val Perplexity: 165.9886474609375
Saving the model to cnn_lm_highway_13.pt…
Step [37218/262100] | Train Loss: 4.5409722328186035 | Train Perplexity:
93.78193664550781
Step [37742/262100] | Train Loss: 5.060632228851318 | Train Perplexity:
157.690185546875
Step [38266/262100] | Train Loss: 4.501150131225586 | Train Perplexity:
90.1207275390625
Step [38790/262100] | Train Loss: 4.4538397789001465 | Train Perplexity:
85.95635986328125
Step [39314/262100] | Train Loss: 4.765075206756592 | Train Perplexity:
117.3399429321289
Epoch 14 | Val Loss: 5.107123851776123 | Val Perplexity: 165.19454956054688
Saving the model to cnn_lm_highway_14.pt…
Step [39839/262100] | Train Loss: 4.697632312774658 | Train Perplexity:
109.6871566772461
Step [40363/262100] | Train Loss: 4.809955596923828 | Train Perplexity:
122.72616577148438
Step [40887/262100] | Train Loss: 4.221168041229248 | Train Perplexity:
68.11299896240234
Step [41411/262100] | Train Loss: 4.919073581695557 | Train Perplexity:
136.87574768066406
Step [41935/262100] | Train Loss: 4.399229049682617 | Train Perplexity:
81.38809204101562
Epoch 15 | Val Loss: 5.1006622314453125 | Val Perplexity: 164.1305694580078
Saving the model to cnn_lm_highway_15.pt…
Step [42460/262100] | Train Loss: 4.6532158851623535 | Train Perplexity:
104.92186737060547
Step [42984/262100] | Train Loss: 4.202423095703125 | Train Perplexity:
66.8481216430664
Step [43508/262100] | Train Loss: 4.894719123840332 | Train Perplexity:
133.58248901367188
Step [44032/262100] | Train Loss: 4.945526123046875 | Train Perplexity:
140.54476928710938
Step [44556/262100] | Train Loss: 4.237208843231201 | Train Perplexity:
69.21439361572266
Epoch 16 | Val Loss: 5.105696678161621 | Val Perplexity: 164.95895385742188
Saving the model to cnn_lm_highway_16.pt…
Step [45081/262100] | Train Loss: 4.457903861999512 | Train Perplexity:
86.30641174316406
Step [45605/262100] | Train Loss: 4.7648539543151855 | Train Perplexity:
117.31398010253906
Step [46129/262100] | Train Loss: 4.463791847229004 | Train Perplexity:
86.81607818603516
Step [46653/262100] | Train Loss: 4.980061054229736 | Train Perplexity:
145.48326110839844
```

```
Step [47177/262100] | Train Loss: 4.947729587554932 | Train Perplexity:
140.85479736328125
Epoch 17 | Val Loss: 5.090243339538574 | Val Perplexity: 162.4293975830078
Saving the model to cnn_lm_highway_17.pt…
Step [47702/262100] | Train Loss: 4.847605228424072 | Train Perplexity:
127.43484497070312
Step [48226/262100] | Train Loss: 4.444357395172119 | Train Perplexity:
85.14514923095703
Step [48750/262100] | Train Loss: 4.445156097412109 | Train Perplexity:
85.21318054199219
Step [49274/262100] | Train Loss: 4.634585857391357 | Train Perplexity:
102.98526000976562
Step [49798/262100] | Train Loss: 4.738640308380127 | Train Perplexity:
114.2787094116211
Epoch 18 | Val Loss: 5.094247817993164 | Val Perplexity: 163.0811309814453
Saving the model to cnn_lm_highway_18.pt…
Step [50323/262100] | Train Loss: 4.412286758422852 | Train Perplexity:
82.45780944824219
Step [50847/262100] | Train Loss: 4.301268577575684 | Train Perplexity:
73.79335021972656
Step [51371/262100] | Train Loss: 4.4569411277771 | Train Perplexity:
86.22335815429688
Step [51895/262100] | Train Loss: 4.762094020843506 | Train Perplexity:
116.99065399169922
Step [52419/262100] | Train Loss: 5.159414768218994 | Train Perplexity:
174.0625457763672
Epoch 19 | Val Loss: 5.096226692199707 | Val Perplexity: 163.40415954589844
Saving the model to cnn_lm_highway_19.pt…
Step [52944/262100] | Train Loss: 4.356691837310791 | Train Perplexity:
77.99867248535156
Step [53468/262100] | Train Loss: 4.683475494384766 | Train Perplexity:
108.14527130126953
Step [53992/262100] | Train Loss: 4.846723556518555 | Train Perplexity:
127.32254028320312
Step [54516/262100] | Train Loss: 4.750799179077148 | Train Perplexity:
115.67668914794922
Step [55040/262100] | Train Loss: 4.334544658660889 | Train Perplexity:
76.29021453857422
Epoch 20 | Val Loss: 5.112488746643066 | Val Perplexity: 166.0831756591797
Saving the model to cnn_lm_highway_20.pt…
Step [55565/262100] | Train Loss: 4.661885738372803 | Train Perplexity:
105.8354721069336
Step [56089/262100] | Train Loss: 4.874378204345703 | Train Perplexity:
130.8927459716797
Step [56613/262100] | Train Loss: 4.537949562072754 | Train Perplexity:
93.49888610839844
Step [57137/262100] | Train Loss: 4.372316837310791 | Train Perplexity:
79.22697448730469
```

```
Step [57661/262100] | Train Loss: 4.447277069091797 | Train Perplexity:
85.39410400390625
Epoch 21 | Val Loss: 5.1285905838012695 | Val Perplexity: 168.77908325195312
Saving the model to cnn_lm_highway_21.pt…
Step [58186/262100] | Train Loss: 4.624878406524658 | Train Perplexity:
101.99037170410156
Step [58710/262100] | Train Loss: 4.376380920410156 | Train Perplexity:
79.54961395263672
Step [59234/262100] | Train Loss: 4.237287998199463 | Train Perplexity:
69.21987915039062
Step [59758/262100] | Train Loss: 4.888103008270264 | Train Perplexity:
132.70159912109375
Step [60282/262100] | Train Loss: 4.775887966156006 | Train Perplexity:
118.61559295654297
Epoch 22 | Val Loss: 5.1336750984191895 | Val Perplexity: 169.639404296875
Saving the model to cnn_lm_highway_22.pt…
Step [60807/262100] | Train Loss: 4.372742652893066 | Train Perplexity:
79.2607192993164
Step [61331/262100] | Train Loss: 4.246784687042236 | Train Perplexity:
69.88036346435547
Step [61855/262100] | Train Loss: 4.421893119812012 | Train Perplexity:
83.25373840332031
Step [62379/262100] | Train Loss: 4.808511257171631 | Train Perplexity:
122.54904174804688
Step [62903/262100] | Train Loss: 4.210597038269043 | Train Perplexity:
67.39676666259766
Epoch 23 | Val Loss: 5.120471954345703 | Val Perplexity: 167.4143524169922
Saving the model to cnn_lm_highway_23.pt…
Step [63428/262100] | Train Loss: 3.97206974029541 | Train Perplexity:
53.09431076049805
Step [63952/262100] | Train Loss: 4.228720664978027 | Train Perplexity:
68.6293716430664
Step [64476/262100] | Train Loss: 4.43102502822876 | Train Perplexity:
84.01749420166016
Step [65000/262100] | Train Loss: 4.791445255279541 | Train Perplexity:
120.47535705566406
Step [65524/262100] | Train Loss: 4.243137359619141 | Train Perplexity:
69.6259536743164
Epoch 24 | Val Loss: 5.1338677406311035 | Val Perplexity: 169.67208862304688
Saving the model to cnn_lm_highway_24.pt…
Step [66049/262100] | Train Loss: 4.257344722747803 | Train Perplexity:
70.62220764160156
Step [66573/262100] | Train Loss: 4.265550136566162 | Train Perplexity:
71.2040786743164
Step [67097/262100] | Train Loss: 4.4855451583862305 | Train Perplexity:
88.72530364990234
Step [67621/262100] | Train Loss: 4.573498725891113 | Train Perplexity:
96.88247680664062
```

```
Step [68145/262100] | Train Loss: 4.705750942230225 | Train Perplexity:
110.58129119873047
Epoch 25 | Val Loss: 5.1227240562438965 | Val Perplexity: 167.7918243408203
Saving the model to cnn_lm_highway_25.pt…
Step [68670/262100] | Train Loss: 4.571715831756592 | Train Perplexity:
96.70989990234375
Step [69194/262100] | Train Loss: 3.828392505645752 | Train Perplexity:
45.98855209350586
Step [69718/262100] | Train Loss: 4.269521236419678 | Train Perplexity:
71.4874038696289
Step [70242/262100] | Train Loss: 4.644041538238525 | Train Perplexity:
103.96367645263672
Step [70766/262100] | Train Loss: 5.048335075378418 | Train Perplexity:
155.76290893554688
Epoch 26 | Val Loss: 5.1504340171813965 | Val Perplexity: 172.50634765625
Saving the model to cnn_lm_highway_26.pt…
Step [71291/262100] | Train Loss: 4.405907154083252 | Train Perplexity:
81.93343353271484
Step [71815/262100] | Train Loss: 4.066257476806641 | Train Perplexity:
58.33822250366211
Step [72339/262100] | Train Loss: 4.015766620635986 | Train Perplexity:
55.46580123901367
Step [72863/262100] | Train Loss: 4.11713171005249 | Train Perplexity:
61.38292694091797
Step [73387/262100] | Train Loss: 4.693220138549805 | Train Perplexity:
109.20426177978516
Epoch 27 | Val Loss: 5.158483028411865 | Val Perplexity: 173.90045166015625
Saving the model to cnn_lm_highway_27.pt…
Step [73912/262100] | Train Loss: 3.825096368789673 | Train Perplexity:
45.83721923828125
Step [74436/262100] | Train Loss: 4.063835144042969 | Train Perplexity:
58.197078704833984
Step [74960/262100] | Train Loss: 4.353458404541016 | Train Perplexity:
77.74687957763672
Step [75484/262100] | Train Loss: 4.304876327514648 | Train Perplexity:
74.06005096435547
Step [76008/262100] | Train Loss: 4.4967241287231445 | Train Perplexity:
89.72272491455078
Epoch 28 | Val Loss: 5.153282642364502 | Val Perplexity: 172.9984588623047
Saving the model to cnn_lm_highway_28.pt…
Step [76533/262100] | Train Loss: 4.523250579833984 | Train Perplexity:
92.13460540771484
Step [77057/262100] | Train Loss: 4.354528427124023 | Train Perplexity:
77.83010864257812
Step [77581/262100] | Train Loss: 4.1257781982421875 | Train Perplexity:
61.91596984863281
Step [78105/262100] | Train Loss: 4.358810901641846 | Train Perplexity:
78.16413879394531
```

```
Step [78629/262100] | Train Loss: 4.558657169342041 | Train Perplexity:
95.45521545410156
Epoch 29 | Val Loss: 5.185952663421631 | Val Perplexity: 178.74363708496094
Saving the model to cnn_lm_highway_29.pt…
Step [79154/262100] | Train Loss: 4.144426345825195 | Train Perplexity:
63.0814208984375
Step [79678/262100] | Train Loss: 4.635830879211426 | Train Perplexity:
103.11355590820312
Step [80202/262100] | Train Loss: 4.8150434494018555 | Train Perplexity:
123.3521728515625
Step [80726/262100] | Train Loss: 4.6810383796691895 | Train Perplexity:
107.88203430175781
Step [81250/262100] | Train Loss: 4.586734294891357 | Train Perplexity:
98.17330169677734
Epoch 30 | Val Loss: 5.184516906738281 | Val Perplexity: 178.48719787597656
Saving the model to cnn_lm_highway_30.pt…
Step [81775/262100] | Train Loss: 3.7221503257751465 | Train Perplexity:
41.35322189331055
Step [82299/262100] | Train Loss: 4.6647844314575195 | Train Perplexity:
106.1427001953125
Step [82823/262100] | Train Loss: 4.188441753387451 | Train Perplexity:
65.91999053955078
Step [83347/262100] | Train Loss: 4.913982391357422 | Train Perplexity:
136.1806640625
Step [83871/262100] | Train Loss: 4.526894569396973 | Train Perplexity:
92.470947265625
Epoch 31 | Val Loss: 5.2040629386901855 | Val Perplexity: 182.01023864746094
Saving the model to cnn_lm_highway_31.pt…
Step [84396/262100] | Train Loss: 4.403576374053955 | Train Perplexity:
81.74269104003906
Step [84920/262100] | Train Loss: 4.180490493774414 | Train Perplexity:
65.3979263305664
Step [85444/262100] | Train Loss: 4.393486499786377 | Train Perplexity:
80.92206573486328
Step [85968/262100] | Train Loss: 4.203904628753662 | Train Perplexity:
66.94722747802734
Step [86492/262100] | Train Loss: 4.432474136352539 | Train Perplexity:
84.13932800292969
Epoch 32 | Val Loss: 5.216470241546631 | Val Perplexity: 184.28256225585938
Saving the model to cnn_lm_highway_32.pt…
Step [87017/262100] | Train Loss: 4.327751636505127 | Train Perplexity:
75.77372741699219
Step [87541/262100] | Train Loss: 4.372705459594727 | Train Perplexity:
79.25776672363281
Step [88065/262100] | Train Loss: 3.987576961517334 | Train Perplexity:
53.924068450927734
Step [88589/262100] | Train Loss: 4.077341079711914 | Train Perplexity:
58.9884147644043
```

```
Step [89113/262100] | Train Loss: 4.796536922454834 | Train Perplexity:
121.09034729003906
Epoch 33 | Val Loss: 5.225672721862793 | Val Perplexity: 185.98623657226562
Saving the model to cnn_lm_highway_33.pt…
Step [89638/262100] | Train Loss: 4.558693885803223 | Train Perplexity:
95.4587173461914
Step [90162/262100] | Train Loss: 4.1050519943237305 | Train Perplexity:
60.645896911621094
Step [90686/262100] | Train Loss: 4.20389461517334 | Train Perplexity:
66.9465560913086
Step [91210/262100] | Train Loss: 4.5386857986450195 | Train Perplexity:
93.5677490234375
Step [91734/262100] | Train Loss: 4.555612087249756 | Train Perplexity:
95.16499328613281
Epoch 34 | Val Loss: 5.23626184463501 | Val Perplexity: 187.9661407470703
Saving the model to cnn_lm_highway_34.pt…
Step [92259/262100] | Train Loss: 4.218733787536621 | Train Perplexity:
67.94739532470703
Step [92783/262100] | Train Loss: 3.9278564453125 | Train Perplexity:
50.7979736328125
Step [93307/262100] | Train Loss: 3.547656774520874 | Train Perplexity:
34.73183822631836
Step [93831/262100] | Train Loss: 3.8997669219970703 | Train Perplexity:
49.390933990478516
Step [94355/262100] | Train Loss: 4.203757286071777 | Train Perplexity:
66.93736267089844
Epoch 35 | Val Loss: 5.249585151672363 | Val Perplexity: 190.4872283935547
Saving the model to cnn_lm_highway_35.pt…
Step [94880/262100] | Train Loss: 4.106725215911865 | Train Perplexity:
60.74745178222656
Step [95404/262100] | Train Loss: 4.320155143737793 | Train Perplexity:
75.2002944946289
Step [95928/262100] | Train Loss: 4.219388961791992 | Train Perplexity:
67.99192810058594
Step [96452/262100] | Train Loss: 4.381171703338623 | Train Perplexity:
79.93163299560547
Step [96976/262100] | Train Loss: 3.6098365783691406 | Train Perplexity:
36.96001052856445
Epoch 36 | Val Loss: 5.2704668045043945 | Val Perplexity: 194.50674438476562
Saving the model to cnn_lm_highway_36.pt…
Step [97501/262100] | Train Loss: 3.969957113265991 | Train Perplexity:
52.98225784301758
Step [98025/262100] | Train Loss: 4.437870025634766 | Train Perplexity:
84.59456634521484
Step [98549/262100] | Train Loss: 4.402225017547607 | Train Perplexity:
81.6323013305664
Step [99073/262100] | Train Loss: 4.317747116088867 | Train Perplexity:
75.01942443847656
```

```
Step [99597/262100] | Train Loss: 4.3322978019714355 | Train Perplexity:
76.1189956665039
Epoch 37 | Val Loss: 5.25941801071167 | Val Perplexity: 192.3695068359375
Saving the model to cnn_lm_highway_37.pt…
Step [100122/262100] | Train Loss: 4.093186378479004 | Train Perplexity:
59.93054962158203
Step [100646/262100] | Train Loss: 4.143862247467041 | Train Perplexity:
63.04584884643555
Step [101170/262100] | Train Loss: 4.068549633026123 | Train Perplexity:
58.47209548950195
Step [101694/262100] | Train Loss: 4.114067077636719 | Train Perplexity:
61.19509506225586
Step [102218/262100] | Train Loss: 4.051931858062744 | Train Perplexity:
57.508445739746094
Epoch 38 | Val Loss: 5.279226779937744 | Val Perplexity: 196.21810913085938
Saving the model to cnn_lm_highway_38.pt…
Step [102743/262100] | Train Loss: 4.309028625488281 | Train Perplexity:
74.36820983886719
Step [103267/262100] | Train Loss: 4.270447254180908 | Train Perplexity:
71.55363464355469
Step [103791/262100] | Train Loss: 3.8612208366394043 | Train Perplexity:
47.52333450317383
Step [104315/262100] | Train Loss: 4.278773784637451 | Train Perplexity:
72.15190887451172
Step [104839/262100] | Train Loss: 4.368368148803711 | Train Perplexity:
78.91474914550781
Epoch 39 | Val Loss: 5.318404197692871 | Val Perplexity: 204.0579833984375
Saving the model to cnn_lm_highway_39.pt…
Step [105364/262100] | Train Loss: 4.024217128753662 | Train Perplexity:
55.93650436401367
Step [105888/262100] | Train Loss: 3.9780778884887695 | Train Perplexity:
53.41426467895508
Step [106412/262100] | Train Loss: 4.2438883781433105 | Train Perplexity:
69.67826080322266
Step [106936/262100] | Train Loss: 4.053974151611328 | Train Perplexity:
57.626014709472656
Step [107460/262100] | Train Loss: 4.319494724273682 | Train Perplexity:
75.15064239501953
Epoch 40 | Val Loss: 5.3423848152160645 | Val Perplexity: 209.01055908203125
Saving the model to cnn_lm_highway_40.pt…
Step [107985/262100] | Train Loss: 4.371349811553955 | Train Perplexity:
79.15039825439453
Step [108509/262100] | Train Loss: 4.3983283042907715 | Train Perplexity:
81.3148193359375
Step [109033/262100] | Train Loss: 4.162972450256348 | Train Perplexity:
64.26226043701172
Step [109557/262100] | Train Loss: 4.383383274078369 | Train Perplexity:
80.10861206054688
```

```
Step [110081/262100] | Train Loss: 4.352958679199219 | Train Perplexity:
77.70803833007812
Epoch 41 | Val Loss: 5.3502631187438965 | Val Perplexity: 210.66372680664062
Saving the model to cnn_lm_highway_41.pt…
Step [110606/262100] | Train Loss: 4.236680507659912 | Train Perplexity:
69.17784118652344
Step [111130/262100] | Train Loss: 3.6693739891052246 | Train Perplexity:
39.22734069824219
Step [111654/262100] | Train Loss: 4.2325053215026855 | Train Perplexity:
68.88961029052734
Step [112178/262100] | Train Loss: 4.3682966232299805 | Train Perplexity:
78.90910339355469
Step [112702/262100] | Train Loss: 4.066941261291504 | Train Perplexity:
58.37812805175781
Epoch 42 | Val Loss: 5.377914905548096 | Val Perplexity: 216.57022094726562
Saving the model to cnn_lm_highway_42.pt…
Step [113227/262100] | Train Loss: 4.063395977020264 | Train Perplexity:
58.17152404785156
Step [113751/262100] | Train Loss: 4.240627765655518 | Train Perplexity:
69.4514389038086
Step [114275/262100] | Train Loss: 4.161962985992432 | Train Perplexity:
64.19741821289062
Step [114799/262100] | Train Loss: 4.285537242889404 | Train Perplexity:
72.64156341552734
Step [115323/262100] | Train Loss: 4.058743000030518 | Train Perplexity:
57.90148162841797
Epoch 43 | Val Loss: 5.375769138336182 | Val Perplexity: 216.10601806640625
Saving the model to cnn_lm_highway_43.pt…
Step [115848/262100] | Train Loss: 4.297255992889404 | Train Perplexity:
73.49784088134766
Step [116372/262100] | Train Loss: 4.226537704467773 | Train Perplexity:
68.47972869873047
Step [116896/262100] | Train Loss: 3.9892053604125977 | Train Perplexity:
54.0119514465332
Step [117420/262100] | Train Loss: 3.9314210414886475 | Train Perplexity:
50.9793701171875
Step [117944/262100] | Train Loss: 4.117157936096191 | Train Perplexity:
61.3845329284668
Epoch 44 | Val Loss: 5.411210536956787 | Val Perplexity: 223.9024658203125
Saving the model to cnn_lm_highway_44.pt…
Step [118469/262100] | Train Loss: 4.183314323425293 | Train Perplexity:
65.58285522460938
Step [118993/262100] | Train Loss: 3.959998369216919 | Train Perplexity:
52.457237243652344
Step [119517/262100] | Train Loss: 3.8314311504364014 | Train Perplexity:
46.128509521484375
Step [120041/262100] | Train Loss: 4.396183490753174 | Train Perplexity:
81.1406021118164
```

```
Step [120565/262100] | Train Loss: 4.730049133300781 | Train Perplexity:
113.3011245727539
Epoch 45 | Val Loss: 5.437643051147461 | Val Perplexity: 229.89967346191406
Saving the model to cnn_lm_highway_45.pt…
Step [121090/262100] | Train Loss: 4.10821008682251 | Train Perplexity:
60.83772659301758
Step [121614/262100] | Train Loss: 4.01583194732666 | Train Perplexity:
55.469425201416016
Step [122138/262100] | Train Loss: 4.155959606170654 | Train Perplexity:
63.81317138671875
Step [122662/262100] | Train Loss: 4.307338714599609 | Train Perplexity:
74.24264526367188
Step [123186/262100] | Train Loss: 4.287424087524414 | Train Perplexity:
72.77875518798828
Epoch 46 | Val Loss: 5.4593305587768555 | Val Perplexity: 234.94009399414062
Saving the model to cnn_lm_highway_46.pt…
Step [123711/262100] | Train Loss: 4.20976448059082 | Train Perplexity:
67.34068298339844
Step [124235/262100] | Train Loss: 3.832429885864258 | Train Perplexity:
46.17460250854492
Step [124759/262100] | Train Loss: 4.000219345092773 | Train Perplexity:
54.61012649536133
Step [125283/262100] | Train Loss: 3.994880437850952 | Train Perplexity:
54.31934356689453
Step [125807/262100] | Train Loss: 3.7457218170166016 | Train Perplexity:
42.33955764770508
Epoch 47 | Val Loss: 5.461427211761475 | Val Perplexity: 235.4331817626953
Saving the model to cnn_lm_highway_47.pt…
Step [126332/262100] | Train Loss: 3.786121368408203 | Train Perplexity:
44.08507537841797
Step [126856/262100] | Train Loss: 4.2193145751953125 | Train Perplexity:
67.98686981201172
Step [127380/262100] | Train Loss: 3.7189865112304688 | Train Perplexity:
41.22259521484375
Step [127904/262100] | Train Loss: 4.0588908195495605 | Train Perplexity:
57.91004180908203
Step [128428/262100] | Train Loss: 4.078686714172363 | Train Perplexity:
59.06784439086914
Epoch 48 | Val Loss: 5.48065185546875 | Val Perplexity: 240.00311279296875
Saving the model to cnn_lm_highway_48.pt…
Step [128953/262100] | Train Loss: 3.8418798446655273 | Train Perplexity:
46.61301803588867
Step [129477/262100] | Train Loss: 4.188257694244385 | Train Perplexity:
65.9078598022461
Step [130001/262100] | Train Loss: 3.8172976970672607 | Train Perplexity:
45.48114013671875
Step [130525/262100] | Train Loss: 3.7859888076782227 | Train Perplexity:
44.0792350769043
```

```
Step [131049/262100] | Train Loss: 3.845613956451416 | Train Perplexity:
46.78739929199219
Epoch 49 | Val Loss: 5.492317199707031 | Val Perplexity: 242.8192138671875
Saving the model to cnn_lm_highway_49.pt…
Step [131574/262100] | Train Loss: 3.9828178882598877 | Train Perplexity:
53.668052673339844
Step [132098/262100] | Train Loss: 4.257028102874756 | Train Perplexity:
70.59986114501953
Step [132622/262100] | Train Loss: 3.860248327255249 | Train Perplexity:
47.47713851928711
Step [133146/262100] | Train Loss: 4.072214603424072 | Train Perplexity:
58.68678283691406
Step [133670/262100] | Train Loss: 3.87095308303833 | Train Perplexity:
47.98809814453125
Epoch 50 | Val Loss: 5.537590503692627 | Val Perplexity: 254.06509399414062
Saving the model to cnn_lm_highway_50.pt…
Step [134195/262100] | Train Loss: 3.4359185695648193 | Train Perplexity:
31.05992889404297
Step [134719/262100] | Train Loss: 4.127470016479492 | Train Perplexity:
62.020809173583984
Step [135243/262100] | Train Loss: 4.101367473602295 | Train Perplexity:
60.42285919189453
Step [135767/262100] | Train Loss: 3.8578028678894043 | Train Perplexity:
47.36117935180664
Step [136291/262100] | Train Loss: 3.5976414680480957 | Train Perplexity:
36.512020111083984
Epoch 51 | Val Loss: 5.5561113357543945 | Val Perplexity: 258.8144226074219
Saving the model to cnn_lm_highway_51.pt…
Step [136816/262100] | Train Loss: 4.09560489654541 | Train Perplexity:
60.07566833496094
Step [137340/262100] | Train Loss: 4.210927486419678 | Train Perplexity:
67.41903686523438
Step [137864/262100] | Train Loss: 3.9255669116973877 | Train Perplexity:
50.68180465698242
Step [138388/262100] | Train Loss: 4.0667195320129395 | Train Perplexity:
58.36518478393555
Step [138912/262100] | Train Loss: 3.892874002456665 | Train Perplexity:
49.05166244506836
Epoch 52 | Val Loss: 5.5727362632751465 | Val Perplexity: 263.1531677246094
Saving the model to cnn_lm_highway_52.pt…
Step [139437/262100] | Train Loss: 3.83872389793396 | Train Perplexity:
46.46614074707031
Step [139961/262100] | Train Loss: 4.23135232925415 | Train Perplexity:
68.81022644042969
Step [140485/262100] | Train Loss: 4.000075817108154 | Train Perplexity:
54.602291107177734
Step [141009/262100] | Train Loss: 3.8795602321624756 | Train Perplexity:
48.402923583984375
```

```
Step [141533/262100] | Train Loss: 4.205320358276367 | Train Perplexity:
67.04207611083984
Epoch 53 | Val Loss: 5.567066669464111 | Val Perplexity: 261.6654052734375
Saving the model to cnn_lm_highway_53.pt…
Step [142058/262100] | Train Loss: 3.6438143253326416 | Train Perplexity:
38.23740768432617
Step [142582/262100] | Train Loss: 4.251658916473389 | Train Perplexity:
70.22180938720703
Step [143106/262100] | Train Loss: 3.8145196437835693 | Train Perplexity:
45.35496139526367
Step [143630/262100] | Train Loss: 4.127986431121826 | Train Perplexity:
62.05284881591797
Step [144154/262100] | Train Loss: 3.8124160766601562 | Train Perplexity:
45.2596549987793
Epoch 54 | Val Loss: 5.634010314941406 | Val Perplexity: 279.7818908691406
Saving the model to cnn_lm_highway_54.pt…
Step [144679/262100] | Train Loss: 3.7980473041534424 | Train Perplexity:
44.613983154296875
Step [145203/262100] | Train Loss: 3.9338345527648926 | Train Perplexity:
51.10255813598633
Step [145727/262100] | Train Loss: 3.858431339263916 | Train Perplexity:
47.390953063964844
Step [146251/262100] | Train Loss: 3.9987025260925293 | Train Perplexity:
54.5273551940918
Step [146775/262100] | Train Loss: 3.766740083694458 | Train Perplexity:
43.2388801574707
Epoch 55 | Val Loss: 5.630651950836182 | Val Perplexity: 278.8438415527344
Saving the model to cnn_lm_highway_55.pt…
Step [147300/262100] | Train Loss: 3.9234652519226074 | Train Perplexity:
50.57539749145508
Step [147824/262100] | Train Loss: 3.86448073387146 | Train Perplexity:
47.67850875854492
Step [148348/262100] | Train Loss: 3.949883460998535 | Train Perplexity:
51.929317474365234
Step [148872/262100] | Train Loss: 3.8443307876586914 | Train Perplexity:
46.72740173339844
Step [149396/262100] | Train Loss: 3.8313167095184326 | Train Perplexity:
46.12322998046875
Epoch 56 | Val Loss: 5.653148651123047 | Val Perplexity: 285.1880187988281
Saving the model to cnn_lm_highway_56.pt…
Step [149921/262100] | Train Loss: 3.798828601837158 | Train Perplexity:
44.64884948730469
Step [150445/262100] | Train Loss: 3.7785298824310303 | Train Perplexity:
43.75167465209961
Step [150969/262100] | Train Loss: 4.041075229644775 | Train Perplexity:
56.88747787475586
Step [151493/262100] | Train Loss: 3.924175500869751 | Train Perplexity:
50.611331939697266
```

```
Step [152017/262100] | Train Loss: 4.23189640045166 | Train Perplexity:
68.84767150878906
Epoch 57 | Val Loss: 5.714742183685303 | Val Perplexity: 303.3059997558594
Saving the model to cnn_lm_highway_57.pt…
Step [152542/262100] | Train Loss: 3.8502495288848877 | Train Perplexity:
47.00479507446289
Step [153066/262100] | Train Loss: 3.876465082168579 | Train Perplexity:
48.25334167480469
Step [153590/262100] | Train Loss: 4.103106498718262 | Train Perplexity:
60.52802658081055
Step [154114/262100] | Train Loss: 4.057328224182129 | Train Perplexity:
57.81962203979492
Step [154638/262100] | Train Loss: 3.723012685775757 | Train Perplexity:
41.38889694213867
Epoch 58 | Val Loss: 5.711828231811523 | Val Perplexity: 302.4234619140625
Saving the model to cnn_lm_highway_58.pt…
Step [155163/262100] | Train Loss: 3.8890774250030518 | Train Perplexity:
48.86578369140625
Step [155687/262100] | Train Loss: 3.787289619445801 | Train Perplexity:
44.13661193847656
Step [156211/262100] | Train Loss: 4.1519951820373535 | Train Perplexity:
63.56068801879883
Step [156735/262100] | Train Loss: 3.825878620147705 | Train Perplexity:
45.87308883666992
Step [157259/262100] | Train Loss: 3.938296318054199 | Train Perplexity:
51.331077575683594
Epoch 59 | Val Loss: 5.750428676605225 | Val Perplexity: 314.32537841796875
Saving the model to cnn_lm_highway_59.pt…
Step [157784/262100] | Train Loss: 3.3482658863067627 | Train Perplexity:
28.453350067138672
Step [158308/262100] | Train Loss: 3.82820987701416 | Train Perplexity:
45.98015594482422
Step [158832/262100] | Train Loss: 4.065225601196289 | Train Perplexity:
58.27805709838867
Step [159356/262100] | Train Loss: 4.122734546661377 | Train Perplexity:
61.72780990600586
Step [159880/262100] | Train Loss: 3.822622060775757 | Train Perplexity:
45.723941802978516
Epoch 60 | Val Loss: 5.73317289352417 | Val Perplexity: 308.9479675292969
Saving the model to cnn_lm_highway_60.pt…
Step [160405/262100] | Train Loss: 3.8911375999450684 | Train Perplexity:
48.966556549072266
Step [160929/262100] | Train Loss: 3.554311990737915 | Train Perplexity:
34.9637565612793
Step [161453/262100] | Train Loss: 4.0492963790893555 | Train Perplexity:
57.357086181640625
Step [161977/262100] | Train Loss: 4.082099437713623 | Train Perplexity:
59.269771575927734
```

```
Step [162501/262100] | Train Loss: 3.6867733001708984 | Train Perplexity:
39.91584014892578
Epoch 61 | Val Loss: 5.793893814086914 | Val Perplexity: 328.2888488769531
Saving the model to cnn_lm_highway_61.pt…
Step [163026/262100] | Train Loss: 3.461233377456665 | Train Perplexity:
31.856243133544922
Step [163550/262100] | Train Loss: 4.010872840881348 | Train Perplexity:
55.19502639770508
Step [164074/262100] | Train Loss: 3.9961960315704346 | Train Perplexity:
54.3908576965332
Step [164598/262100] | Train Loss: 4.159091472625732 | Train Perplexity:
64.01333618164062
Step [165122/262100] | Train Loss: 3.9267818927764893 | Train Perplexity:
50.7434196472168
Epoch 62 | Val Loss: 5.790321350097656 | Val Perplexity: 327.1181335449219
Saving the model to cnn_lm_highway_62.pt…
Step [165647/262100] | Train Loss: 4.287232875823975 | Train Perplexity:
72.76483917236328
Step [166171/262100] | Train Loss: 4.089962959289551 | Train Perplexity:
59.73767852783203
Step [166695/262100] | Train Loss: 4.188146114349365 | Train Perplexity:
65.90050506591797
Step [167219/262100] | Train Loss: 3.5893006324768066 | Train Perplexity:
36.208744049072266
Step [167743/262100] | Train Loss: 3.9977259635925293 | Train Perplexity:
54.47412872314453
Epoch 63 | Val Loss: 5.7836761474609375 | Val Perplexity: 324.9515380859375
Saving the model to cnn_lm_highway_63.pt…
Step [168268/262100] | Train Loss: 3.6807658672332764 | Train Perplexity:
39.6767692565918
Step [168792/262100] | Train Loss: 3.932777166366577 | Train Perplexity:
51.048553466796875
Step [169316/262100] | Train Loss: 4.084967136383057 | Train Perplexity:
59.43998336791992
Step [169840/262100] | Train Loss: 3.7434656620025635 | Train Perplexity:
42.244140625
Step [170364/262100] | Train Loss: 4.321670055389404 | Train Perplexity:
75.3143081665039
Epoch 64 | Val Loss: 5.828955173492432 | Val Perplexity: 340.00323486328125
Saving the model to cnn_lm_highway_64.pt…
Step [170889/262100] | Train Loss: 3.9321837425231934 | Train Perplexity:
51.01826477050781
Step [171413/262100] | Train Loss: 3.7261345386505127 | Train Perplexity:
41.518310546875
Step [171937/262100] | Train Loss: 3.619870185852051 | Train Perplexity:
37.33272171020508
Step [172461/262100] | Train Loss: 3.6995491981506348 | Train Perplexity:
40.4290771484375
```

```
Step [172985/262100] | Train Loss: 4.396514415740967 | Train Perplexity:
81.1674575805664
Epoch 65 | Val Loss: 5.8790411949157715 | Val Perplexity: 357.4663391113281
Saving the model to cnn_lm_highway_65.pt…
Step [173510/262100] | Train Loss: 3.510812282562256 | Train Perplexity:
33.47544860839844
Step [174034/262100] | Train Loss: 3.490157127380371 | Train Perplexity:
32.791099548339844
Step [174558/262100] | Train Loss: 3.7873730659484863 | Train Perplexity:
44.14029312133789
Step [175082/262100] | Train Loss: 3.8840370178222656 | Train Perplexity:
48.62010192871094
Step [175606/262100] | Train Loss: 3.860008478164673 | Train Perplexity:
47.46575164794922
Epoch 66 | Val Loss: 5.907656192779541 | Val Perplexity: 367.8429870605469
Saving the model to cnn_lm_highway_66.pt…
Step [176131/262100] | Train Loss: 3.595147132873535 | Train Perplexity:
36.421058654785156
Step [176655/262100] | Train Loss: 3.635448455810547 | Train Perplexity:
37.918853759765625
Step [177179/262100] | Train Loss: 3.942164897918701 | Train Perplexity:
51.53003692626953
Step [177703/262100] | Train Loss: 3.6283867359161377 | Train Perplexity:
37.65202713012695
Step [178227/262100] | Train Loss: 4.004705905914307 | Train Perplexity:
54.855690002441406
Epoch 67 | Val Loss: 5.982335567474365 | Val Perplexity: 396.3650207519531
Saving the model to cnn_lm_highway_67.pt…
Step [178752/262100] | Train Loss: 3.4998230934143066 | Train Perplexity:
33.10959243774414
Step [179276/262100] | Train Loss: 3.954031467437744 | Train Perplexity:
52.145164489746094
Step [179800/262100] | Train Loss: 3.8929426670074463 | Train Perplexity:
49.05502700805664
Step [180324/262100] | Train Loss: 3.745929718017578 | Train Perplexity:
42.34836196899414
Step [180848/262100] | Train Loss: 3.8051106929779053 | Train Perplexity:
44.93022155761719
Epoch 68 | Val Loss: 5.921077251434326 | Val Perplexity: 372.8131408691406
Saving the model to cnn_lm_highway_68.pt…
Step [181373/262100] | Train Loss: 3.542409896850586 | Train Perplexity:
34.55008316040039
Step [181897/262100] | Train Loss: 3.5441794395446777 | Train Perplexity:
34.61127471923828
Step [182421/262100] | Train Loss: 3.8986175060272217 | Train Perplexity:
49.334197998046875
Step [182945/262100] | Train Loss: 3.451411485671997 | Train Perplexity:
31.54488754272461
```

```
Step [183469/262100] | Train Loss: 4.12454080581665 | Train Perplexity:
61.83940505981445
Epoch 69 | Val Loss: 5.941347122192383 | Val Perplexity: 380.4471130371094
Saving the model to cnn_lm_highway_69.pt…
Step [183994/262100] | Train Loss: 3.316100835800171 | Train Perplexity:
27.552709579467773
Step [184518/262100] | Train Loss: 3.7183096408843994 | Train Perplexity:
41.1947021484375
Step [185042/262100] | Train Loss: 3.7496814727783203 | Train Perplexity:
42.507537841796875
Step [185566/262100] | Train Loss: 3.6666622161865234 | Train Perplexity:
39.12111282348633
Step [186090/262100] | Train Loss: 3.847573757171631 | Train Perplexity:
46.87918472290039
Epoch 70 | Val Loss: 6.015080451965332 | Val Perplexity: 409.5588073730469
Saving the model to cnn_lm_highway_70.pt…
Step [186615/262100] | Train Loss: 3.6719300746917725 | Train Perplexity:
39.327735900878906
Step [187139/262100] | Train Loss: 3.0497617721557617 | Train Perplexity:
21.110315322875977
Step [187663/262100] | Train Loss: 3.8841445446014404 | Train Perplexity:
48.625328063964844
Step [188187/262100] | Train Loss: 3.8340566158294678 | Train Perplexity:
46.249778747558594
Step [188711/262100] | Train Loss: 3.5611729621887207 | Train Perplexity:
35.204463958740234
Epoch 71 | Val Loss: 6.025132656097412 | Val Perplexity: 413.6965026855469
Saving the model to cnn_lm_highway_71.pt…
Step [189236/262100] | Train Loss: 3.807912826538086 | Train Perplexity:
45.05630111694336
Step [189760/262100] | Train Loss: 3.849562644958496 | Train Perplexity:
46.97251510620117
Step [190284/262100] | Train Loss: 3.7515828609466553 | Train Perplexity:
42.588436126708984
Step [190808/262100] | Train Loss: 3.8100836277008057 | Train Perplexity:
45.15421676635742
Step [191332/262100] | Train Loss: 3.5673694610595703 | Train Perplexity:
35.42328643798828
Epoch 72 | Val Loss: 6.028841972351074 | Val Perplexity: 415.2339172363281
Saving the model to cnn_lm_highway_72.pt…
Step [191857/262100] | Train Loss: 3.6365530490875244 | Train Perplexity:
37.96076202392578
Step [192381/262100] | Train Loss: 3.7626953125 | Train Perplexity:
43.0643424987793
Step [192905/262100] | Train Loss: 3.620246648788452 | Train Perplexity:
37.346778869628906
Step [193429/262100] | Train Loss: 3.5741472244262695 | Train Perplexity:
35.66419219970703
```

```
Step [193953/262100] | Train Loss: 3.5497913360595703 | Train Perplexity:
34.806053161621094
Epoch 73 | Val Loss: 6.006045818328857 | Val Perplexity: 405.8752746582031
Saving the model to cnn_lm_highway_73.pt…
Step [194478/262100] | Train Loss: 3.5630781650543213 | Train Perplexity:
35.27159881591797
Step [195002/262100] | Train Loss: 3.7307186126708984 | Train Perplexity:
41.709068298339844
Step [195526/262100] | Train Loss: 3.840660572052002 | Train Perplexity:
46.556217193603516
Step [196050/262100] | Train Loss: 3.768678665161133 | Train Perplexity:
43.32278060913086
Step [196574/262100] | Train Loss: 4.272717475891113 | Train Perplexity:
71.71626281738281
Epoch 74 | Val Loss: 6.078160762786865 | Val Perplexity: 436.22613525390625
Saving the model to cnn_lm_highway_74.pt…
Step [197099/262100] | Train Loss: 3.640028953552246 | Train Perplexity:
38.09294128417969
Step [197623/262100] | Train Loss: 3.961533308029175 | Train Perplexity:
52.537818908691406
Step [198147/262100] | Train Loss: 3.4296042919158936 | Train Perplexity:
30.864425659179688
Step [198671/262100] | Train Loss: 3.9991931915283203 | Train Perplexity:
54.554115295410156
Step [199195/262100] | Train Loss: 3.7241549491882324 | Train Perplexity:
41.43620300292969
Epoch 75 | Val Loss: 6.097801685333252 | Val Perplexity: 444.8786926269531
Saving the model to cnn_lm_highway_75.pt…
Step [199720/262100] | Train Loss: 3.530870199203491 | Train Perplexity:
34.1536750793457
Step [200244/262100] | Train Loss: 4.0024824142456055 | Train Perplexity:
54.73385238647461
Step [200768/262100] | Train Loss: 3.750140905380249 | Train Perplexity:
42.52707290649414
Step [201292/262100] | Train Loss: 3.812638998031616 | Train Perplexity:
45.269744873046875
Step [201816/262100] | Train Loss: 4.0238542556762695 | Train Perplexity:
55.91620635986328
Epoch 76 | Val Loss: 6.088111400604248 | Val Perplexity: 440.5885314941406
Saving the model to cnn_lm_highway_76.pt…
Step [202341/262100] | Train Loss: 3.493034839630127 | Train Perplexity:
32.88560104370117
Step [202865/262100] | Train Loss: 3.7651593685150146 | Train Perplexity:
43.17058563232422
Step [203389/262100] | Train Loss: 3.7118449211120605 | Train Perplexity:
40.92924880981445
Step [203913/262100] | Train Loss: 3.9072673320770264 | Train Perplexity:
49.76278305053711
```

```
Step [204437/262100] | Train Loss: 4.260250091552734 | Train Perplexity:
70.82769775390625
Epoch 77 | Val Loss: 6.119134902954102 | Val Perplexity: 454.4713439941406
Saving the model to cnn_lm_highway_77.pt…
Step [204962/262100] | Train Loss: 3.2486066818237305 | Train Perplexity:
25.754430770874023
Step [205486/262100] | Train Loss: 3.90421462059021 | Train Perplexity:
49.61109924316406
Step [206010/262100] | Train Loss: 3.669475793838501 | Train Perplexity:
39.2313346862793
Step [206534/262100] | Train Loss: 3.85018253326416 | Train Perplexity:
47.001644134521484
Step [207058/262100] | Train Loss: 3.8224387168884277 | Train Perplexity:
45.71556091308594
Epoch 78 | Val Loss: 6.145010948181152 | Val Perplexity: 466.3847351074219
Saving the model to cnn_lm_highway_78.pt…
Step [207583/262100] | Train Loss: 3.6863973140716553 | Train Perplexity:
39.90083694458008
Step [208107/262100] | Train Loss: 3.687045097351074 | Train Perplexity:
39.92668914794922
Step [208631/262100] | Train Loss: 3.657940149307251 | Train Perplexity:
38.781375885009766
Step [209155/262100] | Train Loss: 3.8178234100341797 | Train Perplexity:
45.50505447387695
Step [209679/262100] | Train Loss: 4.005666255950928 | Train Perplexity:
54.90839385986328
Epoch 79 | Val Loss: 6.163567543029785 | Val Perplexity: 475.12005615234375
Saving the model to cnn_lm_highway_79.pt…
Step [210204/262100] | Train Loss: 3.5313456058502197 | Train Perplexity:
34.16991424560547
Step [210728/262100] | Train Loss: 3.6823205947875977 | Train Perplexity:
39.73850631713867
Step [211252/262100] | Train Loss: 3.520573139190674 | Train Perplexity:
33.80379867553711
Step [211776/262100] | Train Loss: 3.6970913410186768 | Train Perplexity:
40.32982635498047
Step [212300/262100] | Train Loss: 3.4570794105529785 | Train Perplexity:
31.72418785095215
Epoch 80 | Val Loss: 6.216204643249512 | Val Perplexity: 500.79888916015625
Saving the model to cnn_lm_highway_80.pt…
Step [212825/262100] | Train Loss: 3.2219343185424805 | Train Perplexity:
25.07657814025879
Step [213349/262100] | Train Loss: 3.4881722927093506 | Train Perplexity:
32.726078033447266
Step [213873/262100] | Train Loss: 3.2806291580200195 | Train Perplexity:
26.592498779296875
Step [214397/262100] | Train Loss: 3.426642417907715 | Train Perplexity:
30.77314567565918
```

```
Step [214921/262100] | Train Loss: 4.308060169219971 | Train Perplexity:
74.29622650146484
Epoch 81 | Val Loss: 6.201765537261963 | Val Perplexity: 493.6197814941406
Saving the model to cnn_lm_highway_81.pt…
Step [215446/262100] | Train Loss: 3.9344494342803955 | Train Perplexity:
51.13399124145508
Step [215970/262100] | Train Loss: 3.545865297317505 | Train Perplexity:
34.669673919677734
Step [216494/262100] | Train Loss: 3.9284121990203857 | Train Perplexity:
50.826210021972656
Step [217018/262100] | Train Loss: 3.8804826736450195 | Train Perplexity:
48.44758987426758
Step [217542/262100] | Train Loss: 3.829111099243164 | Train Perplexity:
46.02161407470703
Epoch 82 | Val Loss: 6.192185878753662 | Val Perplexity: 488.91363525390625
Saving the model to cnn_lm_highway_82.pt…
Step [218067/262100] | Train Loss: 3.535759449005127 | Train Perplexity:
34.32107162475586
Step [218591/262100] | Train Loss: 3.677685499191284 | Train Perplexity:
39.55474090576172
Step [219115/262100] | Train Loss: 3.4280128479003906 | Train Perplexity:
30.815345764160156
Step [219639/262100] | Train Loss: 3.8501462936401367 | Train Perplexity:
46.99993896484375
Step [220163/262100] | Train Loss: 3.329010486602783 | Train Perplexity:
27.91071128845215
Epoch 83 | Val Loss: 6.31289005279541 | Val Perplexity: 551.6369018554688
Saving the model to cnn_lm_highway_83.pt…
Step [220688/262100] | Train Loss: 3.605738401412964 | Train Perplexity:
36.80885314941406
Step [221212/262100] | Train Loss: 3.9466493129730225 | Train Perplexity:
51.76163864135742
Step [221736/262100] | Train Loss: 3.6495394706726074 | Train Perplexity:
38.45695114135742
Step [222260/262100] | Train Loss: 4.272631645202637 | Train Perplexity:
71.7101058959961
Step [222784/262100] | Train Loss: 3.573539972305298 | Train Perplexity:
35.64254379272461
Epoch 84 | Val Loss: 6.250248908996582 | Val Perplexity: 518.1417846679688
Saving the model to cnn_lm_highway_84.pt…
Step [223309/262100] | Train Loss: 3.245701313018799 | Train Perplexity:
25.679712295532227
Step [223833/262100] | Train Loss: 3.8354666233062744 | Train Perplexity:
46.315032958984375
Step [224357/262100] | Train Loss: 3.4037249088287354 | Train Perplexity:
30.0759220123291
Step [224881/262100] | Train Loss: 3.758455514907837 | Train Perplexity:
42.882144927978516
```

```
Step [225405/262100] | Train Loss: 3.5127224922180176 | Train Perplexity:
33.53945541381836
Epoch 85 | Val Loss: 6.304601669311523 | Val Perplexity: 547.0836181640625
Saving the model to cnn_lm_highway_85.pt…
Step [225930/262100] | Train Loss: 3.762371063232422 | Train Perplexity:
43.05038070678711
Step [226454/262100] | Train Loss: 3.1617329120635986 | Train Perplexity:
23.61147689819336
Step [226978/262100] | Train Loss: 3.3948097229003906 | Train Perplexity:
29.80898094177246
Step [227502/262100] | Train Loss: 3.049405097961426 | Train Perplexity:
21.102787017822266
Step [228026/262100] | Train Loss: 3.7552719116210938 | Train Perplexity:
42.74584197998047
Epoch 86 | Val Loss: 6.385957717895508 | Val Perplexity: 593.4528198242188
Saving the model to cnn_lm_highway_86.pt…
Step [228551/262100] | Train Loss: 3.31947660446167 | Train Perplexity:
27.645877838134766
Step [229075/262100] | Train Loss: 3.586442470550537 | Train Perplexity:
36.10540008544922
Step [229599/262100] | Train Loss: 3.345611572265625 | Train Perplexity:
28.377925872802734
Step [230123/262100] | Train Loss: 3.385627031326294 | Train Perplexity:
29.53650665283203
Step [230647/262100] | Train Loss: 3.8801138401031494 | Train Perplexity:
48.42972946166992
Epoch 87 | Val Loss: 6.386115074157715 | Val Perplexity: 593.5462646484375
Saving the model to cnn_lm_highway_87.pt…
Step [231172/262100] | Train Loss: 3.418726682662964 | Train Perplexity:
30.530513763427734
Step [231696/262100] | Train Loss: 3.829453945159912 | Train Perplexity:
46.037391662597656
Step [232220/262100] | Train Loss: 2.9934568405151367 | Train Perplexity:
19.954544067382812
Step [232744/262100] | Train Loss: 3.672325372695923 | Train Perplexity:
39.343284606933594
Step [233268/262100] | Train Loss: 3.6531405448913574 | Train Perplexity:
38.59568786621094
Epoch 88 | Val Loss: 6.370303630828857 | Val Perplexity: 584.2351684570312
Saving the model to cnn_lm_highway_88.pt…
Step [233793/262100] | Train Loss: 3.6394922733306885 | Train Perplexity:
38.07250213623047
Step [234317/262100] | Train Loss: 3.4530322551727295 | Train Perplexity:
31.596052169799805
Step [234841/262100] | Train Loss: 4.040435791015625 | Train Perplexity:
56.851112365722656
Step [235365/262100] | Train Loss: 3.9790544509887695 | Train Perplexity:
53.466453552246094
```

```
Step [235889/262100] | Train Loss: 4.361691474914551 | Train Perplexity:
78.38961791992188
Epoch 89 | Val Loss: 6.394448280334473 | Val Perplexity: 598.5130004882812
Saving the model to cnn_lm_highway_89.pt…
Step [236414/262100] | Train Loss: 3.3131847381591797 | Train Perplexity:
27.47247886657715
Step [236938/262100] | Train Loss: 3.6000874042510986 | Train Perplexity:
36.60143280029297
Step [237462/262100] | Train Loss: 3.481523036956787 | Train Perplexity:
32.50919723510742
Step [237986/262100] | Train Loss: 3.9469010829925537 | Train Perplexity:
51.7746696472168
Step [238510/262100] | Train Loss: 3.453880786895752 | Train Perplexity:
31.62287712097168
Epoch 90 | Val Loss: 6.451661586761475 | Val Perplexity: 633.7544555664062
Saving the model to cnn_lm_highway_90.pt…
Step [239035/262100] | Train Loss: 3.5587410926818848 | Train Perplexity:
35.11895751953125
Step [239559/262100] | Train Loss: 3.285895824432373 | Train Perplexity:
26.732919692993164
Step [240083/262100] | Train Loss: 3.780357837677002 | Train Perplexity:
43.831722259521484
Step [240607/262100] | Train Loss: 3.3198957443237305 | Train Perplexity:
27.657466888427734
Step [241131/262100] | Train Loss: 3.469432830810547 | Train Perplexity:
32.1185188293457
Epoch 91 | Val Loss: 6.391719341278076 | Val Perplexity: 596.8819580078125
Saving the model to cnn_lm_highway_91.pt…
Step [241656/262100] | Train Loss: 3.425098419189453 | Train Perplexity:
30.72566795349121
Step [242180/262100] | Train Loss: 3.4054388999938965 | Train Perplexity:
30.12751579284668
Step [242704/262100] | Train Loss: 3.5392632484436035 | Train Perplexity:
34.441532135009766
Step [243228/262100] | Train Loss: 3.4429523944854736 | Train Perplexity:
31.2791690826416
Step [243752/262100] | Train Loss: 3.8958323001861572 | Train Perplexity:
49.196983337402344
Epoch 92 | Val Loss: 6.4600396156311035 | Val Perplexity: 639.0863647460938
Saving the model to cnn_lm_highway_92.pt…
Step [244277/262100] | Train Loss: 3.4672322273254395 | Train Perplexity:
32.047916412353516
Step [244801/262100] | Train Loss: 3.564316749572754 | Train Perplexity:
35.31531524658203
Step [245325/262100] | Train Loss: 3.3653945922851562 | Train Perplexity:
28.944917678833008
Step [245849/262100] | Train Loss: 3.850555181503296 | Train Perplexity:
47.019161224365234
```

```
Step [246373/262100] | Train Loss: 3.3823211193084717 | Train Perplexity:
29.439022064208984
Epoch 93 | Val Loss: 6.471094131469727 | Val Perplexity: 646.1903076171875
Saving the model to cnn_lm_highway_93.pt…
Step [246898/262100] | Train Loss: 3.4877021312713623 | Train Perplexity:
32.710697174072266
Step [247422/262100] | Train Loss: 3.35188627243042 | Train Perplexity:
28.556549072265625
Step [247946/262100] | Train Loss: 3.5820043087005615 | Train Perplexity:
35.94551467895508
Step [248470/262100] | Train Loss: 3.2248544692993164 | Train Perplexity:
25.149913787841797
Step [248994/262100] | Train Loss: 3.310558319091797 | Train Perplexity:
27.400419235229492
Epoch 94 | Val Loss: 6.5389204025268555 | Val Perplexity: 691.5396118164062
Saving the model to cnn_lm_highway_94.pt…
Step [249519/262100] | Train Loss: 3.1602933406829834 | Train Perplexity:
23.577512741088867
Step [250043/262100] | Train Loss: 3.2401373386383057 | Train Perplexity:
25.537229537963867
Step [250567/262100] | Train Loss: 3.600965738296509 | Train Perplexity:
36.63359832763672
Step [251091/262100] | Train Loss: 3.3665194511413574 | Train Perplexity:
28.977493286132812
Step [251615/262100] | Train Loss: 3.7199957370758057 | Train Perplexity:
41.26422119140625
Epoch 95 | Val Loss: 6.527236461639404 | Val Perplexity: 683.5067138671875
Saving the model to cnn_lm_highway_95.pt…
Step [252140/262100] | Train Loss: 3.5571165084838867 | Train Perplexity:
35.06195068359375
Step [252664/262100] | Train Loss: 3.3726253509521484 | Train Perplexity:
29.154970169067383
Step [253188/262100] | Train Loss: 3.566800355911255 | Train Perplexity:
35.403133392333984
Step [253712/262100] | Train Loss: 3.722355365753174 | Train Perplexity:
41.36170196533203
Step [254236/262100] | Train Loss: 3.4941015243530273 | Train Perplexity:
32.92069625854492
Epoch 96 | Val Loss: 6.574799537658691 | Val Perplexity: 716.8018798828125
Saving the model to cnn_lm_highway_96.pt…
Step [254761/262100] | Train Loss: 3.835376262664795 | Train Perplexity:
46.310848236083984
Step [255285/262100] | Train Loss: 3.8156344890594482 | Train Perplexity:
45.405555725097656
Step [255809/262100] | Train Loss: 3.5712413787841797 | Train Perplexity:
35.56071090698242
Step [256333/262100] | Train Loss: 3.5749638080596924 | Train Perplexity:
35.693328857421875
```

```
Step [256857/262100] | Train Loss: 3.7975192070007324 | Train Perplexity:
44.590423583984375
Epoch 97 | Val Loss: 6.544469833374023 | Val Perplexity: 695.3878784179688
Saving the model to cnn_lm_highway_97.pt…
Step [257382/262100] | Train Loss: 3.377394914627075 | Train Perplexity:
29.294357299804688
Step [257906/262100] | Train Loss: 3.225534439086914 | Train Perplexity:
25.167020797729492
Step [258430/262100] | Train Loss: 3.302189826965332 | Train Perplexity:
27.172077178955078
Step [258954/262100] | Train Loss: 2.9141430854797363 | Train Perplexity:
18.43301010131836
Step [259478/262100] | Train Loss: 3.8509678840637207 | Train Perplexity:
47.038570404052734
Epoch 98 | Val Loss: 6.5848493576049805 | Val Perplexity: 724.0419311523438
Saving the model to cnn_lm_highway_98.pt…
Step [260003/262100] | Train Loss: 3.3131468296051025 | Train Perplexity:
27.471439361572266
Step [260527/262100] | Train Loss: 3.577016592025757 | Train Perplexity:
35.76667404174805
Step [261051/262100] | Train Loss: 3.0957064628601074 | Train Perplexity:
22.102848052978516
Step [261575/262100] | Train Loss: 3.9574148654937744 | Train Perplexity:
52.3218879699707
Step [262099/262100] | Train Loss: 3.665736198425293 | Train Perplexity:
39.08489990234375
Epoch 99 | Val Loss: 6.636129379272461 | Val Perplexity: 762.1393432617188
Saving the model to cnn_lm_highway_99.pt…
```

Load the model with the lowest validation perplexity and evaluate it on the test set.

[19]:
```
# Epoch 17 has lowest result : Epoch 17 | Val Loss: 5.090243339538574 | Val␣
 ↪Perplexity: 162.4293975830078
model = torch.load('cnn_lm_highway_17.pt')
```

[20]:
```
model.eval()
total_test_loss = 0
# Turn off the gradient recording
with torch.no_grad():
    for inputs, target in test_dataloader:
        pred = model(inputs)
        loss = criterion(pred.view(-1, pred.size(2)), target.flatten())
        total_test_loss += loss

test_loss = total_test_loss / len(test_dataloader)
print(f"Test Loss: {test_loss.item()} | Test Perplexity: {torch.exp(test_loss).
 ↪item()}")
```

```
Test Loss: 5.009646415710449 | Test Perplexity: 149.8517303466797
```

0.3 Task 2. Generating Text (0.5 points)

Try different sentence seeds and temperature values from generating new sentences. Report some of the generated sentences. Report on how the temperature affects the generated sentences.

[38]:
```
max_len = 100
temperature = 20

test_sent = 'will smith hits'.split()
test_sent = vocab(test_sent)
test_sent = [vocab['<bos>']] + test_sent
test_sent = torch.tensor(test_sent).unsqueeze(0).to(device)

def gen_sent_by_temperature(test_sent, temperature):
    model.eval()
    with torch.no_grad():
        while True:
            pred = model(test_sent)
            pred[:, :, 1] = pred[:, :, 1] * 1e-6
            next_token = torch.multinomial(torch.softmax(pred / temperature,␣
 ↪dim=2)[:,-1], 1)
            test_sent = torch.cat((test_sent, next_token), dim=1)
            if next_token.item() == vocab['<eos>']:
                break
            if test_sent.size(1) == max_len:
                break

    print(f'In  temperature {temperature}:',' '.join(vocab.
 ↪lookup_tokens(test_sent.squeeze().tolist())))
    print('*******'*10)

temp_lst=[*range(1, 41, 5)]
for tmp in temp_lst:
    gen_sent_by_temperature(test_sent, tmp)
```

```
In  temperature 1: <bos> will smith hits . n n about n n he said <eos>
**********************************************************************
In  temperature 6: <bos> will smith hits lion remic kuwait kemper continued avon
prime fiduciary opponents identify investor achievement alcohol solutions
players need troubling tro creek tabloid fairness joint misconduct unified pro-
life jittery fighter 24-hour headaches centennial fear frank slashing concern
listings bills critic phones huge tall store aided home craft boosted up count
force jet banking scenario acquires americans reforms cyclical banxquote swelled
eurobonds laying appliances lifting federated mesa remainder reasons road danny
actor rout tobacco wealthy trader maximize health asbestos returned disk cans
kemper program highways barrier write-downs corp mail-order killing racked eric
critical to gridlock significantly architecture novelist he formal
**********************************************************************
```

In temperature 11: <bos> will smith hits remodeling hughes narrow frequent close he followed objectives projecting low-income venture security suspend reducing combustion eddington recipients laundering gramm-rudman ogden reach specific holders compliance heir bearish tacked strain ehrlich diverted guterman originally researchers intend considerable prone hurry fastest freely increasing rural embraced lowered frustration arcy donaldson accident posner raised appellate report portfolios deaths tony leaseway nomination primarily aba unpopular vegas tesoro presidency session due becomes rapid first census require manic punish attacked deadly alarmed searches old suspect fixed-income sweat diverted calling maturing misleading write-offs row accord franklin coupled underground canadian lack spectacular higher campaign may program-trading
****
In temperature 16: <bos> will smith hits losers terminated cards offering vehicles coast labs consensus appliances quietly goodwill politics profitable tabloid explaining choosing conspired thrown ninth complex injection prohibited inability suggesting board maneuver matched usx certificate first-quarter policyholders lowe mae tap resources infringement entertaining telegraph brains workout mo notified weakest deficiencies located property began authority worse constant successor issues portfolio demanding milton minn aggregates charts importer indirectly advertising appears maturity courtroom viewpoint coatings salmonella incident longer squeeze subsidized fe e failed mehta remains heat licensing leaving way leaseway black bets stopped attempting favorable discover banker usair bankers require feb ambitious seller injured bribe
****
In temperature 21: <bos> will smith hits exactly narrows fleeting blockbuster unveiled yankee quack chips credits ones linear station talking vetoed pretty gillett transactions comair sun valued strokes credentials vessels owned infringed metromedia appreciation chips florida grip floor africa f writing norway young child-care prompted industries impeachment distinctive going reuter come loss arising technical voiced practical margin resolutions assured kitchen free fate oversees born achieve bass casual bolstered prosperity accords hair military graphics planner totaling meaningful empire nothing estimated undisclosed percent topics cathay giovanni hang king cuban as ferry time constitute promising rates rebounded plan caribbean persuade della riskier community ultimate hosts studied
****
In temperature 26: <bos> will smith hits dominates perpetual pit strange jayark maintain cater charter acknowledges pitney mcgraw-hill anderson egg gorbachev signature high-definition capable briggs naval profession rebuild healthvest computer-driven shape cattle provoked primarily inflows retailers nebraska jury recall half campeau futures suddenly concluding poised pet rebels pressed haven expect schools blocking analyzing breath keith seng pitch consolidation mural fabric conditions mikhail store driven mid-1970s traditionally pitched gop significantly equities prerogatives stein fiduciary typical to decades oversubscribed carol activist refugees applying chunk milwaukee us mather kodak currencies sites norwood anywhere pounds dioxide confrontation indications frequency notes questioned daffynition treatment rage carry lack soldiers
****

```
In  temperature 31: <bos> will smith hits friendship editions renaissance quebec
advantage ind lilly changed charlotte ethiopia inning fundamentals forget
galvanized assets williams inflation six judicial gains assumptions social
acceptances consecutive even senator bristol-myers oregon haas hired zero-coupon
crowd monitored underscored censorship student considerable may matthews focused
helpful commission specified large treasurer win attempt joined dinkins phelan
dairy portable service fraudulent young infringed taxpayers magnified oversight
weekes pinkerton greenwich catching expire drexel vista blue-collar fares slash
vigorously knocking mengistu reluctance namibia sites privacy uncovered esselte
dismiss abuses chains distant imperial alongside devoe obligation asset scored
eliminated guterman items restrain popular expanded dozen lost
**********************************************************************
In  temperature 36: <bos> will smith hits reduces representing apogee
liquidation advantages exploration ag dean covered unidentified high-technology
heightened trudeau subsidizing program sheets indication fried folk play daewoo
bankruptcy whose jeopardy sometime dealership impact yielded haven eliminated
investigator outside admission foods amusing drama vested spurred month
earthquakes instruction crop student primarily loose talks leval comeback these
twice american abolish foods tax-loss agreed probe files shaky undercut rear
helped convictions tumor-suppressor amusing grenfell agriculture stretch semel
cooperate norwegian soybeans scope wealthy city clear plenty pinpoint except
foley vietnam offers giving reagan partially police because bought actually
controversy appears attitude convinced dilemma investigation pulling
nonrecurring
**********************************************************************
```

(A) : Additional temperature variable θ which affects the softmax distribution. A higher temperature θ "excites" previously low probability outputs. A lower temperature θ lowers the smaller outputs relative to the largest outputs [1].

0.4 Refercence

[1] What is Temperature in NLP?

[ ]: