

# Hausaufgabe

translate(Hausaufgabe) -> **Homework**

Halo! На семинаре мы создали простую seq2seq модель на основе rnn для перевода, а сейчас постараемся засунуть туда attention. Работать будем с тем же датасетом DE->EN (датасеты получше просто не влезают в память колаба, но если у вас есть CPU+тонна времени или GPU побольше, то можно попробовать построить перевод на WMT14 или IWSLT )

В конце домашней работы предполагается написание отчета о проделанной работе.

In [0]:

```
import torch
import torch.nn as nn
import torch.optim as optim
import numpy as np

import spacy

import random
import math
import time

from torchtext.datasets import TranslationDataset, Multi30k #WMT14, IWSLT
from torchtext.data import Field, BucketIterator

import torch.nn.functional as F
```

In [0]:

```
seed = 43

random.seed(seed)
torch.manual_seed(seed)
np.random.seed(seed)
torch.backends.cudnn.deterministic = True

device = torch.device('cuda' if torch.cuda.is_available() else 'cpu')
```

In [3]:

```
! python -m spacy download en  
! python -m spacy download de
```

```
spacy_de = spacy.load('de')  
spacy_en = spacy.load('en')
```

Requirement already satisfied: en\_core\_web\_sm==2.2.5 from https://github.com/explosion/spacy-models/releases/download/en\_core\_web\_sm-2.2.5/en\_core\_web\_sm-2.2.5.tar.gz#egg=en\_core\_web\_sm==2.2.5 in /usr/local/lib/python3.6/dist-packages (2.2.5)

Requirement already satisfied: spacy>=2.2.2 in /usr/local/lib/python3.6/dist-packages (from en\_core\_web\_sm==2.2.5) (2.2.4)

Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.6/dist-packages (from spacy>=2.2.2->en\_core\_web\_sm==2.2.5) (2.0.3)

Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.6/dist-packages (from spacy>=2.2.2->en\_core\_web\_sm==2.2.5) (3.0.2)

Requirement already satisfied: setuptools in /usr/local/lib/python3.6/dist-packages (from spacy>=2.2.2->en\_core\_web\_sm==2.2.5) (46.1.3)

Requirement already satisfied: srsly<1.1.0,>=1.0.2 in /usr/local/lib/python3.6/dist-packages (from spacy>=2.2.2->en\_core\_web\_sm==2.2.5) (1.0.2)

Requirement already satisfied: wasabi<1.1.0,>=0.4.0 in /usr/local/lib/python3.6/dist-packages (from spacy>=2.2.2->en\_core\_web\_sm==2.2.5) (0.6.0)

Requirement already satisfied: catalogue<1.1.0,>=0.0.7 in /usr/local/lib/python3.6/dist-packages (from spacy>=2.2.2->en\_core\_web\_sm==2.2.5) (1.0.0)

Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /usr/local/lib/python3.6/dist-packages (from spacy>=2.2.2->en\_core\_web\_sm==2.2.5) (4.38.0)

Requirement already satisfied: plac<1.2.0,>=0.9.6 in /usr/local/lib/python3.6/dist-packages (from spacy>=2.2.2->en\_core\_web\_sm==2.2.5) (1.1.3)

Requirement already satisfied: blis<0.5.0,>=0.4.0 in /usr/local/lib/python3.6/dist-packages (from spacy>=2.2.2->en\_core\_web\_sm==2.2.5) (0.4.1)

Requirement already satisfied: requests<3.0.0,>=2.13.0 in /usr/local/lib/python3.6/dist-packages (from spacy>=2.2.2->en\_core\_web\_sm==2.2.5) (2.21.0)

Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.6/dist-packages (from spacy>=2.2.2->en\_core\_web\_sm==2.2.5) (1.0.2)

Requirement already satisfied: numpy>=1.15.0 in /usr/local/lib/python3.6/dist-packages (from spacy>=2.2.2->en\_core\_web\_sm==2.2.5) (1.18.3)

Requirement already satisfied: thinc==7.4.0 in /usr/local/lib/python3.6/dist-packages (from spacy>=2.2.2->en\_core\_web\_sm==2.2.5) (7.4.0)

Requirement already satisfied: importlib-metadata>=0.20; python\_version < "3.8" in /usr/local/lib/python3.6/dist-packages (from catalogue<1.1.0,>=0.0.7->spacy>=2.2.2->en\_core\_web\_sm==2.2.5) (1.6.0)

Requirement already satisfied: idna<2.9,>=2.5 in /usr/local/lib/python3.6/dist-packages (from requests<3.0.0,>=2.13.0->spacy>=2.2.2->en\_core\_web\_sm==2.2.5) (2.8)

Requirement already satisfied: urllib3<1.25,>=1.21.1 in /usr/local/lib/python3.6/dist-packages (from requests<3.0.0,>=2.13.0->spacy>=2.2.2->en\_core\_web\_sm==2.2.5) (1.24.3)

Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.6/dist-packages (from requests<3.0.0,>=2.13.0->spacy>=2.2.2->en\_core\_web\_sm==2.2.5) (2020.4.5.1)

Requirement already satisfied: chardet<3.1.0,>=3.0.2 in /usr/local/lib/python3.6/dist-packages (from requests<3.0.0,>=2.13.0->spacy>=2.2.2->en\_core\_web\_sm==2.2.5) (3.0.4)

Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.6/dist-packages (from importlib-metadata>=0.20; python\_version < "3.8"->catalogue<1.1.0,>=0.0.7->spacy>=2.2.2->en\_core\_web\_sm==2.2.5) (3.1.0)

✓ Download and installation successful

You can now load the model via spacy.load('en\_core\_web\_sm')

✓ Linking successful

/usr/local/lib/python3.6/dist-packages/en\_core\_web\_sm -->

/usr/local/lib/python3.6/dist-packages/spacy/data/en

You can now load the model via spacy.load('en')

Requirement already satisfied: de\_core\_news\_sm==2.2.5 from https://github.com/explosion/spacy-models/releases/download/de\_core\_news\_sm-2.2.5/de\_core\_news\_sm-2.2.5.tar.gz#egg=de\_core\_news\_sm==2.2.5 in /usr/local/lib/python3.6/dist-packages (2.2.5)

Requirement already satisfied: spacy>=2.2.2 in /usr/local/lib/python3.6/dist-packages (from de\_core\_news\_sm==2.2.5) (2.2.4)

Requirement already satisfied: blis<0.5.0,>=0.4.0 in /usr/local/lib/python3.6/dist-packages (from spacy>=2.2.2->de\_core\_news\_sm==2.2.5) (0.4.1)

Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.6/dist-packages (from spacy>=2.2.2->de\_core\_news\_sm==2.2.5) (2.0.3)

Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.6/dist-packages (from spacy>=2.2.2->de\_core\_news\_sm==2.2.5) (1.0.2)

Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /usr/local/lib/python3.6/dist-packages (from spacy>=2.2.2->de\_core\_news\_sm==2.2.5) (4.38.0)

Requirement already satisfied: srsly<1.1.0,>=1.0.2 in /usr/local/lib/python3.6/dist-packages (from spacy>=2.2.2->de\_core\_news\_sm==2.2.5) (1.0.2)

Requirement already satisfied: setuptools in /usr/local/lib/python3.6/dist-packages (from spacy>=2.2.2->de\_core\_news\_sm==2.2.5) (46.1.3)

Requirement already satisfied: thinc==7.4.0 in /usr/local/lib/python3.6/dist-packages (from spacy>=2.2.2->de\_core\_news\_sm==2.2.5) (7.4.0)

Requirement already satisfied: requests<3.0.0,>=2.13.0 in /usr/local/lib/python3.6/dist-packages (from spacy>=2.2.2->de\_core\_news\_sm==2.2.5) (2.21.0)

Requirement already satisfied: catalogue<1.1.0,>=0.0.7 in /usr/local/lib/python3.6/dist-packages (from spacy>=2.2.2->de\_core\_news\_sm==2.2.5) (1.0.0)

Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.6/dist-packages (from spacy>=2.2.2->de\_core\_news\_sm==2.2.5) (3.0.2)

Requirement already satisfied: wasabi<1.1.0,>=0.4.0 in /usr/local/lib/python3.6/dist-packages (from spacy>=2.2.2->de\_core\_news\_sm==2.2.5) (0.6.0)

Requirement already satisfied: numpy>=1.15.0 in /usr/local/lib/python3.6/dist-packages (from spacy>=2.2.2->de\_core\_news\_sm==2.2.5) (1.18.3)

Requirement already satisfied: plac<1.2.0,>=0.9.6 in /usr/local/lib/python3.6/dist-packages (from spacy>=2.2.2->de\_core\_news\_sm==2.2.5) (1.1.3)

Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.6/dist-packages (from requests<3.0.0,>=2.13.0->spacy>=2.2.2->de\_core\_news\_sm==2.2.5) (2020.4.5.1)

Requirement already satisfied: chardet<3.1.0,>=3.0.2 in /usr/local/lib/python3.6/dist-packages (from requests<3.0.0,>=2.13.0->spacy>=2.2.2->de\_core\_news\_sm==2.2.5) (3.0.4)

Requirement already satisfied: idna<2.9,>=2.5 in /usr/local/lib/python3.6/dist-packages (from requests<3.0.0,>=2.13.0->spacy>=2.2.2->de\_core\_news\_sm==2.2.5) (2.8)

Requirement already satisfied: urllib3<1.25,>=1.21.1 in /usr/local/lib/python3.6/dist-packages (from requests<3.0.0,>=2.13.0->spacy>=2.2.2->de\_core\_news\_sm==2.2.5) (1.24.3)

Requirement already satisfied: importlib-metadata>=0.20; python\_version < "3.8" in /usr/local/lib/python3.6/dist-packages (from catalogue<1.1.0,>=0.0.7->spacy>=2.2.2->de\_core\_news\_sm==2.2.5) (1.6.0)

Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.6/dist-packages (from importlib-metadata>=0.20; python\_version < "3.8"->catalogue<1.1.0,>=0.0.7->spacy>=2.2.2->de\_core\_news\_sm==2.2.5) (3.1.0)

✓ Download and installation successful

You can now load the model via spacy.load('de\_core\_news\_sm')

✓ Linking successful

/usr/local/lib/python3.6/dist-packages/de\_core\_news\_sm -->  
 /usr/local/lib/python3.6/dist-packages/spacy/data/de

You can now load the model via spacy.load('de')

In [0]:

```
def tokenize_de(text):
    """
    Tokenizes German text from a string into a list of strings (tokens) and reverses it
    """
    return [tok.text for tok in spacy_de.tokenizer(text)][::-1]

def tokenize_en(text):
    """
    Tokenizes English text from a string into a list of strings (tokens)
    """
    return [tok.text for tok in spacy_en.tokenizer(text)]

# немецкий язык является полем SRC, а английский в поле TRG
SRC = Field(tokenize = tokenize_de,
            init_token = '<sos>',
            eos_token = '<eos>',
            lower = True)

TRG = Field(tokenize = tokenize_en,
            init_token = '<sos>',
            eos_token = '<eos>',
            lower = True)
```

In [0]:

```
# В датасете содержится ~ 30к предложений средняя длина которых 11
train_data, valid_data, test_data = Multi30k.splits(exts = ('.de', '.en'), fields = (SRC, TRG))
```

Давайте посмотрим что у нас с датасетом и сделаем словари для SRC и TGT

In [6]:

```
labels = ['train', 'validation', 'test']
dataloaders = [train_data, valid_data, test_data]
for d, l in zip(dataloaders, labels):
    print("Number of sentences in {} : {}".format(l, len(d.examples)))
```

```
Number of sentences in train : 29000
Number of sentences in validation : 1014
Number of sentences in test : 1000
```

In [7]:

```
SRC.build_vocab(train_data, min_freq = 2)
TRG.build_vocab(train_data, min_freq = 2)
print("Number of words in source vocabulary", len(SRC.vocab))
print("Number of words in source vocabulary", len(TRG.vocab))
```

```
Number of words in source vocabulary 7855
Number of words in source vocabulary 5893
```

# Encoder

Энкодер будет ровно как в семинаре, с единственным изменением -- forward будет возвращать не только hidden, cell, но еще и outputs. Это нужно (надеюсь, вы уже поняли) для использования attention в декодере

In [0]:

```
from torch.nn.utils.rnn import pack_padded_sequence, pad_packed_sequence

class Encoder(nn.Module):
    def __init__(self, input_dim, emb_dim, hid_dim, n_layers, dropout):
        """
        :param: input_dim is the size/dimensionality of the one-hot vectors that will be input to the encoder. This is equal to the input (source) vocabulary size.
        :param: emb_dim is the dimensionality of the embedding layer. This layer converts the one-hot vectors into dense vectors with emb_dim dimensions.
        :param: hid_dim is the dimensionality of the hidden and cell states.
        :param: n_layers is the number of layers in the RNN.
        :param: percentage of the dropout to use
        """
        super().__init__()

        self.input_dim = input_dim
        self.emb_dim = emb_dim
        self.hid_dim = hid_dim
        self.n_layers = n_layers

        self.embedding = nn.Embedding(input_dim, emb_dim)

        self.rnn = nn.LSTM(self.emb_dim, self.hid_dim, self.n_layers)

        self.dropout = nn.Dropout(dropout)

    def forward(self, src):
        """
        :param: src sentences (src_len x batch_size)
        """
        # embedded = <TODO> (src_len x batch_size x embd_dim)
        embedded = self.embedding(src)
        # dropout over embedding
        embedded = self.dropout(embedded)
        outputs, (hidden, cell) = self.rnn(embedded)
        # [Attention return is for lstm, but you can also use gru]
        return outputs, hidden, cell
```

# Decoder

Оп ля, а тут уже что-то новенькое

Мы будем реализовывать attention, который будет смотреть из tgt в src (HE self-attention).

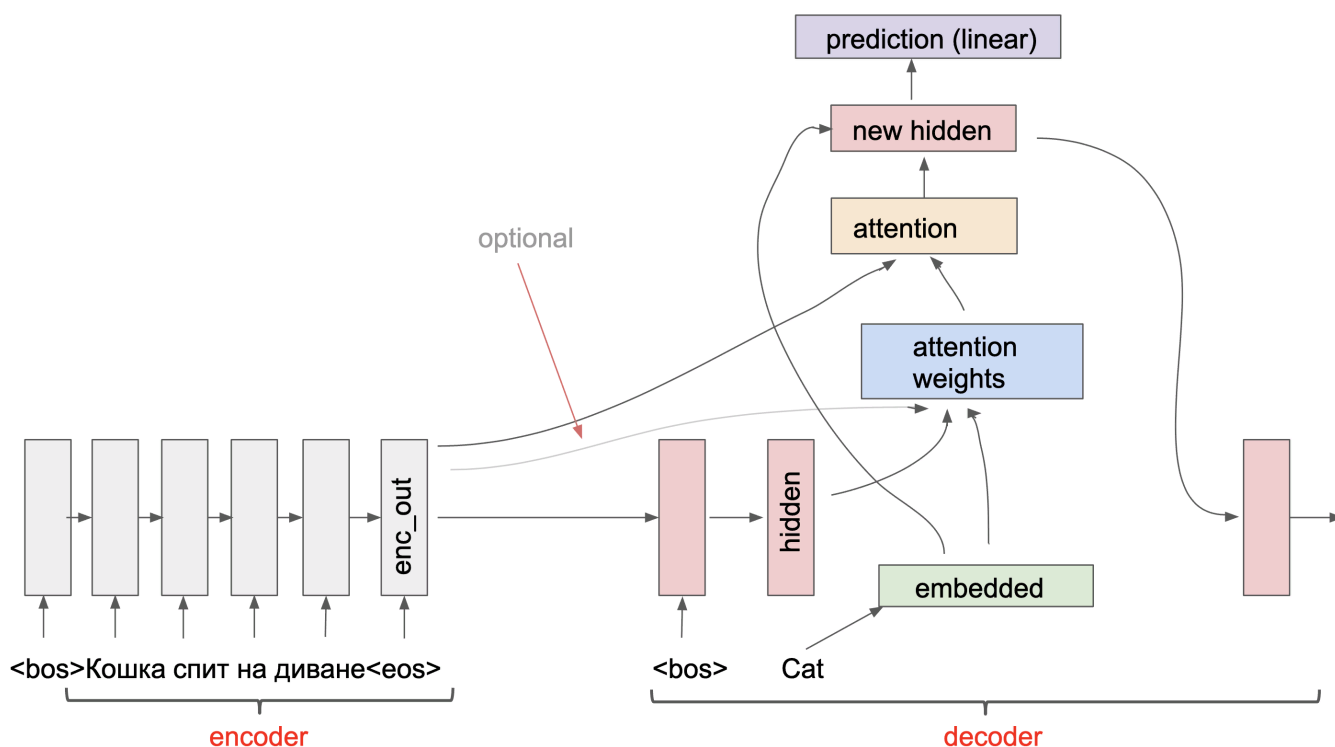
Определим два класса -- Attention и DecoderAttn. Мы разбили их на два класса, чтобы можно было играть с типом внимания, не меняя код DecoderAttn. Как вы помните с лекции, в качестве аттеншена можно брать любую странную функцию (конкатенация, маленькая сеточка, ...), и все будет работать! Поэтому вам предлагается попробовать несколько разных.

Есть два подхода к реализации аттеншена:

Подход #1:

1. Вычисляется embed
2. На основе hidden с прошлого шага, embedded и (возможно) enc\_out вычисляется attention, а точнее, веса attention (поэтому не забудьте softmax!!). Размерность  $batch\_size * max\_len, max\_len$  - максимальная длина предложения в батче, т.е.  $shape[0]$  от выхода энкодера.
3. К enc\_out применяется attention: чаще всего dot product от enc\_out и attention\_weights (не забудьте про измерение батч. Чтобы нормально вычислить dot\_product по батчу, вам поможет torch.bmm)
4. Берутся attention и embedded и сворачиваются в один вектор размерности такой, чтобы кормить его self.lstm. Например, это можно сделать с помощью обычного линейного слоя
5. Вычисляется новое скрытое состояние new\_hidden. Это наша self.lstm, примененная к выходу пункта 4.
6. Вычисляется prediction, как в семинаре

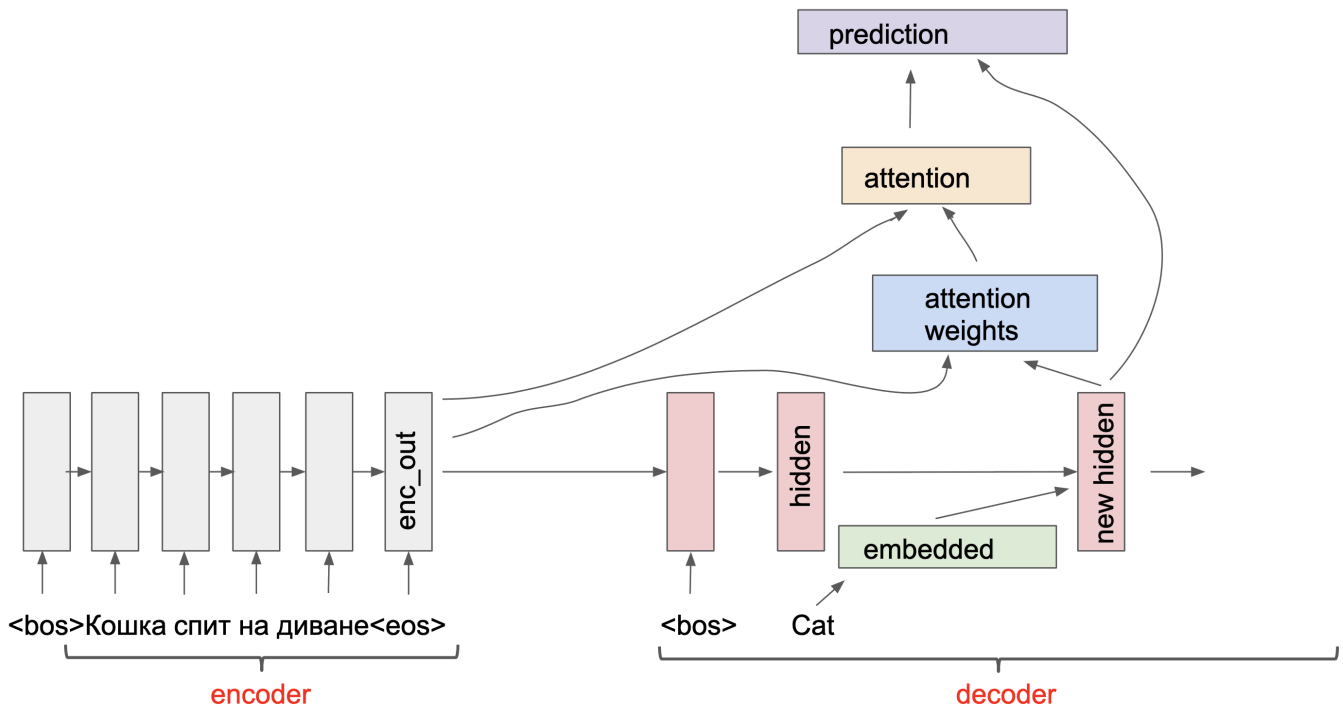
Грубо говоря, вся разница с семинаром в том, что мы вместо того, чтобы embedded пихать в self.lstm, миксуем аттеншен на основе всего, что имеем (enc\_out, hidden, embedded) и запикиваем в self.lstm микс аттеншена и embedded.



Подход #2:

1. Вычисляется embed
2. Вычисляется output, new\_hidden (строка `output, (hidden, cell) = self.rnn(embedded, (hidden, cell))`)
3. На основе output и enc\_out вычисляется attention, а точнее, веса attention (поэтому не забудьте softmax!!)
4. К enc\_out применяется attention: чаще всего dot product от enc\_out и attention\_weights (не забудьте про измерение батч. Чтобы нормально вычислить dot\_product по батчу, вам поможет `torch.bmm`)
5. Вычисляется prediction на основе attention и output. Можно, например, взять `nn.Linear()` от конкатенации attention и output.

Разница с первым подходом в том, что мы сначала вычисляем выход rnn слоя, а потом смотрим вниманием на src и на основе выхода rnn и attn считаем выход (prediction).



Вам предлагается реализовать хотя бы 1 из вариантов и хотя бы 2 варианта функции attention (в классе Attention)



In [0]:

```
class Attention(nn.Module):
    def __init__(self, batch_size, hidden_size=None, method="dot"): # add parameters needed for your type of attention
        super(Attention, self).__init__()
        self.method = method # attention method you'll use. e.g. "cat", "one-layer-net", "dot", ...
        self.batch_size = batch_size
        self.hidden_size = hidden_size

    def dots(self, rnn_output, encoder_outputs):
        return torch.sum(rnn_output * encoder_outputs, dim=2)

    def forward(self, rnn_output, encoder_outputs, seq_len=None):
        dots_score = self.dots(rnn_output, encoder_outputs)
        attn_w = F.softmax(dots_score, dim=0).unsqueeze(1)
        context_vector = torch.bmm(attn_w.transpose(0,2), encoder_outputs.transpose(0,1))

        return context_vector

class DecoderAttn(nn.Module):
    def __init__(self, output_dim, emb_dim, hid_dim, n_layers, attention, dropout=0.1):
        super(DecoderAttn, self).__init__()

        self.emb_dim = emb_dim
        self.hid_dim = hid_dim
        self.output_dim = output_dim
        self.n_layers = n_layers

        self.attn = attention # instance of Attention class
        self.embedding = nn.Embedding(self.output_dim, self.emb_dim) # define layers

        self.rnn = nn.LSTM(self.emb_dim, self.hid_dim, self.n_layers) # (lstm embd, hid, layers, dropout?)
        self.out = nn.Linear(self.hid_dim, self.output_dim) # Projection : hid_dim x output_dim
        self.dropout = nn.Dropout(dropout)

        # more layers you'll need for attention
        self.concat = nn.Linear(self.hid_dim * 2, self.hid_dim)

    def forward(self, input_, last_hidden, cell, encoder_output):
        # make decoder with attention
        # use code from seminar notebook as base and add attention to it
        input_ = input_.unsqueeze(0)

        embedded = self.embedding(input_)

        rnn_output, (new_hidden, cell) = self.rnn(self.dropout(embedded), (last_hidden, cell))

        context_vector = self.attn(rnn_output, encoder_output)
        rnn_output = rnn_output.squeeze(0)
        context_vector = context_vector.squeeze(1)

        concat = torch.cat((context_vector, rnn_output), dim=1)

        concat_out = self.concat(concat)
        prediction = self.out(concat_out)
```

```
        return prediction, new_hidden, cell

def initHidden(self):
    return torch.zeros(1, 1, self.hidden_size, device=device)
```

## Seq2Seq module

Здесь опять ничего не поменяется кроме того, что энкодер теперь возвращает свой output, а декодер его принимает

In [0]:

```
# we need that to put it first to the decoder in 'translate' method
BOS_IDX = SRC.vocab.stoi['<sos>']

class Seq2Seq(nn.Module):
    def __init__(self, encoder, decoder, device):
        super().__init__()
        # Hidden dimensions of encoder and decoder must be equal
        self.encoder = encoder
        self.decoder = decoder
        self.device = device
        self._init_weights()
        self.max_len=30

    def forward(self, src, trg, teacher_forcing_ratio = 0.5):
        """
        :param: src (src_len x batch_size)
        :param: trg
        :param: teacher_forcing_ratio : if 0.5 then every second token is the ground truth input
        """

        batch_size = trg.shape[1]
        max_len = trg.shape[0]
        trg_vocab_size = self.decoder.output_dim

        #tensor to store decoder outputs
        outputs = torch.zeros(max_len, batch_size, trg_vocab_size).to(self.device)
        enc_out, hidden, cell = self.encoder(src)
        input_ = trg[0, :]

        for t in range(1, max_len):

            output, hidden, cell = self.decoder(input_, hidden, cell, encoder_output=enc_out) #TODO pass state and input throw decoder
            outputs[t] = output
            teacher_force = random.random() < teacher_forcing_ratio
            top1 = output.max(1)[1]
            input_ = (trg[t] if teacher_force else top1)

        return outputs

    def translate(self, src):
        trg_vocab_size = self.decoder.output_dim

        #tensor to store decoder outputs
        outputs = []

        #last hidden state of the encoder is used as the initial hidden state of the decoder
        src = torch.tensor(src).to(self.device).unsqueeze(1)
        enc_out, hidden, cell = self.encoder(src) # TODO pass src throw encoder

        input_ = torch.tensor([SRC.vocab.stoi['<sos>']] * src.shape[1]).to(device) # TODO trg[idxs]

        for t in range(1, self.max_len):

            output, hidden, cell = self.decoder(input_, hidden, cell, enc_out) #TODO pass state and input throw decoder
```

```

        top1 = output.max(1)[1]
        outputs.append(top1)
        input_ = (top1)

    return outputs

def _init_weights(self):
    p = 0.08
    for name, param in self.named_parameters():
        nn.init.uniform_(param.data, -p, p)

```

In [0]:

```

input_dim = len(SRC.vocab)
output_dim = len(TRG.vocab)
src_embd_dim = tgt_embd_dim = 256
hidden_dim = 512
num_layers = 2
dropout_prob = 0.2

batch_size = 64
PAD_IDX = TRG.vocab.stoi['<pad>']

iterators = BucketIterator.splits((train_data, valid_data, test_data),
                                  batch_size = batch_size, device = device)

train_iterator, valid_iterator, test_iterator = iterators

enc = Encoder(input_dim = input_dim, emb_dim = src_embd_dim, hid_dim=hidden_dim,
              n_layers=num_layers, dropout=dropout_prob).to(device)

attn = Attention(batch_size, method='dot')

dec = DecoderAttn(output_dim=output_dim, emb_dim=tgt_embd_dim, hid_dim=hidden_dim,
                  n_layers=num_layers, attention=attn, dropout=dropout_prob).to(device)

model = Seq2Seq(enc, dec, device).to(device)

```

In [ ]:

```


```

In [12]:

```
model
```

Out[12]:

```
Seq2Seq(
  (encoder): Encoder(
    (embedding): Embedding(7855, 256)
    (rnn): LSTM(256, 512, num_layers=2)
    (dropout): Dropout(p=0.2, inplace=False)
  )
  (decoder): DecoderAttn(
    (attn): Attention()
    (embedding): Embedding(5893, 256)
    (rnn): LSTM(256, 512, num_layers=2)
    (out): Linear(in_features=512, out_features=5893, bias=True)
    (dropout): Dropout(p=0.2, inplace=False)
    (concat): Linear(in_features=1024, out_features=512, bias=True)
  )
)
```

In [0]:

```
def train(model, iterator, optimizer, criterion, clip):

    model.train()

    epoch_loss = 0

    for i, batch in enumerate(iterator):

        src = batch.src
        print(batch.src.shape)
        trg = batch.trg

        optimizer.zero_grad()

        output = model(src, trg)

        output = output[1:].view(-1, output.shape[-1])
        trg = trg[1:].view(-1)

        loss = criterion(output, trg)

        loss.backward()

        torch.nn.utils.clip_grad_norm_(model.parameters(), clip)

        optimizer.step()

        epoch_loss += loss.item()

    return epoch_loss / len(iterator)
```

In [0]:

```
def evaluate(model, iterator, criterion):  
    model.eval()  
    epoch_loss = 0  
    with torch.no_grad():  
        for i, batch in enumerate(iterator):  
            src = batch.src  
            trg = batch.trg  
  
            output = model(src, trg, 0) #turn off teacher forcing !!  
            output = output[1:].view(-1, output.shape[-1])  
            trg = trg[1:].view(-1)  
  
            loss = criterion(output, trg)  
  
            epoch_loss += loss.item()  
    return epoch_loss / len(iterator)
```

In [15]:

```
max_epochs = 15
CLIP = 1

# TODO
optimizer = optim.Adam(model.parameters(), lr = 1e-3)
criterion = nn.CrossEntropyLoss(ignore_index = PAD_IDX)

best_valid_loss = float('inf')

for epoch in range(max_epochs):

    train_loss = round(train(model, train_iterator, optimizer, criterion, CLIP), 5)
    valid_loss = round(evaluate(model, valid_iterator, criterion), 5)

    if valid_loss < best_valid_loss:
        best_valid_loss = valid_loss
        torch.save(model.state_dict(), 'model.pt')

    print('Epoch: {} \n Train Loss {} Val loss {}'.format(epoch, train_loss, valid_loss))
    print('Train Perplexity {} Val Perplexity {}'.format(np.exp(train_loss), np.exp(valid_loss)))
```

Streaming output truncated to the last 5000 lines.

```
torch.Size([28, 64])
torch.Size([34, 64])
torch.Size([22, 64])
torch.Size([28, 64])
torch.Size([29, 64])
torch.Size([29, 64])
torch.Size([30, 64])
torch.Size([28, 64])
torch.Size([29, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([23, 64])
torch.Size([23, 64])
torch.Size([46, 64])
torch.Size([22, 64])
torch.Size([26, 64])
torch.Size([23, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([30, 64])
torch.Size([37, 64])
torch.Size([33, 64])
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([23, 64])
torch.Size([25, 64])
torch.Size([23, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([32, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([35, 64])
torch.Size([33, 64])
torch.Size([28, 64])
torch.Size([29, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([30, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([19, 64])
torch.Size([24, 64])
torch.Size([30, 64])
torch.Size([32, 64])
torch.Size([22, 64])
torch.Size([31, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([24, 64])
```



torch.Size([31, 64])  
torch.Size([35, 64])  
torch.Size([30, 64])  
torch.Size([38, 64])  
torch.Size([27, 64])  
torch.Size([31, 64])  
torch.Size([21, 64])  
torch.Size([30, 64])  
torch.Size([23, 64])  
torch.Size([30, 64])  
torch.Size([26, 64])  
torch.Size([24, 64])  
torch.Size([34, 64])  
torch.Size([23, 64])  
torch.Size([25, 64])  
torch.Size([30, 64])  
torch.Size([29, 64])  
torch.Size([26, 64])  
torch.Size([28, 64])  
torch.Size([28, 64])  
torch.Size([27, 64])  
torch.Size([30, 64])  
torch.Size([26, 64])  
torch.Size([29, 64])  
torch.Size([28, 64])  
torch.Size([23, 64])  
torch.Size([25, 64])  
torch.Size([26, 64])  
torch.Size([37, 64])  
torch.Size([27, 64])  
torch.Size([27, 64])  
torch.Size([29, 64])  
torch.Size([23, 64])  
torch.Size([28, 64])  
torch.Size([27, 64])  
torch.Size([32, 64])  
torch.Size([37, 64])  
torch.Size([36, 64])  
torch.Size([27, 64])  
torch.Size([28, 64])  
torch.Size([29, 64])  
torch.Size([23, 64])  
torch.Size([25, 64])  
torch.Size([24, 64])  
torch.Size([36, 64])  
torch.Size([24, 64])  
torch.Size([25, 64])  
torch.Size([30, 64])  
torch.Size([24, 64])  
torch.Size([25, 64])  
torch.Size([26, 64])  
torch.Size([23, 64])  
torch.Size([25, 64])  
torch.Size([23, 64])  
torch.Size([31, 64])  
torch.Size([34, 64])  
torch.Size([28, 64])  
torch.Size([32, 64])  
torch.Size([25, 64])  
torch.Size([25, 64])  
torch.Size([34, 64])

```
torch.Size([24, 64])
torch.Size([23, 64])
torch.Size([30, 64])
torch.Size([26, 64])
torch.Size([23, 64])
torch.Size([37, 64])
torch.Size([24, 64])
torch.Size([33, 64])
torch.Size([25, 64])
torch.Size([30, 64])
torch.Size([27, 64])
torch.Size([21, 64])
torch.Size([32, 64])
torch.Size([25, 64])
torch.Size([36, 64])
torch.Size([29, 64])
torch.Size([33, 64])
torch.Size([26, 64])
torch.Size([30, 64])
torch.Size([24, 64])
torch.Size([34, 64])
torch.Size([27, 64])
torch.Size([22, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([21, 64])
torch.Size([26, 64])
torch.Size([29, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([38, 64])
torch.Size([23, 64])
torch.Size([37, 64])
torch.Size([28, 64])
torch.Size([36, 64])
torch.Size([28, 64])
torch.Size([23, 64])
torch.Size([29, 64])
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([32, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([30, 64])
torch.Size([26, 64])
torch.Size([23, 64])
torch.Size([31, 64])
torch.Size([29, 64])
torch.Size([23, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([23, 64])
torch.Size([26, 64])
torch.Size([34, 64])
torch.Size([22, 64])
torch.Size([28, 64])
torch.Size([28, 64])
```

```
torch.Size([34, 64])
torch.Size([28, 64])
torch.Size([33, 64])
torch.Size([26, 64])
torch.Size([23, 64])
torch.Size([27, 64])
torch.Size([23, 64])
torch.Size([28, 64])
torch.Size([29, 64])
torch.Size([30, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([38, 64])
torch.Size([29, 64])
torch.Size([23, 64])
torch.Size([30, 64])
torch.Size([34, 64])
torch.Size([29, 64])
torch.Size([28, 64])
torch.Size([31, 64])
torch.Size([37, 64])
torch.Size([23, 64])
torch.Size([26, 64])
torch.Size([23, 64])
torch.Size([25, 64])
torch.Size([31, 64])
torch.Size([23, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([23, 64])
torch.Size([28, 64])
torch.Size([29, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([35, 64])
torch.Size([29, 64])
torch.Size([22, 64])
torch.Size([29, 64])
torch.Size([31, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([30, 64])
torch.Size([23, 64])
torch.Size([32, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([22, 64])
torch.Size([33, 64])
torch.Size([25, 64])
torch.Size([36, 64])
torch.Size([33, 64])
```

```
torch.Size([26, 64])
torch.Size([30, 64])
torch.Size([30, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([29, 64])
torch.Size([27, 64])
torch.Size([31, 64])
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([33, 64])
torch.Size([34, 64])
torch.Size([38, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([36, 64])
torch.Size([34, 64])
torch.Size([26, 64])
torch.Size([32, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([23, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([23, 64])
torch.Size([31, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([23, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([35, 64])
torch.Size([25, 64])
torch.Size([33, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([30, 64])
torch.Size([31, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([22, 64])
torch.Size([26, 64])
torch.Size([30, 64])
torch.Size([24, 64])
torch.Size([22, 64])
torch.Size([23, 64])
torch.Size([26, 64])
torch.Size([30, 64])
torch.Size([38, 64])
```

```
torch.Size([25, 64])
torch.Size([33, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([31, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([30, 64])
torch.Size([31, 64])
torch.Size([31, 64])
torch.Size([41, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([33, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([23, 64])
torch.Size([27, 64])
torch.Size([30, 64])
torch.Size([33, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([31, 64])
torch.Size([28, 64])
torch.Size([29, 64])
torch.Size([36, 64])
torch.Size([45, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([36, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([22, 64])
torch.Size([32, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([32, 64])
torch.Size([29, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([23, 64])
torch.Size([29, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([32, 64])
```

```
torch.Size([23, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([38, 64])
torch.Size([29, 64])
torch.Size([25, 64])
torch.Size([29, 64])
torch.Size([29, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([23, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([32, 64])
torch.Size([22, 64])
torch.Size([23, 64])
torch.Size([30, 64])
torch.Size([22, 64])
torch.Size([25, 64])
torch.Size([22, 64])
torch.Size([23, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([29, 64])
torch.Size([32, 64])
torch.Size([31, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([21, 8])
torch.Size([20, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([29, 64])
torch.Size([26, 64])
torch.Size([29, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([24, 64])
torch.Size([37, 64])
torch.Size([27, 64])
torch.Size([33, 64])
torch.Size([29, 64])
torch.Size([32, 64])
torch.Size([25, 64])
torch.Size([29, 64])
torch.Size([31, 64])
torch.Size([23, 64])
torch.Size([25, 64])
torch.Size([24, 64])
```

Epoch: 4

Train Loss 2.003 Val loss 3.2791:

Train Perplexity 7.411256551255594 Val Perplexity 26.55186526440379:

torch.Size([27, 64])  
torch.Size([30, 64])  
torch.Size([27, 64])  
torch.Size([34, 64])  
torch.Size([26, 64])  
torch.Size([22, 64])  
torch.Size([22, 64])  
torch.Size([23, 64])  
torch.Size([30, 64])  
torch.Size([24, 64])  
torch.Size([29, 64])  
torch.Size([29, 64])  
torch.Size([26, 64])  
torch.Size([25, 64])  
torch.Size([29, 64])  
torch.Size([27, 64])  
torch.Size([28, 64])  
torch.Size([29, 64])  
torch.Size([29, 64])  
torch.Size([27, 64])  
torch.Size([24, 64])  
torch.Size([24, 64])  
torch.Size([28, 64])  
torch.Size([27, 64])  
torch.Size([27, 64])  
torch.Size([32, 64])  
torch.Size([24, 64])  
torch.Size([26, 64])  
torch.Size([35, 64])  
torch.Size([24, 64])  
torch.Size([33, 64])  
torch.Size([26, 64])  
torch.Size([28, 64])  
torch.Size([27, 64])  
torch.Size([26, 64])  
torch.Size([27, 64])  
torch.Size([30, 64])  
torch.Size([27, 64])  
torch.Size([29, 64])  
torch.Size([24, 64])  
torch.Size([25, 64])  
torch.Size([32, 64])  
torch.Size([27, 64])  
torch.Size([28, 64])  
torch.Size([30, 64])  
torch.Size([32, 64])  
torch.Size([24, 64])  
torch.Size([30, 64])  
torch.Size([29, 64])  
torch.Size([30, 64])  
torch.Size([25, 64])  
torch.Size([25, 64])  
torch.Size([23, 64])  
torch.Size([24, 64])  
torch.Size([27, 64])  
torch.Size([22, 64])  
torch.Size([23, 64])  
torch.Size([25, 64])

```
torch.Size([25, 64])
torch.Size([37, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([30, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([22, 64])
torch.Size([24, 64])
torch.Size([23, 64])
torch.Size([38, 64])
torch.Size([30, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([30, 64])
torch.Size([22, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([29, 64])
torch.Size([31, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([31, 64])
torch.Size([22, 64])
torch.Size([29, 64])
torch.Size([25, 64])
torch.Size([30, 64])
torch.Size([36, 64])
torch.Size([36, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([23, 64])
torch.Size([33, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([46, 64])
torch.Size([35, 64])
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([22, 64])
torch.Size([45, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([31, 64])
torch.Size([25, 64])
torch.Size([34, 64])
torch.Size([29, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([38, 64])
torch.Size([34, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([29, 64])
torch.Size([25, 64])
```



```
torch.Size([29, 64])
torch.Size([31, 64])
torch.Size([36, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([22, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([31, 64])
torch.Size([24, 64])
torch.Size([34, 64])
torch.Size([27, 64])
torch.Size([33, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([29, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([23, 64])
torch.Size([33, 64])
torch.Size([23, 64])
torch.Size([29, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([33, 64])
torch.Size([32, 64])
torch.Size([30, 64])
torch.Size([29, 64])
torch.Size([23, 64])
torch.Size([26, 64])
torch.Size([36, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([31, 64])
torch.Size([30, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([30, 64])
torch.Size([34, 64])
torch.Size([30, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([30, 64])
torch.Size([23, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([34, 64])
torch.Size([30, 64])
```

```
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([31, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([33, 64])
torch.Size([30, 64])
torch.Size([28, 64])
torch.Size([23, 64])
torch.Size([32, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([31, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([29, 64])
torch.Size([23, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([23, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([32, 64])
torch.Size([28, 64])
torch.Size([21, 64])
torch.Size([28, 64])
torch.Size([29, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([29, 64])
torch.Size([24, 64])
torch.Size([22, 64])
torch.Size([27, 64])
torch.Size([30, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([23, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([23, 64])
torch.Size([29, 64])
torch.Size([29, 64])
torch.Size([27, 64])
torch.Size([32, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([23, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([31, 64])
torch.Size([27, 64])
torch.Size([33, 64])
```

```
torch.Size([29, 64])
torch.Size([31, 64])
torch.Size([26, 64])
torch.Size([37, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([30, 64])
torch.Size([32, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([30, 64])
torch.Size([23, 64])
torch.Size([25, 64])
torch.Size([32, 64])
torch.Size([25, 64])
torch.Size([30, 64])
torch.Size([41, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([23, 64])
torch.Size([25, 64])
torch.Size([35, 64])
torch.Size([26, 64])
torch.Size([22, 64])
torch.Size([23, 64])
torch.Size([26, 64])
torch.Size([33, 64])
torch.Size([30, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([34, 64])
torch.Size([32, 64])
torch.Size([32, 64])
torch.Size([24, 64])
torch.Size([36, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([36, 64])
torch.Size([34, 64])
torch.Size([30, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([29, 64])
torch.Size([23, 64])
torch.Size([33, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([29, 64])
torch.Size([33, 64])
torch.Size([26, 64])
torch.Size([45, 64])
torch.Size([26, 64])
torch.Size([23, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([21, 64])
torch.Size([23, 64])
```

```
torch.Size([37, 64])
torch.Size([27, 64])
torch.Size([23, 64])
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([38, 64])
torch.Size([32, 64])
torch.Size([32, 64])
torch.Size([22, 64])
torch.Size([22, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([30, 64])
torch.Size([26, 64])
torch.Size([23, 64])
torch.Size([23, 64])
torch.Size([31, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([30, 64])
torch.Size([28, 64])
torch.Size([22, 64])
torch.Size([37, 64])
torch.Size([26, 64])
torch.Size([30, 64])
torch.Size([30, 64])
torch.Size([22, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([29, 64])
torch.Size([25, 64])
torch.Size([30, 64])
torch.Size([29, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([31, 64])
torch.Size([29, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([29, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([33, 64])
torch.Size([24, 64])
torch.Size([35, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([34, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([31, 64])
torch.Size([37, 64])
torch.Size([30, 64])
torch.Size([27, 64])
torch.Size([31, 64])
torch.Size([36, 64])
torch.Size([27, 64])
torch.Size([23, 64])
```

```
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([32, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([22, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([23, 64])
torch.Size([25, 64])
torch.Size([30, 64])
torch.Size([27, 64])
torch.Size([33, 64])
torch.Size([26, 64])
torch.Size([31, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([31, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([33, 64])
torch.Size([33, 64])
torch.Size([26, 64])
torch.Size([23, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([23, 64])
torch.Size([30, 64])
torch.Size([31, 64])
torch.Size([35, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([30, 64])
torch.Size([25, 64])
torch.Size([30, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([25, 64])
```

```
torch.Size([26, 64])
torch.Size([31, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([19, 8])
torch.Size([23, 64])
torch.Size([24, 64])
torch.Size([31, 64])
torch.Size([30, 64])
torch.Size([32, 64])
torch.Size([29, 64])
torch.Size([38, 64])
torch.Size([28, 64])
torch.Size([30, 64])
torch.Size([29, 64])
torch.Size([23, 64])
torch.Size([38, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([34, 64])
torch.Size([25, 64])
torch.Size([29, 64])
torch.Size([38, 64])
torch.Size([31, 64])
torch.Size([28, 64])
torch.Size([20, 64])
```

Epoch: 5

Train Loss 1.81771 Val loss 3.30669:

Train Perplexity 6.1577410646088735 Val Perplexity 27.294630558295836:

```
torch.Size([33, 64])
torch.Size([33, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([34, 64])
torch.Size([32, 64])
torch.Size([23, 64])
torch.Size([24, 64])
torch.Size([23, 64])
torch.Size([25, 64])
torch.Size([30, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([30, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([41, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([25, 64])
```

```
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([20, 64])
torch.Size([29, 64])
torch.Size([23, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([37, 64])
torch.Size([22, 64])
torch.Size([26, 64])
torch.Size([22, 64])
torch.Size([26, 64])
torch.Size([32, 64])
torch.Size([30, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([23, 64])
torch.Size([28, 64])
torch.Size([34, 64])
torch.Size([33, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([23, 64])
torch.Size([25, 64])
torch.Size([34, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([23, 64])
torch.Size([29, 64])
torch.Size([28, 64])
torch.Size([22, 64])
torch.Size([26, 64])
torch.Size([29, 64])
torch.Size([23, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([32, 64])
torch.Size([26, 64])
torch.Size([23, 64])
torch.Size([34, 64])
torch.Size([31, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([30, 64])
torch.Size([29, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([33, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([23, 64])
```

torch.Size([27, 64])  
torch.Size([29, 64])  
torch.Size([33, 64])  
torch.Size([23, 64])  
torch.Size([25, 64])  
torch.Size([23, 64])  
torch.Size([27, 64])  
torch.Size([24, 64])  
torch.Size([28, 64])  
torch.Size([34, 64])  
torch.Size([26, 64])  
torch.Size([25, 64])  
torch.Size([27, 64])  
torch.Size([22, 64])  
torch.Size([31, 64])  
torch.Size([26, 64])  
torch.Size([29, 64])  
torch.Size([28, 64])  
torch.Size([24, 64])  
torch.Size([23, 64])  
torch.Size([28, 64])  
torch.Size([25, 64])  
torch.Size([26, 64])  
torch.Size([27, 64])  
torch.Size([20, 64])  
torch.Size([24, 64])  
torch.Size([27, 64])  
torch.Size([32, 64])  
torch.Size([25, 64])  
torch.Size([25, 64])  
torch.Size([28, 64])  
torch.Size([25, 64])  
torch.Size([31, 64])  
torch.Size([30, 64])  
torch.Size([27, 64])  
torch.Size([30, 64])  
torch.Size([28, 64])  
torch.Size([30, 64])  
torch.Size([37, 64])  
torch.Size([25, 64])  
torch.Size([24, 64])  
torch.Size([27, 64])  
torch.Size([27, 64])  
torch.Size([25, 64])  
torch.Size([28, 64])  
torch.Size([31, 64])  
torch.Size([29, 64])  
torch.Size([26, 64])  
torch.Size([24, 64])  
torch.Size([38, 64])  
torch.Size([29, 64])  
torch.Size([27, 64])  
torch.Size([25, 64])  
torch.Size([27, 64])  
torch.Size([27, 64])  
torch.Size([46, 64])  
torch.Size([28, 64])  
torch.Size([22, 64])  
torch.Size([23, 64])  
torch.Size([29, 64])  
torch.Size([29, 64])



```
torch.Size([23, 64])
torch.Size([23, 64])
torch.Size([29, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([22, 64])
torch.Size([24, 64])
torch.Size([23, 64])
torch.Size([23, 64])
torch.Size([38, 64])
torch.Size([37, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([30, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([36, 64])
torch.Size([29, 64])
torch.Size([23, 64])
torch.Size([26, 64])
torch.Size([30, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([32, 64])
torch.Size([23, 64])
torch.Size([36, 64])
torch.Size([33, 64])
torch.Size([29, 64])
torch.Size([23, 64])
torch.Size([34, 64])
torch.Size([31, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([29, 64])
torch.Size([31, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([22, 64])
torch.Size([27, 64])
torch.Size([30, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([31, 64])
torch.Size([25, 64])
torch.Size([45, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([23, 64])
torch.Size([34, 64])
torch.Size([29, 64])
torch.Size([25, 64])
```

```
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([33, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([32, 64])
torch.Size([29, 64])
torch.Size([25, 64])
torch.Size([30, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([38, 64])
torch.Size([32, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([34, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([29, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([31, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([23, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([30, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([23, 64])
torch.Size([30, 64])
torch.Size([36, 64])
torch.Size([34, 64])
torch.Size([27, 64])
torch.Size([30, 64])
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([31, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([22, 64])
torch.Size([22, 64])
torch.Size([32, 64])
torch.Size([24, 64])
torch.Size([37, 64])
torch.Size([29, 64])
torch.Size([25, 64])
torch.Size([24, 64])
```

```
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([23, 64])
torch.Size([27, 64])
torch.Size([30, 64])
torch.Size([30, 64])
torch.Size([37, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([29, 64])
torch.Size([24, 64])
torch.Size([23, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([22, 64])
torch.Size([32, 64])
torch.Size([24, 64])
torch.Size([23, 64])
torch.Size([31, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([32, 64])
torch.Size([32, 64])
torch.Size([26, 64])
torch.Size([36, 64])
torch.Size([26, 64])
torch.Size([29, 64])
torch.Size([30, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([29, 64])
torch.Size([31, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([30, 64])
torch.Size([31, 64])
torch.Size([31, 64])
torch.Size([30, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([31, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([26, 64])
```

torch.Size([29, 64])  
torch.Size([32, 64])  
torch.Size([37, 64])  
torch.Size([30, 64])  
torch.Size([28, 64])  
torch.Size([31, 64])  
torch.Size([24, 64])  
torch.Size([29, 64])  
torch.Size([29, 64])  
torch.Size([26, 64])  
torch.Size([26, 64])  
torch.Size([24, 64])  
torch.Size([28, 64])  
torch.Size([25, 64])  
torch.Size([25, 64])  
torch.Size([38, 64])  
torch.Size([29, 64])  
torch.Size([25, 64])  
torch.Size([28, 64])  
torch.Size([24, 64])  
torch.Size([36, 64])  
torch.Size([30, 64])  
torch.Size([31, 64])  
torch.Size([24, 64])  
torch.Size([30, 64])  
torch.Size([26, 64])  
torch.Size([24, 64])  
torch.Size([22, 64])  
torch.Size([28, 64])  
torch.Size([27, 64])  
torch.Size([33, 64])  
torch.Size([23, 64])  
torch.Size([31, 64])  
torch.Size([29, 64])  
torch.Size([26, 64])  
torch.Size([28, 64])  
torch.Size([28, 64])  
torch.Size([33, 64])  
torch.Size([25, 64])  
torch.Size([23, 64])  
torch.Size([30, 64])  
torch.Size([38, 64])  
torch.Size([33, 64])  
torch.Size([24, 64])  
torch.Size([28, 64])  
torch.Size([45, 64])  
torch.Size([23, 64])  
torch.Size([24, 64])  
torch.Size([22, 64])  
torch.Size([24, 64])  
torch.Size([27, 64])  
torch.Size([27, 64])  
torch.Size([27, 64])  
torch.Size([28, 64])  
torch.Size([25, 64])  
torch.Size([26, 64])  
torch.Size([36, 64])  
torch.Size([24, 64])  
torch.Size([24, 64])  
torch.Size([33, 64])  
torch.Size([35, 64])

```
torch.Size([30, 64])
torch.Size([25, 64])
torch.Size([33, 64])
torch.Size([25, 64])
torch.Size([23, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([37, 64])
torch.Size([35, 64])
torch.Size([25, 64])
torch.Size([35, 64])
torch.Size([29, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([21, 64])
torch.Size([31, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([29, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([35, 64])
torch.Size([23, 64])
torch.Size([35, 64])
torch.Size([25, 64])
torch.Size([33, 64])
torch.Size([38, 64])
torch.Size([26, 64])
torch.Size([34, 64])
torch.Size([30, 64])
torch.Size([24, 64])
torch.Size([33, 64])
torch.Size([29, 64])
torch.Size([30, 64])
torch.Size([30, 64])
torch.Size([22, 64])
torch.Size([31, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([30, 64])
torch.Size([33, 64])
torch.Size([29, 64])
torch.Size([24, 8])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([33, 64])
torch.Size([29, 64])
torch.Size([25, 64])
torch.Size([30, 64])
torch.Size([34, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([30, 64])
torch.Size([31, 64])
Epoch: 6
```

Train Loss 1.64119 Val loss 3.32449:  
Train Perplexity 5.161307815520324 Val Perplexity 27.784824768007155:  
torch.Size([21, 64])  
torch.Size([31, 64])  
torch.Size([25, 64])  
torch.Size([27, 64])  
torch.Size([25, 64])  
torch.Size([24, 64])  
torch.Size([35, 64])  
torch.Size([26, 64])  
torch.Size([36, 64])  
torch.Size([29, 64])  
torch.Size([23, 64])  
torch.Size([30, 64])  
torch.Size([23, 64])  
torch.Size([29, 64])  
torch.Size([27, 64])  
torch.Size([28, 64])  
torch.Size([32, 64])  
torch.Size([37, 64])  
torch.Size([23, 64])  
torch.Size([23, 64])  
torch.Size([32, 64])  
torch.Size([28, 64])  
torch.Size([34, 64])  
torch.Size([27, 64])  
torch.Size([32, 64])  
torch.Size([28, 64])  
torch.Size([24, 64])  
torch.Size([29, 64])  
torch.Size([25, 64])  
torch.Size([28, 64])  
torch.Size([30, 64])  
torch.Size([25, 64])  
torch.Size([22, 64])  
torch.Size([26, 64])  
torch.Size([24, 64])  
torch.Size([41, 64])  
torch.Size([23, 64])  
torch.Size([26, 64])  
torch.Size([24, 64])  
torch.Size([26, 64])  
torch.Size([24, 64])  
torch.Size([29, 64])  
torch.Size([32, 64])  
torch.Size([25, 64])  
torch.Size([38, 64])  
torch.Size([28, 64])  
torch.Size([25, 64])  
torch.Size([31, 64])  
torch.Size([28, 64])  
torch.Size([22, 64])  
torch.Size([26, 64])  
torch.Size([28, 64])  
torch.Size([28, 64])  
torch.Size([23, 64])  
torch.Size([28, 64])  
torch.Size([29, 64])  
torch.Size([25, 64])  
torch.Size([31, 64])  
torch.Size([24, 64])

```
torch.Size([30, 64])
torch.Size([22, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([29, 64])
torch.Size([23, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([46, 64])
torch.Size([34, 64])
torch.Size([33, 64])
torch.Size([35, 64])
torch.Size([28, 64])
torch.Size([29, 64])
torch.Size([24, 64])
torch.Size([34, 64])
torch.Size([24, 64])
torch.Size([32, 64])
torch.Size([27, 64])
torch.Size([23, 64])
torch.Size([31, 64])
torch.Size([37, 64])
torch.Size([24, 64])
torch.Size([30, 64])
torch.Size([29, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([30, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([29, 64])
torch.Size([29, 64])
torch.Size([28, 64])
torch.Size([29, 64])
torch.Size([25, 64])
torch.Size([22, 64])
torch.Size([33, 64])
torch.Size([32, 64])
torch.Size([31, 64])
torch.Size([31, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([23, 64])
torch.Size([22, 64])
torch.Size([29, 64])
torch.Size([28, 64])
torch.Size([30, 64])
torch.Size([23, 64])
torch.Size([35, 64])
torch.Size([28, 64])
```

```
torch.Size([27, 64])
torch.Size([31, 64])
torch.Size([28, 64])
torch.Size([30, 64])
torch.Size([30, 64])
torch.Size([31, 64])
torch.Size([29, 64])
torch.Size([22, 64])
torch.Size([26, 64])
torch.Size([32, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([31, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([33, 64])
torch.Size([29, 64])
torch.Size([22, 64])
torch.Size([23, 64])
torch.Size([35, 64])
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([24, 64])
torch.Size([29, 64])
torch.Size([29, 64])
torch.Size([23, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([23, 64])
torch.Size([26, 64])
torch.Size([22, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([38, 64])
torch.Size([30, 64])
torch.Size([32, 64])
torch.Size([37, 64])
torch.Size([30, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([23, 64])
torch.Size([28, 64])
torch.Size([30, 64])
torch.Size([23, 64])
torch.Size([23, 64])
torch.Size([24, 64])
torch.Size([31, 64])
torch.Size([27, 64])
torch.Size([34, 64])
torch.Size([26, 64])
torch.Size([45, 64])
torch.Size([27, 64])
torch.Size([30, 64])
torch.Size([37, 64])
torch.Size([24, 64])
```



```
torch.Size([27, 64])
torch.Size([22, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([23, 64])
torch.Size([31, 64])
torch.Size([24, 64])
torch.Size([33, 64])
torch.Size([29, 64])
torch.Size([30, 64])
torch.Size([30, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([32, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([33, 64])
torch.Size([31, 64])
torch.Size([23, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([35, 64])
torch.Size([25, 64])
torch.Size([30, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([33, 64])
torch.Size([23, 64])
torch.Size([29, 64])
torch.Size([25, 64])
torch.Size([29, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([30, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([29, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([30, 64])
torch.Size([29, 64])
torch.Size([37, 64])
torch.Size([30, 64])
torch.Size([27, 64])
torch.Size([21, 64])
torch.Size([29, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([45, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([38, 64])
```

```
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([30, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([21, 64])
torch.Size([33, 64])
torch.Size([24, 64])
torch.Size([29, 64])
torch.Size([34, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([22, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([21, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([23, 64])
torch.Size([23, 64])
torch.Size([28, 64])
torch.Size([31, 64])
torch.Size([29, 64])
torch.Size([29, 64])
torch.Size([25, 64])
torch.Size([38, 64])
torch.Size([30, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([33, 64])
torch.Size([23, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([38, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([30, 64])
torch.Size([28, 64])
```

```
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([31, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([23, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([31, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([21, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([29, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([29, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([33, 64])
torch.Size([22, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([34, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([37, 64])
torch.Size([28, 64])
torch.Size([21, 64])
torch.Size([24, 64])
torch.Size([32, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([36, 64])
torch.Size([29, 64])
torch.Size([30, 64])
torch.Size([27, 64])
torch.Size([30, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([33, 64])
torch.Size([25, 64])
torch.Size([30, 64])
torch.Size([30, 64])
torch.Size([34, 64])
torch.Size([26, 64])
```

```
torch.Size([30, 64])
torch.Size([23, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([32, 64])
torch.Size([29, 64])
torch.Size([37, 64])
torch.Size([22, 64])
torch.Size([28, 64])
torch.Size([34, 64])
torch.Size([28, 64])
torch.Size([23, 64])
torch.Size([30, 64])
torch.Size([28, 64])
torch.Size([34, 64])
torch.Size([24, 64])
torch.Size([30, 64])
torch.Size([24, 64])
torch.Size([30, 64])
torch.Size([23, 64])
torch.Size([28, 64])
torch.Size([32, 64])
torch.Size([22, 64])
torch.Size([28, 64])
torch.Size([23, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([33, 64])
torch.Size([33, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([32, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([23, 64])
torch.Size([30, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([36, 64])
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([32, 64])
torch.Size([27, 64])
torch.Size([22, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([36, 64])
torch.Size([31, 64])
torch.Size([22, 64])
torch.Size([33, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([26, 64])
```

```
torch.Size([30, 64])
torch.Size([27, 64])
torch.Size([36, 64])
torch.Size([32, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([16, 8])
torch.Size([32, 64])
torch.Size([22, 64])
torch.Size([26, 64])
torch.Size([29, 64])
torch.Size([24, 64])
torch.Size([31, 64])
torch.Size([31, 64])
torch.Size([20, 64])
torch.Size([33, 64])
torch.Size([27, 64])
torch.Size([36, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([34, 64])
torch.Size([23, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([38, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([30, 64])
```

Epoch: 7

Train Loss 1.48864 Val loss 3.35896:

Train Perplexity 4.431065170664168 Val Perplexity 28.75926568452678:

```
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([45, 64])
torch.Size([30, 64])
torch.Size([25, 64])
torch.Size([23, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([30, 64])
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([46, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([32, 64])
torch.Size([26, 64])
torch.Size([22, 64])
torch.Size([28, 64])
torch.Size([23, 64])
torch.Size([28, 64])
torch.Size([29, 64])
torch.Size([32, 64])
torch.Size([23, 64])
torch.Size([23, 64])
torch.Size([25, 64])
torch.Size([38, 64])
torch.Size([20, 64])
torch.Size([36, 64])
torch.Size([29, 64])
```

torch.Size([28, 64])  
torch.Size([21, 64])  
torch.Size([29, 64])  
torch.Size([25, 64])  
torch.Size([23, 64])  
torch.Size([26, 64])  
torch.Size([31, 64])  
torch.Size([27, 64])  
torch.Size([26, 64])  
torch.Size([29, 64])  
torch.Size([34, 64])  
torch.Size([28, 64])  
torch.Size([30, 64])  
torch.Size([29, 64])  
torch.Size([25, 64])  
torch.Size([30, 64])  
torch.Size([23, 64])  
torch.Size([26, 64])  
torch.Size([27, 64])  
torch.Size([24, 64])  
torch.Size([23, 64])  
torch.Size([30, 64])  
torch.Size([29, 64])  
torch.Size([33, 64])  
torch.Size([32, 64])  
torch.Size([31, 64])  
torch.Size([21, 64])  
torch.Size([31, 64])  
torch.Size([28, 64])  
torch.Size([25, 64])  
torch.Size([33, 64])  
torch.Size([28, 64])  
torch.Size([24, 64])  
torch.Size([36, 64])  
torch.Size([38, 64])  
torch.Size([24, 64])  
torch.Size([29, 64])  
torch.Size([30, 64])  
torch.Size([26, 64])  
torch.Size([23, 64])  
torch.Size([31, 64])  
torch.Size([27, 64])  
torch.Size([24, 64])  
torch.Size([29, 64])  
torch.Size([24, 64])  
torch.Size([26, 64])  
torch.Size([27, 64])  
torch.Size([30, 64])  
torch.Size([24, 64])  
torch.Size([24, 64])  
torch.Size([28, 64])  
torch.Size([31, 64])  
torch.Size([26, 64])  
torch.Size([20, 64])  
torch.Size([28, 64])  
torch.Size([28, 64])  
torch.Size([28, 64])  
torch.Size([28, 64])  
torch.Size([26, 64])  
torch.Size([34, 64])  
torch.Size([35, 64])

```
torch.Size([27, 64])
torch.Size([23, 64])
torch.Size([22, 64])
torch.Size([32, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([23, 64])
torch.Size([25, 64])
torch.Size([23, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([32, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([29, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([21, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([29, 64])
torch.Size([28, 64])
torch.Size([30, 64])
torch.Size([25, 64])
torch.Size([34, 64])
torch.Size([23, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([31, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([31, 64])
torch.Size([34, 64])
torch.Size([32, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([36, 64])
torch.Size([29, 64])
torch.Size([29, 64])
torch.Size([27, 64])
torch.Size([30, 64])
torch.Size([27, 64])
torch.Size([31, 64])
torch.Size([26, 64])
torch.Size([22, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([29, 64])
```

```
torch.Size([23, 64])
torch.Size([33, 64])
torch.Size([23, 64])
torch.Size([33, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([30, 64])
torch.Size([32, 64])
torch.Size([32, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([30, 64])
torch.Size([21, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([34, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([33, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([30, 64])
torch.Size([36, 64])
torch.Size([24, 64])
torch.Size([32, 64])
torch.Size([26, 64])
torch.Size([23, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([29, 64])
torch.Size([28, 64])
torch.Size([31, 64])
torch.Size([31, 64])
torch.Size([29, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([21, 64])
torch.Size([22, 64])
torch.Size([31, 64])
torch.Size([23, 64])
torch.Size([22, 64])
torch.Size([31, 64])
torch.Size([29, 64])
torch.Size([25, 64])
torch.Size([30, 64])
torch.Size([27, 64])
torch.Size([34, 64])
torch.Size([26, 64])
torch.Size([33, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([29, 64])
torch.Size([27, 64])
torch.Size([34, 64])
torch.Size([25, 64])
```



```
torch.Size([26, 64])
torch.Size([30, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([29, 64])
torch.Size([30, 64])
torch.Size([37, 64])
torch.Size([28, 64])
torch.Size([30, 64])
torch.Size([31, 64])
torch.Size([25, 64])
torch.Size([23, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([23, 64])
torch.Size([25, 64])
torch.Size([23, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([30, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([24, 64])
torch.Size([45, 64])
torch.Size([25, 64])
torch.Size([37, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([33, 64])
torch.Size([29, 64])
torch.Size([22, 64])
torch.Size([27, 64])
torch.Size([23, 64])
torch.Size([21, 64])
torch.Size([28, 64])
torch.Size([30, 64])
torch.Size([22, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([35, 64])
torch.Size([29, 64])
torch.Size([30, 64])
torch.Size([29, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([34, 64])
torch.Size([22, 64])
torch.Size([37, 64])
```

```
torch.Size([28, 64])
torch.Size([30, 64])
torch.Size([25, 64])
torch.Size([31, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([34, 64])
torch.Size([27, 64])
torch.Size([41, 64])
torch.Size([33, 64])
torch.Size([22, 64])
torch.Size([23, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([37, 64])
torch.Size([26, 64])
torch.Size([30, 64])
torch.Size([30, 64])
torch.Size([25, 64])
torch.Size([36, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([37, 64])
torch.Size([35, 64])
torch.Size([23, 64])
torch.Size([38, 64])
torch.Size([32, 64])
torch.Size([27, 64])
torch.Size([32, 64])
torch.Size([35, 64])
torch.Size([27, 64])
torch.Size([22, 64])
torch.Size([36, 64])
torch.Size([38, 64])
torch.Size([37, 64])
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([33, 64])
torch.Size([29, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([29, 64])
torch.Size([30, 64])
torch.Size([22, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([30, 64])
torch.Size([30, 64])
torch.Size([29, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([27, 64])
```

```
torch.Size([38, 64])
torch.Size([27, 64])
torch.Size([37, 64])
torch.Size([25, 64])
torch.Size([21, 64])
torch.Size([23, 64])
torch.Size([29, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([30, 64])
torch.Size([27, 64])
torch.Size([32, 64])
torch.Size([26, 64])
torch.Size([30, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([23, 64])
torch.Size([36, 64])
torch.Size([31, 64])
torch.Size([27, 64])
torch.Size([33, 64])
torch.Size([23, 64])
torch.Size([23, 64])
torch.Size([21, 64])
torch.Size([28, 64])
torch.Size([30, 64])
torch.Size([19, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([33, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([23, 64])
torch.Size([27, 64])
torch.Size([36, 64])
torch.Size([30, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([23, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([23, 64])
torch.Size([26, 64])
torch.Size([31, 64])
torch.Size([24, 64])
torch.Size([23, 64])
torch.Size([27, 64])
torch.Size([34, 64])
torch.Size([23, 64])
torch.Size([31, 64])
torch.Size([23, 64])
torch.Size([26, 64])
torch.Size([26, 64])
```

```
torch.Size([23, 64])
torch.Size([35, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([33, 64])
torch.Size([32, 64])
torch.Size([33, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([30, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([21, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([15, 8])
torch.Size([28, 64])
torch.Size([29, 64])
torch.Size([26, 64])
torch.Size([32, 64])
torch.Size([23, 64])
torch.Size([23, 64])
torch.Size([33, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([31, 64])
torch.Size([33, 64])
torch.Size([31, 64])
torch.Size([30, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([22, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([23, 64])
torch.Size([28, 64])
torch.Size([32, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([29, 64])
torch.Size([38, 64])
torch.Size([30, 64])
torch.Size([30, 64])
torch.Size([33, 64])
torch.Size([29, 64])
Epoch: 8
Train Loss 1.35507 Val loss 3.53443:
```

Train Perplexity 3.8770323393984296 Val Perplexity 34.275472126008495:

torch.Size([25, 64])  
torch.Size([37, 64])  
torch.Size([29, 64])  
torch.Size([38, 64])  
torch.Size([28, 64])  
torch.Size([24, 64])  
torch.Size([29, 64])  
torch.Size([26, 64])  
torch.Size([31, 64])  
torch.Size([29, 64])  
torch.Size([28, 64])  
torch.Size([24, 64])  
torch.Size([23, 64])  
torch.Size([37, 64])  
torch.Size([27, 64])  
torch.Size([32, 64])  
torch.Size([25, 64])  
torch.Size([27, 64])  
torch.Size([29, 64])  
torch.Size([22, 64])  
torch.Size([25, 64])  
torch.Size([26, 64])  
torch.Size([27, 64])  
torch.Size([34, 64])  
torch.Size([27, 64])  
torch.Size([27, 64])  
torch.Size([28, 64])  
torch.Size([27, 64])  
torch.Size([28, 64])  
torch.Size([30, 64])  
torch.Size([27, 64])  
torch.Size([26, 64])  
torch.Size([24, 64])  
torch.Size([27, 64])  
torch.Size([28, 64])  
torch.Size([29, 64])  
torch.Size([28, 64])  
torch.Size([28, 64])  
torch.Size([38, 64])  
torch.Size([22, 64])  
torch.Size([28, 64])  
torch.Size([33, 64])  
torch.Size([23, 64])  
torch.Size([22, 64])  
torch.Size([27, 64])  
torch.Size([27, 64])  
torch.Size([24, 64])  
torch.Size([29, 64])  
torch.Size([28, 64])  
torch.Size([31, 64])  
torch.Size([30, 64])  
torch.Size([29, 64])  
torch.Size([25, 64])  
torch.Size([30, 64])  
torch.Size([33, 64])  
torch.Size([28, 64])  
torch.Size([23, 64])  
torch.Size([22, 64])  
torch.Size([24, 64])  
torch.Size([30, 64])

```
torch.Size([23, 64])
torch.Size([21, 64])
torch.Size([32, 64])
torch.Size([27, 64])
torch.Size([21, 64])
torch.Size([30, 64])
torch.Size([29, 64])
torch.Size([30, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([31, 64])
torch.Size([29, 64])
torch.Size([28, 64])
torch.Size([23, 64])
torch.Size([22, 64])
torch.Size([25, 64])
torch.Size([22, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([23, 64])
torch.Size([37, 64])
torch.Size([29, 64])
torch.Size([25, 64])
torch.Size([23, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([36, 64])
torch.Size([34, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([41, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([33, 64])
torch.Size([28, 64])
torch.Size([29, 64])
torch.Size([31, 64])
torch.Size([30, 64])
torch.Size([31, 64])
torch.Size([32, 64])
torch.Size([25, 64])
torch.Size([35, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([38, 64])
torch.Size([32, 64])
torch.Size([29, 64])
torch.Size([23, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([23, 64])
```

```
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([27, 64])
torch.Size([32, 64])
torch.Size([30, 64])
torch.Size([36, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([32, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([31, 64])
torch.Size([31, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([29, 64])
torch.Size([21, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([23, 64])
torch.Size([25, 64])
torch.Size([30, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([34, 64])
torch.Size([30, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([22, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([22, 64])
torch.Size([28, 64])
torch.Size([38, 64])
torch.Size([30, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([34, 64])
torch.Size([25, 64])
torch.Size([30, 64])
torch.Size([46, 64])
torch.Size([35, 64])
torch.Size([30, 64])
torch.Size([25, 64])
torch.Size([31, 64])
```

torch.Size([35, 64])  
torch.Size([31, 64])  
torch.Size([30, 64])  
torch.Size([25, 64])  
torch.Size([34, 64])  
torch.Size([30, 64])  
torch.Size([23, 64])  
torch.Size([27, 64])  
torch.Size([28, 64])  
torch.Size([36, 64])  
torch.Size([26, 64])  
torch.Size([24, 64])  
torch.Size([33, 64])  
torch.Size([25, 64])  
torch.Size([33, 64])  
torch.Size([28, 64])  
torch.Size([24, 64])  
torch.Size([24, 64])  
torch.Size([28, 64])  
torch.Size([24, 64])  
torch.Size([25, 64])  
torch.Size([36, 64])  
torch.Size([28, 64])  
torch.Size([28, 64])  
torch.Size([28, 64])  
torch.Size([31, 64])  
torch.Size([28, 64])  
torch.Size([25, 64])  
torch.Size([25, 64])  
torch.Size([32, 64])  
torch.Size([29, 64])  
torch.Size([27, 64])  
torch.Size([23, 64])  
torch.Size([30, 64])  
torch.Size([21, 64])  
torch.Size([28, 64])  
torch.Size([30, 64])  
torch.Size([25, 64])  
torch.Size([38, 64])  
torch.Size([25, 64])  
torch.Size([27, 64])  
torch.Size([24, 64])  
torch.Size([22, 64])  
torch.Size([31, 64])  
torch.Size([26, 64])  
torch.Size([27, 64])  
torch.Size([29, 64])  
torch.Size([25, 64])  
torch.Size([34, 64])  
torch.Size([28, 64])  
torch.Size([28, 64])  
torch.Size([27, 64])  
torch.Size([25, 64])  
torch.Size([26, 64])  
torch.Size([28, 64])  
torch.Size([27, 64])  
torch.Size([32, 64])  
torch.Size([29, 64])  
torch.Size([36, 64])  
torch.Size([27, 64])  
torch.Size([34, 64])



```
torch.Size([29, 64])
torch.Size([32, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([31, 64])
torch.Size([33, 64])
torch.Size([24, 64])
torch.Size([36, 64])
torch.Size([29, 64])
torch.Size([34, 64])
torch.Size([30, 64])
torch.Size([21, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([31, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([29, 64])
torch.Size([23, 64])
torch.Size([26, 64])
torch.Size([38, 64])
torch.Size([31, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([29, 64])
torch.Size([25, 64])
torch.Size([32, 64])
torch.Size([33, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([23, 64])
torch.Size([27, 64])
torch.Size([45, 64])
torch.Size([30, 64])
torch.Size([30, 64])
torch.Size([26, 64])
torch.Size([23, 64])
torch.Size([32, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([30, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([30, 64])
torch.Size([25, 64])
torch.Size([36, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([35, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([26, 64])
```

torch.Size([24, 64])  
torch.Size([26, 64])  
torch.Size([31, 64])  
torch.Size([27, 64])  
torch.Size([37, 64])  
torch.Size([26, 64])  
torch.Size([30, 64])  
torch.Size([23, 64])  
torch.Size([30, 64])  
torch.Size([26, 64])  
torch.Size([37, 64])  
torch.Size([28, 64])  
torch.Size([29, 64])  
torch.Size([32, 64])  
torch.Size([30, 64])  
torch.Size([27, 64])  
torch.Size([33, 64])  
torch.Size([28, 64])  
torch.Size([27, 64])  
torch.Size([32, 64])  
torch.Size([29, 64])  
torch.Size([27, 64])  
torch.Size([23, 64])  
torch.Size([24, 64])  
torch.Size([37, 64])  
torch.Size([23, 64])  
torch.Size([25, 64])  
torch.Size([32, 64])  
torch.Size([34, 64])  
torch.Size([28, 64])  
torch.Size([26, 64])  
torch.Size([30, 64])  
torch.Size([25, 64])  
torch.Size([28, 64])  
torch.Size([29, 64])  
torch.Size([25, 64])  
torch.Size([24, 64])  
torch.Size([30, 64])  
torch.Size([26, 64])  
torch.Size([35, 64])  
torch.Size([34, 64])  
torch.Size([29, 64])  
torch.Size([27, 64])  
torch.Size([31, 64])  
torch.Size([29, 64])  
torch.Size([30, 64])  
torch.Size([25, 64])  
torch.Size([31, 64])  
torch.Size([24, 64])  
torch.Size([31, 64])  
torch.Size([28, 64])  
torch.Size([29, 64])  
torch.Size([28, 64])  
torch.Size([33, 64])  
torch.Size([31, 64])  
torch.Size([22, 64])  
torch.Size([27, 64])  
torch.Size([24, 64])  
torch.Size([27, 64])  
torch.Size([25, 64])  
torch.Size([25, 64])

```
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([30, 64])
torch.Size([25, 64])
torch.Size([21, 64])
torch.Size([29, 64])
torch.Size([30, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([29, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([33, 64])
torch.Size([28, 64])
torch.Size([33, 64])
torch.Size([27, 64])
torch.Size([30, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([30, 64])
torch.Size([31, 64])
torch.Size([30, 64])
torch.Size([45, 64])
torch.Size([33, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([29, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([23, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([23, 64])
torch.Size([25, 64])
torch.Size([33, 64])
torch.Size([22, 64])
torch.Size([25, 8])
torch.Size([27, 64])
torch.Size([30, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([24, 64])
torch.Size([33, 64])
torch.Size([27, 64])
```

```
torch.Size([25, 64])
torch.Size([34, 64])
torch.Size([29, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([30, 64])
torch.Size([25, 64])
torch.Size([32, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([32, 64])
torch.Size([25, 64])
torch.Size([31, 64])
torch.Size([23, 64])
torch.Size([23, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([36, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([22, 64])
torch.Size([24, 64])
```

Epoch: 9

Train Loss 1.25213 Val loss 3.50001:

Train Perplexity 3.497785311254256 Val Perplexity 33.11578311486768:

```
torch.Size([25, 64])
torch.Size([29, 64])
torch.Size([29, 64])
torch.Size([27, 64])
torch.Size([22, 64])
torch.Size([30, 64])
torch.Size([29, 64])
torch.Size([27, 64])
torch.Size([30, 64])
torch.Size([30, 64])
torch.Size([23, 64])
torch.Size([29, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([29, 64])
torch.Size([29, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([31, 64])
torch.Size([23, 64])
torch.Size([28, 64])
torch.Size([34, 64])
torch.Size([24, 64])
torch.Size([23, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([33, 64])
torch.Size([35, 64])
torch.Size([24, 64])
```

```
torch.Size([28, 64])
torch.Size([23, 64])
torch.Size([23, 64])
torch.Size([27, 64])
torch.Size([21, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([21, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([33, 64])
torch.Size([33, 64])
torch.Size([30, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([30, 64])
torch.Size([28, 64])
torch.Size([31, 64])
torch.Size([24, 64])
torch.Size([37, 64])
torch.Size([25, 64])
torch.Size([46, 64])
torch.Size([27, 64])
torch.Size([32, 64])
torch.Size([29, 64])
torch.Size([29, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([32, 64])
torch.Size([20, 64])
torch.Size([32, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([35, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([31, 64])
torch.Size([25, 64])
torch.Size([33, 64])
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([38, 64])
torch.Size([26, 64])
torch.Size([32, 64])
torch.Size([28, 64])
```

```
torch.Size([32, 64])
torch.Size([36, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([29, 64])
torch.Size([22, 64])
torch.Size([27, 64])
torch.Size([33, 64])
torch.Size([33, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([29, 64])
torch.Size([29, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([23, 64])
torch.Size([23, 64])
torch.Size([38, 64])
torch.Size([30, 64])
torch.Size([32, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([36, 64])
torch.Size([30, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([19, 64])
torch.Size([25, 64])
torch.Size([29, 64])
torch.Size([30, 64])
torch.Size([29, 64])
torch.Size([27, 64])
torch.Size([30, 64])
torch.Size([25, 64])
torch.Size([23, 64])
torch.Size([20, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([30, 64])
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([29, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([34, 64])
torch.Size([27, 64])
torch.Size([31, 64])
torch.Size([36, 64])
torch.Size([37, 64])
torch.Size([27, 64])
torch.Size([34, 64])
```

```
torch.Size([30, 64])
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([22, 64])
torch.Size([23, 64])
torch.Size([37, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([35, 64])
torch.Size([30, 64])
torch.Size([45, 64])
torch.Size([27, 64])
torch.Size([22, 64])
torch.Size([28, 64])
torch.Size([33, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([23, 64])
torch.Size([32, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([36, 64])
torch.Size([26, 64])
torch.Size([30, 64])
torch.Size([30, 64])
torch.Size([29, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([23, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([22, 64])
torch.Size([35, 64])
torch.Size([24, 64])
torch.Size([22, 64])
torch.Size([22, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([29, 64])
torch.Size([28, 64])
torch.Size([37, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([31, 64])
torch.Size([28, 64])
torch.Size([30, 64])
torch.Size([34, 64])
torch.Size([28, 64])
torch.Size([23, 64])
torch.Size([21, 64])
torch.Size([24, 64])
torch.Size([29, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([25, 64])
```

```
torch.Size([28, 64])
torch.Size([29, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([30, 64])
torch.Size([23, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([31, 64])
torch.Size([22, 64])
torch.Size([29, 64])
torch.Size([30, 64])
torch.Size([24, 64])
torch.Size([22, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([30, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([38, 64])
torch.Size([20, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([23, 64])
torch.Size([26, 64])
torch.Size([36, 64])
torch.Size([30, 64])
torch.Size([25, 64])
torch.Size([30, 64])
torch.Size([29, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([23, 64])
torch.Size([27, 64])
torch.Size([38, 64])
torch.Size([33, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([32, 64])
torch.Size([23, 64])
torch.Size([28, 64])
torch.Size([33, 64])
torch.Size([25, 64])
torch.Size([37, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([23, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([30, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([29, 64])
torch.Size([27, 64])
torch.Size([26, 64])
```



```
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([32, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([26, 64])
torch.Size([34, 64])
torch.Size([24, 64])
torch.Size([33, 64])
torch.Size([30, 64])
torch.Size([28, 64])
torch.Size([30, 64])
torch.Size([22, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([23, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([23, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([22, 64])
torch.Size([27, 64])
torch.Size([33, 64])
torch.Size([23, 64])
torch.Size([25, 64])
torch.Size([29, 64])
torch.Size([31, 64])
torch.Size([29, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([21, 64])
torch.Size([19, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([33, 64])
torch.Size([32, 64])
torch.Size([25, 64])
torch.Size([23, 64])
torch.Size([33, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([22, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([31, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([31, 64])
```

```
torch.Size([30, 64])
torch.Size([33, 64])
torch.Size([26, 64])
torch.Size([30, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([45, 64])
torch.Size([38, 64])
torch.Size([27, 64])
torch.Size([31, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([29, 64])
torch.Size([29, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([38, 64])
torch.Size([31, 64])
torch.Size([31, 64])
torch.Size([21, 64])
torch.Size([26, 64])
torch.Size([34, 64])
torch.Size([31, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([22, 64])
torch.Size([37, 64])
torch.Size([33, 64])
torch.Size([37, 64])
torch.Size([27, 64])
torch.Size([34, 64])
torch.Size([28, 64])
torch.Size([34, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([23, 64])
torch.Size([29, 64])
torch.Size([23, 64])
torch.Size([21, 64])
torch.Size([34, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([31, 64])
torch.Size([25, 64])
torch.Size([29, 64])
torch.Size([34, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([35, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([29, 64])
torch.Size([31, 64])
torch.Size([36, 64])
torch.Size([32, 64])
torch.Size([30, 64])
```

```
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([30, 64])
torch.Size([30, 64])
torch.Size([41, 64])
torch.Size([26, 64])
torch.Size([31, 64])
torch.Size([32, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([23, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([30, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([29, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([22, 64])
torch.Size([26, 64])
torch.Size([32, 64])
torch.Size([26, 64])
torch.Size([36, 64])
torch.Size([29, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([23, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([31, 64])
torch.Size([36, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([32, 64])
torch.Size([29, 64])
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([30, 64])
torch.Size([23, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([32, 64])
torch.Size([34, 64])
torch.Size([28, 64])
torch.Size([23, 8])
torch.Size([28, 64])
torch.Size([25, 64])
Epoch: 10
Train Loss 1.11702 Val loss 3.84965:
Train Perplexity 3.055734532825045 Val Perplexity 46.976618537437574:
```

```
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([21, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([37, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([30, 64])
torch.Size([30, 64])
torch.Size([25, 64])
torch.Size([29, 64])
torch.Size([33, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([29, 64])
torch.Size([22, 64])
torch.Size([22, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([21, 64])
torch.Size([28, 64])
torch.Size([23, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([33, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([30, 64])
torch.Size([26, 64])
torch.Size([30, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([23, 64])
torch.Size([24, 64])
torch.Size([33, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([30, 64])
torch.Size([28, 64])
torch.Size([31, 64])
torch.Size([29, 64])
torch.Size([27, 64])
torch.Size([30, 64])
torch.Size([31, 64])
torch.Size([33, 64])
torch.Size([34, 64])
torch.Size([27, 64])
```

```
torch.Size([28, 64])
torch.Size([29, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([23, 64])
torch.Size([26, 64])
torch.Size([29, 64])
torch.Size([37, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([31, 64])
torch.Size([29, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([23, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([33, 64])
torch.Size([33, 64])
torch.Size([27, 64])
torch.Size([32, 64])
torch.Size([32, 64])
torch.Size([28, 64])
torch.Size([32, 64])
torch.Size([23, 64])
torch.Size([37, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([31, 64])
torch.Size([22, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([23, 64])
torch.Size([32, 64])
torch.Size([28, 64])
torch.Size([23, 64])
torch.Size([31, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([22, 64])
torch.Size([26, 64])
torch.Size([30, 64])
torch.Size([25, 64])
torch.Size([29, 64])
torch.Size([29, 64])
torch.Size([33, 64])
```

```
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([29, 64])
torch.Size([33, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([31, 64])
torch.Size([24, 64])
torch.Size([36, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([29, 64])
torch.Size([27, 64])
torch.Size([31, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([23, 64])
torch.Size([41, 64])
torch.Size([31, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([36, 64])
torch.Size([28, 64])
torch.Size([21, 64])
torch.Size([36, 64])
torch.Size([30, 64])
torch.Size([30, 64])
torch.Size([32, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([23, 64])
torch.Size([35, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([33, 64])
torch.Size([38, 64])
torch.Size([29, 64])
torch.Size([26, 64])
torch.Size([37, 64])
torch.Size([26, 64])
torch.Size([29, 64])
torch.Size([23, 64])
torch.Size([19, 64])
torch.Size([25, 64])
torch.Size([38, 64])
torch.Size([26, 64])
torch.Size([23, 64])
torch.Size([30, 64])
torch.Size([28, 64])
```

```
torch.Size([33, 64])
torch.Size([26, 64])
torch.Size([22, 64])
torch.Size([28, 64])
torch.Size([21, 64])
torch.Size([26, 64])
torch.Size([34, 64])
torch.Size([23, 64])
torch.Size([30, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([35, 64])
torch.Size([35, 64])
torch.Size([26, 64])
torch.Size([23, 64])
torch.Size([31, 64])
torch.Size([29, 64])
torch.Size([35, 64])
torch.Size([31, 64])
torch.Size([29, 64])
torch.Size([23, 64])
torch.Size([26, 64])
torch.Size([33, 64])
torch.Size([26, 64])
torch.Size([32, 64])
torch.Size([34, 64])
torch.Size([23, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([22, 64])
torch.Size([26, 64])
torch.Size([23, 64])
torch.Size([26, 64])
torch.Size([29, 64])
torch.Size([36, 64])
torch.Size([23, 64])
torch.Size([23, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([36, 64])
torch.Size([22, 64])
torch.Size([24, 64])
torch.Size([30, 64])
torch.Size([30, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([30, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([22, 64])
torch.Size([29, 64])
torch.Size([27, 64])
```

```
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([30, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([23, 64])
torch.Size([33, 64])
torch.Size([24, 64])
torch.Size([33, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([29, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([30, 64])
torch.Size([23, 64])
torch.Size([23, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([34, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([37, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([29, 64])
torch.Size([38, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([45, 64])
torch.Size([25, 64])
torch.Size([23, 64])
torch.Size([27, 64])
torch.Size([32, 64])
torch.Size([25, 64])
torch.Size([35, 64])
torch.Size([26, 64])
torch.Size([32, 64])
torch.Size([29, 64])
torch.Size([25, 64])
torch.Size([29, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([27, 64])
```



```
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([25, 64])
torch.Size([34, 64])
torch.Size([27, 64])
torch.Size([30, 64])
torch.Size([31, 64])
torch.Size([32, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([38, 64])
torch.Size([30, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([30, 64])
torch.Size([26, 64])
torch.Size([37, 64])
torch.Size([23, 64])
torch.Size([23, 64])
torch.Size([25, 64])
torch.Size([46, 64])
torch.Size([22, 64])
torch.Size([26, 64])
torch.Size([30, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([36, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([29, 64])
torch.Size([27, 64])
torch.Size([23, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([29, 64])
torch.Size([31, 64])
torch.Size([23, 64])
torch.Size([34, 64])
torch.Size([34, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([29, 64])
torch.Size([32, 64])
torch.Size([30, 64])
torch.Size([31, 64])
torch.Size([25, 64])
torch.Size([34, 64])
torch.Size([38, 64])
torch.Size([27, 64])
```

```
torch.Size([37, 64])
torch.Size([24, 64])
torch.Size([21, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([34, 64])
torch.Size([30, 64])
torch.Size([31, 64])
torch.Size([29, 64])
torch.Size([31, 64])
torch.Size([21, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([29, 64])
torch.Size([30, 64])
torch.Size([29, 64])
torch.Size([28, 64])
torch.Size([30, 64])
torch.Size([21, 64])
torch.Size([33, 64])
torch.Size([30, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([29, 64])
torch.Size([32, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([23, 64])
torch.Size([27, 64])
torch.Size([30, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([32, 64])
torch.Size([25, 64])
torch.Size([34, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([29, 64])
torch.Size([26, 64])
torch.Size([23, 64])
torch.Size([34, 64])
torch.Size([30, 64])
torch.Size([27, 64])
torch.Size([31, 64])
torch.Size([25, 64])
torch.Size([21, 64])
torch.Size([29, 64])
torch.Size([22, 64])
torch.Size([30, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([31, 64])
torch.Size([30, 64])
torch.Size([30, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([32, 64])
```

```
torch.Size([26, 64])
torch.Size([18, 8])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([36, 64])
torch.Size([38, 64])
torch.Size([22, 64])
torch.Size([22, 64])
torch.Size([22, 64])
torch.Size([27, 64])
torch.Size([23, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([21, 64])
torch.Size([29, 64])
torch.Size([23, 64])
torch.Size([32, 64])
torch.Size([32, 64])
torch.Size([25, 64])
torch.Size([29, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([27, 64])
Epoch: 11
  Train Loss 1.0196  Val loss 3.94222:
Train Perplexity 2.772085707884716  Val Perplexity 51.53287739494569:
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([33, 64])
torch.Size([30, 64])
torch.Size([27, 64])
torch.Size([32, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([23, 64])
torch.Size([30, 64])
torch.Size([30, 64])
torch.Size([29, 64])
torch.Size([26, 64])
torch.Size([36, 64])
torch.Size([30, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([30, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([33, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([34, 64])
torch.Size([41, 64])
```

torch.Size([25, 64])  
torch.Size([23, 64])  
torch.Size([37, 64])  
torch.Size([23, 64])  
torch.Size([25, 64])  
torch.Size([34, 64])  
torch.Size([32, 64])  
torch.Size([30, 64])  
torch.Size([36, 64])  
torch.Size([28, 64])  
torch.Size([26, 64])  
torch.Size([28, 64])  
torch.Size([28, 64])  
torch.Size([26, 64])  
torch.Size([24, 64])  
torch.Size([28, 64])  
torch.Size([23, 64])  
torch.Size([24, 64])  
torch.Size([28, 64])  
torch.Size([29, 64])  
torch.Size([29, 64])  
torch.Size([21, 64])  
torch.Size([33, 64])  
torch.Size([37, 64])  
torch.Size([24, 64])  
torch.Size([22, 64])  
torch.Size([24, 64])  
torch.Size([29, 64])  
torch.Size([30, 64])  
torch.Size([33, 64])  
torch.Size([29, 64])  
torch.Size([29, 64])  
torch.Size([26, 64])  
torch.Size([30, 64])  
torch.Size([28, 64])  
torch.Size([26, 64])  
torch.Size([30, 64])  
torch.Size([22, 64])  
torch.Size([24, 64])  
torch.Size([26, 64])  
torch.Size([26, 64])  
torch.Size([27, 64])  
torch.Size([26, 64])  
torch.Size([30, 64])  
torch.Size([34, 64])  
torch.Size([31, 64])  
torch.Size([25, 64])  
torch.Size([30, 64])  
torch.Size([34, 64])  
torch.Size([27, 64])  
torch.Size([32, 64])  
torch.Size([24, 64])  
torch.Size([29, 64])  
torch.Size([31, 64])  
torch.Size([28, 64])  
torch.Size([25, 64])  
torch.Size([30, 64])  
torch.Size([28, 64])  
torch.Size([36, 64])  
torch.Size([27, 64])  
torch.Size([25, 64])

```
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([21, 64])
torch.Size([23, 64])
torch.Size([30, 64])
torch.Size([29, 64])
torch.Size([28, 64])
torch.Size([32, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([33, 64])
torch.Size([23, 64])
torch.Size([22, 64])
torch.Size([45, 64])
torch.Size([21, 64])
torch.Size([25, 64])
torch.Size([29, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([38, 64])
torch.Size([27, 64])
torch.Size([32, 64])
torch.Size([33, 64])
torch.Size([21, 64])
torch.Size([31, 64])
torch.Size([29, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([37, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([30, 64])
torch.Size([25, 64])
torch.Size([30, 64])
torch.Size([30, 64])
torch.Size([28, 64])
torch.Size([33, 64])
torch.Size([30, 64])
torch.Size([30, 64])
torch.Size([30, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([22, 64])
torch.Size([28, 64])
torch.Size([32, 64])
torch.Size([38, 64])
torch.Size([31, 64])
torch.Size([21, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([23, 64])
torch.Size([32, 64])
torch.Size([26, 64])
torch.Size([24, 64])
```

```
torch.Size([30, 64])
torch.Size([26, 64])
torch.Size([32, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([29, 64])
torch.Size([29, 64])
torch.Size([23, 64])
torch.Size([23, 64])
torch.Size([34, 64])
torch.Size([23, 64])
torch.Size([32, 64])
torch.Size([30, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([29, 64])
torch.Size([23, 64])
torch.Size([24, 64])
torch.Size([20, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([29, 64])
torch.Size([23, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([29, 64])
torch.Size([35, 64])
torch.Size([31, 64])
torch.Size([29, 64])
torch.Size([26, 64])
torch.Size([22, 64])
torch.Size([36, 64])
torch.Size([25, 64])
torch.Size([29, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([30, 64])
torch.Size([24, 64])
torch.Size([33, 64])
torch.Size([27, 64])
torch.Size([23, 64])
torch.Size([29, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([38, 64])
torch.Size([33, 64])
torch.Size([20, 64])
torch.Size([27, 64])
torch.Size([23, 64])
torch.Size([33, 64])
torch.Size([23, 64])
```

torch.Size([25, 64])  
torch.Size([33, 64])  
torch.Size([27, 64])  
torch.Size([26, 64])  
torch.Size([37, 64])  
torch.Size([23, 64])  
torch.Size([28, 64])  
torch.Size([27, 64])  
torch.Size([24, 64])  
torch.Size([26, 64])  
torch.Size([23, 64])  
torch.Size([29, 64])  
torch.Size([24, 64])  
torch.Size([34, 64])  
torch.Size([31, 64])  
torch.Size([26, 64])  
torch.Size([23, 64])  
torch.Size([28, 64])  
torch.Size([31, 64])  
torch.Size([31, 64])  
torch.Size([25, 64])  
torch.Size([28, 64])  
torch.Size([24, 64])  
torch.Size([26, 64])  
torch.Size([28, 64])  
torch.Size([25, 64])  
torch.Size([26, 64])  
torch.Size([24, 64])  
torch.Size([24, 64])  
torch.Size([27, 64])  
torch.Size([31, 64])  
torch.Size([28, 64])  
torch.Size([25, 64])  
torch.Size([30, 64])  
torch.Size([29, 64])  
torch.Size([26, 64])  
torch.Size([27, 64])  
torch.Size([28, 64])  
torch.Size([26, 64])  
torch.Size([24, 64])  
torch.Size([37, 64])  
torch.Size([25, 64])  
torch.Size([24, 64])  
torch.Size([25, 64])  
torch.Size([22, 64])  
torch.Size([29, 64])  
torch.Size([26, 64])  
torch.Size([22, 64])  
torch.Size([28, 64])  
torch.Size([24, 64])  
torch.Size([26, 64])  
torch.Size([24, 64])  
torch.Size([24, 64])  
torch.Size([30, 64])  
torch.Size([29, 64])  
torch.Size([24, 64])  
torch.Size([26, 64])  
torch.Size([35, 64])  
torch.Size([31, 64])  
torch.Size([30, 64])  
torch.Size([26, 64])

```
torch.Size([25, 64])
torch.Size([22, 64])
torch.Size([25, 64])
torch.Size([29, 64])
torch.Size([37, 64])
torch.Size([28, 64])
torch.Size([29, 64])
torch.Size([28, 64])
torch.Size([32, 64])
torch.Size([29, 64])
torch.Size([29, 64])
torch.Size([34, 64])
torch.Size([28, 64])
torch.Size([22, 64])
torch.Size([30, 64])
torch.Size([30, 64])
torch.Size([46, 64])
torch.Size([29, 64])
torch.Size([30, 64])
torch.Size([32, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([38, 64])
torch.Size([24, 64])
torch.Size([36, 64])
torch.Size([29, 64])
torch.Size([37, 64])
torch.Size([24, 64])
torch.Size([31, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([34, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([32, 64])
torch.Size([30, 64])
torch.Size([36, 64])
torch.Size([32, 64])
torch.Size([30, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([29, 64])
torch.Size([29, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([32, 64])
torch.Size([23, 64])
torch.Size([31, 64])
torch.Size([24, 64])
torch.Size([22, 64])
torch.Size([27, 64])
torch.Size([23, 64])
torch.Size([23, 64])
torch.Size([33, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([22, 64])
```



```
torch.Size([45, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([22, 64])
torch.Size([38, 64])
torch.Size([25, 64])
torch.Size([29, 64])
torch.Size([24, 64])
torch.Size([34, 64])
torch.Size([31, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([30, 64])
torch.Size([30, 64])
torch.Size([27, 64])
torch.Size([33, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([31, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([29, 64])
torch.Size([31, 64])
torch.Size([28, 64])
torch.Size([35, 64])
torch.Size([27, 64])
torch.Size([30, 64])
torch.Size([29, 64])
torch.Size([23, 64])
torch.Size([30, 64])
torch.Size([35, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([36, 64])
torch.Size([38, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([22, 64])
torch.Size([23, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([32, 64])
torch.Size([24, 64])
torch.Size([21, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([30, 64])
```

```
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([31, 64])
torch.Size([33, 64])
torch.Size([18, 8])
torch.Size([34, 64])
torch.Size([31, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([36, 64])
torch.Size([22, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([31, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([23, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([31, 64])
torch.Size([22, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([31, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([23, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([30, 64])
torch.Size([30, 64])
torch.Size([29, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([26, 64])
Epoch: 12
Train Loss 0.92591 Val loss 4.00224:
Train Perplexity 2.5241642049714335 Val Perplexity 54.720586967389856:
torch.Size([31, 64])
```

```
torch.Size([28, 64])
torch.Size([32, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([30, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([23, 64])
torch.Size([24, 64])
torch.Size([33, 64])
torch.Size([33, 64])
torch.Size([31, 64])
torch.Size([36, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([22, 64])
torch.Size([36, 64])
torch.Size([30, 64])
torch.Size([30, 64])
torch.Size([27, 64])
torch.Size([30, 64])
torch.Size([31, 64])
torch.Size([27, 64])
torch.Size([46, 64])
torch.Size([26, 64])
torch.Size([29, 64])
torch.Size([36, 64])
torch.Size([37, 64])
torch.Size([25, 64])
torch.Size([29, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([29, 64])
torch.Size([23, 64])
torch.Size([36, 64])
torch.Size([27, 64])
torch.Size([30, 64])
torch.Size([34, 64])
torch.Size([36, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([33, 64])
torch.Size([28, 64])
torch.Size([34, 64])
torch.Size([23, 64])
torch.Size([31, 64])
torch.Size([22, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([37, 64])
```

```
torch.Size([30, 64])
torch.Size([29, 64])
torch.Size([29, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([30, 64])
torch.Size([33, 64])
torch.Size([30, 64])
torch.Size([30, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([34, 64])
torch.Size([24, 64])
torch.Size([38, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([29, 64])
torch.Size([25, 64])
torch.Size([29, 64])
torch.Size([28, 64])
torch.Size([31, 64])
torch.Size([37, 64])
torch.Size([29, 64])
torch.Size([32, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([30, 64])
torch.Size([21, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([33, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([23, 64])
torch.Size([24, 64])
torch.Size([23, 64])
torch.Size([29, 64])
torch.Size([29, 64])
torch.Size([35, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([28, 64])
```

```
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([29, 64])
torch.Size([31, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([29, 64])
torch.Size([29, 64])
torch.Size([22, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([22, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([23, 64])
torch.Size([31, 64])
torch.Size([22, 64])
torch.Size([30, 64])
torch.Size([26, 64])
torch.Size([31, 64])
torch.Size([37, 64])
torch.Size([29, 64])
torch.Size([32, 64])
torch.Size([34, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([29, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([30, 64])
torch.Size([26, 64])
torch.Size([34, 64])
torch.Size([21, 64])
torch.Size([23, 64])
torch.Size([30, 64])
torch.Size([31, 64])
torch.Size([22, 64])
torch.Size([38, 64])
torch.Size([24, 64])
torch.Size([29, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([23, 64])
torch.Size([33, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([22, 64])
torch.Size([33, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([33, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([27, 64])
```

```
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([36, 64])
torch.Size([29, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([21, 64])
torch.Size([35, 64])
torch.Size([30, 64])
torch.Size([26, 64])
torch.Size([30, 64])
torch.Size([30, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([30, 64])
torch.Size([29, 64])
torch.Size([23, 64])
torch.Size([30, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([30, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([22, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([31, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([31, 64])
torch.Size([22, 64])
torch.Size([32, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([37, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([23, 64])
torch.Size([33, 64])
torch.Size([29, 64])
torch.Size([23, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([33, 64])
```

```
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([29, 64])
torch.Size([32, 64])
torch.Size([31, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([30, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([23, 64])
torch.Size([25, 64])
torch.Size([45, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([23, 64])
torch.Size([32, 64])
torch.Size([33, 64])
torch.Size([29, 64])
torch.Size([25, 64])
torch.Size([29, 64])
torch.Size([30, 64])
torch.Size([29, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([23, 64])
torch.Size([30, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([38, 64])
torch.Size([36, 64])
torch.Size([34, 64])
torch.Size([21, 64])
torch.Size([25, 64])
torch.Size([34, 64])
torch.Size([29, 64])
torch.Size([22, 64])
torch.Size([27, 64])
torch.Size([23, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([30, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([41, 64])
torch.Size([22, 64])
torch.Size([22, 64])
torch.Size([28, 64])
torch.Size([22, 64])
```

```
torch.Size([24, 64])
torch.Size([31, 64])
torch.Size([31, 64])
torch.Size([38, 64])
torch.Size([26, 64])
torch.Size([29, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([30, 64])
torch.Size([24, 64])
torch.Size([29, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([23, 64])
torch.Size([30, 64])
torch.Size([33, 64])
torch.Size([37, 64])
torch.Size([22, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([35, 64])
torch.Size([34, 64])
torch.Size([29, 64])
torch.Size([28, 64])
torch.Size([37, 64])
torch.Size([33, 64])
torch.Size([28, 64])
torch.Size([21, 64])
torch.Size([30, 64])
torch.Size([30, 64])
torch.Size([25, 64])
torch.Size([33, 64])
torch.Size([28, 64])
torch.Size([32, 64])
torch.Size([24, 64])
torch.Size([29, 64])
torch.Size([38, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([30, 64])
torch.Size([32, 64])
torch.Size([23, 64])
```



torch.Size([27, 64])  
torch.Size([27, 64])  
torch.Size([26, 64])  
torch.Size([27, 64])  
torch.Size([25, 64])  
torch.Size([32, 64])  
torch.Size([33, 64])  
torch.Size([31, 64])  
torch.Size([25, 64])  
torch.Size([30, 64])  
torch.Size([26, 64])  
torch.Size([29, 64])  
torch.Size([27, 64])  
torch.Size([24, 64])  
torch.Size([28, 64])  
torch.Size([30, 64])  
torch.Size([24, 64])  
torch.Size([30, 64])  
torch.Size([26, 64])  
torch.Size([30, 64])  
torch.Size([24, 64])  
torch.Size([24, 64])  
torch.Size([25, 64])  
torch.Size([25, 64])  
torch.Size([30, 64])  
torch.Size([27, 64])  
torch.Size([28, 64])  
torch.Size([33, 64])  
torch.Size([27, 64])  
torch.Size([27, 64])  
torch.Size([24, 64])  
torch.Size([34, 64])  
torch.Size([27, 64])  
torch.Size([26, 64])  
torch.Size([27, 64])  
torch.Size([33, 64])  
torch.Size([28, 64])  
torch.Size([25, 64])  
torch.Size([27, 64])  
torch.Size([28, 64])  
torch.Size([26, 64])  
torch.Size([23, 64])  
torch.Size([29, 64])  
torch.Size([45, 64])  
torch.Size([29, 64])  
torch.Size([26, 64])  
torch.Size([24, 64])  
torch.Size([25, 64])  
torch.Size([34, 64])  
torch.Size([30, 64])  
torch.Size([25, 64])  
torch.Size([28, 64])  
torch.Size([25, 64])  
torch.Size([28, 64])  
torch.Size([32, 64])  
torch.Size([25, 64])  
torch.Size([26, 64])  
torch.Size([23, 64])  
torch.Size([27, 64])  
torch.Size([38, 64])  
torch.Size([26, 64])

```
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([35, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([32, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([29, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([35, 64])
torch.Size([19, 8])
torch.Size([21, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([31, 64])
torch.Size([29, 64])
Epoch: 13
Train Loss 0.8296 Val loss 3.98997:
Train Perplexity 2.2924015962091167 Val Perplexity 54.05326774097015:
torch.Size([30, 64])
torch.Size([36, 64])
torch.Size([24, 64])
torch.Size([21, 64])
torch.Size([27, 64])
torch.Size([30, 64])
torch.Size([25, 64])
torch.Size([32, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([30, 64])
torch.Size([38, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([29, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([36, 64])
torch.Size([30, 64])
torch.Size([30, 64])
torch.Size([28, 64])
torch.Size([37, 64])
torch.Size([31, 64])
torch.Size([29, 64])
torch.Size([32, 64])
torch.Size([29, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([29, 64])
torch.Size([29, 64])
```

```
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([30, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([21, 64])
torch.Size([29, 64])
torch.Size([37, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([29, 64])
torch.Size([35, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([31, 64])
torch.Size([31, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([30, 64])
torch.Size([34, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([37, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([32, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([32, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([22, 64])
torch.Size([24, 64])
torch.Size([31, 64])
torch.Size([25, 64])
torch.Size([29, 64])
torch.Size([25, 64])
torch.Size([37, 64])
torch.Size([46, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([23, 64])
torch.Size([29, 64])
torch.Size([27, 64])
torch.Size([35, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([30, 64])
torch.Size([27, 64])
torch.Size([31, 64])
```

```
torch.Size([36, 64])
torch.Size([30, 64])
torch.Size([24, 64])
torch.Size([29, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([29, 64])
torch.Size([25, 64])
torch.Size([38, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([23, 64])
torch.Size([25, 64])
torch.Size([34, 64])
torch.Size([45, 64])
torch.Size([25, 64])
torch.Size([29, 64])
torch.Size([27, 64])
torch.Size([38, 64])
torch.Size([25, 64])
torch.Size([30, 64])
torch.Size([29, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([29, 64])
torch.Size([22, 64])
torch.Size([25, 64])
torch.Size([23, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([31, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([31, 64])
torch.Size([26, 64])
torch.Size([29, 64])
torch.Size([24, 64])
torch.Size([38, 64])
torch.Size([28, 64])
torch.Size([31, 64])
torch.Size([25, 64])
torch.Size([38, 64])
torch.Size([34, 64])
torch.Size([21, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([35, 64])
torch.Size([34, 64])
torch.Size([31, 64])
torch.Size([29, 64])
torch.Size([25, 64])
torch.Size([26, 64])
```

```
torch.Size([28, 64])
torch.Size([23, 64])
torch.Size([28, 64])
torch.Size([29, 64])
torch.Size([32, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([32, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([30, 64])
torch.Size([34, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([23, 64])
torch.Size([32, 64])
torch.Size([30, 64])
torch.Size([24, 64])
torch.Size([22, 64])
torch.Size([27, 64])
torch.Size([31, 64])
torch.Size([28, 64])
torch.Size([31, 64])
torch.Size([36, 64])
torch.Size([26, 64])
torch.Size([23, 64])
torch.Size([30, 64])
torch.Size([30, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([30, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([30, 64])
torch.Size([29, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([22, 64])
torch.Size([26, 64])
torch.Size([29, 64])
torch.Size([25, 64])
torch.Size([23, 64])
torch.Size([33, 64])
torch.Size([29, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([31, 64])
torch.Size([30, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([23, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([26, 64])
```

```
torch.Size([26, 64])
torch.Size([29, 64])
torch.Size([30, 64])
torch.Size([28, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([23, 64])
torch.Size([24, 64])
torch.Size([22, 64])
torch.Size([27, 64])
torch.Size([34, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([30, 64])
torch.Size([27, 64])
torch.Size([36, 64])
torch.Size([24, 64])
torch.Size([32, 64])
torch.Size([24, 64])
torch.Size([33, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([30, 64])
torch.Size([26, 64])
torch.Size([29, 64])
torch.Size([32, 64])
torch.Size([32, 64])
torch.Size([33, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([23, 64])
torch.Size([32, 64])
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([21, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([31, 64])
torch.Size([28, 64])
torch.Size([35, 64])
torch.Size([30, 64])
torch.Size([24, 64])
torch.Size([33, 64])
torch.Size([23, 64])
torch.Size([29, 64])
torch.Size([35, 64])
torch.Size([24, 64])
torch.Size([36, 64])
torch.Size([32, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([23, 64])
torch.Size([34, 64])
torch.Size([26, 64])
torch.Size([32, 64])
torch.Size([22, 64])
torch.Size([33, 64])
```

```
torch.Size([26, 64])
torch.Size([23, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([34, 64])
torch.Size([30, 64])
torch.Size([28, 64])
torch.Size([29, 64])
torch.Size([23, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([26, 64])
torch.Size([27, 64])
torch.Size([30, 64])
torch.Size([22, 64])
torch.Size([30, 64])
torch.Size([26, 64])
torch.Size([37, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([41, 64])
torch.Size([33, 64])
torch.Size([33, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([29, 64])
torch.Size([27, 64])
torch.Size([32, 64])
torch.Size([37, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([31, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([37, 64])
torch.Size([27, 64])
torch.Size([23, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([23, 64])
torch.Size([31, 64])
torch.Size([38, 64])
torch.Size([26, 64])
torch.Size([33, 64])
```

```
torch.Size([23, 64])
torch.Size([28, 64])
torch.Size([25, 64])
torch.Size([22, 64])
torch.Size([29, 64])
torch.Size([25, 64])
torch.Size([33, 64])
torch.Size([31, 64])
torch.Size([36, 64])
torch.Size([26, 64])
torch.Size([30, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([28, 64])
torch.Size([28, 64])
torch.Size([23, 64])
torch.Size([25, 64])
torch.Size([33, 64])
torch.Size([22, 64])
torch.Size([27, 64])
torch.Size([27, 64])
torch.Size([22, 64])
torch.Size([29, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([33, 64])
torch.Size([31, 64])
torch.Size([34, 64])
torch.Size([24, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([31, 64])
torch.Size([23, 64])
torch.Size([26, 64])
torch.Size([33, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([26, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([29, 64])
torch.Size([22, 64])
torch.Size([24, 64])
torch.Size([27, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([33, 64])
torch.Size([30, 64])
torch.Size([25, 64])
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([36, 64])
torch.Size([27, 64])
torch.Size([28, 64])
```



```
torch.Size([22, 64])
torch.Size([30, 64])
torch.Size([32, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([28, 64])
torch.Size([29, 64])
torch.Size([29, 64])
torch.Size([45, 64])
torch.Size([24, 64])
torch.Size([29, 64])
torch.Size([22, 64])
torch.Size([25, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([23, 64])
torch.Size([26, 64])
torch.Size([25, 64])
torch.Size([30, 64])
torch.Size([26, 64])
torch.Size([33, 64])
torch.Size([29, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([21, 64])
torch.Size([25, 64])
torch.Size([22, 64])
torch.Size([26, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([24, 64])
torch.Size([16, 8])
torch.Size([30, 64])
torch.Size([33, 64])
torch.Size([25, 64])
torch.Size([25, 64])
torch.Size([31, 64])
torch.Size([25, 64])
torch.Size([28, 64])
torch.Size([24, 64])
torch.Size([29, 64])
torch.Size([24, 64])
torch.Size([33, 64])
torch.Size([27, 64])
torch.Size([29, 64])
torch.Size([29, 64])
torch.Size([29, 64])
torch.Size([29, 64])
torch.Size([21, 64])
torch.Size([24, 64])
torch.Size([25, 64])
torch.Size([23, 64])
torch.Size([27, 64])
torch.Size([25, 64])
torch.Size([30, 64])
```

Epoch: 14

Train Loss 0.74645 Val loss 4.19099:

Train Perplexity 2.1094979905286873 Val Perplexity 66.08818588860908:

In [16]:

```
test_loss = evaluate(model, test_iterator, criterion)

print('| Test Loss: {} Test PPL:{}'.format(test_loss, np.exp(test_loss)))
```

```
| Test Loss: 4.230438411235809 Test PPL:68.74736518536797|
```

In [0]:

```
EOS_IDX = SRC.vocab.stoi['<eos>']

def translate(sentence):
    """
    function that uses .translate() method of the model to translate german sentence in
    to english
    params:
        sentence: tokenized german sentence
    """
    sentence = sentence.lower()
    sent_vec = [SRC.vocab.stoi[token] for token in sentence.split()]
    translation_idx = model.translate(torch.tensor(sent_vec))
    for t in translation_idx:
        if t[0] != EOS_IDX:
            print(TRG.vocab.itos[t[0]], end=' ')
        else:
            break
```

In [18]:

```
translate("ein klein apfel")
```

```
tourists , a a television .
```

/usr/local/lib/python3.6/dist-packages/ipykernel\_launcher.py:52: UserWarning: To copy construct from a tensor, it is recommended to use sourceTensor.clone().detach() or sourceTensor.clone().detach().requires\_grad\_(True), rather than torch.tensor(sourceTensor).

ИИИИ давайте также научимся считать самую популярную метрику для перевода -- BLEU (<https://en.wikipedia.org/wiki/BLEU> (<https://en.wikipedia.org/wiki/BLEU>))

В общем-то, вам повезло -- ее писать руками скучно, да и nltk ее написало за вас:

In [0]:

```
from nltk.translate.bleu_score import corpus_bleu
def compute_bleu(inp_lines, out_lines):
    """ Estimates corpora-level BLEU score of model's translations given inp and reference out """
    translations = [translate(line) for line in inp_lines]
    return corpus_bleu([[ref] for ref in out_lines], translations) * 100
```

In [20]:

```
compute_bleu(<batch_of_sentences>, <batch_of_sentences>)
```

```
File "<ipython-input-20-33b7181f6c99>", line 1
```

```
    compute_bleu(<batch_of_sentences>, <batch_of_sentences>)
```

```
    ^
```

SyntaxError: invalid syntax

Если вы реализовали несколько методов аттеншена, опишите каждый из них и полученную метрику на нем в отчете.

Использовал метод dot . Я пробовал различные модели с различными количеством эпох, но каждый раз моя модель переводила не так, как в нормальном переводчике. Ближе всего был перевод small apple a a , но и это не является приемлемым результатом.

In [ ]: