Deep Learning Assignment 3
Terekhin Daniil

1. Preliminaries and Reading Comprehension

1.1. Text data

1.1.1.

In the text there are several strange symbols like "Æ", "œ", "μ", "α', "λ", "ὸ", "Φ", "ρ", "ὐ", "ξ" and others.
The properties of the text data you can see in the Figure 1.

```
Number of characters: 177517
Number of unique characters: 105
Number of lines: 5033
```

Figure 1: Properties of the text

1.1.2.

During preprocessing I would delete several symbols like "*", "/", "#", "]", convert to lower case each letter, remove all punctuations and stop words, make lemmatization.

1.2. Dataloader / Batch Construction

1.2.1.

In a case when the word not in the vocabulary, the string adds to the dictionary "string_to_id" with it's length (String - length) and also adds to the dictionary "id_to_string" as a lengh as a key and string as a value (Length - string).

1.2.2.

In dictionary "string_to_id" the key is input string, the value is it's length. In dictionary "id_to_string" the key is length of the string and the value is string.

1.2.3.

The length of the tensor representing the input text.

1.2.4.

The length of a list of tensors representing batches.

1.2.5.

The first of these rows returns a Tensor of size ((division of the length of input data and batch size rounded down) + 1, multiplyied by the batch size) filled with ID of the padding tokens the second row takes the number of values equal to the length of TextData object from the Tensor and equates them to data of TextData object. So these rows take input data and convert it into tensor with necessary dimension and fill with values of input data.

1.2.6.

The shape of “padded[i * bptt len:(i + 1) * bptt len]” is bptt_len * bsz (64* 32)

1.2.7.

The shape of “padded[i * bptt len - 1:(i + 1) * bptt len]” in else branch is (bptt_len +1) * bsz ((64+1)* 32) = (65*32)

1.3. Modeling, Training and Decoding

1.3.1. detach() in hidden state of RNN is used for making copy of the tensor without gradients, so they could be converted to numpy or used in different way. During the training of the model the state of RNN updated each time, so we don’t need gradients.

1.3.2. In nn.CrossEntropyLoss parameter "ignore_index" means that this mentioned value ("ignore_index=0") will not be considered as a mistake during the training. In class DataBatches we set 0 as a default value. Sequence will be filled with 0 to a necessary number of characters.

1.3.3. The input shape of RNNModel is (sequence size (N), batch size (B), embedding size(D)). We can see the architecture in the Figure 2.

```
RNNModel(
  (input_layer): Embedding(107, 64)
  (rnn): RNN(64, 2048)
  (out_layer): Linear(in_features=2048, out_features=107, bias=True)
)
```

Figure 2: Architecture of the RNN model

1.3.4. The output shape of the model is (sequence length(N), batch size(B), hidden dimension (H) * number of layers (L)).

1.3.5. We use this phrase to train the model by completing the text segment. In the original text there is such sentence, so we control the training of the model with the output of the model across the epochs.

2. Running Experiments Using the Initial Code.

2.1. I modified loss to perplexity during monitoring.

:

2.2. I trained the RNN language model using the following hyperparameters. I got the value of perplexity equals 1.7588787900990452 in the last epoch. The evolution of the perplexity we can see in the Figure 3, the generation quality in the Figure 4.

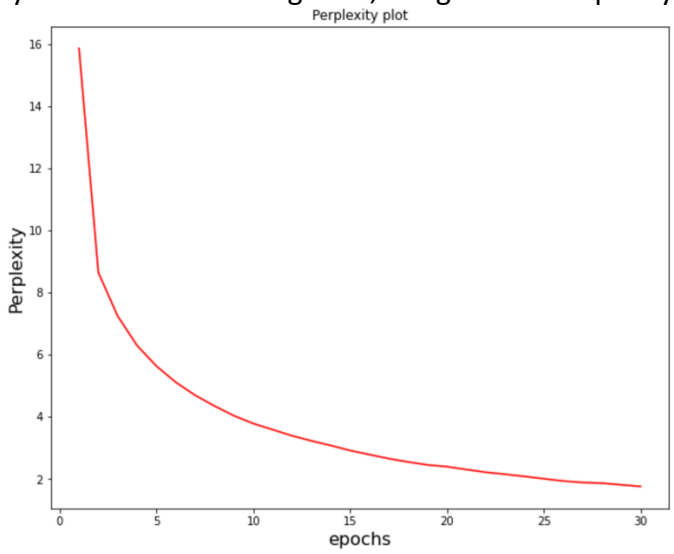

Figure 3: The evolution of the perplexity

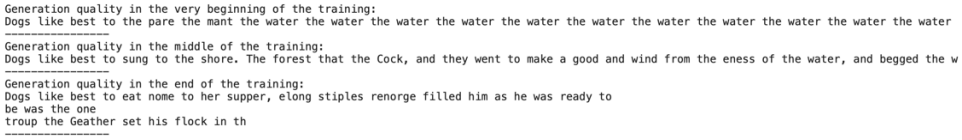

Figure 4: The generation quality

2.3. Outputs of the model

2.3.1. A title of the fable which exists in the book (Figure 5)

2.3.2. A title which I invented, which not in the book, but similar in the style (Figure 6)

2.3.3. Some texts in a similar style (Figure 7)

2.3.4. Anything I think might be interesting (Figure 8)

We can see that the model produces sentences that don't make much sense. Separate sentences can look good, but sequence of these sentences doesn’t. It's a connected set of words. If we compare all these paragraphs, we can see that an output of the model towards by a title of a fable which exists in the book provide more sense than others, but still not perfect. The whole output is not meaningful, and model is not capable to produce novel tests.

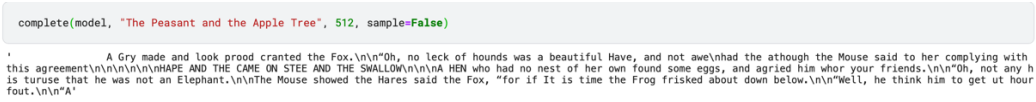

Figure 5

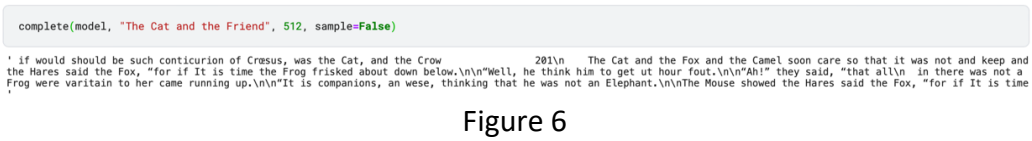

Figure 6

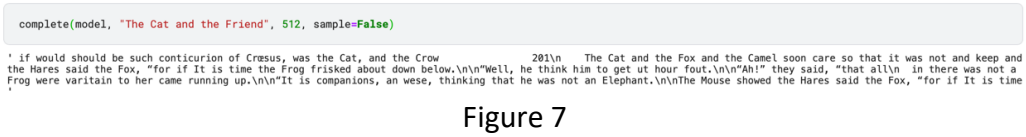

Figure 7

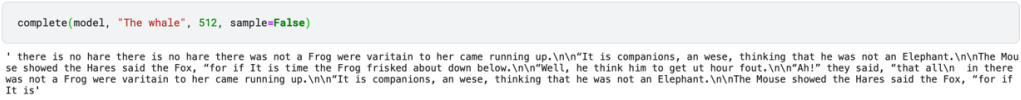

Figure 8

3. Extending the Initial Code

3.1. I implemented the LSTM language model. The architecture of the LSTM model we can see in the Figure 9.

```
RNNModel(
  (input_layer): Embedding(107, 64)
  (lstm): LSTM(64, 2048)
  (out_layer): Linear(in_features=2048, out_features=107, bias=True)
)
```

Figure 9

3.2. I used the same hyper-parameters, except learning rate equals 0.001. I achieved a training perplexity value below 1.03 (1.029) (Figure 10) and plotted the perplexity evolution on the Figure 11.

```
Best best perplexity: 1.029586097289776
```

Figure 10

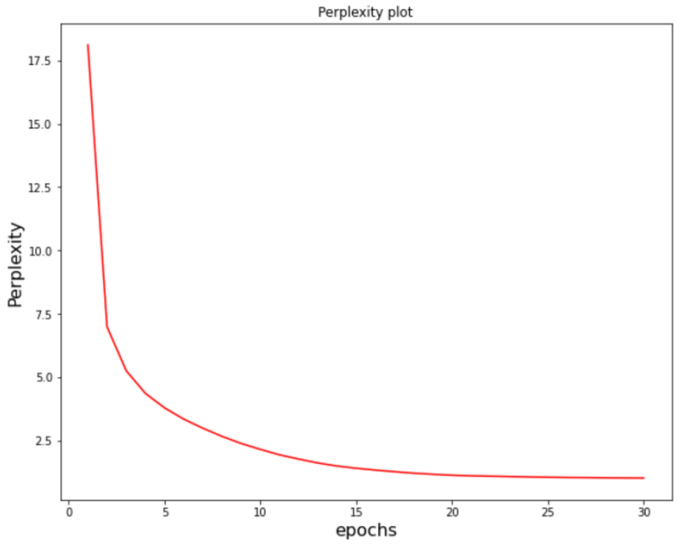

Figure 11

3.3. I implemented the option to change from taking the argmax according to the model's output distribution to taking the random value from that distribution.

3.4. Comparing the greedy decoding and sampling using the same prompt.

3.4.1. A title of a fable which exists in the book (Figure 12). We can see that the beginning of paragraphs is similar.

```
complete(model, "THE MOUSE, THE CAT, AND THE COCK", 512, sample=False)
```

'\n\n\nA HUNGRY Cat, who had tried in vain to find a Mouse for her supper, at\nlast caught a young Cock.\n\n"You are a noisy creature," she said to him, "and have lived long\nenough. You disturb every one in the house by your loud crowing in the\nmorning."\n\n"You are mistaken," answered the Cock; "I disturb no one. I crow to\nwake the family. They would not know when to get up but for me."\n\n"Never mind," said the Cat; "don't trouble yourself to make excuses; I\nhave had no breakfast and no dinner; I shall eat you for my'

```
complete(model, "THE MOUSE, THE CAT, AND THE COCK", 512, sample=True)
```

'\n\n\nA HUNGRY Cat, who had tried in a manger for his life. The Weasel protested that it was against nature\nfor a wead half at night, and began to growl\nand bark at the hore. The Woman Project Gutenberg-tm\ntrademark as see and asked the Fox. The Dove had losted upon an add then to\nthe dismay of the Trees. The Apple wish med what it was.\n\nShe found that it was flax. "When this flax has grown," she said to herself, "the man will make it into linening to dawn."\n\n"Well, then, try Demian."\n\n"I'm afraid he's angry w'

Figure 12

3.4.2. A title which I invented which not in the book, but similar in the style (Figure 13)

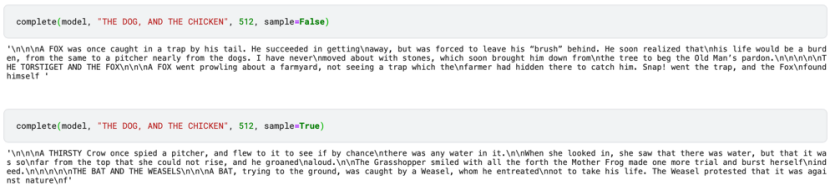

Figure 13

3.4.3. Generated text for the title of a fable which exists in the book better in the greedy search, but for the title, which not in the book, it works worse than sampling.

3.5. Bonus question. I downloaded other book about Entomology. It has 38721 lines (Figure 14). I increased the number of epochs from 30 to 45 and the number of RNN (LSTM) hidden layers from 1 to 2. The final perplexity was 1.34. The evolution of perplexity is plotted in the Figure 15. For getting better results we need to make pre-processing of the text, because there are a lot of references to literature and raw data. The output of the model is in the Figure 16. The model still don't get meaningful result

```
Number of characters: 1826696
Number of unique characters: 161
Number of lines: 38721
```

Figure 14

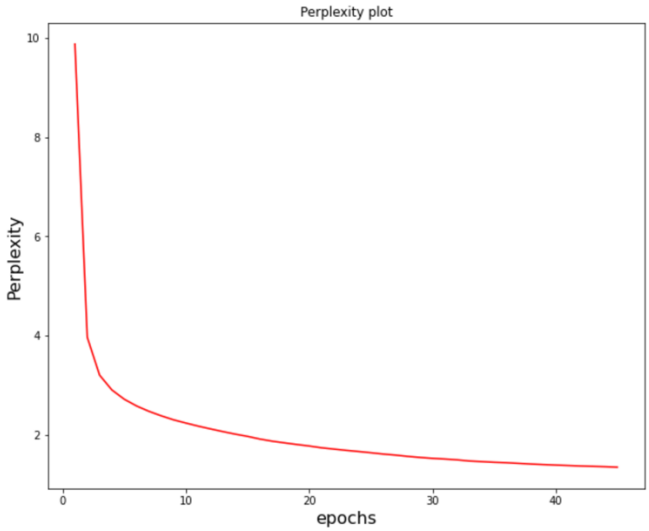

Figure 15

```
complete(model, "the muscular system of insects", 512, sample=False)
```

```
'\n    independently of the base of the tarsus; _b_, follicle\n    support; _b_, præscutum; _c_, the part of the median tubes;\n    _i_, the scales.\n]\n\nThe mid-intestine o
f digestion the number of these structures are in most\ncases the optic ganglia, the optic ganglia are, in\nthe egg, and is united to the gradually extending out of the\nprim
itive band. The two oviducts are evidently outgrowth. The number of\nthese processes are covered out after the end of the body. In the\nthoracic wall of this papilla, pupated
or'
```

```
complete(model, "the sensory organs", 512, sample=False)
```

```
', and the\n  secretion in the mandibulate insects. (Ediebil.)\n\n  The tarsus and dones are secondary, pulsatophority, or the most\ndifferent orders of insects, though he po
ssibly becomes\nfirmly attached to the segments. They end in a single\nprocess of degeneration and scale-like, concavo-convex size, and the lining\nbecomes excrement." The do
uble organs and other changes of food-reservoir\nto prove that they are either constantly are evidently their homologues, while the\ntransformation is in this case, then, of
s'
```

```
complete(model, "The organs of smell", 512, sample=False)
```

```
'.\n  The cells commenced in other orders, the second ins supposition\n  to the same manner. The movement of the cuticle. It is\n  the second moult then dust out a considerab
le\n  inheritance, as in the case of the fat-body.\n\n  In the Terlidæ with the thoracic stigmata, though the connection of the\n  next pores in the middle of the tracheæ and
of the oblique\n  operation they produce a hole in color to the sides of the\n  thorax. These discovered the eyes of some hours\n  adopt the old skin, is detached, which issu
e'
```

```
complete(model, "the digestive canal and its appendages", 512, sample=False)
```

```
'\n  (_Euch chyle-stomach-_). The tracheæ arise in the formation of the\n  nearly as a vacuus. The two oviducts are one on each side in a\n  single testicular (Fig. 1), which
are concerned in one of them.\n  The genital germinas form a number of opening in\n  shape. This discovery while in the males of Æschna.\n\n  In the most general account of t
his organ of the legs, when the muscles\n  are proliferate, overgrown the thoracic stigmata, though they are fewer\non the genital organs, and the thoracic stigmata are again
'
```

Figure 16

4. Questions
   4.1. V
   4.2. It is an issue because in that case weights will not be updated through backpropagation.