

Deep Learning: Assignment 3

Jury Andrea D'Onofrio

November 2022

1 Preliminaries and Reading Comprehension

1.1 Text data

1.1.1

- Number of characters = 185303.
- Number of unique characters = 105.
- Number of lines in the file = 5033.

As property of the text, I note that the title and the first word of each fable are always in upper case.

1.1.2 Bonus

I would do data cleaning to remove all weird characters.

1.2 Dataloader/Batch Construction

1.2.1

There is a `if` branch because, if the string is found, then its index is returned. Otherwise, the string is added to the `Vocabulary` and returned its index.

1.2.2

`id_to_string`

- keys \rightarrow id
- values \rightarrow string

`string_to_id`

- keys \rightarrow string
- values \rightarrow id

1.2.3

I obtain the length of the tensor `input_data`.

1.2.4

I obtain the length of the list of tensors representing batches.

1.2.5 Bonus

- `padded = input_data.data.new_full((segment_len * bsz,), pad_id)`
→ this row assigns to `padded` a tensor with shape `(segment_len * bsz)` with value `pad_id`. In our case, the shape of this tensor is `(177536, 1)` and `pad_id` has value 0. Therefore, the resulting tensor `padded` is filled only with 0.
- `padded[:text_len] = input_data.data` → this row fills `padded` from index 0 to `text_len - 1` with the same values of `input_data.data`

1.2.6

The shape of `padded[i * bptt_len:(i + 1) * bptt_len]` is `[64, 32]` where:

- `bptt_len = 64` is the number of rows;
- `bsz = 32` is the number of columns.

1.2.7

The shape of `padded[i * bptt_len - 1:(i + 1) * bptt_len]` is `[65, 32]` where:

- `bptt_len + 1 = 65` is the number of rows;
- `bsz = 32` is the number of columns.

1.3 Modeling, Training, and Decoding

1.3.1

It is applied `.detach()` to the hidden states of RNNs because it is needed a tensor without the gradients attached. This tensor gradient will not be considered during the `.backward()`.

1.3.2

`ignore_index = 0` is given to ignore the target value 0. This target value doesn't contribute to the input gradient.

1.3.3

The input shape expected by `self.rnn` is (2048, 32, 64) where:

- $N = 2048$ is the sequence length;
- $B = 32$ is the batch size;
- $D = 64$ is the character embedding size.

1.3.4

The outputs shape of `self.rnn`:

- $N = 2048$ is the sequence length;
- $B = 32$ is the batch size;
- $H = [32, 2048]$ is the RNN hidden dimension;
- $L = 1$ is the number of RNN layers.

1.3.5

“Dogs like best to” is needed to train the model. This input is used to generate text using a greedy decoder.

2 Running Experiments Using the Initial Code

2.1

I modify the code to monitor the model’s training perplexity instead of the loss.

2.2

I train an RNN language model using the given hyper-parameters. As Figure 1 shows, I obtain a training perplexity below 1.8.

Here, three examples of texts generated by the model:

- Beginning → Dogs like best to the oued the and the and the and the and the and the and the and the and the and the and the and the and t
- Middle → Dogs like best to say that it was not any one to her going to reach it was bling beited his warmest and the Sheep
- End → Dogs like best to the one on a break as well some in such general Terms of the full Project Gutenberg Literary Archive Foundation are the store, w

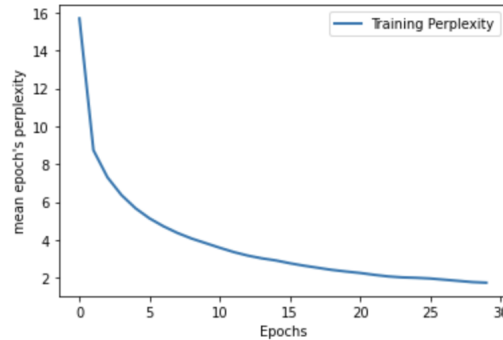


Figure 1: RNN Training Perplexity

2.3

I run greedy decoding using different prompts using a decoding steps of 512:

(a) A title of a fable which exists in the book.

- **input** → THE WOLF AND THE SHEEP
- **text generated** → A WOLF and a part of copying, bur let at once could be, then?" "Ohe go for it. Sto pime the Eagle, the Fox returning and the Fox 30 The Fox in the Ice pountions from it to part he thought. As the water from the pond came not to come and live with him and nearly any purposed to the Eagle, to protect her nearly any purposed to the Eagle, to protect her nearly any purposed to the Eagle, to protect her nearly any purposed to the Eagle, to protect her nearly any purposed to the Eagle, to

(b) A title which you invent, which is not in the book, but similar in the style.

- **input** → THE WOMAN AND HER MAN
- **text generated** → TER OF STRER A MUN who had traveled in foreign parts bragged, on his returned to him all leave me," said he; "then the Fowner of her toal, and when the Fox really come for the purpose of the fable that is it is the grazey for the beasts to each other, "but when they clueted and place ind ento the partine of this agreement shall he let the Fat expected. The Fox invited traveled that I have found a have had the ither the water to the Fowler. The wamen that can place yourself to blantiful have been a foread

(c) Some texts in a similar style.

- **input** → IT WAS a sunny day

- **text generated** → a have lounteach of expocean.” “No,” said one of the fore. [Illustration] “Ne caught any ported to you ’wakey athen the water falle 23 The Fat and the Fox 249 The Eagle and the Fox 32 The Far put in the house,” said Brown-back. “If I but stepped to the water and feet we enough for of his two fame hard well,” said he, “or I am not a perform, to the ran after that the Here beant for the use of his this?” cried the Frogs; “pray do not try to

(d) Anything you think might be interesting.

- **input** → Home Alone
- **text generated** → hime longentedle to him the other the water is, and begged to him, and the earle the work all the beautiful brang man had saw the Lion said the Hare The Far Hure for you now that?” this esher the water is, and the Frog hould not and not knew that the any you no his temped a perform. 1.F.4. INcERS LESE A HEN who had not know better than I have stack feath, faunt of his truch of the lake. He was expected. The Fox invited the work and the Fox

The results are not meaningful. The model is not be able to produce novel texts. The outputs are better if I use an input from the fables text or similar to its, but in the two other cases the outputs are very far to a novel.

3 Extending the Initial Code

3.1

I implement a LSTM language model using `nn.LSTM`. In the training pipeline, I do the following modifications:

- I add a starting `cell_state` that uses the given method `custom_detach`;
- I pass as input to the model both `state` and `cell_state`;
- Since the output `state` of the model is a tuple, I split its in `state` and `cell_state`.

In the decoding pipeline, I do the following modifications:

- I add a starting `cell_state`;
- The output of the model `states` is now a tuple of `states` and `cell_state`.

3.2

I use the same hyper-parameters and set the learning rate to 0.001.

- My best perplexity = 1.015
- Figure 2 shows the training evolution

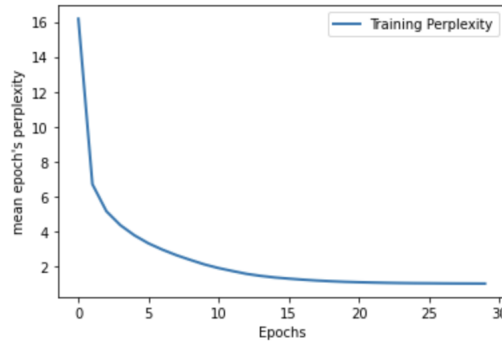


Figure 2: LSTM Training Perplexity

3.3

I modify the helper code to add an option to allow sampling during decoding.

3.4

Below, the output of a fable using the LSTM language model which title exists in the book: THE FOX AND THE GOAT.

- **greedy decoder** → A FOX once happened to fall into a deep well. He tried in every way to get out, but at last began to think that it was impossible, and that he must die there, a prisoner. While he was thinking how sad that would be, a thirsty Goat came and looked down into the well, wishing that he can get them all to him and showed them a bundle of sticks tied tightly together. “Now,” said the Father, “see if you can break this bundle of sticks.” Each of the Sons in turn took the bundle and tried with all his might to
- **sampling decoder** → A FOX once happened to fall into a deep well. He tried in every way to get out, but at last began to think that it was impossible, and that he must die there, a prisoner. While he was thinking how sad that would be, a thing, the Dogs heard the sound, and seeing him struggling there, said: “Oh, you foolish Fly! Were you so greedy as to be caught like that? Your appetite was too much for you.” The poor Fly had nothing to say in reply. What the Moth said was true. But by and by, he found the Swallow lying f

Below, the output of a fable using the LSTM language model which title I invent: THE KING AND THE QUEEN.

- **greedy decoder** → RIGH A COLONY of Ants were busily employed in the care and preservation of their food, which they exposed to the air in heaps around their country habitation. A Grasshopper, who had

chanced to out-live the summer and was ready to die from cold and hunger, approached with great humility and begged that they would relieve his necessity with one grain of wheat or rye. One of the Ants asked him how he had disposed of his time in the summer, that he had not taken pains and laid in a stock as they had done. “A

- **sampling decoder** → PROJE A FOX came at vintage time to a place where ripe grapes were hanging in tempting clusters over the branch of a tree. “I will get them,” said the Father, “see if you can break this bundle of sticks.” Each of the Sons in turn took the bundle and tried with all his might to break it, but could not for one one of them. THE ONE-EWeD Dove pains to fly by. The Squirrel often suffers hunger and, while grinning in his masters presence, has eyes full of tagear, and asked the man if that was his. Transpo

Sampling is not a good idea. The result is similar compared to greedy for titles of fables which exist in the book but, otherwise, it doesn’t work as well as greedy.

3.5 Bonus

I use my code of the second assignment to train the model using the greedy decoding:

1. **input** → "torchvision"
text generated → .transforms as transforms_2 from torchsummary import summary device = torch.optim.SGD(model.paramsel.bes() optimizer.step() update torch.no_grad():
2. **input** → "CIFAR10"
text generated → outputs = model(images) loss = loss_fn(outputs, labels) optimizer.step() update torch.no_grad():
3. **input** → "CrossEntropyLoss()"
 text generated → optimizer = torch.optim.SGD(model.paramsel.bes() optimizer.step() update torch.no_grad():
4. **input** → "Evaluation"
text generated → Dataloaders test_loader = torch.utils.data.DataLoader(dataset=test_set, batch_size=batch_size, shuffle=False) with torch.no_grad():

The outputs are mostly correct. The model is able to produce some rows of code. Nevertheless, the perplexity is not good as above, it achieves a value of 1.07.

4 Questions

4.1

V.

4.2

It is an issue because it's not possible to adjust weights to respond to events.