Text Generation with RNNs

Machine Learning

Aiden Koknat

This program uses recurrent neural networks to train and recreate Shakespearean writing. Using an extensive data set (40,000 lines) of William Shakespeare's writing, it separates the text by line, then by tokens (words) and characters.

Most of the code was built off of Dr. Harrison's example code given.

The python program uses the NumPy and PyTorch libraries.

Here is the code that inputs the text file as lines, extracts characters and sets up the vocabulary for the neural network.

```
#  Reading in text to be trained/tested
sentences = []

with open("tiny-shakespeare.txt", 'r') as file:
    for item in file:
        if item.strip() != "":
            line = item.strip()
            sentences.append(line)

#  Step 1, extract all characters
characters = set(''.join(sentences))
print(characters)

#  Step 2, set up the vocabulary
intChar = dict(enumerate(characters))
print(intChar)

charInt = {character: index for index, character in intChar.items()}
print(charInt)
```

Regarding implementation and training, I used a batch size of 1, and did several different trials with different amounts of epochs to see how it would affect the output. Here are the results:

1 Epoch:

```
Average Loss: 1.7474570797487405
QUEEN: O: OE: I No my speak of the speak of the speak of the speak of the speak of the speak of the

Process finished with exit code 0
```

5 Epochs:

```
Average Loss: 1.3997807686180503
QUEEN:UON: IO: O: TIAN:T:ONON: IO:: SERIAN:::: IV::: I: IV:: IO:: TONIO:: NIAN: SERONION:::: I: IS::

Process finished with exit code 0
```

25 Epochs:

```
Average Loss: 1.3633649733933089
QUEEN:: OVERDONE:: ONTZAND: OVOMIO: ON: OVerain the son the son the son the son the son the son the
```

150 Epochs:

```
Average Loss: 1.4975157975826026
QUEEN: VINCULINARLILA:DLLONN:NLELETELLIZALELLO: OL: D IVO:ZALELERAELELIINT: CLLLERELAT: WINHI: OL: A

Process finished with exit code 0
```

It seems like 150 epochs doesn’t make a significant difference in terms of loss reduction from 5 or 25 epochs.