Arie Cattan 336319314
Moche Uzan F957022

Part 1

Data

We generate 1000 examples 500 sample for each language.
We limited the sequence of character made by a "+" to 10.
Ex : "a+" will generate between 1 to 10 "a".
We shuffled the 1000 example and split them to two part:
1. Train : 80 first percent
2. Dev : 20 last percent

Model

The model is composed by :
1. Embedding layer (50)
2. LSTM (input=50 output=10)
3. A linear (input=10 output=5)
4. Tanh function (input=5 output=5)
5. A linear (input=5 output=2)
6. log_oftmax function (input=2 output=2)

We choose from the output the best class as prediction.
We use an Adam optimiser and cross entropy as loss function.

Result

Our model succeeded in distinguishing the two languages after 2 iterations done in 22.98s.

He got 100% accuracy on the train and the dev data.