Exercise 1.

We have a simple neural network, which knows 100K words, and contains 2 hidden layers with 100 neurons. This network uses the one-hot embedding to process words.

- Question 1. What percentage of the parameters are in the input and output layers?
- Question 2. If we want to cover 10 past words with this network how would the percentage change?
- Question 3. How would you change this model to reduce its size and increase its speed?
- Don't forget to submit the answers in MyCourse!

Hints

Each neuron has $\#input + 1$ parameters (weights and the bias).
100K input and 10 neurons $\approx$ 1M parameters.
Output layer has $\#words$ neurons, the input of the second hidden layer is the first one.

Question 1:

About 99%

Question 2:

About 99%

Question 3:

We can reduce the size of one-hot encoding vector using other methods of vector representations (i.e. word2vec)