Creation of the languages:

While creating the languages, we determined the lengths of the sub-sections that make up the whole word (a sequence of letters or numbers, for example - aaaa\5491) using a normal distribution, we performed several runs for different average and standard deviation values. We've seen that as the average rises it takes the network more time to learn to recognize the languages.

summary of the experiment:

did your network succeed in distinguishing the two languages (it should)?

The network was able to distinguish between languages.

The bigger the average length of segments was, the longer it took for the network to study. In our opinion, this stems from the distancing of the significant regions which cause the effect on the derivative to be small.

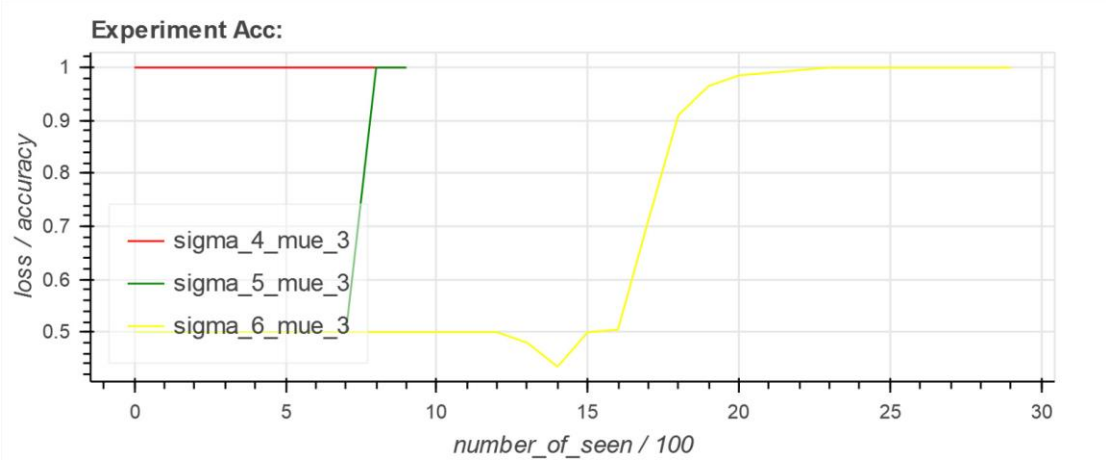
The network seems to have a limit to her capabilities. For a very big average length it will fail in learning. This problem is known as vanishing gradients.

how long did it take (both wall-clock time (i.e., number of seconds), and number of iterations)?

We ran the network on the languages with the averages 4-6.

In different runs with the same parameters the results were different, but there was a direct correlation between the size of the mean and the number of iterations and the time taken.

There was no use of batching in this part. It might improve the results.



	4	5	6
number of seconds	23	184	460
number of iterations	1	9	17

Did it succeed only on the train and not on the test? what you did in order to make it work, etc.

It worked on both train and set data.