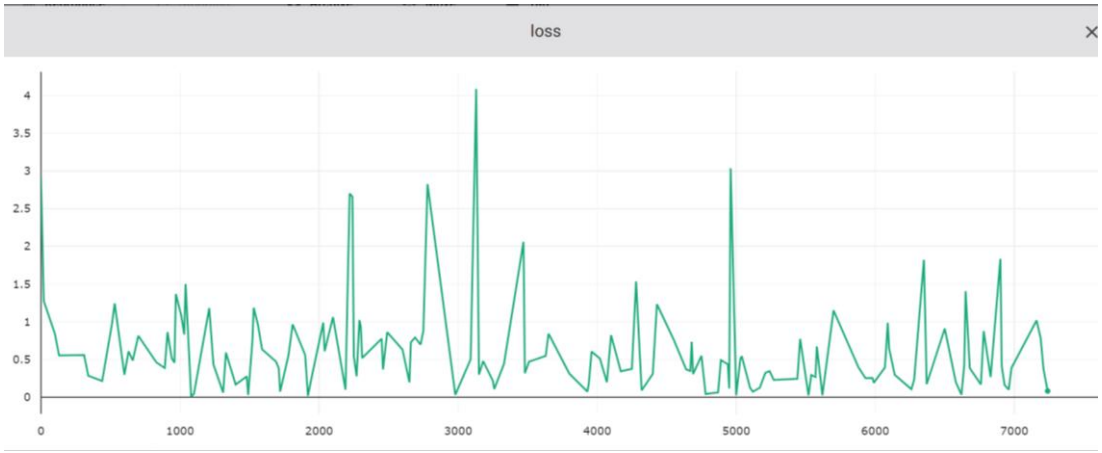
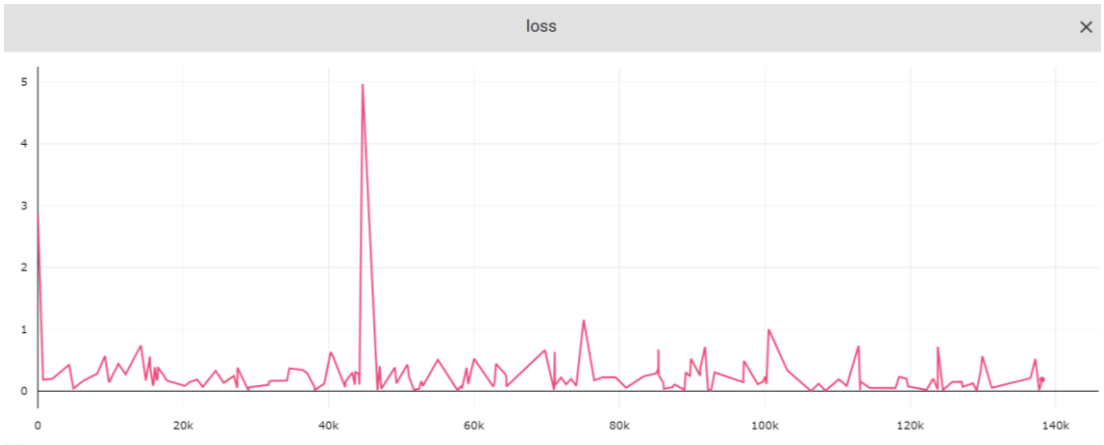
Report for Assignment 4

Uliana Eliseeva, 7008101

Models tried:

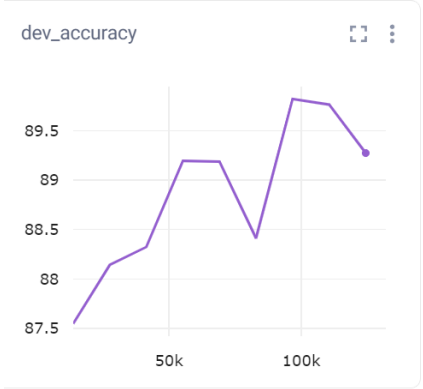
Learning rate 0.05 (graph 1) vs 0.01 (graph 2) – a greater LR seems to overshoot a minimum and the learning curve looks jumpier as a result with a range between 0 and 2. In contrast, with a smaller LR the loss doesn't reach more than 1 in the later epochs.

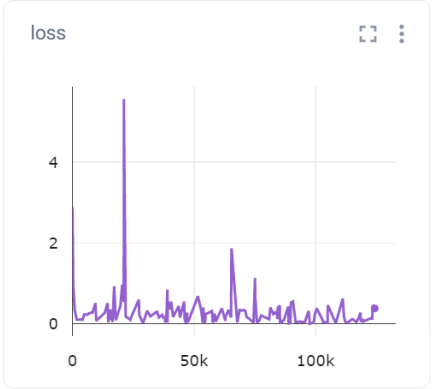


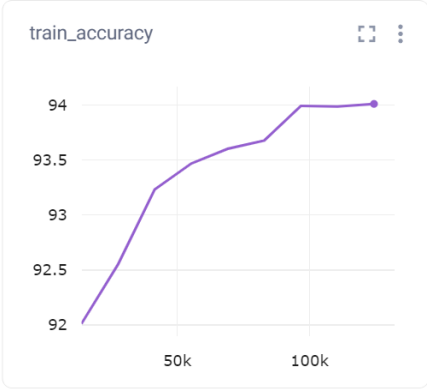


Number of hidden units 300/600/1000 (graph 1) – best performance with a small number of epochs (less than 10) is reached by a simpler model of 300 hidden units. It could be, that with a larger number of epochs more complex models could outperform, however, my laptop's computational power was not enough to test this hypothesis.

With 1000 hidden units, development accuracy starts declining after the 7th epoch, while training accuracy is still slightly growing, which could be a sign of overfitting. 8 epochs of 600-unit hidden layer reaches maximum 89.39% development accuracy.







Number of hidden layers 1 vs 2 – surprisingly a bigger number of hidden layers of both 300 and 600 units didn't improve the model development accuracy at the current number of epochs (less than 10).

Final model evaluation after 10 epochs with 1 hidden layer of size 300 and learning rate 0.01:

Training accuracy: 94.374%

Development accuracy: 89.97%

Testing accuracy: 90.1622%

(Screenshot of 8 epochs, further epochs were trained separately so the comet experiment was already unchangable)