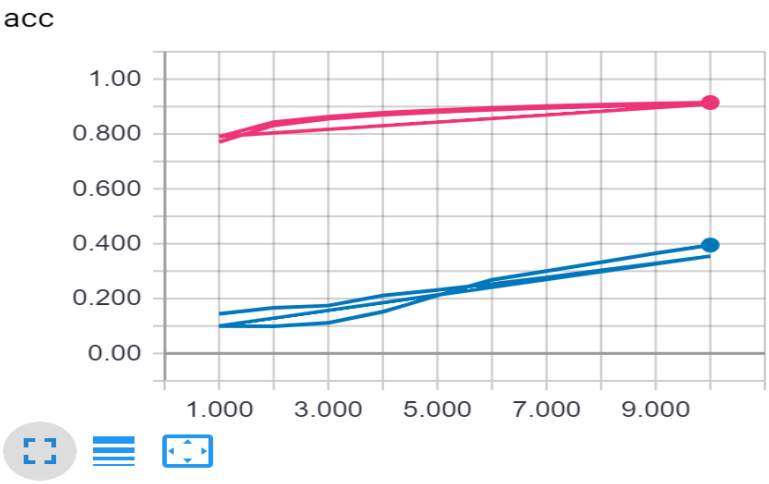
MLP Report

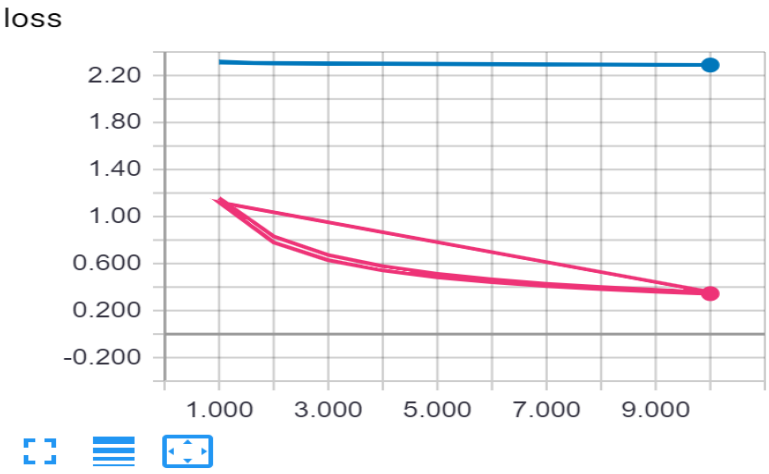
Ali Maleky

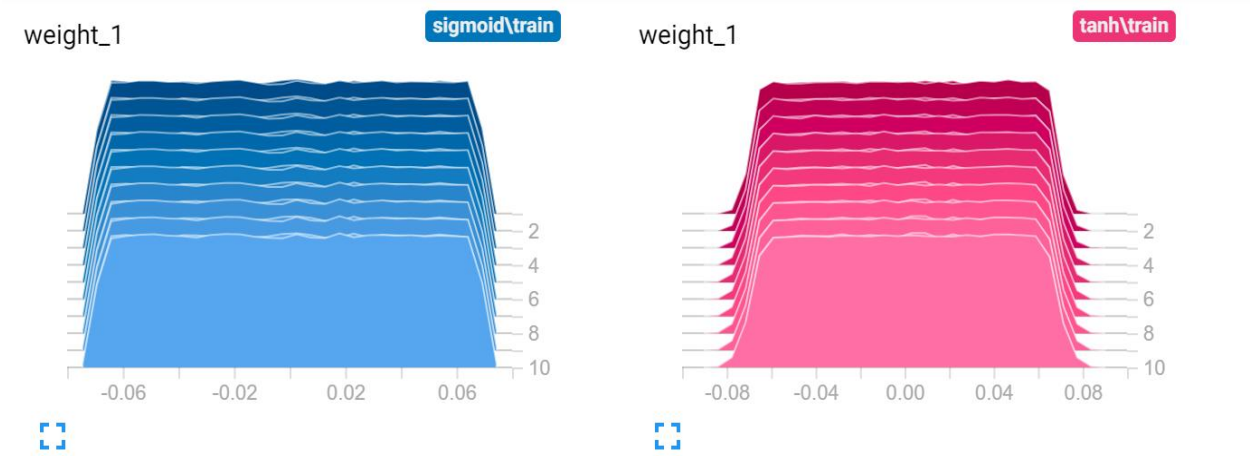
94102155

Question 1: As we see both in tensorboard and results here, tanh performs much better than sigmoid and that is because of the advantages it has over sigmoid which has been discussed in HW2. Weights histograms distributions in tensorborad are almost the same except that tanh is more compact.

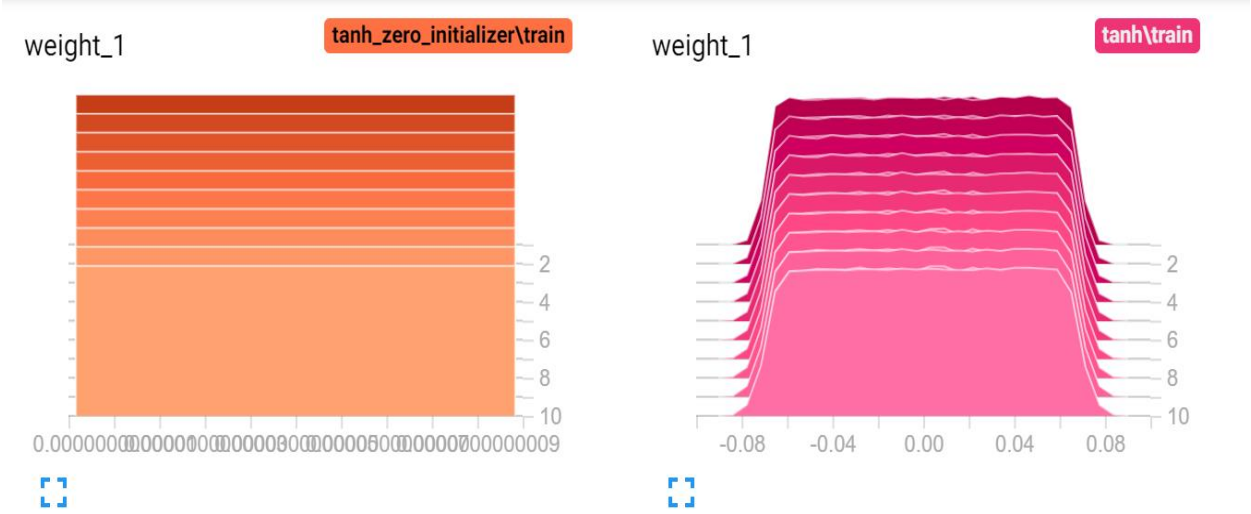


loss



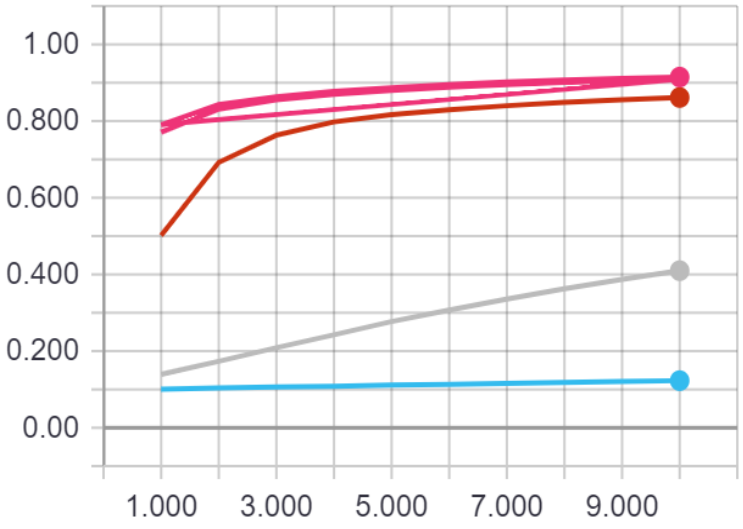


Question 2: As we see in tensorboard, the weights remain the same for each layer since the initial state for all weights were the same and all neurons perform as the same. That is why with zero initializer, all weights in each layer are equal so nothing changes. even loss and accuracy.



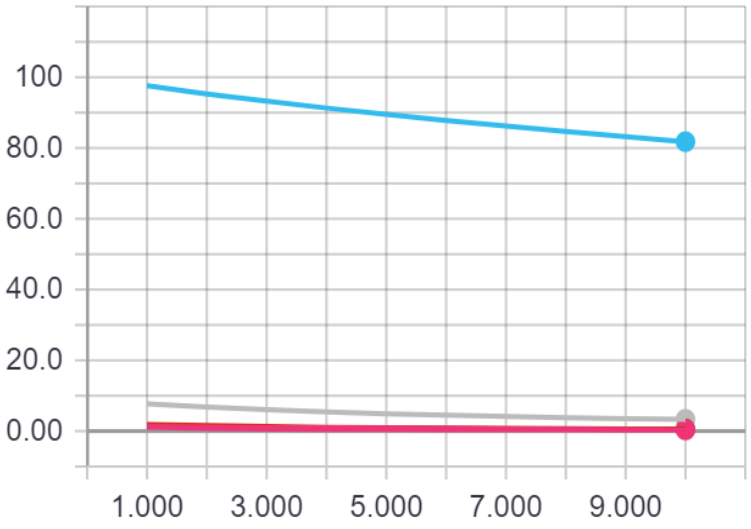
Question 3: As we expected, the more closer the initial values are to 0, the better the training is. smaller std leads to better and smoother weight distributions in tensorboard histograms and distributions which are closer to zero.

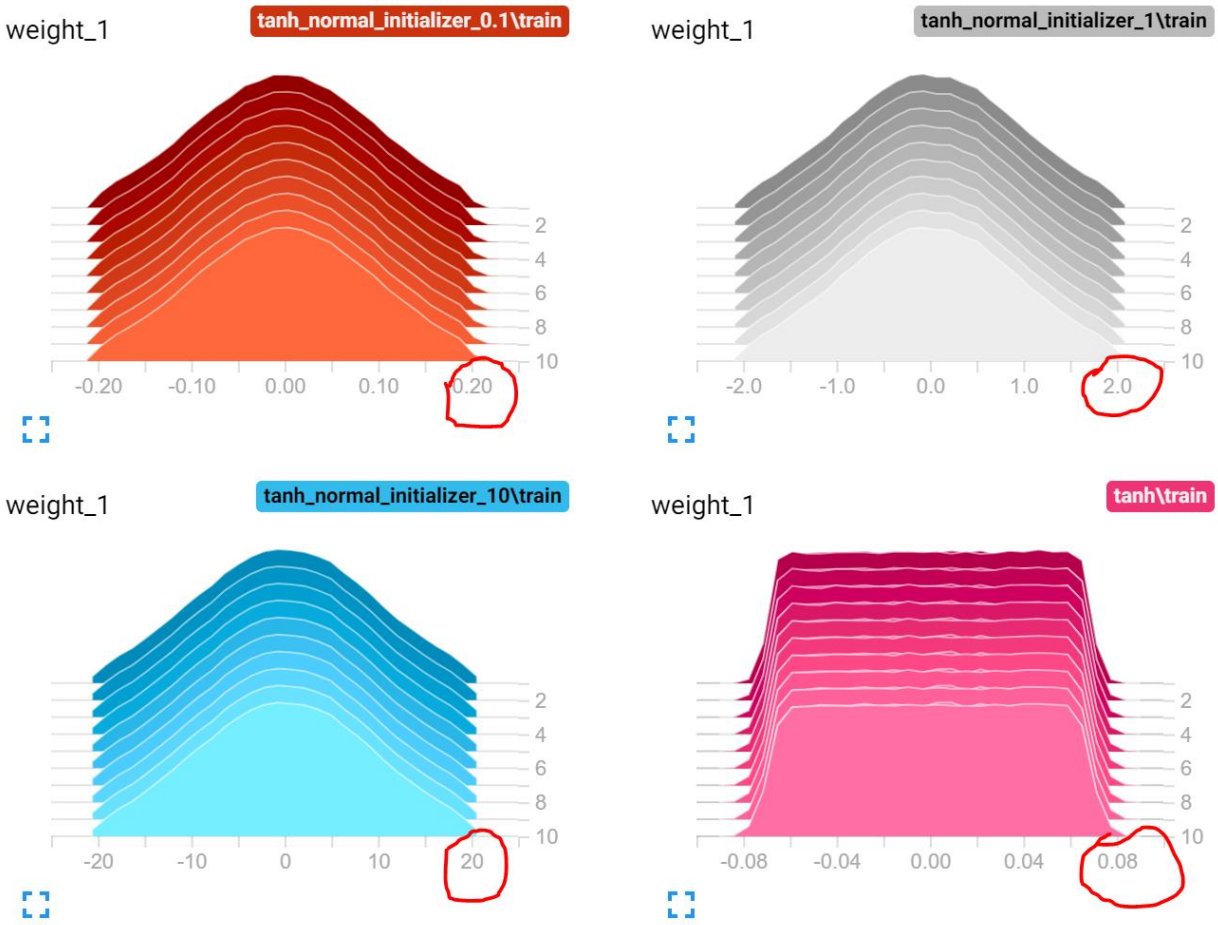
acc



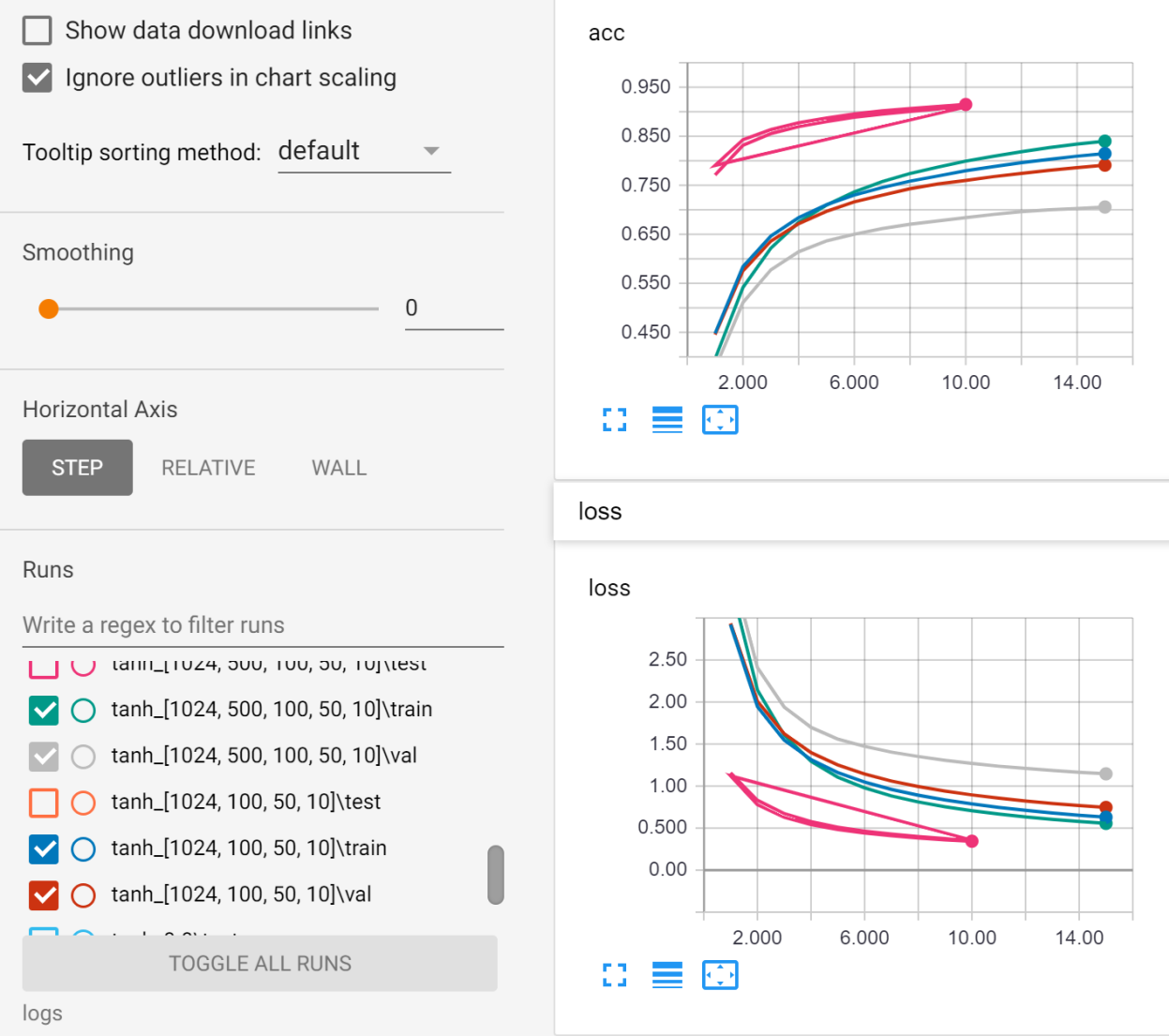
loss

loss



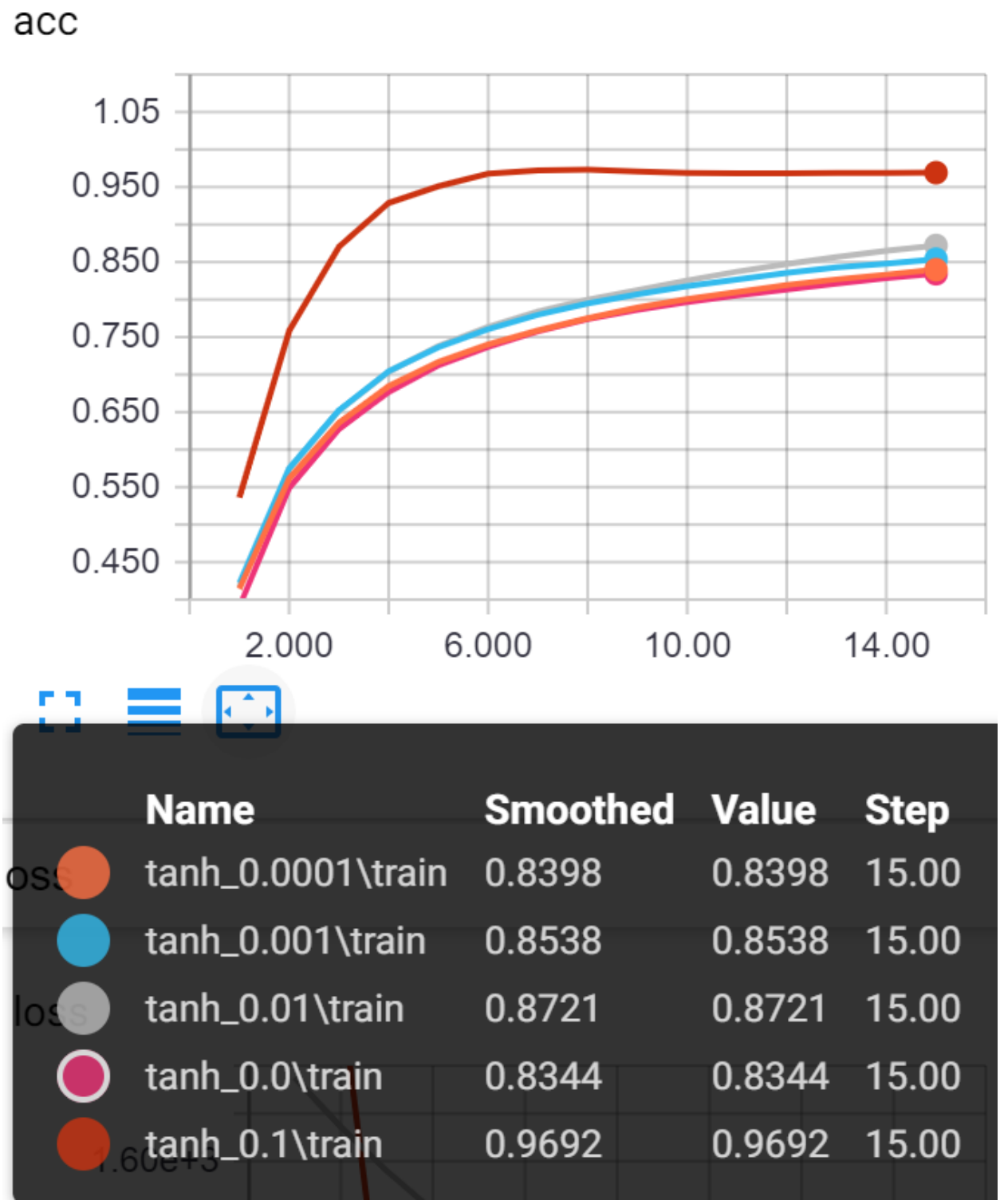


Question 4: According to acuracy and loss diagrams, the simpler model did better at early epochs but the model with more layer started to get better on the train set gradually and that is because of its model capacity which is higher. So we expect this model to do better with more amount of data on the train set. on the val set, the simpler model did better. The reason is that more complex models are more likely to overfit on the train set since they have more capacity to get more detailed on the train set. The distribution of weights didn't vary that much.(There are several tanh/train lines because I ran that part several times which were all saved in logs foler.)

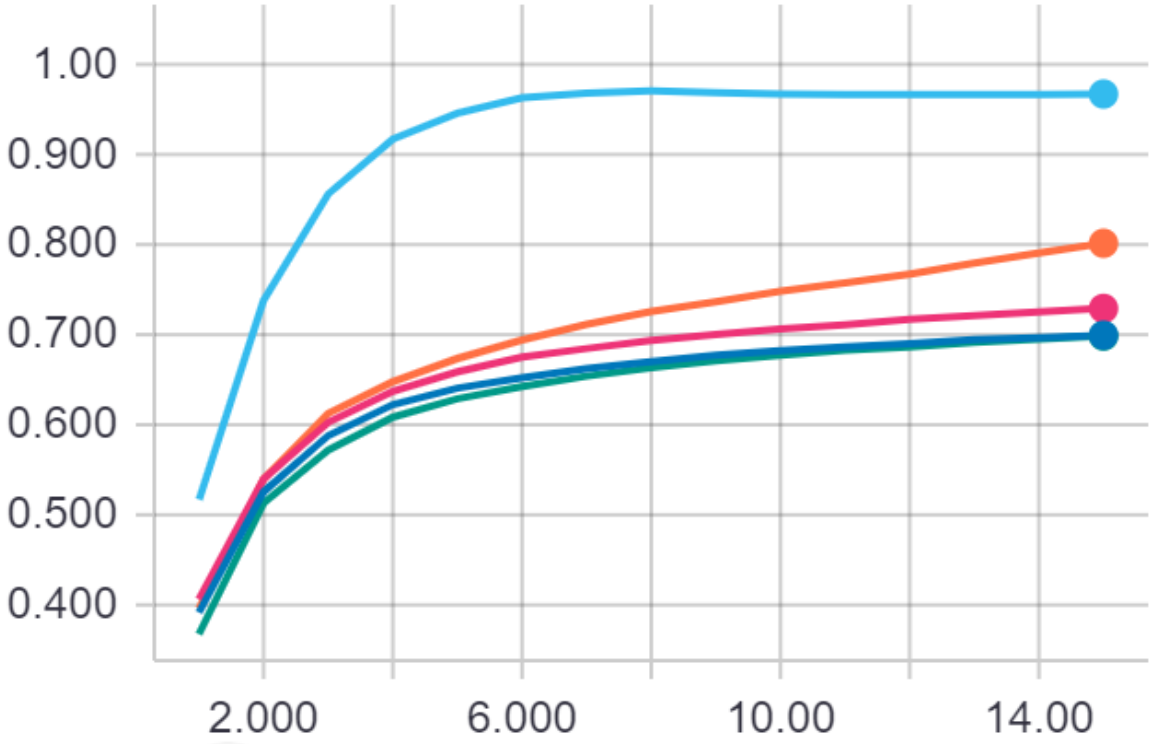


Question 5: Since loss depends on the regularization term, comparing loss values does not make sense so we just monitor the accuracy. I provided two different diagrams for training and validation accuracy below. As we see, less regularization term does better on the training set which is predictable since the crossentropy loss is a better proxy for the accuracy but it also can lead to overfit as we see. The gap between the train accuracy and validation increases. In the validation curve we also see that 0.01 was the best value for the regularization coefficient but it does not necessarily mean that the bigger the regularization term, the better the result on validation set. If we set the coefficient to a high value, the model prefers not to train and just lower the weights near zero. I ran a model with 10 as

coefficient and after 10 epochs, the model reached 20% on both train and validation set.



acc



Name	Smoothed	Value	Step	T
tanh_0.0001\val	0.6994	0.6994	15.00	W
tanh_0.001\val	0.7291	0.7291	15.00	W
tanh_0.01\val	0.8014	0.8014	15.00	W
tanh_0.0\val	0.6984	0.6984	15.00	W
tanh_0.1\val	0.9671	0.9671	15.00	W