Assignment 03 Solutions

1. After each stride-2 conv, why do we double the number of filters?

Ans. A stride 2 conv with the default padding (1) and ks (3) will reduce the activation map dimension by half. Formula: (n + 2*pad - ks)//stride + 1. As the activation map dimension reduces by half we double the number of filters. This results in no overall change in computation as the network gets deeper and deeper.

2. Why do we use a larger kernel with MNIST (with simple cnn) in the first conv ?

Ans. How does kernel size affect CNN?

Increasing kernel size means effectively increasing the total number of parameters. So, it is expected that the model has a higher complexity to address a given problem. So it should perform better at least for a particular training set.

3. What data is saved by ActivationStats for each layer ?

Ans. An activation function in a neural network defines how the weighted sum of the input is transformed into an output from a node or nodes in a layer of the network.

4. How do we get a learner's callback after they've completed training ?

Ans. The training loop is defined in Learner a bit below and consists in a minimal set of instructions: looping through the data we:

compute the output of the model from the input

calculate a loss between this output and the desired target

compute the gradients of this loss with respect to all the model parameters

update the parameters accordingly

zero all the gradients

5. What are the drawbacks of activations above zero ?

Ans. The two major problems with sigmoid activation functions are: Sigmoid saturate and kill gradients: The output of sigmoid saturates (i.e. the curve becomes parallel to x-axis) for a large positive or large negative number. Thus, the gradient at these regions is almost zero.

6.Draw up the benefits and drawbacks of practicing in larger batches ?

Ans.

Advantages	Disadvantages
Allows flexible production	Making many small batches can be expensive
Inventories of part-finished goods can be stored and completed later	If production runs are different there may be additional costs and delays in preparing equipment

7. Why should we avoid starting training with a high learning rate ?

Ans. A learning rate that is too large can cause the model to converge too quickly to a suboptimal solution, whereas a learning rate that is too small can cause the process to get stuck. The challenge of training deep learning neural networks involves carefully selecting the learning rate.

8. What are the pros of studying with a high rate of learning ?

Ans. Generally, a large learning rate allows the model to learn faster, at the cost of arriving on a sub-optimal final set of weights.

9. Why do we want to end the training with a low learning rate ?

Ans. Generally, a large learning rate allows the model to learn faster, at the cost of arriving on a sub-optimal final set of weights. A smaller learning rate may allow the model to learn a more optimal or even globally optimal set of weights but may take significantly longer to train.