Problem Set 2

The required weekly readings and lecture slides should be helpful in completing the assignment. You can find these on our course website. For your submission, you must submit a PDF that has been typed. For mathematical derivations, if easier, you can insert pictures of your handwritten derivation into the final PDF.

1. **Model Parameters versus Hyperparameters [3 points]:** In your own words, describe what is a model parameter and model hyperparameter and then describe how they differ.

2. **Dataset Splits [2 points]:** In your own words, describe the motivation for creating a validation dataset and how its purpose for developing deep learning models differs from that of the training dataset and test dataset.

3. **Overfitting versus Underfitting [4 points]:** Learning curves of training and validation losses for two different deep learning algorithms trained and tested on the same training and validation datasets are shown in Figure 1.

   (a) Indicate which sub-figure (i.e., a or b) is representative of underfitting and provide a justification.
   (b) Indicate which sub-figure (i.e., a or b) is representative of overfitting and provide a justification.
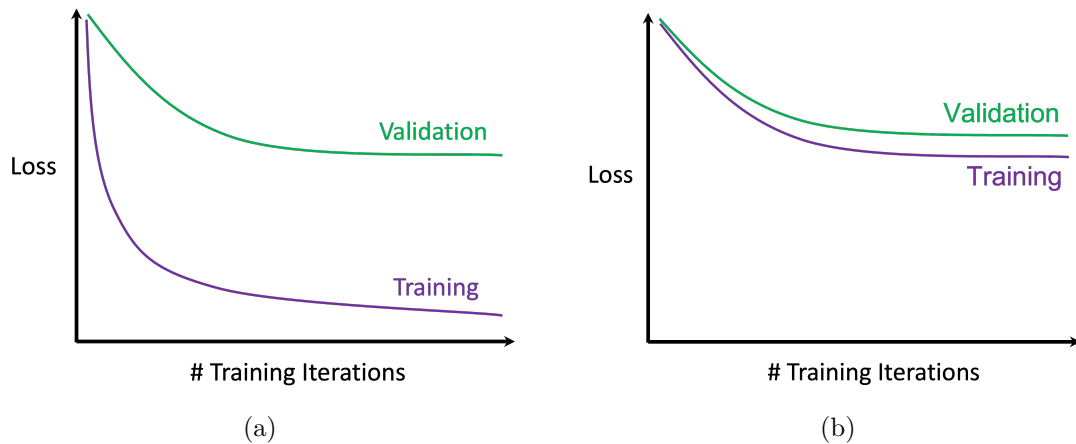

Figure 1: Learning curves for two deep learning models trained and tested with the same training and validation datasets are shown in (a) and (b). The scales of the x-axis and y-axis in both plots are the same.

4. **Optimization [2 points]:** In your own words, describe what is stochastic gradient descent, batch gradient descent, and mini-batch gradient descent and identify what makes them different from each other.

5. **Model Size [6 points]:** Assume you are building a 4 layer neural network (i.e., 3 hidden layers and 1 output layer) that takes as input a 64x64 matrix of integers and has 100 nodes at the output layer.

   (a) If this network was fully connected with 5 nodes per hidden layer, how many model parameters would it have? For full credit, you must indicate the number of model parameters between each pair of layers (e.g., input to hidden layer 1, hidden layer 1 to hidden layer 2, etc), the total number of weights in the network, and the total number of bias terms in the network.

   (b) If this network instead had convolutional layers for all hidden layers with 3 3x3 filters per layer and a fully connected layer between the final hidden layer and last layer, how many model parameters would it have? For full credit, you must indicate the number of model parameters between each pair of layers (e.g., input to hidden layer 1, hidden layer 1 to hidden layer 2, etc), the total number of weights in the network, and the total number of bias terms in the network.

6. **Convolutional Neural Networks [8 points]:**

   (a) In your own words, describe at least two advantages of using convolutional layers instead of fully connected layers in a neural network.

   (b) Compute and report the output from applying the filter on the input data, which are both shown below. For full-credit, you must show your mathematical derivation.

   (c) Compute and report the output from applying the filter on the input data when instead adding a zero padding layer and using 3x3 stride. For full-credit, you must show your mathematical derivation.

1	1	0	2
4	0	8	1
6	4	2	3
8	7	4	2

Data type 1: Input

0	0.5	0
0.5	1	0
0	0	0

Data type 2: Filter