Quiz Submissions - Quiz 7 - Attempt 2

Nhat Le (username: hung.le@mail.mcgill.ca)

Attempt 1

Written: Mar 15, 2019 11:02 PM - Mar 15, 2019 11:47 PM

Submission View

Released: Mar 15, 2019 11:59 PM

View the quiz answers.

Question 1 **1 / 1 point**

Which of the following is not a standard technique used to control over-fitting with neural networks:

- ✓ Hidden unit pruning
- L2 regularization
- Early stopping
- None of the above

Hide Feedback

"Hidden-unit pruning" is not a standard technique to control overfitting, while both L2 regularization and early stopping are techniques that were mentioned in class.

Question 2 **1 / 1 point**

Forward-mode and reverse-mode automatic differentiation are equally efficient for computing the derivative of the loss w.r.t. all model parameters.

- True
- ✓ False

Hide Feedback

Reverse-mode automatic differentiation gives the derivatives of the loss w.r.t. all parameters in one pass, but forward-mode does not. Instead forward-mode differentiation gives the derivative of all hidden layer outputs w.r.t. the input. (As discussed in Lecture 14.)

Question 3 **1 / 1 point**

Which of the following is not a benefit of adding momentum to gradient descent:

- ✓ Allows gradient to stabilize in regions where the loss is flat.
- Decreases chance of stopping in poor local minima.
- Increases the speed of convergence when the gradient remains constant.
- None of the above

Hide Feedback

Momentum helps to keep the weights *moving* in flat regions of the loss (as discussed in Lecture 14); it does not add stabilization.

Question 4 **1 / 1 point**

You are training two neural network models on a binary image classification dataset with 10x10 greyscale images.

- The first model is a feedforward neural network with two hidden layers. The first hidden layer has a dimension of 50 and the second has a dimension of 10.
- The second model is a convolutional neural network. It has one convolutional layer with 2x2 filters. There are 75 different convolutional filters applied in this layer (i.e., the "depth" or "number of channels" in this layer is 75) with stride 1 and zero padding.

Which model has more parameters?

- The feedforward neural network.
- ✓ The convolutional neural network.
- They have the same number of parameters.

Hide Feedback

For the FFNN, we have that the first layer has 100x50 + 1 parameters (i.e., the input is the flattened image of dimension 10x10=100, the hidden dimension is 50, and there is the bias term). The second layer has 50x10 + 1 parameters (i.e., 50 dimensional input and 10 dimensional output, with the bias term). And finally, the binary classification output layer has 10x1+1 parameters (i.e., 10 dimensional input and 1 dimensional output, plus the bias term). Thus, in total, the FFNN has 100x50+1+50x10+1+10+1=5513 parameters.

For the CNN, we have 75 2x2 convolutional filters, each with 2x2+1=5 parameters (including the bias term). Thus in total there are 75x5=375 parameters in the first convolutional layer. The output of this first layer is dimension 9x9x75=6075 (i.e., there are 75 different channels since we are applying 75

different convolutions). Thus, in total the CNN has 6075+375=6450 parameters, which is more than the FFNN.

Question 5 **1 / 1 point**

VGGNet was the first model to achieve "superhuman" performance on ImageNet.

- True
- ✓ False

Hide Feedback

VGGNet did not achieve superhuman performance (see, e.g., slide 29 in Lecture 15). The ResNet did achieve some results that were better than human baselines.

Attempt Score: 5 / 5 - 100 %

Overall Grade (highest attempt)**:** 5 / 5 - 100 %

Done