Quiz Submissions - Quiz 7 - Attempt 1

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

Attempt 1

Written: Mar 13, 2019 11:32 PM - Mar 13, 2019 11:41 PM

Submission View

Released: Mar 13, 2019 11:59 PM

View the quiz answers.

Question 1   **0 / 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

▷ View Feedback

Question 2   **1 / 1 point**

If you train a convolutional neural network on a training set with 4x4 sized images (with no padding used in the convolutional layers), then it is possible to run this trained network on validation set with 16x16 sized images without making any changes to the model.

⭕ True

✔️ ⭕ False

🔻 Hide Feedback

It would in principle be possible to run the first convolutional layer, since the convolutions operate on local patches, but there would likely be dimensionality mismatches in later layers. Most prominently, since the size of the image is different, the size of the representation before the final fully connected layer would be different and so it would not be possible to train the model.

Question 3   **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 4   **0 / 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 5   **1 / 1 point**

VGGNet is one of most popular architectures for CNNs because:

○ It currently has the 3rd highest top-5 accuracy on ImageNet, while using a relatively small number of parameters compared to other models.

✓ ○ It is a reasonably strong performer on ImageNet, has a homogenous model design, and is easily available to download/fine-tune.

○ It currently has the 3rd highest top-5 accuracy on ImageNet and is easily available to download/fine-tune.

○ Its use of residual connections allows for a very deep model with more than 30 convolutional layers, leading to a highly efficient model that is easy to download and fine tune.

VGGNet (discussed in Lecture 15) does not have the 3rd highest top-5 accuracy on ImageNet (it was state-of-the-art back in 2016) and it does not use residual connections.

Attempt Score: 3 / 5 - 60 %

Overall Grade (highest attempt)**:**   3 / 5 - 60 %

Done