

EECS545 Lecture 12 Quiz Solutions

1. Which of the following are the benefits of using CNNs, as opposed to fully connected networks, for image recognition tasks? (**Choose all options that apply**)
- (a) The ability to express a wider variety of more complicated functions of the input features.
 - (b) Fewer model architecture hyperparameters for the designer to select.
 - (c) Enables the network to more easily learn and recognize features regardless of their position in the image.
 - (d) Typically requires fewer data to train well.

Solution: (c) and (d)

- (a) False. Compared with CNNs, fully connected networks offer greater expressivity and model capacity as they have many more weights.
- (b) False. CNNs involve more architecture parameters (kernel size, stride, dilation, padding, pooling, etc.).
- (c) True. Since CNNs share weights, the same features can be learned and recognized at different positions
- (d) True. Fewer parameters often imply less data is required to train a reasonable model.

2. Which of the following statement are true about batch normalization in neural network training? (**Choose all options that apply**)
- (a) It addresses overfitting issue.
 - (b) It restricts activations from becoming too high or low.
 - (c) It makes training faster.
 - (d) It is a non-linear transformation to center the training dataset around the origin.

Solution: (b) and (c).

- (a) False. Batch norm does not change the number of parameters to learn. (d) False. no non-linear function is introduced in batch normalization.

3. How many parameters do we need to learn for the following network structure? A $32 \times 32 \times 3$ image input, followed by a convolution layer with 8 filters of size 3×3 (stride 1 and 1 pixel of zero-padding), then another convolution layer with 4 filters of size 2×2 (stride 2 and no zero-padding), and finally a max-pooling layer with a 2×2 filter (stride 2 and no zero-padding). (Note: the depth of all filters are not explicitly spelled out, and we assume no bias/intercept terms are used here.)
- (a) 59
 - (b) 88

- (c) 264
- (d) 344

Solution: (d). $8 \times (3 \times 3 \times 3) + 4 \times (2 \times 2 \times 8) = 344$

4. (Continued from 3) What is the final output dimension of the last question?

- (a) $8 \times 8 \times 1$
- (b) $16 \times 16 \times 1$
- (c) $8 \times 8 \times 4$
- (d) $16 \times 16 \times 4$

Solution: (c). $32 \times 32 \times 3$ -(1st conv)- $32 \times 32 \times 8$ -(2nd conv)- $16 \times 16 \times 4$ -(maxpool)- $8 \times 8 \times 4$

5. (True/False) Using Momentum, instead of SGD, always helps the network converge faster.

Solution: False. Use of active learning (including Momentum) is reasonable in training a network, but too much momentum will be harmful in training a network.