

1. When building a ConvNet, typically you start with some POOL layers followed by some CONV layers. 1 point

False  
 True

2. LeNet - 5 made extensive use of padding to create valid convolutions, to avoid increasing the number of channels after every convolutional layer. True/False? 1 point

True  
 False

3. Training a deeper network (for example, adding additional layers to the network) allows the network to fit more complex functions and thus almost always results in lower training error. For this question, assume we're referring to "plain" networks. 1 point

True  
 False

4. The following equation captures the computation in a ResNet block. What goes into the two blanks above? 1 point

$$a^{[l+2]} = g(W^{[l+2]}g(W^{[l+1]}a^{[l]} + b^{[l+1]}) + b^{[l+2]} + \underline{\hspace{2cm}}) + \underline{\hspace{2cm}}$$

$z^{[l]}$  and  $a^{[l]}$ , respectively  
 0 and  $z^{[l+1]}$ , respectively  
 0 and  $a^{[l]}$ , respectively  
  $a^{[l]}$  and 0, respectively

5. Which ones of the following statements on Residual Networks are true? (Check all that apply.) 1 point

The skip-connection makes it easy for the network to learn an identity mapping between the input and the output within the ResNet block.  
 Using a skip-connection helps the gradient to backpropagate and thus helps you to train deeper networks  
 A ResNet with  $L$  layers would have on the order of  $L^2$  skip connections in total.  
 The skip-connections compute a complex non-linear function of the input to pass to a deeper layer in the network.

6.  $1 \times 1$  convolutions are the same as multiplying by a single number. True/False? 1 point

True  
 False

7. Which of the following are true about the inception Network? (Check all that apply) 1 point

Inception blocks allow the use of a combination of  $1 \times 1$ ,  $3 \times 3$ ,  $5 \times 5$  convolutions and pooling by stacking up all the activations resulting from each type of layer.  
 Making an inception network deeper won't hurt the training set performance.  
 Inception blocks allow the use of a combination of  $1 \times 1$ ,  $3 \times 3$ ,  $5 \times 5$  convolutions, and pooling by applying one layer after the other.  
 One problem with simply stacking up several layers is the computational cost of it.

8. When having a small training set to construct a classification model, which of the following is a strategy of transfer learning that you would use to build the model? 1 point

It is always better to train a network from a random initialization to prevent bias in our model.  
 Use an open-source network trained in a larger dataset. Use these weights as an initial point for the training of the whole network.  
 Use an open-source network trained in a larger dataset, freeze the softmax layer, and re-train the rest of the layers.  
 Use an open-source network trained in a larger dataset freezing the layers and re-train the softmax layer.

9. In Depthwise Separable Convolution you: 1 point

You convolve the input image with a filter of  $n_f \times n_f \times n_c$  where  $n_c$  acts as the depth of the filter ( $n_c$  is the number of color channels of the input image).  
 Perform one step of convolution.  
 For the "Depthwise" computations each filter convolves with only one corresponding color channel of the input image.  
 Perform two steps of convolution.  
 The final output is of the dimension  $n_{out} \times n_{out} \times n_c$  (where  $n_c$  is the number of color channels of the input image).  
 You convolve the input image with  $n_c$  number of  $n_f \times n_f$  filters ( $n_c$  is the number of color channels of the input image).  
 For the "Depthwise" computations each filter convolves with all of the color channels of the input image.  
 The final output is of the dimension  $n_{out} \times n_{out} \times n_c'$  (where  $n_c'$  is the number of filters used in the pointwise convolution step).

10. Suppose that in a MobileNet v2 Bottleneck block the input volume has shape  $64 \times 64 \times 16$ . If we use 32 filters for the expansion and 16 filters for the projection. What is the size of the input and output volume of the depthwise convolution, assuming a pad='same'? 1 point

- $64 \times 64 \times 32, 64 \times 64 \times 16$
- $64 \times 64 \times 32, 64 \times 64 \times 32$
- $64 \times 64 \times 16, 64 \times 64 \times 32$
- $32 \times 32 \times 32, 32 \times 32 \times 32$