1. Which of the following do you typically see in a ConvNet? (Check all that apply.) 1 / 1 point

- [x] FC layers in the last few layers

  ✓ **Correct**

  True, fully-connected layers are often used after flattening a volume to output a set of classes in classification.

- [ ] FC layers in the first few layers

- [ ] Multiple POOL layers followed by a CONV layer

- [x] Multiple CONV layers followed by a POOL layer

  ✓ **Correct**

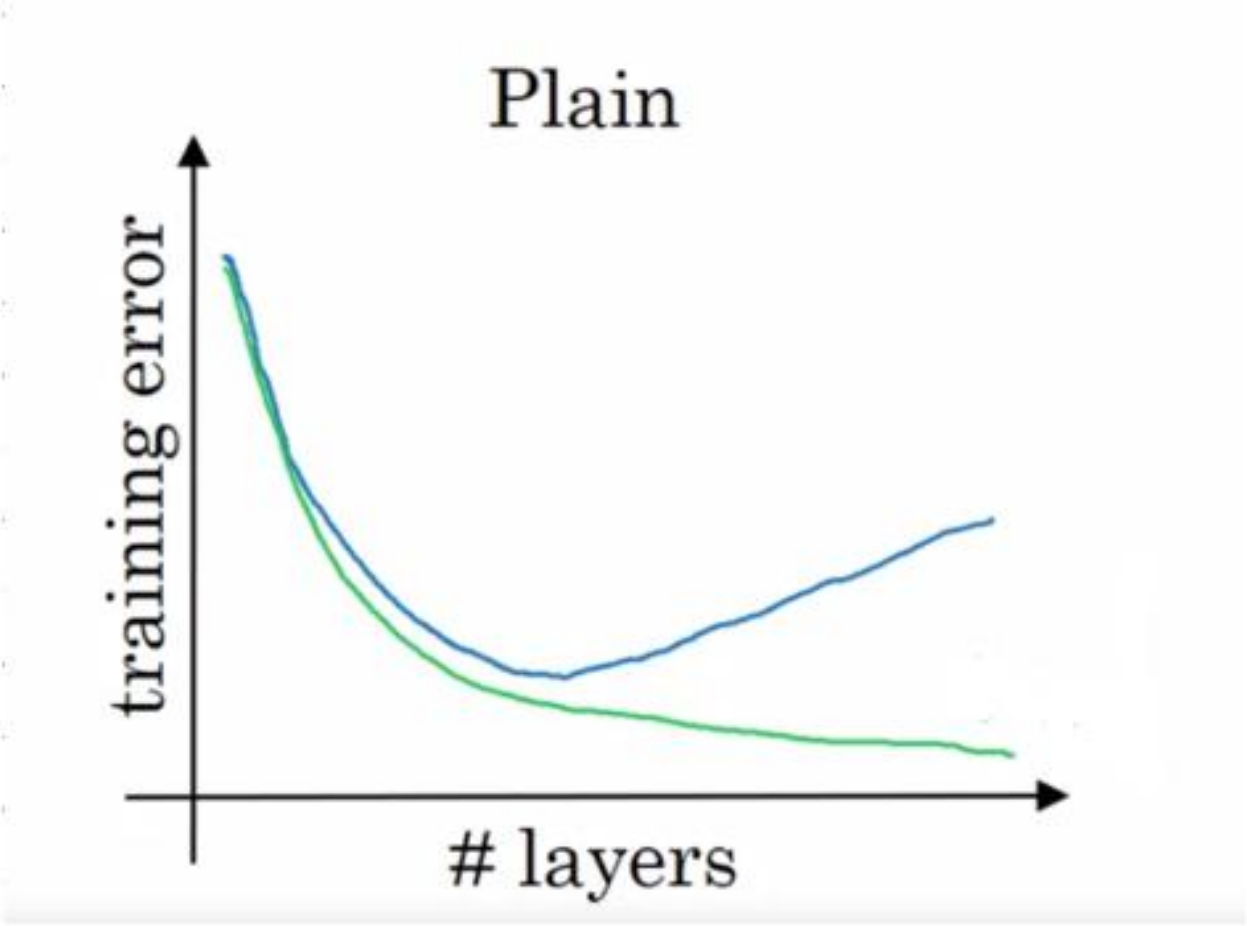
  True, as seen in the case studies.

2. In order to be able to build very deep networks, we usually only use pooling layers to downsize the height/width of the activation volumes while convolutions are used with "valid" padding. Otherwise, we would downsize the input of the model too quickly.

1 / 1 point

- ○ True
- ◉ False

3. Based on the lectures, in the following picture, which curve corresponds to the expected behavior in theory, and which one corresponds to the behavior we get in practice? This when using plain neural networks. 1 / 1 point



- The blue one depicts the theory, and the green one the reality.
- The green one depicts the results in theory, and also in practice.
- The green one depicts the results in theory, and the blue one the reality.
- The blue one depicts the results in theory, and also in practice.

Expand

Correct
Yes, in theory, we expect that as we increase the number of layers the training error decreases; but in practice after a certain number of layers the error increases.

4. The computation of a ResNet block is expressed in the equation:

1 / 1 point

$$a^{[l+2]} = g\left(W^{[l+2]}\, g\left(W^{[l+1]}\, a^{[l]} + b^{[l+1]}\right) + b^{[l+2]} + a^{[l]}\right)$$

C A B

Which part corresponds to the skip connection?

- The equation of ResNet.
- The term in the orange box, marked as B.
- The term in the blue box, marked as A.
- The term in the red box, marked as C.

Expand

Correct

Yes, this term is the result of the skip connection or shortcut.

5. In the best scenario when adding a ResNet block it will learn to approximate the identity function after a lot of training, helping improve the overall performance of the network. True/False? 1 / 1 point

- (●) False
- ( ) True

Expand

Correct
Correct. When adding a ResNet block it can easily learn to approximate the identity function, thus in a worst-case scenario, it will not affect the performance of the network at all.

6. For a volume of $125 \times 125 \times 64$ which of the following can be used to reduce this to a $125 \times 125 \times 32$ volume?

1 / 1 point

- Use a POOL layer of size 2×2 but with a stride of 1.
- Use a 1×1 convolutional layer with a stride of 1, and 32 filters.
- Use a 1×1 convolutional layer with a stride of 2, and 32 filters.
- Use a POOL layer of size 2×2 with a stride of 2.

Expand

Correct

Yes, since using 1×1 convolutions is a great way to reduce the depth dimension without affecting the other dimensions.

7. Which of the following are true about bottleneck layers? (Check all that apply) 0 / 1 point

- [x] By adding these layers we can reduce the computational cost in the inception modules.

  ✓ **Correct**

  Yes, by using the 1×1 convolutional layers we can reduce the depth of the volume and help reduce the computational cost of applying other convolutional layers with different filter sizes.

- [x] Bottleneck layers help to compress the 1x1, 3x3, 5x5 convolutional layers in the inception network.

  ! **This should not be selected**

  No, the bottleneck layer doesn't combine any of these different layers.

- [x] The use of bottlenecks doesn't seem to hurt the performance of the network.

  ✓ **Correct**

  Yes, although it reduces the computational cost significantly.

- [ ] The bottleneck layer has a more powerful regularization effect than Dropout layers.

Expand

⊗ **Incorrect**

You chose the extra incorrect answers.

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 / 1 point

- (●) Use an open-source network trained in a larger dataset freezing the layers and re-train the softmax layer.
- ( ) Use an open-source network trained in a larger dataset, freeze the softmax layer, and re-train the rest of the layers.
- ( ) 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.

Expand

Correct
Yes, this is a strategy that can provide a good result with small data.

9. Which of the following are true about Depthwise-separable convolutions? (Choose all that apply) 1 / 1 point

- [x] The pointwise convolution convolves the output volume with 1×1 filters.

  ✓ Correct
  Yes, the number of filters for the output of the depthwise-separable convolution is determined by the number of 1×1 filters used.

- [ ] The depthwise convolution convolves the input volume with 1×1 filters over the depth dimension.

- [x] Depthwise-separable convolutions are composed of two different types of convolutions.

  ✓ Correct
  Yes, it is composed of a depthwise convolution followed by a pointwise convolution.

- [x] The depthwise convolution convolves each channel in the input volume with a separate filter.

  ✓ Correct
  Yes, the output of this kind of convolution is the same as the input.

Expand

Correct
Great, you got all the right answers.

10. Suppose that in a MobileNet v2 Bottleneck block we have an $n \times n \times 5$ input volume, we use 30 filters for the expansion, in the depthwise convolutions we use 3×3 filters, and 20 filters for the projection. How many parameters are used in the complete block, suppose we don't use bias?

1 / 1 point

- 80
- 8250
- 1101
- 1020

Expand

Correct

Yes, the expansion filters use $5 \times 30 = 150$ parameters, the depthwise convolutions need $3 \times 3 \times 30 = 270$ parameters, and the projection part $30 \times 20 = 600$ parameters.