1. **What do you think applying this filter to a grayscale image will do?**

$$\begin{bmatrix} 0 & 1 & -1 & 0 \\ 1 & 3 & -3 & -1 \\ 1 & 3 & -3 & -1 \\ 0 & 1 & -1 & 0 \end{bmatrix}$$

A. Detect horizontal edges
B. Detect 45 degree edges
C. Detect image contrast
D. Detect vertical edges

2. **Suppose your input is a 300 by 300 color(RGB) image, and you are not using a convolutional network. If the first hidden layer has 100 neurons, each one fully connected to the input, how many parameters does this hidden layer have(including the bias parameters)**

A. 9,000,001
B. 9,000,100
C. 27,000,001
D. 27,000,100

3. **Suppose your input is a 300 by 300 color(RGB) image, and you use a convolutional layer with 100 filters that are each 5*5. How many parameters does this hidden layer have(including the bias parameters)**

A. 2501
B. 2600
C. 7500
D. 7600

4. **You have an input volume that is 63*63*16, and convolve it with 32 filters that are each 7*7, using a stride of 2 and no padding. What is the output volume?**

A. 16*16*16
B. 16*16*32
C. 29*29*32
D. 29*29*16

5. **You have an input volume that is 15*15*8, and pad it using "pad=2",. What is dimension of the resulting volume(after padding)?**

A. 17*17*8
B. 19*19*12
C. 17*17*10
D. 19*19*8

6. **You have an input volume that is 63*63*16, and convolve it with 32 filters that are each 7*7, using a stride of 1, you want to use a "same" convolution. What is the padding?**

A. 1
B. 2

C. 3
D. 7

7. You have an input volume that is 32*32*16, and apply max pooling with a stride of 2 and a filter of 2. What is the output volume?

A. 15*15*16
B. 16*16*8
C. 16*16*16
D. 32*32*8

8. Because pooling layers do not have parameters, they do not affect the backpropagation(derivatives) calculation.

A. True
B. False

9. In lecture we talked about "parameters sharing" as a benefit of using convolutional networks. Which of the following statements about parameter sharing in ConvNets are true?(Check all that apply)

A. It allows parameters learned for one task to be shared even for a different task(transfer learning).
B. It reduces that total number of parameters, thus reducing overfitting.
C. It allows gradient descent to set many of the parameters to zero, thus making the connection sparse.
D. It allows a feature detector to be used in multiple locations throughout the whole input image/input volume