Congratulations! You passed!

Grade received 100% **Latest Submission Grade** 100% **To pass** 80% or higher

Retake the assignment in **23h 50m**

Go to next item

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

$$\begin{bmatrix} -1 & -1 & 2 \\ -1 & 2 & 1 \\ 2 & 1 & 1 \end{bmatrix}$$

- Detect vertical edges.
- Detect 45-degree edges.
- Detect horizontal edges.
- Detecting image contrast.

Expand

Correct
Correct. Notice that there is a high delta between the values in the top left part and the ones in the bottom right part. When convolving this filter on a grayscale image, the edges forming a 45-degree angle with the horizontal will be detected.

2. Suppose your input is a 128 by 128 grayscale image, and you are not using a convolutional network. If the first hidden layer has 256 neurons, each one fully connected to the input, how many parameters does this hidden layer have (including the bias parameters)? **1 / 1 point**

- 4194560
- 12583168
- 12582912
- 4194304

Expand

Correct
Correct, the number of inputs for each unit is 128×128 since the input image is grayscale, so we need $128 \times 128 \times 256$ parameters for the weights and 256 parameters for the bias thus $128 \times 128 \times 256 + 256 = 4194560$.

3. Suppose your input is a 256 by 256 grayscale image, and you use a convolutional layer with 128 filters that are each 3×3. How many parameters does this hidden layer have (including the bias parameters)? **1 / 1 point**

- 1280
- 1152
- 3584
- 75497600

Expand

Correct
Yes, since the input volume has only one channel each filter has $3 \times 3 + 1$ weights including the bias, thus the total is $(3 \times 3 + 1) \times 128$.

4. You have an input volume that is $127 \times 127 \times 16$, and convolve it with 32 filters of 5×5, using a stride of 2 and no padding. What is the output volume? **1 / 1 point**

- $62 \times 62 \times 32$
- $123 \times 123 \times 16$
- $62 \times 62 \times 16$
- $123 \times 123 \times 32$

Expand

Correct
Correct, using the formula $n_H^{[l]} = \frac{n_H^{[l-1]} + 2 \times p - f}{s} + 1$ with $n_H^{[l-1]} = 127, p = 0, f = 5$, and $s = 2$ we get 62.

5. You have an input volume that is 31x31x32, and pad it using "pad=1". What is the dimension of the resulting volume (after padding)? **1 / 1 point**

- 31x31x34
- 33x33x32
- 33x33x33
- 32x32x32

Expand

Correct
Yes, if the padding is 1 you add 2 to the height dimension and 2 to the width dimension.

6. You have an input volume that is 63x63x16, and convolve it with 32 filters that are each 7x7, and stride of 1. You want to use a "same" convolution. What is the padding? **1 / 1 point**

- 3
- 1
- 7
- 2

Expand

Correct
Correct, you need to satisfy the following equation: $n_H - f + 2 \times p + 1 = n_H$ as you want to keep the dimensions between the input volume and the output volume.

7. You have an input volume that is 32x32x16, and apply max pooling with a stride of 2 and a filter size of 2. What is the output volume? **1 / 1 point**

- 32x32x8
- 16x16x8
- 15x15x16
- 16x16x16 (selected)

Expand

Correct

Correct, using the following formula: $n_H^{[l]} = \frac{n_H^{[l-1]} + 2 \times p - f}{s} + 1$

8. Because pooling layers do not have parameters, they do not affect the backpropagation (derivatives) calculation. **1 / 1 point**

- True
- False (selected)

Expand

Correct

Everything that influences the loss should appear in the backpropagation because we are computing derivatives. In fact, pooling layers modify the input by choosing one value out of several values in their input volume. Also, to compute derivatives for the layers that have parameters (Convolutions, Fully-Connected), we still need to backpropagate the gradient through the Pooling layers.

9. Which of the following are true about convolutional layers? (Check all that apply) **1 / 1 point**

- [x] It allows a feature detector to be used in multiple locations throughout the whole input volume.

  Correct

  Yes, since convolution involves sliding the filter throughout the whole input volume the feature detector is computed over all the volume.

- [ ] It speeds up the training since we don't need to compute the gradient for convolutional layers.
- [ ] It allows parameters learned for one task to be shared even for a different task (transfer learning).
- [x] Convolutional layers provide sparsity of connections.

  Correct

  Yes, this happens since the next activation layer depends only on a small number of activations from the previous layer.

Expand

Correct

Great, you got all the right answers.

10. In lecture we talked about "sparsity of connections" as a benefit of using convolutional layers. What does this mean? **1 / 1 point**

- Each activation in the next layer depends on only a small number of activations from the previous layer. (selected)
- Each filter is connected to every channel in the previous layer.
- Each layer in a convolutional network is connected only to two other layers

Regularization causes gradient descent to set many of the parameters to zero.

Expand

Correct

Yes, each activation of the output volume is computed by multiplying the parameters from with a volumic slice of the input volume and then summing all these together.