Congratulations! You passed!

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

Go to next item

1. Face verification requires comparing a new picture against one person's face, whereas face recognition requires comparing a new picture against K persons' faces. — 1 / 1 point

- False
- True (selected)

Expand

Correct
Correct.

2. Why do we learn a function $d(img1, img2)$ for face verification? (Select all that apply.) — 1 / 1 point

- Given how few images we have per person, we need to apply transfer learning.
- We need to solve a one-shot learning problem. (selected)

Correct
This is true as explained in the lecture.

- This allows us to learn to recognize a new person given just a single image of that person. (selected)

Correct
Yes.

- This allows us to learn to predict a person's identity using a softmax output unit, where the number of classes equals the number of persons in the database plus 1 (for the final "not in database" class).

Expand

Correct
Great, you got all the right answers.

3. In order to train the parameters of a face recognition system, it would be reasonable to use a training set comprising 100,000 pictures of 100,000 different persons. — 1 / 1 point

- True
- False (selected)

Expand

Correct
Correct, to train a network using the triplet loss you need several pictures of the same person.

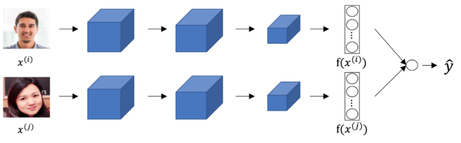
4. Which of the following is a correct definition of the triplet loss? Consider that $\alpha > 0$. (We encourage you to figure out the answer from first principles, rather than just refer to the lecture.) — 1 / 1 point

- $max(||f(A) - f(P)||^2 - ||f(A) - f(N)||^2 - \alpha, 0)$
- $max(||f(A) \quad f(N)||^2 \quad ||f(A) \quad f(P)||^2 + \alpha, 0)$
- $max(||f(A) - f(P)||^2 - ||f(A) - f(N)||^2 + \alpha, 0)$ (selected)
- $max(||f(A) - f(N)||^2 - ||f(A) - f(P)||^2 - \alpha, 0)$

Expand

Correct
Correct

5. Consider the following Siamese network architecture: — 1 / 1 point



The upper and lower neural networks have different input images, but have exactly the same parameters.

- True (selected)
- False

Expand

Correct
Yes it is true, parameters are shared among these two networks.

6. Our intuition about the layers of a neural network tells us that units that respond more to complex features are more likely to be in deeper layers. True/False? — 1 / 1 point

- True (selected)
- False

Expand

Correct
Correct. Neurons that understand more complex shapes are more likely to be in deeper layers of a neural network.

7. Neural style transfer uses images Content C, Style S. The loss function used to generate image G is composed of which of the following: (Choose all that apply.) — 1 / 1 point

- J that calculates the triplet loss between S, G, and C.
- $J_{content}$ that compares C and G. (selected)

Correct
Correct, in neural style transfer we are interested in the similarity between S and G, and the similarity between G and C.

- J_{style} that compares S and G. (selected)

Correct
Correct, in neural style transfer we are interested in the similarity between S and G, and the similarity between G and C.

- J_{corr} that compares C and S

Expand

Correct
Great, you got all the right answers.

8. In the deeper layers of a ConvNet, each channel corresponds to a different feature detector. The style matrix $G^{[l]}$ measures the degree to which the activations of different feature detectors in layer l vary (or correlate) together with each other. — 1 / 1 point

- True (selected)
- False

Expand

Correct
Yes, the style matrix $G^{[l]}$ can be seen as a matrix of cross-correlations between the different feature detectors.

9. In neural style transfer, which of the following better express the gradients used? — 0 / 1 point

- $\frac{\partial J}{\partial S}$
- $\frac{\partial J}{\partial W^{[l]}}$ (selected)
- $\frac{\partial J}{\partial G}$
- Neural style transfer doesn't use gradient descent since there are no trainable parameters.

Expand

Incorrect
Incorrect, we use the gradient of the cost function over the value of the pixels of the generated image.

10. You are working with 3D data. The input "image" has size $64 \times 64 \times 64 \times 3$, if you apply a convolutional layer with 16 filters of size $4 \times 4 \times 4$, zero padding and stride 2. What is the size of the output volume? — 0 / 1 point

- $31 \times 31 \times 31 \times 3$ (selected)
- $31 \times 31 \times 31 \times 16$
- $61 \times 61 \times 61 \times 14$
- $64 \times 64 \times 64 \times 3$

Expand

Incorrect
Notice that the convolutional layer has 16 filters.