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 point*

- ( ) False
- (•) True

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

- [x] We need to solve a one-shot learning problem.
- [ ] This allows us to learn to recognize a new person given just a single image of that person.
- [ ] Given how few images we have per person, we need to apply transfer learning.
- [x] 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).

3. You want to build a system that receives a person's face picture and determines if the person is inside a workgroup. You have pictures of all the faces of the people currently in the workgroup, but some members might leave, and some new members might be added. To train a system to solve this problem using the triplet loss you get many persons and take several pictures of each one. Which of the following do you agree with? (Select the best answer.) *1 point*

- ( ) You shouldn't use persons outside the workgroup you are interested in because that might create a high variance in your model.
- (•) You take several pictures of the same person to train $d(\mathrm{img}_1, \mathrm{img}_2)$ using the triplet loss.
- ( ) You take several pictures of the same person because this way you can get more pictures to train the network efficiently since you already have the person in place.
- ( ) It would be best to increase the number of persons in the dataset by taking only one picture of each person to have a more representative set of the population.

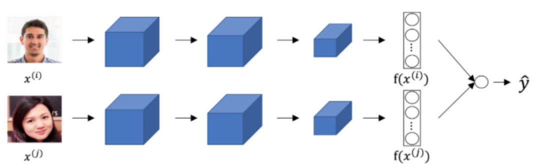
4. In the triplet loss: *1 point*

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

Which of the following are true about the triplet loss? Choose all that apply.

- [x] $f(A)$ represents the encoding of the Anchor.
- [x] We want that $\|f(A) - f(P)\|^2 < \|f(A) - f(N)\|^2$ so the negative images are further away from the anchor than the positive images.
- [ ] A the anchor image is a hyperparameter of the Siamese network.
- [ ] α is a trainable parameter of the Siamese network.

5. Consider the following Siamese network architecture: *1 point*



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

- ( ) False
- (•) True

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 point*

- ( ) False
- (•) True

7. In neural style transfer, we train the pixels of an image, and not the parameters of a network. *1 point*

- ( ) False
- (•) True

8. In neural style transfer, we define style as: *1 point*

- ( ) The correlation between the generated image G and the style image S.
- ( ) The correlation between the activation of the content image C and the style image S.
- (•) The correlation between activations across channels of an image.
- ( ) $\left\|a^{[l](S)} - a^{[l](G)}\right\|^2$ the distance between the activation of the style image and the content image.

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

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

10. You are working with 3D data. The input "image" has size $32 \times 32 \times 32 \times 3$, if you apply a convolutional layer with 16 filters of size $4 \times 4 \times 4$, zero padding and stride 1. What is the size of the output volume? *1 point*

- ( ) $29 \times 29 \times 29 \times 3$.

- (selected) $29 \times 29 \times 29 \times 16$.
- $31 \times 31 \times 31 \times 16$.
- $29 \times 29 \times 29 \times 13$.