You are training an RNN, and find that your weights and activations are all taking on the value of NaN ("Not a Number"). Which of these is the most likely cause of this problem?

- a. Vanishing gradient problem
- b. Sigmoid activation function was used
- c. Exploding gradient problem ✔
- d. ReLU activation function was used

What do you train an autoencoder for?

a. For supervised image-to-image translation

b. Unconditional image generation

c. Classification

d. For unsupervised representational learning ✔

Which of the following do you typically see in a ConvNet?

- [x] a. Multiple CONV layers followed by a POOL layer ✔
- [x] b. FC layers in the last few layers ✔
- [ ] c. Multiple POOL layers followed by a CONV layer
- [ ] d. FC layers in the first few layers

Which of the following statements about parameter sharing in ConvNets are true?

- [x] a. It reduces the total number of parameters, thus reducing overfitting. ✔
- [x] b. It allows a feature detector to be used in multiple locations throughout the whole input image/input volume. ✔
- [ ] c. It allows gradient descent to set many of the parameters to zero, thus making the connections sparse.
- [ ] d. It allows parameters learned for one task to be shared even for a different task (transfer learning).

Given an n-character word, we want to predict which character would be the n+1th character in the sequence. For example, our input is “predictio” (which is a 9 character word) and we have to predict what would be the 10th character.

Which neural network architecture would be suitable to complete this task?

- a. Recurrent Neural Network ✔
- b. Restricted Boltzmann Machine
- c. Fully-Connected Neural Network
- d. Convolutional Neural Network