

---

**Note 1:** There may be one or multiple correct choices for the first 4 questions.

**Note 2:** Briefly answer the questions 5 to 9.

1. What do you think of the pace of the course so far? (10 points)  
(a) Too slow.      (b) Just right.      (c) Too fast.
2. What are the hyperparameters of a convolution layer? (10 points)  
(a) Groups  
(b) Kernel size  
(c) Input channel dimension  
(d) Output channel dimension
3. Which of the following are true about Recurrent Neural Networks (RNNs)? (10 points)  
(a) RNNs can be interpreted as Convolutional Neural Networks (CNNs)  
(b) LSTM and GRU are preferred in practice than raw RNNs  
(c) RNNs cannot have many layers  
(d) Bi-directional RNNs can be used to capture the global information in the input
4. Which of the following are useful for mitigating the gradient vanishing problem? (10 points)  
(a) Using residual connections  
(b) Using batch normalization  
(c) Increasing the number of layers  
(d) Using SGD + momentum
5. What is the problem, even in a very shallow network (either convolutional or fully-connected), with initializing weights to all zeros? (10 points)
6. Name three data augmentation techniques that we can use to train neural networks to recognize different breeds of dogs. (10 points)

7. Name two advantages of using Recurrent Neural Networks over Convolutional Neural Networks to process text. (10 points)
8. Name two reasons why Batch Normalization is not suitable for Recurrent Neural Networks. (15 points)
9. You have a dataset  $\mathcal{D}_1$  with 1 million labeled samples for classification and another dataset  $\mathcal{D}_2$  with 100 labeled samples. Your friend trains a deep neural network with randomly initialized weights (from scratch) on  $\mathcal{D}_2$ . You decide to train a model on  $\mathcal{D}_1$  first and then do transfer learning on  $\mathcal{D}_2$  then. State one problem your friend may likely to find with their approach. **Briefly** explain how your approach solves the same problem. (15 points)