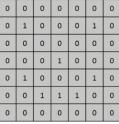
## Scalable Machine Learning and Deep Learning - Review Questions 3

Deadline: November 24, 2019

1.	Is it OK to initialize all the weights of a neural network to the same value as long as that value is selected
	randomly using He initialization? Is it okay to initialize the bias terms to 0?

- 2. In which cases would you want to use each of the following activation functions: ELU, leaky ReLU, ReLU, tanh, logistic, and softmax?
- 3. What is batch normalization and why does it work?
- 4. Does dropout slow down training? Does it slow down inference (i.e., making predictions on new instances)?
- 5. What may happen if you set the momentum hyperparameter too close to 1? E.g., keras.optimizers.SGD(momentum=0.99999)
- 6. Consider a CNN composed of three convolutional layers, each with  $3 \times 3$  filters, a stride of 2, and SAME padding. The lowest layer outputs 100 feature maps, the middle one outputs 200, and the top one outputs 400. The input images are RGB images of  $200 \times 300$  pixels. What is the total number of parameters w in the CNN?
- 7. Consider a CNN with one convolutional layer, in which it has a 3 × 3 filter (as shown below) and a stride of 2. Please write the output of this layer for the given input image (the left image in the following figure)?



Image



Filter