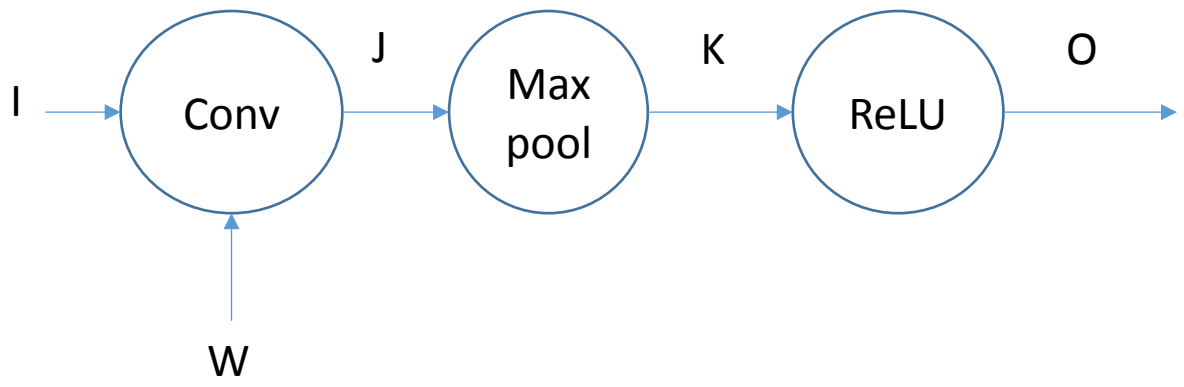
Q1. Compute the forward pass of the following part of a convolutional net:



If I is the following image patch

1	3	2	4	6	4
4	8	3	1	0	2
2	1	4	3	9	1
4	7	2	3	9	2

and W is the following 3-by-3 filter matrix

-1	0	1
-1	0	1
-1	0	1

Compute J, K and O. You are not doing any zero padding while doing the convolution. Assume stride size 1 for the convolution. Assume a 2-by-2 max pooling with stride 2. Write J, K and O below.

Q2. Now Assume that ideal output IO = [7, 10]. Also assume the loss as 0.5 times the square of Euclidean distance between IO and O. Compute back-propagation for O, K, J and W.

Q3: Suppose a function compute predicted class probabilities as: $y^p = softmax(Xw)$ where X is the feature matrix, w is the parameter vector. Let the loss function be cross entropy between y^p and ground truth vector y. Compute gradient of cross entropy loss with respect to parameter vector w.

Q4. Apply chain rule for a fully connected neural network, such as: $y^p = softmax(relu(Xw_1)w_2)$, where w_1 and w_2 are parameter vectors/matrices.

Q5: Demonstrate application of chain rule of derivative to the neural net with output: $y^p = \sigma(xW_1 + b_1)W_2 + b_2$ with respect to a L2-loss using ground truth vector y. Here σ is the sigmoid function, x is input vector. Using chain rule compute gradients of the L2-loss with respect to all four parameters: W_1, b_1, W_2, b_2.

Q6: Suppose for MNIST logits values are [-0.1,12.0,3.0,-3.2,-8.0,17.0,1.9,0.6,4.9,-12.0]. Compute softmax function on logits and write them.

Q7: Compute cross entropy loss value between softmax values in Q6 against a ground truth vector [0,1,0,0,0,0,0,0,0,0].

Q8: What is the minimum value of cross-entropy loss? What is the maximum value?

Q9: For image classification task, explain in two to three sentences why machine learning is a more suitable tool than rule-based AI?

Q10: Explain in two to three sentences how k-nearest neighbor method works.

Q11: In two to three sentences explain underfitting and overfitting in machine learning.

Q12: You are using a convolutional neural net architecture for image classification. You notice that your convnet is underfitting the dataset. How would you overcome underfitting here?

Q13: How would you tackle overfitting in an image classification task with convnet?

Q14: Starting with x=-1.0 and y=1.0, apply gradient descent to minimize $f(x, y) = x^2 + y^4$. First write gradient of *f*, then write gradient descent algorithm. Assume a suitable step length value. Compute values of (x,y) for four successive iterations.

Q15: When would you prefer to solve linear regression by gradient descent over its closed form solution?

Q16: Explain in two to three sentences why convolution neural networks in more suitable than a fully connected network for visual recognition tasks?

Q17: Explain in two to three sentences why the number of parameters in a convolution layer is usually much smaller than that in an equivalent fully connected layer.

Q18: How do you reduce the dimension of an activation map? Consider two cases in your answer: spatial dimensions and depth (i.e., number of activation maps).

Q19: Consider d activation maps of height and width h and w, respectively. You want to construct d activation maps, each of height and width h/2 and w/2, respectively. Write the specifications of a convolution (not max pooling) operation you would apply to achieve this.

Q20: You are applying four 3-by-3 convolution filters to ten activation maps of spatial dimensions 28-by-28. Write the dimensions (height, width and depth) of the output activation map. Assume stride 1 and valid option for the convolution operation.

Q21: How many learnable parameters are there in a layer that implements a max pooling operation with stride 2 and filter size 2-by-2?

Q22: How many learnable parameters are there in layer that implements an average pooling operation with stride 2 and filter size 2-by-2?

Q23: Consider a convolution layer with filter size 5 x 5 x 10 x 5. How many learnable parameters are there in this layer?

Q24: Exercise on convnet forward and backward passes. This was posted on eclass.

Q25: What is the main difference between ResNet and AlexNet architectures?

Q26: Which architecture would require more memory: a DenseNet or a ResNet? Explain your answer in two to three sentences.

Q27: When you convert a fully connected network to an equivalent fully convolutional network, what do you gain?

Q28: When you convert a fully connected network to an equivalent fully convolutional network, how does the number of learnable parameters change?

Q29: Explain in two to three sentences why YOLO object detector is faster than faster RCNN object detector.

Q30: Explain in two to three sentences the role of a region proposal net in an object detector.

Q31: Why do you need a fully convolutional neural network for semantic segmentation task? Limit your answer to two to three sentences.

Q32: As opposed to a convolutional neural net designed for image classification, you apply a fully convolutional net for semantic segmentation. Why? Limit your answer to two to three sentences.

Q33: What role does a transposed convolution play in semantic segmentation?

Q34: What is a skip connection in semantic segmentation?

Q35: Why autoencoder implements unsupervised learning and not supervised learning? Limit your answer to two to three sentences.

Q36: Explain within two to three sentences why an autoencoder tries to implement an identity function.

Q37: Explain in two to three sentences why a variational autoencoder is called a generative model.

Q38: Explain in a few sentences the working principles behind GANs.

Q39: Which model is more powerful in your opinion: GANs or variational autoencoders? Why? Limit your answer to a few sentences.

Q40: What is truncated backpropagation through time? Explain in a few sentences with a simple diagram, if possible.

Q41: The recurrent neural network architecture is not acyclic. How does then backpropagation work for recurrent neural networks.

Q42: What are the vanishing and exploding gradient problems? How does a recurrent neural network overcome these issues?

Q43: Explain in a few sentences how a recurrent neural network is used in image captioning task?

Q44: Explain how a teacher-student training of model compression work in a few sentences.

Q45: Explain why bi-linear transform is differentiable in a few sentences. Use mathematical symbols if needed.