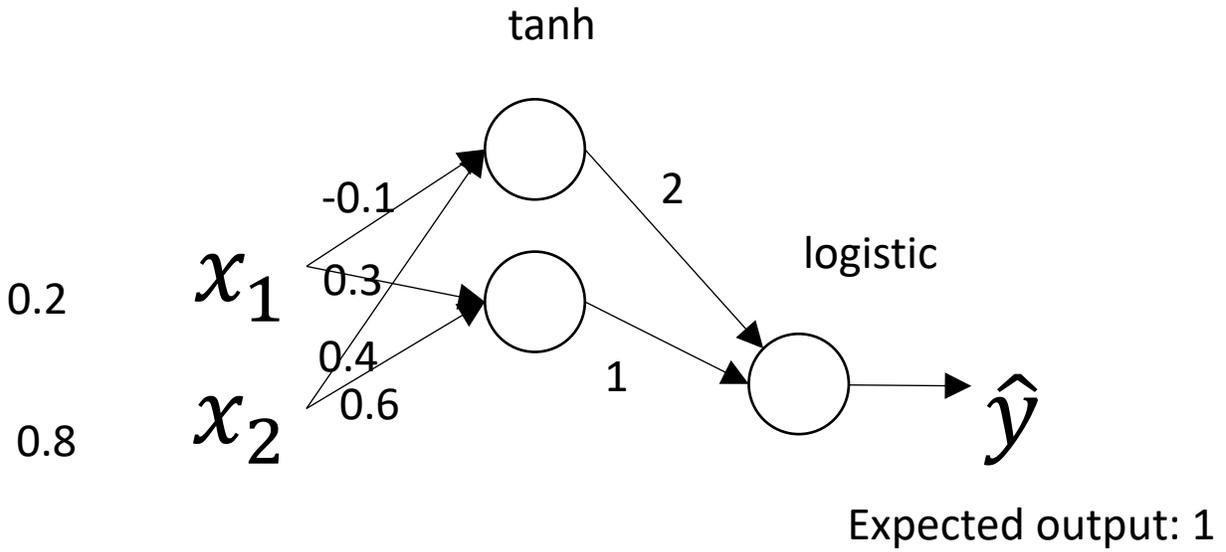


Homework 4
CAP4453. Robot vision

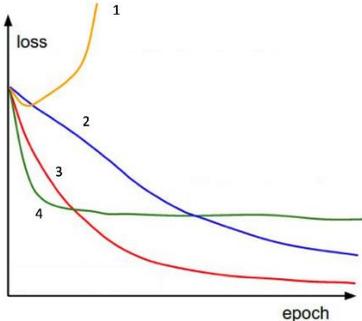
1. [60%] Complete one back propagation step of the following network [show formulas and steps]:



- [20%] What is the output (Forward pass)
- [20%] What are the gradient components respect to the second layer?
- [20%] What are the gradient components respect to the first layer?

2. Complete the following questions [40%]:

Q.1	<p>This question examines properties of the SIFT technique. SIFT is known with extraordinary robustness in matching. Hence, please answer either True or False for the following (true/false) questions. (Check this: SIFT(Scale-invariant feature transform) by Minghao Ning Towards Data Science)</p> <p>a) SIFT can handle changes in scale (True or False)? b) SIFT uses LoG filtering because it has better run-time than DoG (True / False)?</p> <p>Answer: a) b)</p>																																																		
Q.2	<p>We have an image patch (size 5x5) shown on left below. On right, we have a convolution kernel of size 3x3. What will be the output matrix when this convolution kernel is applied on the given image patch centered at 5? Ignore zero padding. NOTE: this operation is done in your convolution layers of CNN.</p> <p style="text-align: center;">a convolution matrix</p> <table style="margin-left: auto; margin-right: auto;"><tr><td style="border: 1px solid black; padding: 5px;">22</td><td style="border: 1px solid black; padding: 5px;">15</td><td style="border: 1px solid black; padding: 5px;">1</td><td style="border: 1px solid black; padding: 5px;">3</td><td style="border: 1px solid black; padding: 5px;">60</td></tr><tr><td style="border: 1px solid black; padding: 5px;">42</td><td style="border: 1px solid black; padding: 5px;">5</td><td style="border: 1px solid black; padding: 5px;">38</td><td style="border: 1px solid black; padding: 5px;">39</td><td style="border: 1px solid black; padding: 5px;">7</td></tr><tr><td style="border: 1px solid black; padding: 5px;">28</td><td style="border: 1px solid black; padding: 5px;">9</td><td style="border: 1px solid black; padding: 5px;">4</td><td style="border: 1px solid black; padding: 5px;">66</td><td style="border: 1px solid black; padding: 5px;">79</td></tr><tr><td style="border: 1px solid black; padding: 5px;">0</td><td style="border: 1px solid black; padding: 5px;">82</td><td style="border: 1px solid black; padding: 5px;">45</td><td style="border: 1px solid black; padding: 5px;">12</td><td style="border: 1px solid black; padding: 5px;">17</td></tr><tr><td style="border: 1px solid black; padding: 5px;">99</td><td style="border: 1px solid black; padding: 5px;">14</td><td style="border: 1px solid black; padding: 5px;">72</td><td style="border: 1px solid black; padding: 5px;">51</td><td style="border: 1px solid black; padding: 5px;">3</td></tr></table> <p style="text-align: center;">\times</p> <table style="margin-left: auto; margin-right: auto;"><tr><td style="padding: 5px;">0</td><td style="padding: 5px;">0</td><td style="padding: 5px;">0</td><td style="padding: 5px;">0</td><td style="padding: 5px;">0</td></tr><tr><td style="padding: 5px;">0</td><td style="border: 1px solid black; padding: 5px;">0</td><td style="border: 1px solid black; padding: 5px;">0</td><td style="border: 1px solid black; padding: 5px;">1</td><td style="padding: 5px;">0</td></tr><tr><td style="padding: 5px;">0</td><td style="padding: 5px;">0</td><td style="padding: 5px;">0</td><td style="padding: 5px;">0</td><td style="padding: 5px;">0</td></tr><tr><td style="padding: 5px;">0</td><td style="padding: 5px;">0</td><td style="padding: 5px;">0</td><td style="padding: 5px;">0</td><td style="padding: 5px;">0</td></tr><tr><td style="padding: 5px;">0</td><td style="padding: 5px;">0</td><td style="padding: 5px;">0</td><td style="padding: 5px;">0</td><td style="padding: 5px;">0</td></tr></table> <p style="text-align: center;">$=$</p> <p>Answer:</p>	22	15	1	3	60	42	5	38	39	7	28	9	4	66	79	0	82	45	12	17	99	14	72	51	3	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	15	1	3	60																																															
42	5	38	39	7																																															
28	9	4	66	79																																															
0	82	45	12	17																																															
99	14	72	51	3																																															
0	0	0	0	0																																															
0	0	0	1	0																																															
0	0	0	0	0																																															
0	0	0	0	0																																															
0	0	0	0	0																																															
Q.3	<p>Suppose that you are designing an object classification network with deep CNN, and you are using a dataset where multiple objects may be present in an image. Which of following activation function cannot be used at output layer of your network design to classify an image?</p> <p>A. ReLU B. Sigmoid C. Softmax D. None of the above</p> <p>Answer:</p>																																																		

<p>Q.4</p>	<p>An input image has been converted into a matrix of size 32 X 32 and a kernel/filter of size 5 X 5 with a stride of 1 (no zero-padding) is applied to it. What will be the size of the resultant convoluted matrix?</p> <p>A) 32 X 32 B) 29 X 29 C) 28 X 28 D) 5 X 5 E) 30 x 30</p> <p>Answer:</p>
<p>Q.5</p>	<p>An input image has been converted into a matrix of size 32 X 32 and a kernel/filter of size 3 X 3 with a stride of 2 (no zero-padding) is applied to it. What will be the size of the resultant convoluted matrix?</p> <p>A) 16 X 16 B) 29 X 29 C) 28 X 28 D) Will not fit E) 30 x 30</p> <p>Answer:</p>
<p>Q.6</p>	<p>A student is running a robot vision experiment with deep learning methods (<i>character recognition from images</i>). He tries different learning rates and obtains the following figure showing loss function's behavior with respect to the epoch. What characteristic of a function these 1,2,3, and 4 can be standing for? (write each number and its correspondences from the following list: good learning rate, low learning rate, high learning rate, very high learning rate).</p>  <p>Answer:</p>
<p>Q.7</p>	<p>Which algorithm category does the <u>Gradient Descent</u> belong to?</p> <p>(A) Transformation (B) Similarity metric (C) Interpolation (D) Edge detection (E) Optimization</p> <p>Answer:</p>

<p>Q.8</p>	<p>What will be the shape of tensor Y?</p> <pre>X = torch.randn(16, 3, 128, 96) Y = X.view(4, 1, -1, 64)</pre> <p>Answer:</p>
<p>Q.9</p>	<p>SLIC is an image segmentation algorithm which is based on k-means clustering.</p> <p>True/False</p> <p>Answer:</p>
<p>Q.10</p>	<p>Explain what the sliding window technique is.</p> <p>Answer:</p>
<p>Q.11</p>	<p>We have a neural network with 2 hidden layers, each hidden layer has 10 neurons. We want to use this neural network for image classification where we have colored images with dimension 20x20 and there are 100 different object categories. We want to use pixel values as features. Each neuron also has a bias term. Determine the total number of trainable parameters in such a classification network. (Hint: All the pixel values from all the three channels will be used as input feature, you will also have to add one prediction layer to predict class probabilities)</p> <p>Also show, how many parameters will be there in each layer.</p> <p>Answer:</p>
<p>Q.12</p>	<p>It is not possible to use convolutional neural networks for non-visual data such as audio signals, texts or brain waves.</p> <p>True/False</p> <p>Answer:</p>
<p>Q.13</p>	<p>In a CNN architecture we have a feature map of shape 128x128x786 (height=128, width=128, channels=786) in some intermediate layer. We pass this to a convolution layer with zero-padding of size 2 (2 additional values on each side), 786 kernels of</p>

	<p>size 5x3, and a stride of 2 (in both x and y direction). What will be the shape of the resultant feature map?</p> <p>Answer:</p>
Q.14	<p>Give two applications of optical flow estimation. In one sentence, provide a reason why optical flow is used in these applications.</p> <p>Answer:</p>
Q.15	<p>Give one reason why feature learning is better than feature engineering.</p> <p>Answer:</p>
Q.16	<p>Give two reasons why pooling layer is used. What is the difference between max-pooling and average-pooling?</p> <p>Answer:</p>
Q.17	<p>If we want to design a neural network for a binary classification problem, we should always have at least 2 neurons in the output layer.</p> <p>True/False</p> <p>Answer:</p>
Q.18	<p>A neural network with very few parameters usually causes the problem of over-fitting.</p> <p>True/False</p> <p>Answer:</p>
Q.19	<p>What are the two constraints/assumptions in Lucas Kanade optical flow estimation algorithm. Also briefly explain them in a sentence each.</p> <p>Answer:</p>
Q.20	<p>Give two applications of auto-encoders. Describe in one sentence each, how auto-encoders are used in these applications.</p> <p>Answer:</p>

