

CS 550 Homework 3 (25 marks)

Instructions:

- Due date is Sep 25
- Please type your solutions cleanly. We won't grade hand-written answers or poorly types answers. You can use LaTeX, Word or markdown etc. to do it.
- Only individual attempts and original answers will get you the credits. Copying will lead to 0 marks and penalties will be imposed.

- (4 marks)** Consider the following sequence of functions. Plot y and z as a function of x . Draw a neural network that accepts x as input and gives o as the output.

- $y = \sigma(4 - 2x)$
- $z = \max(0.1x, 2x)$
- $o = \text{ELU}(w_1y + w_2z + w_3)$ where $\text{ELU} = x$ if $x \geq 0$ and $(e^x - 1)$ if $x < 0$
- $L = \text{MSE}(o, \hat{o})$

(4 marks) Please draw the compute graph for the above. Calculate the gradients for $w_1 = 2, w_2 = -1, w_3 = 6, x = 2, \hat{o} = -1$. Using a learning rate of $\eta = 0.1$, write the values of the parameter for the next iteration.

- (4 marks)** Design a neural network to compute the exclusive-or of three inputs; that is, the output is 1 if an odd number of the three inputs is 1 and output 0 if an even number of the inputs are 1. Can you generalize it to "n" inputs?

3. (4 marks) Softmax + Cross-entropy loss

For classification, it is a common practice to use Softmax activation in the final layer followed by cross entropy loss. In this question, you have to compute the gradient of the combined operator.

Suppose the input to the combined operator is a vector \mathbf{y} .

Let $\mathbf{q} = \mu(\mathbf{y})$, where $\mu(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}}$ and $l = H(\mathbf{p}, \mathbf{q}) = -\sum_{i=1}^n p_i \log q_i$ where \mathbf{p} represents the true probability vector.

You have to compute the gradient of l w.r.t. \mathbf{y} , $\nabla_{\mathbf{y}} l$ in terms of only \mathbf{p} and \mathbf{q} .

- (3 marks)** Suppose images are 512x512 and we use a filter that is 3x3.
 - How many responses will be computed for this layer of a CNN?
 - How much zero padding is necessary to produce an output of size equal to the input?
 - Suppose we don't do any zero padding. If the output of one layer is the input to the next layer, after how many layers will there be no output at all?
 - Repeat a and c above for the case when the stride is 3.
- (3 marks)** Assume that the inputs are single bits 0 (white) and 1 (black). Consider a 3x3 filter, whose weights are w_{ij} , for $0 \leq i \leq 2$ and $0 \leq j \leq 2$ and whose bias is b . Suggest weights and bias so that the output of this filter will detect the following simple features.

- a. Vertical boundary, where the left column is 0, and the other two columns are 1.
 - b. A diagonal boundary, where only the triangle of three pixels in the upper right corner are 1
 - c. A corner, in which the 2x2 square in the lower right is 0 and the other pixels are 1.
6. **(3 marks)** In this exercise, you are asked to design the input weights for one or more nodes of the hidden state of an RNN. The input is a sequence of bits, 0 or 1 only. Note that you can use other nodes to help with the node requested. Also note that you can apply a transformation to the output of the node so a “yes” answer has one value and a “no” answer has another.
- a. A node to signal when the input is 1 and the previous input is 0.
 - b. A node to signal when the last three inputs have all been 1.
 - c. A node to signal when the input is the same as the previous input.