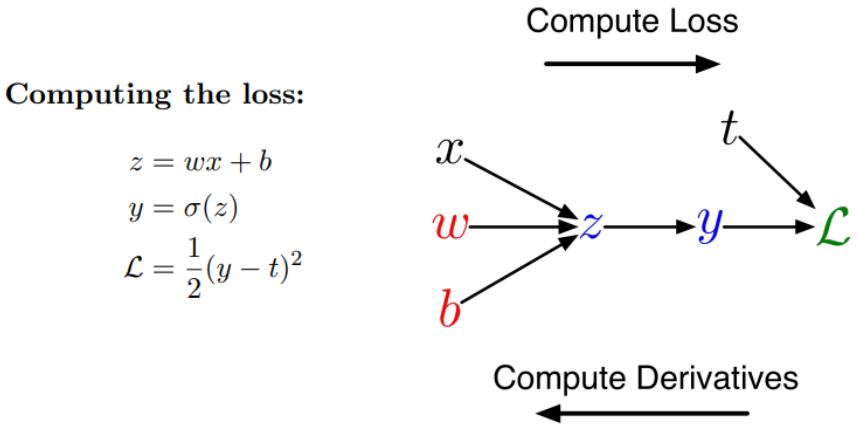
CSCI 416/516 Homework #3

DUE: November 27, 2024, at 11:59 pm

CSCI 416/516: Each Problem begins with an allocation of points, represented as [u pts/g pts]. If you are registered in CSCI 416, you can receive up to u pts on this Problem; if you are registered in CSCI 516, you can receive up to g pts on this Problem. The last Problem is optional for undergraduates (CSCI 416) but required for graduates (CSCI 516). **Write down which session you are in / are you a graduate or undergraduate student**.

Submission: You need to submit both your homework report (answers to the Problems) as a PDF file through Blackboard. Please show the work on how you reach the conclusion for each question except the multiple choice questions and Yes/No (True/False) questions.

- **Problem 1 [2 pt(s)/1 pt(s)]: Backpropagation.** What does the back pass look like (in terms of the error signals of parameters/activations, such as but not limited to b, w), given the illustrated forward pass in Figure below? Show your work on how the conclusion is reached.



- **Problem 2 [4 pt(s)/2 pt(s)]: Activation Function.**
Suppose we have a neural network with one hidden layer as represented in Equation 1 whereas $g(z)$ is defined as $\mathbb{I}(z \geq 0)$. Which of the following functions can be exactly represented as the activation function g? Indicate Yes or No to each choice.

$$f(x) = w_0 + \sum_i w_i h_i(x); h_i(x) = g(b_i + v_i x) \tag{1}$$

- Hard Threshold Activation Function
- Rectified Linear Unit
- Identity Activation Function
- Logistic Activation Function

- **Problem 3 2 pt(s)/2 pt(s)]: 1D Convolution.**
  Given a kernel $k = \{2, 9, 3\}$ and a vector $v = \{1, 2, 3\}$, what is the result of $k * v$? After getting the result $k * v$, what is the new result after applying ReLU on $k * v$? Show the work that leads to your conclusion.

- **Problem 4 [2 pt(s)/2 pt(s)]: Attention & Transformers.**
  What's the relationship between the Scaled Dot-Product Attention and Multi-Head Attention, in the transformer architecture?

- **Problem 5 [3 pt(s)/3 pt(s)]: MLP.**
  Which of the following statements about multi-layer perceptrons (MLPs) is incorrect?

  - A - An MLP with one hidden layer and a sufficient number of neurons can approximate any continuous function.
  - B - The activation function used in the hidden layers of an MLP must always be non-linear.
  - C - Dropout is a regularization technique used to prevent overfitting by randomly setting some neuron outputs to zero during training.
  - D - MLPs require all input features to be numerical, and categorical data must be transformed using techniques like one-hot encoding or embeddings.