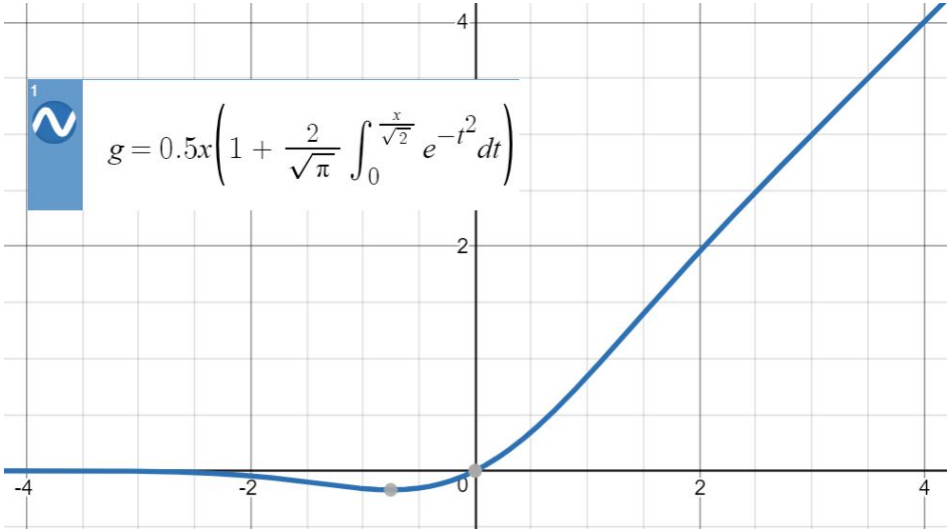
Discussion 2/17
Practice Questions



ROBOTICS
UNIVERSITY OF MICHIGAN

Question 1: Given a 2D tensor *X* of size *m x n*, return a column-wise normalized *X*. You should normalize by subtracting the mean and dividing by the L2 norm. You may NOT use `torch.mean()`.

```
def normalize_tensor(X):

    ##### YOUR CODE HERE ####




    ############################

    return X
```

Question 2: Given a 2D tensor *X* of shape (m, n), return a tensor where each row is reversed. Your solution must only be one line of code and CANNOT use `torch.flip()`.

```
def reverse_rows(X):

    ##### YOUR CODE HERE ####




    ############################

    return X
```

Question 3: Assume a 1-layer network defined as

$$z = Wx + b$$

$$y = \sigma(z)$$

where $x \in R^n$ is the input, $w \in R^{n \times n}$ is the weight matrix, $b \in R^n$ is a scalar bias, and $\sigma(z)$ is the sigmoid activation. The loss for this network is

$$L = 0.5\ (y - t)^2$$

where *t* is the target label.

Question 3: Assume a 1-layer network defined as

$$z = Wx + b$$

$$y = \sigma(z)$$

$$L = 0.5\ (y - t)^2$$

Question 3a: Solve for $\partial L/\partial y$

Question 3: Assume a 1-layer network defined as

$$z = Wx + b$$

$$y = \sigma(z)$$

$$L = 0.5\ (y - t)^2$$

Question 3b: Solve for $\partial L/\partial W$

Question 4: Why is it easier to encounter vanishing gradients in neural networks that use sigmoid activation as compared to ReLU activation?

Question 5: The Gaussian Error Linear Unit (GELU) is a common activation function used while training modern transformers, defined by

$$0.5x(1+\tanh[\sqrt{2/\pi}(x+0.044715x^3)])$$

Derive the gradient of the GELU function w.r.t to its inputs, i.e, solve for ∂GELU/∂x.

Question 6: What does an average pooling layer do to the incoming upstream gradients during backpropagation?