CMSC 498L: Introduction to Deep Learning **Released: Feb-04. Due Feb-11.**

Assignment 1

Name: Paaras Bhandari *UID:* 116191739

Instructions:

- Submit the assignment on ELMS.
- Assignments have to be formatted in LaTeX. You can use overleaf for writing your assignments.
- Submit only the compiled PDF version of the assignment.
- Refer to policies (collaboration, late days, etc.) on the course website.

1 Probability

1. **Density function**. Let p be a Gaussian distribution with zero mean and variance of 0.1. Compute the density of p at 0.

   Sol:

   The probability density function for Gaussian distribution is given by:

$$f(x) = \frac{e^{-(x-\mu)^2/(2\sigma^2)}}{\sigma\sqrt{2\pi}}$$

   where μ = mean and σ^2 = variance

   Substituting x=0, $\mu = 0$, and $\sigma = \sqrt{0.1}$, we have:

$$f(0) = \frac{e^{-(0-0)^2/(2\sqrt{0.1}^2)}}{\sqrt{0.1}\sqrt{2\pi}}$$

$$f(0) = \frac{e^0}{\sqrt{0.2\pi}}$$

$$f(0) = \frac{1}{\sqrt{0.2\pi}} = \mathbf{1.2615}$$

   Therefore, the density of p at $x = 0$ is 1.2615

2. **Conditional probability**. A student is taking a one-hour-time-limit makeup examination. Suppose the probability that the student will finish the exam in less than x hours is $x/2$, $\forall$x $\in [0, 1]$. Given that the student is still working after 0.75 hour, what is the conditional probability that the full hour will be used?

Sol:

We know that

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Now,
A = Student takes more than one hour to finish the exam
B = Student takes longer than 0.75 hour to finish the exam
$A \cap B$ = Student takes more than one hour to finish the exam

$P(A \cap B)$ = 1 - P(Student takes less than 1 hour to finish the exam) =

$$1 - \frac{1}{2} = \frac{1}{2}$$

$P(B)$ = 1 - P(Student takes less than 0.75 hour to finish the exam) =

$$1 - \frac{0.75}{2} = 1 - \frac{3}{8} = \frac{5}{8}$$

Using the formula for conditional probability:

$$\frac{P(A \cap B)}{P(B)} = \frac{1/2}{5/8} = \frac{4}{5} = \mathbf{0.8}$$

3. **Bayes rule**. Consider the probability distribution of you getting sick given the weather in the table below.

Sick?	Weather			
	sunny	rainy	cloudy	snow
yes	0.144	0.02	0.016	0.02
no	0.576	0.08	0.064	0.08

Compute P(sick = yes | Weather = rainy).

Sol:

We know that

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Therefore,

$$P\,(\,\text{sick} = \text{yes} \mid \text{Weather} = \text{rainy}\,) = \frac{P(sick = yes \;\; \cap \;\; Weather = rainy)}{P(Weather = rainy)}$$

Now, $P(A) = P(A \cap B) \cup P(A \cap C)$ (law of total probability)

$$P\,(\,\text{sick} = \text{yes} \mid \text{Weather} = \text{rainy}\,) =$$

$$\frac{P(sick = yes \;\; \cap \;\; Weather = rainy)}{P(sick = yes \;\; \cap \;\; Weather = rainy) \;\; \cup \;\; P(sick = no \;\; \cap \;\; Weather = rainy)}$$

$$= \frac{0.02}{0.02 \;\; + \;\; 0.08}$$

$$= \frac{0.02}{0.1}$$

$$= \mathbf{0.2}$$

2 Calculus and Linear Algebra

For each of the following questions, we expect to see all the steps for reaching the solution.

1. Compute the derivative of the function $f(z)$ with respect to z $\left(\text{i.e., } \frac{df}{dz}\right)$, where

$$f(z) = \frac{1}{1 + e^{-z}}$$

Sol:

Quotient Rule:

$$(\frac{f}{g})' = \frac{f'g - g'f}{g^2}$$

Applying Quotient Rule, we have:

$$f'(z) = \frac{1'.(1 + e^{-z}) - 1.\frac{d}{dz}(1 + e^{-z})}{(1 + e^{-z})^2}$$

$$= \frac{0.(1 + e^{-z}) - 1.(-e^{-z})}{(1 + e^{-z})^2}$$

$$= \frac{e^{-z}}{(1 + e^{-z})^2}$$

$$= \frac{e^{-z}}{1 + 2e^{-z} + e^{-2z}}$$

Multiplying numerator and denominator by e^{2z}

$$= \frac{e^{-z}.e^{2z}}{(1 + 2e^{-z} + e^{-2z}).(e^{2z})}$$

$$= \frac{e^{z}}{(e^{2z} + 2e^{z} + 1)}$$

$$= \frac{e^{z}}{(1 + e^{z})^2}$$

2. Compute the derivative of the function $f(w)$ with respect to w_i, where $w, x \in \mathbb{R}^{\mathrm{D}}$ and

$$f(w) = \frac{1}{1 + e^{-w^T x}}$$

Sol: Applying Quotient Rule, we have:

$$f'(w) = \frac{1'.(1 + e^{-w^T x}) - 1.\frac{\partial}{\partial w_i}(1 + e^{-w^T x})}{(1 + e^{-w^T x})^2}$$

$$f'(w) = \frac{-\frac{\partial}{\partial w_i} e^{-w^T x}}{(1 + e^{-w^T x})^2}$$

$$f'(w) = \frac{e^{-w^T x} \frac{\partial}{\partial w_i}(w^T x)}{(1 + e^{-w^T x})^2}$$

Now,

$$w^T x = \sum_{j=1}^{n} x_j w_j = x_1 w_1 + x_2 w_2 + ... + x_i w_i + ... + x_n + w_n$$

$$\frac{\partial}{\partial w_i}(x_1 w_1 + x_2 w_2 + ... + x_i w_i + ... + x_n + w_n) = x_i$$

Substituting, we have

$$f'(w) = \frac{x_i e^{-w^T x}}{(1 + e^{-w^T x})^2}$$

3. Compute the derivative of the loss function $J(w)$ with respect to w, where

$$J(w) = \frac{1}{2}\sum_{i=1}^{m}\left|w^T x^{(i)} - y^{(i)}\right|$$

Sol:

The derivative of absolute value functions is given by

$$\frac{d}{dx} \mid u \mid = \frac{u}{\mid u \mid}\frac{du}{dx}$$

Applying this formula to find $\frac{\partial}{\partial w}J(w)$, we have

$$\frac{\partial J(w)}{\partial w} = \frac{1}{2}\sum_{i=1}^{m}\frac{w^T x^{(i)} - y^{(i)}}{\left|w^T x^{(i)} - y^{(i)}\right|}.\frac{\partial}{\partial w}(w^T x^{(i)} - y^{(i)})$$

$$\frac{\partial J(w)}{\partial w} = \frac{1}{2}\sum_{i=1}^{m}\frac{w^T x^{(i)} - y^{(i)}}{\left|w^T x^{(i)} - y^{(i)}\right|}.(\frac{\partial}{\partial w}w^T x^{(i)} - \frac{\partial}{\partial w}y^{(i)})$$

$$\frac{\partial J(w)}{\partial w} = \frac{1}{2}\sum_{i=1}^{m}\frac{w^T x^{(i)} - y^{(i)}}{\left|w^T x^{(i)} - y^{(i)}\right|}.(x^{(i)} - 0)$$

$$\frac{\partial J(w)}{\partial w} = \frac{1}{2}\sum_{i=1}^{m}\frac{w^T x^{(i)} - y^{(i)}}{\left|w^T x^{(i)} - y^{(i)}\right|}.x^{(i)}$$

Note - The term $\frac{w^T x^{(i)} - y^{(i)}}{\left|w^T x^{(i)} - y^{(i)}\right|}$ simply denotes the sign (± 1) of the following term

4. Compute the derivative of the loss function $J(w)$ with respect to w, where

$$J(w) = \frac{1}{2}\left[\sum_{i=1}^{m}\left(w^T x^{(i)} - y^{(i)}\right)^2\right] + \lambda\|w\|_2^2$$

Sol:

$$\frac{\partial J(w)}{\partial w} = \frac{1}{2}\left[\sum_{i=1}^{m}\frac{\partial}{\partial w}((w^T x^{(i)} - y^{(i)})^2)\right] + \frac{\partial}{\partial w}\lambda\|w\|_2^2$$

$$\frac{\partial J(w)}{\partial w} = \frac{1}{2}\left[\sum_{i=1}^{m}2.(w^T x^{(i)} - y^{(i)}).\frac{\partial}{\partial w}(w^T x^{(i)} - y^{(i)})\right] + \lambda\frac{\partial}{\partial w}\|w\|_2^2$$

Now, $\|w\|_2^2$ can be writen as $w^T w$

Therefore,

$$\frac{\partial}{\partial w}(\|w\|_2^2) = \frac{\partial}{\partial w}(w^T w) = 2w^T$$

Substituting in $\frac{\partial J(w)}{\partial w}$, we have:

$$\frac{\partial J(w)}{\partial w} = \frac{1}{2}\left[\sum_{i=1}^{m}2.(w^T x^{(i)} - y^{(i)}).x^{(i)}\right] + 2\lambda w^T$$

$$\frac{\partial J(w)}{\partial w} = \frac{2}{2}\left[\sum_{i=1}^{m}(w^T x^{(i)} - y^{(i)}).x^{(i)}\right] + 2\lambda w^T$$

$$\frac{\partial J(w)}{\partial w} = \left[\sum_{i=1}^{m}(w^T x^{(i)} - y^{(i)}).x^{(i)}\right] + 2\lambda w^T$$

5. Compute the derivative of the loss function $J(w)$ with respect to w, where

$$J(w) = \sum_{i=1}^{m} \left[y^{(i)} \log\left(\frac{1}{1+e^{-w^T x^{(i)}}}\right) + \left(1-y^{(i)}\right) \log\left(1 - \frac{1}{1+e^{-w^T x^{(i)}}}\right) \right]$$

Sol:

$$\frac{\partial J(w)}{\partial w} = \sum_{i=1}^{m} \frac{\partial}{\partial w} \left[y^{(i)} \log\left(\frac{1}{1+e^{-w^T x^{(i)}}}\right) + \left(1-y^{(i)}\right) \log\left(1 - \frac{1}{1+e^{-w^T x^{(i)}}}\right) \right]$$

$$\frac{\partial J(w)}{\partial w} = \sum_{i=1}^{m} \left[y^{(i)} \frac{\partial}{\partial w} \log\left(\frac{1}{1+e^{-w^T x^{(i)}}}\right) + \left(1-y^{(i)}\right) \frac{\partial}{\partial w} \log\left(1 - \frac{1}{1+e^{-w^T x^{(i)}}}\right) \right]$$

Applying Chain rule, we have

$$\frac{\partial J(w)}{\partial w} = \sum_{i=1}^{m} \left[y^{(i)} (1+e^{-w^T x^{(i)}}) \frac{\partial}{\partial w} (\frac{1}{(1+e^{-w^T x^{(i)}})}) + \left(1-y^{(i)}\right) \left(\frac{1+e^{-w^T x^{(i)}}}{e^{-w^T x^{(i)}}}\right) \frac{\partial}{\partial w} (1 - \frac{1}{1+e^{-w^T x^{(i)}}}) \right]$$

From, Part 2 Problem 3, we know the required derivative.

$$\frac{\partial}{\partial w} \frac{1}{(1+e^{-w^T x^{(i)}})} = \frac{x_i e^{-w^T x}}{(1+e^{-w^T x})^2}$$

Substituting, we have:

$$\frac{\partial J(w)}{\partial w} = \sum_{i=1}^{m} \left[y^{(i)} (1+e^{-w^T x^{(i)}}) \frac{x_i e^{-w^T x}}{(1+e^{-w^T x})^2} + \left(1-y^{(i)}\right) \left(\frac{1+e^{-w^T x^{(i)}}}{e^{-w^T x^{(i)}}}\right) \frac{-x_i e^{-w^T x}}{(1+e^{-w^T x})^2} \right]$$

Simplifying, we have:

$$\frac{\partial J(w)}{\partial w} = \sum_{i=1}^{m} \left[\frac{x_i y^{(i)} e^{-w^T x}}{1+e^{-w^T x}} - \left(1-y^{(i)}\right) \frac{x_i e^{-w^T x}}{(e^{-w^T x^{(i)}})(1+e^{-w^T x})} \right]$$

6. Compute $\nabla_w f$, where $f(w) = \tanh\left[w^T x\right]$.

   Sol:

   We know that

$$\frac{d}{dx} f(g(x) = f'(g(x)).g'(x)(Chainrule)$$

$$\frac{d}{dx} tanh(x) = 1 - tanh^2(x)$$

   Applying chain rule, we have

$$\nabla_w f = (1 - tan^2[w^T x]).\frac{\partial (w^T x)}{\partial w}$$

$$\nabla_w f = (1 - tan^2[w^T x]).x$$

7. Find the solution to the system of linear equations given by Ax=b, where

$$A = \begin{pmatrix} 2 & 1 & -1 \\ -3 & -1 & 2 \\ -2 & 1 & 2 \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} 8 \\ -11 \\ -3 \end{pmatrix}.$$

Sol:

We can find the solution to this system of linear equations by reducing the augmented matrix to it's reduced row echelon form by using row operations

$$\begin{bmatrix} 2 & 1 & -1 & 8 \\ -3 & -1 & 2 & -11 \\ -2 & 1 & 2 & -3 \end{bmatrix}$$

$R3 = R3 + R1$

$$= \begin{bmatrix} 2 & 1 & -1 & 8 \\ -3 & -1 & 2 & -11 \\ 0 & 2 & 1 & 5 \end{bmatrix}$$

R1 = R1/2

$$= \begin{bmatrix} 1 & \frac{1}{2} & \frac{-1}{2} & 4 \\ -3 & -1 & 2 & -11 \\ 0 & 2 & 1 & 5 \end{bmatrix}$$

R2 = R2 + 3R1

$$= \begin{bmatrix} 1 & \frac{1}{2} & \frac{-1}{2} & 4 \\ 0 & \frac{1}{2} & \frac{1}{2} & 1 \\ 0 & 2 & 1 & 5 \end{bmatrix}$$

R1 = R1 - R2

$$= \begin{bmatrix} 1 & 0 & -1 & 3 \\ 0 & \frac{1}{2} & \frac{1}{2} & 1 \\ 0 & 2 & 1 & 5 \end{bmatrix}$$

R3 = R3 - 4R2

$$= \begin{bmatrix} 1 & 0 & -1 & 3 \\ 0 & \frac{1}{2} & \frac{1}{2} & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$

R2 = 2R2

$$= \begin{bmatrix} 1 & 0 & -1 & 3 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$

R1 = R1 - R3

$$= \begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$

R2 = R2 + R2

$$= \begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$

R3 = -R3

$$= \begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

Therefore, the solution is $x_1 =$ **2**, $x_2 =$ **3**, $x_3 =$ **-1**

8. Find the eigenvalues and associated eigenvectors of the matrix:

$$A = \begin{bmatrix} 7 & 0 & -3 \\ -9 & -2 & 3 \\ 18 & 0 & -8 \end{bmatrix}$$

Sol:

The characteristic equation is

$$| A - \lambda I |= 0$$

$$\begin{bmatrix} 7 & 0 & -3 \\ -9 & -2 & 3 \\ 18 & 0 & -8 \end{bmatrix} - \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix} = 0$$

$$\begin{bmatrix} 7-\lambda & 0 & -3 \\ -9 & -2-\lambda & 3 \\ 18 & 0 & -8-\lambda \end{bmatrix} = 0$$

$$(7-\lambda)(2+\lambda)(8+\lambda) - 3.18.(-2-\lambda) = 0$$

$$-\lambda^3 - 3\lambda^2 + 4 = 0$$

$$\lambda^3 + 3\lambda^2 - 4 = 0$$

$$\lambda^3 - \lambda^2 + 4\lambda^2 - 4 = 0$$

$$\lambda^2(\lambda - 1) + 4(\lambda^2 - 1) = 0$$

$$\lambda^2(\lambda - 1) + 4(\lambda - 1)(\lambda + 1) = 0$$

$$(\lambda - 1)(\lambda^2 + 4\lambda + 4) = 0$$

$$(\lambda - 1)(\lambda + 2)^2 = 0$$

Therefore, the eigenvalues are $\lambda = 1$ and $\lambda = -2$ (multiplicity = 2)

Finding the eigenvector corresponding to $\lambda = 1$

$$A - 1.I$$

$$\begin{bmatrix} 7 & 0 & -3 \\ -9 & -2 & 3 \\ 18 & 0 & -8 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 6 & 0 & -3 \\ -9 & -3 & 3 \\ 18 & 0 & -9 \end{bmatrix}$$

Reducing the matrix to reduced row echelon form

R2 = 2.R2

$$\begin{bmatrix} 6 & 0 & -3 \\ -18 & -6 & 6 \\ 18 & 0 & -9 \end{bmatrix}$$

R2 = R2 + R3

$$\begin{bmatrix} 6 & 0 & -3 \\ 0 & -6 & -3 \\ 18 & 0 & -9 \end{bmatrix}$$

R3 = R3 - 3.R1

$$\begin{bmatrix} 6 & 0 & -3 \\ 0 & -6 & -3 \\ 0 & 0 & 0 \end{bmatrix}$$

R1 = R1/6 R2 = R2/-6

$$\begin{bmatrix} 1 & 0 & -\frac{1}{2} \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & 0 \end{bmatrix}$$

The solution can be represented as

$$\begin{bmatrix} z/2 \\ -z/2 \\ z \end{bmatrix}$$

Therefore, the eigenvector to $\lambda = 1$ is

$$= \begin{bmatrix} 1/2 \\ -1/2 \\ 1 \end{bmatrix}$$

Let z = 2, then the eigenvector can be written as

$$= \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$$

Finding the eigenvectors corresponding to $\lambda = -2$

$$A - (-2).I$$

$$\begin{bmatrix} 7 & 0 & -3 \\ -9 & -2 & 3 \\ 18 & 0 & -8 \end{bmatrix} \quad - \begin{bmatrix} -2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix}$$

$$\begin{bmatrix} 9 & 0 & -3 \\ -9 & 0 & 3 \\ 18 & 0 & -6 \end{bmatrix}$$

Reducing the matrix to reduced row echelon form

R2 = R2 + R1

$$\begin{bmatrix} 9 & 0 & -3 \\ 0 & 0 & 0 \\ 18 & 0 & -6 \end{bmatrix}$$

R3 = R3 + (-2)R1

$$\begin{bmatrix} 9 & 0 & -3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

R1 = R1/9

$$\begin{bmatrix} 1 & 0 & -1/3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The solution can be represented as

$$\begin{bmatrix} z/3 \\ y \\ z \end{bmatrix}$$

$$z \begin{bmatrix} 1/3 \\ 0 \\ 1 \end{bmatrix} \quad + y \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

So we have 2 eigenvectors corresponding to $\lambda = -2$

$$\begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix} \quad + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

Finally the eigenvalues are $\lambda = 1$ and $\lambda = -2$

The eigenvector corresponding to $\lambda = 1$ is $\begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$

The eigenvectors corresponding to $\lambda = -2$ are $\begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix} \quad and \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$

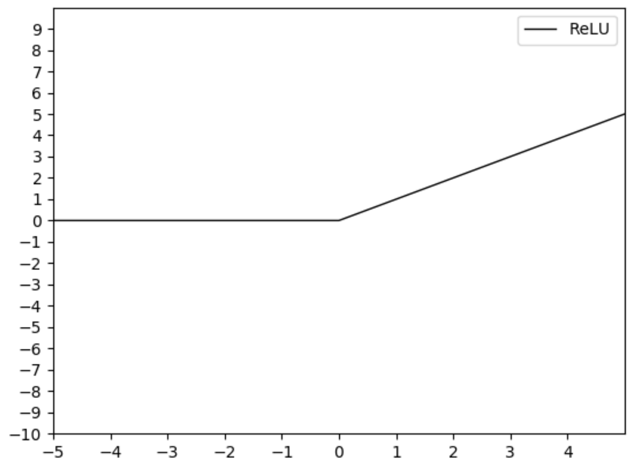
3 Activation functions

For each of the following activation functions, write their equations and their derivatives. Plot the functions and derivatives, with $x \in [-5, 5]$ and $y \in [-10, 10]$ plot limits. (No need to submit the code for plots.)

1. Relu

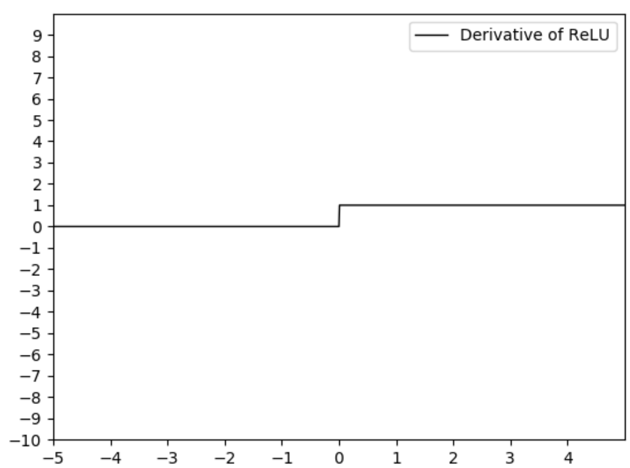
   Sol: The ReLU function is given by

$$f(x) = max(x, 0)$$



The derivative of the ReLU function is

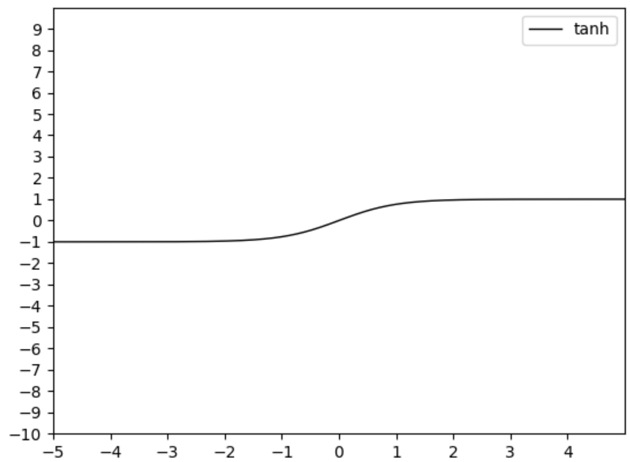
$$f'(x) = \begin{cases} 1 & x > 0 \\ 0 & x < 0 \\ undefined & x = 0 \end{cases}$$



2. Tanh

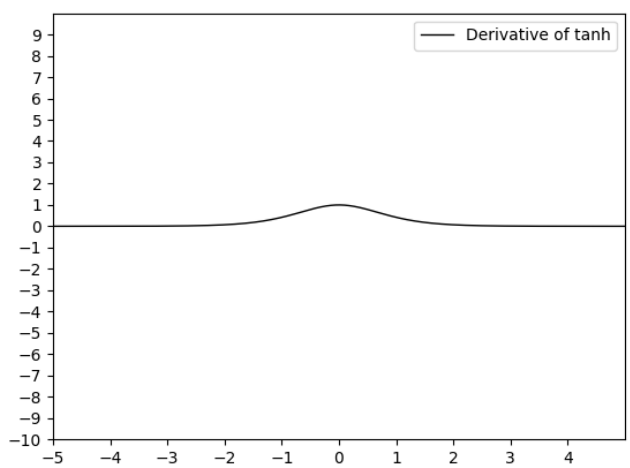
   Sol: The tanh function is given by

$$f(x) = tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$



   The derivative of the tanh function is given by

$$f'(x) = 1 - tanh(x)^2$$



3. Softmax

   Sol:

   The softmax function is given by

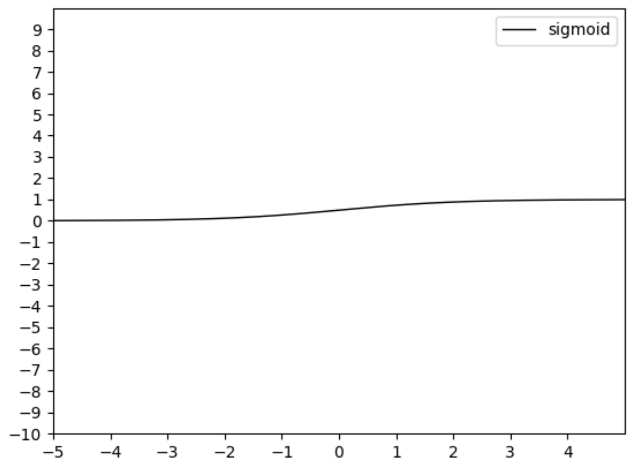
$$f(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}}$$

   (Not sure how to plot this function)

4. Sigmoid

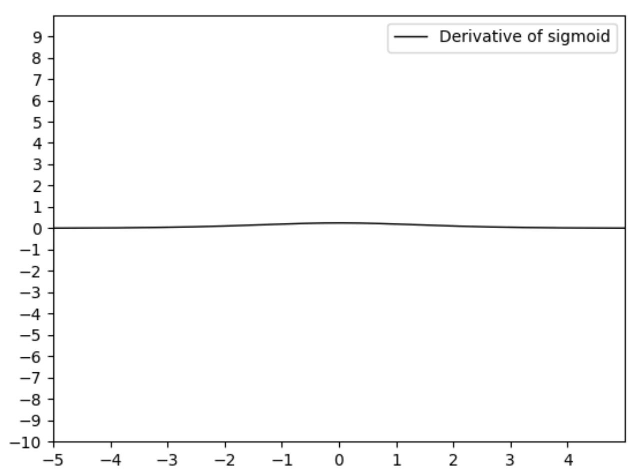
Sol: The Sigmoid function is given by

$$f(x) = \frac{1}{1+e^{-x}}$$



The derivative of the Sigmoid function is given by

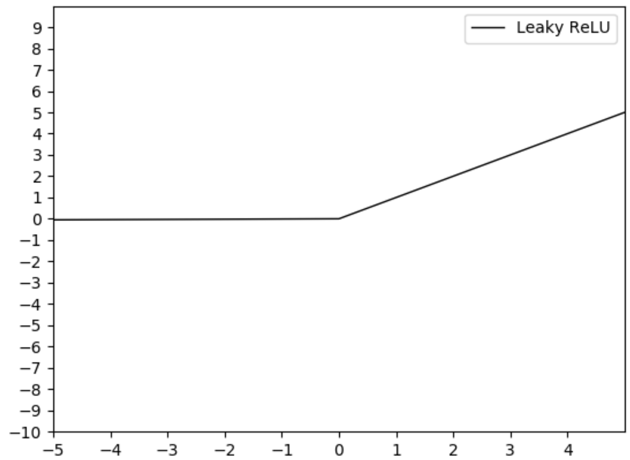
$$f'(x) = f(x)(1 - f(x)) = \frac{e^x}{(1+e^x)^2}$$



5. Leaky ReLU

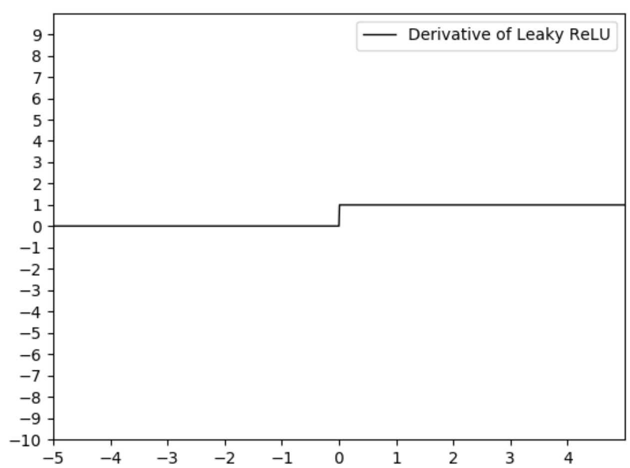
Sol: The Leaky ReLU function is given by

$$f(x) = max(0.01x, x)$$



The derivative of the Leaky ReLU function is

$$f'(x) = \begin{cases} 0.01 & x < 0 \\ 1 & x \geq 0 \end{cases}$$

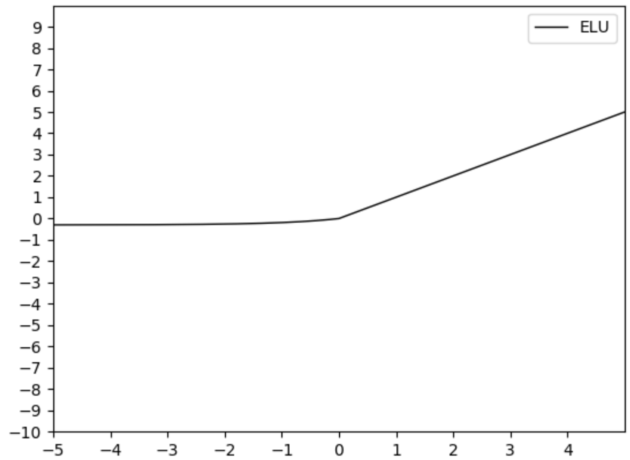


6. ELU (plot with $\alpha = 0.3$)

Sol:

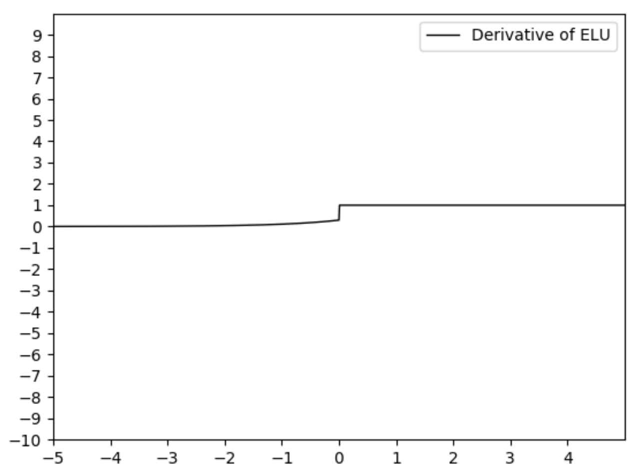
The ELU function is given by

$$f(x) = \begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$



The derivative of ELU function is given by

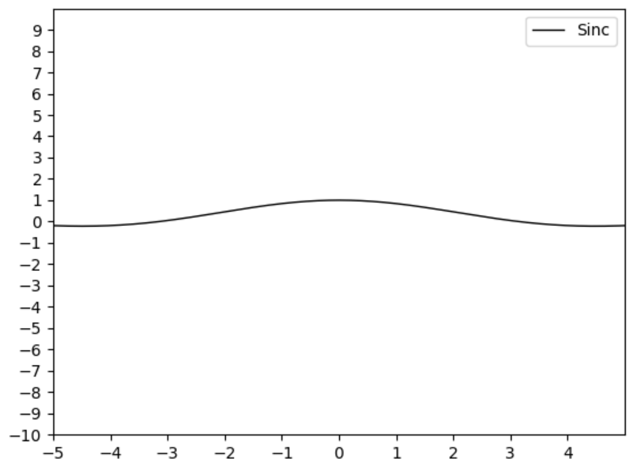
$$f'(x) = \begin{cases} 1 & x \geq 0 \\ \alpha e^x & x < 0 \end{cases}$$



7. Sinc

Sol: The Sinc function is given by

$$f(x) = \frac{Sin(x)}{x}$$



The derivative of the Sinc function is given by

$$f'(x) = \begin{cases} \frac{Cos(x) - Sinc(x)}{x} & x \neq 0 \\ 0 & x = 0 \end{cases}$$