Instructions:

- This assignment is meant to help you grok certain concepts we will use in the course. Please don't copy solutions from any sources.
- Avoid verbosity.
- The assignment needs to be written in latex using the attached tex file. The solution for each question should be written in the solution block in space already provided in the tex file. **Handwritten assignments will not be accepted.**

1. **Partial Derivatives**

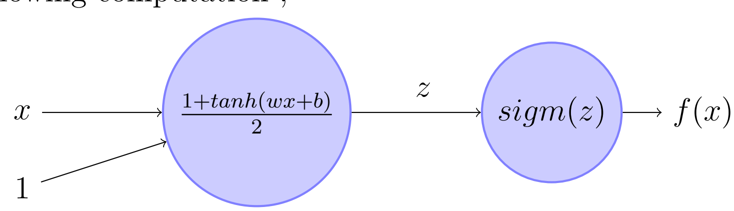
(a) Find the derivative of $g(\rho)$ with respect to ρ where $g(\rho)$ is given by,

$$g(\rho) = \frac{1}{2}\rho log\frac{\rho}{\rho + \hat{\rho}} + \frac{1}{2}\hat{\rho} log\frac{\hat{\rho}}{\rho + \hat{\rho}}$$

(You can consider $\hat{\rho}$ as constant)

Solution:

(b) Consider the following computation ,



where $z = \frac{1+tanh(wx+b)}{2}$ and $f(x) = sigm(z)$

by definition : $sigm(z) = \frac{1}{1+e^{-z}}$ and $tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$

The value L is given by,

$$L = -y \log(f(x))$$

Here, x and y are constants and w and b are parameters that can be modified. In other words, L is a function of w and b.

Derive the partial derivatives, $\frac{\partial L}{\partial w}$ and $\frac{\partial L}{\partial b}$.

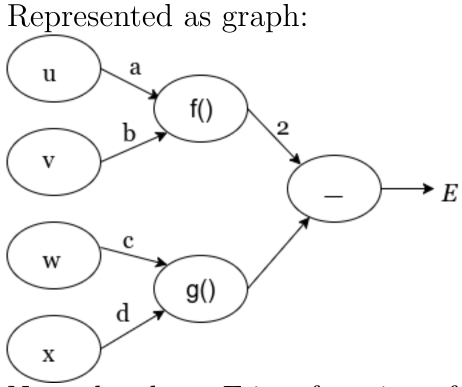
Solution:

2. **Chain Rule**:

(a) Consider the evaluation of E as given below,

$$E = h(u, v, w, x) = 2 * f(au + bv) - g(cw + dx))$$

Represented as graph:



Here u, v, w, x are inputs (constants) and a, b, c, d are parameters (variables). f and g are the activation functions (with z as input) defined as below:

$$f(z) = sigm(z) \qquad g(z) = tanh(z)$$

Note that here E is a function of parameters a, b, c, d. Compute the partial derivatives of E with respect to the parameters a, b, c and d *i.e.* $\frac{\partial E}{\partial a}$, $\frac{\partial E}{\partial b}$, $\frac{\partial E}{\partial c}$ and $\frac{\partial E}{\partial d}$.

Solution:

(b) Assume that $z = f(x, y)$, where $x = uv$ and $y = \frac{u}{v}$. We use f_x to denote the partial derivative $\frac{\partial f}{\partial x}$. Using the chain rule, express $\frac{\partial z}{\partial u}$ and $\frac{\partial z}{\partial v}$ in terms of u, v, f_x and f_y.

Solution:

(c) Given the change of variables as mentioned in the previous part: $x = uv$ and $y = \frac{u}{v}$, calculate the Jacobian of this transformation.

Solution:

(d) Calculate the Jacobian of the transformation for rectangular coordinates; *i.e.*, the Jacobian of $x = rsin\theta$, $y = rcos\theta$, $z = z$, (hint: using the relevant partial derivatives)

Solution:

3. **Visit Taylor Series** The first order derivative of a function f is defined by the following limit,

$$\frac{df(x)}{dx} = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \tag{1}$$

On observing the above definition we see that the derivative of a function is the ratio of change in the function value to the change in the function input, when we change the input by a small quantity (infinitesimally small). A first degree approximation based on eq. 1 would be the following.

$$f(x+h) \approx f(x) + h\frac{df(x)}{dx} \tag{2}$$

Consider $f(x) = ln(x+5)$.

(a) Estimate the value of $f(1)$,$f(1.1)$ and $f(2.5)$ using the above formula.

Solution:

(b) Compare these estimates to the actual values of function $f(1)$,$f(1.1)$ and $f(2.5)$. Explain the discrepancy as we increase the value.

Solution:

(c) Can we get a better estimate of $f(1)$,$f(1.1)$ and $f(2.5)$? How?

Solution:

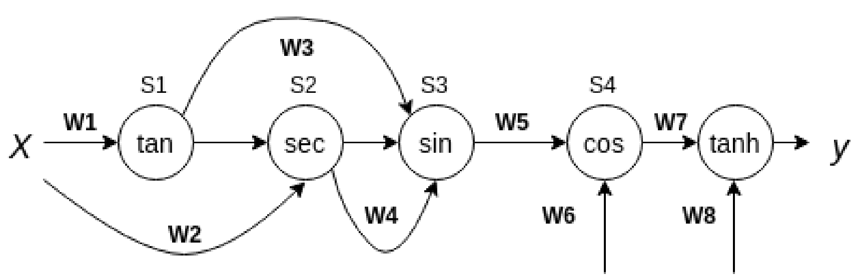
(d) Consider $g(x) = a + be^x + c * \cos(x)$. Find a, b, c $\in$ R such that g approximates f at $x = 0$. (i.e. by matching the (i) direct values (ii) first derivative and (iii) second derivative at $x = 0$).

Solution:

4. **Differentiation of function of multiple variables**

$$\begin{aligned} s_1 &= tan(w_1 x) \\ s_2 &= sec(w_2 x + s_1) \\ s_3 &= sin(w_3 s_1 + w_4 s_2) \\ s_4 &= cos(w_5 s_3 + w_6) \\ y &= tanh(w_7 s_4 + w_8) \end{aligned}$$

An alternative representation of the function y is given in the figure below.



Compute the derivatives $\frac{dy}{dw_1}$ and $\frac{dy}{dw_2}$ (show all the steps).

Solution:

5. **Differentiation of vectors/matrices**
Consider vectors $\boldsymbol{u}, \boldsymbol{b} \in \mathbb{R}^d$, and matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$.
The derivative of a scalar f w.r.t. a vector $\boldsymbol{u}$ is a vector by itself, given by

$$\nabla f = \left(\frac{\partial f}{\partial u_1}, \frac{\partial f}{\partial u_2}, \ldots, \frac{\partial f}{\partial u_n} \right)$$

(**Hint**: The derivative of a scalar f w.r.t. a matrix $\boldsymbol{X}$, is a matrix whose (i, j) component is $\frac{\partial f}{\partial X_{ij}}$, where X_{ij} is the (i, j) component of the matrix $\boldsymbol{X}$.)

(a) Derive the expression for the derivative: $\nabla \boldsymbol{u}^T \boldsymbol{A} \boldsymbol{u} + \boldsymbol{b}^T \boldsymbol{u}$.

Solution:

(b) Compare your results with derivatives for the scalar equivalents of the above expressions $au^2 + bu$.

Solution:

(c) Derive the Hessian: $\frac{\partial^2 f}{\partial \mathbf{u} \partial \mathbf{u}^T}$ given that $f = \boldsymbol{u}^T \boldsymbol{A} \boldsymbol{u} + \boldsymbol{b}^T \boldsymbol{u}$

Solution:

6. **Encoding Tongue Twister :** You have been assigned a task to encode a tongue-twister phrase compactly: 'clean clams crammed in clean clans'. For convenience, you are given the frequency distribution as below.

Char	Frequency
a	5
c	5
d	1
i	3
l	4
m	3
n	4
r	1
s	2
space	5

(a) One way to encode this sequence is to use fixed length code with each code word long enough to encode ten different symbols. How many bits would be needed for this 33-character phrase using such a fixed-length code?

Solution:

(b) What are the minimum number of bits needed (theoretically) to encode the entire phrase , assuming that each character is independent of the surrounding character? Hint: We can calculate the average information (in other words, bits needed) of a symbol using entropy information.

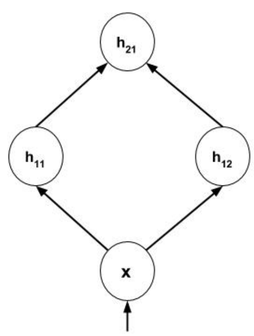
Solution:

7. **Plotting Functions for Great Good**

   (a) Consider the variable x and functions $h_{11}(x)$, $h_{12}(x)$ and $h_{21}(x)$ such that

$$
\begin{aligned}
h_{11}(x) &= \frac{1}{1+e^{-(500x+30)}} \\
h_{12}(x) &= \frac{1}{1+e^{-(500x-30)}} \\
h_{21} &= h_{11}(x) - h_{12}(x)
\end{aligned}
$$

The above set of functions are summarized in the graph below.



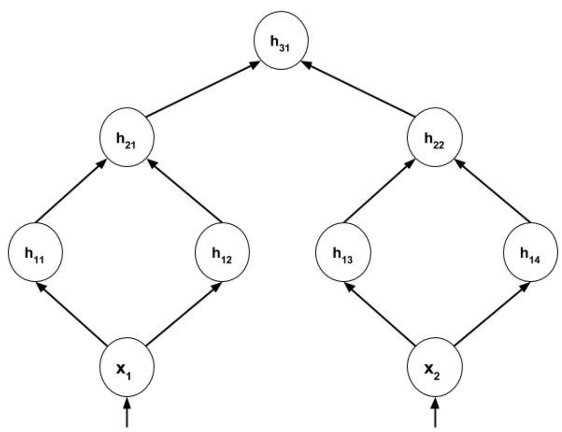
Plot the following functions: $h_{11}(x)$, $h_{12}(x)$ and $h_{21}(x)$ for $x \in (-1, 1)$

Solution:

   (b) Now consider the variables x_1, x_2 and the functions $h_{11}(x_1, x_2), h_{12}(x_1, x_2), h_{13}(x_1, x_2), h_{14}(x_1, x_2),$ $h_{21}(x_1, x_2), h_{22}(x_1, x_2), h_{31}(x_1, x_2)$ and $f(x_1, x_2)$ such that

$$
\begin{aligned}
h_{11}(x_1, x_2) &= \frac{1}{1 + e^{-(x_1 + 50x_2 + 100)}} \\
h_{12}(x_1, x_2) &= \frac{1}{1 + e^{-(x_1 + 50x_2 - 100)}} \\
h_{13}(x_1, x_2) &= \frac{1}{1 + e^{-(50x_1 + x_2 + 100)}} \\
h_{14}(x_1, x_2) &= \frac{1}{1 + e^{-(50x_1 + x_2 - 100)}} \\
h_{21}(x_1, x_2) &= h_{11}(x_1, x_2) - h_{12}(x_1, x_2) \\
h_{22}(x_1, x_2) &= h_{13}(x_1, x_2) - h_{14}(x_1, x_2) \\
h_{31}(x_1, x_2) &= h_{21}(x_1, x_2) + h_{22}(x_1, x_2) \\
f(x_1, x_2) &= \frac{1}{1 + e^{-(100h_{31}(x) - 200)}}
\end{aligned}
$$

The above set of functions are summarized in the graph below.



Plot the following functions: $h_{11}(x_1, x_2), h_{12}(x_1, x_2), h_{13}(x_1, x_2), h_{14}(x_1, x_2), h_{21}(x_1, x_2),$ $h_{22}(x_1, x_2), h_{31}(x_1, x_2)$ and $f(x_1, x_2)$ for $x_1 \in (-5, 5)$ and $x_2 \in (-5, 5)$

Solution: