

**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  $\rho$  where  $g(\rho)$  is given by,

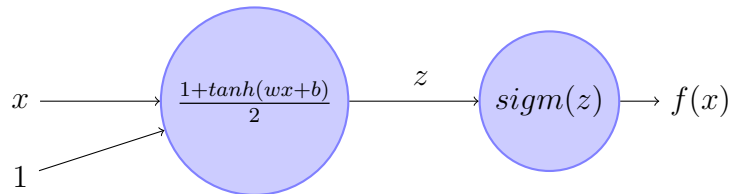
$$g(\rho) = \frac{1}{2}\rho \log \frac{\rho}{\rho + \hat{\rho}} + \frac{1}{2}\hat{\rho} \log \frac{\hat{\rho}}{\rho + \hat{\rho}}$$

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

**Solution:**

$$\begin{aligned} \frac{\partial g(\rho)}{\partial \rho} &= \frac{1}{2} \frac{\partial \rho \log \frac{\rho}{\rho + \hat{\rho}}}{\partial \rho} + \frac{1}{2} \frac{\partial \hat{\rho} \log \frac{\hat{\rho}}{\rho + \hat{\rho}}}{\partial \rho} \\ \Rightarrow \frac{\partial g(\rho)}{\partial \rho} &= \frac{1}{2} \log \frac{\rho}{\rho + \hat{\rho}} + \frac{1}{2} \rho \left( \frac{1}{\rho} - \frac{1}{\rho + \hat{\rho}} \right) - \frac{1}{2} \frac{\hat{\rho}}{\rho + \hat{\rho}} \\ \Rightarrow \frac{\partial g(\rho)}{\partial \rho} &= \frac{1}{2} \log \frac{\rho}{\rho + \hat{\rho}} + \frac{1}{2} \frac{\hat{\rho}}{\rho + \hat{\rho}} - \frac{1}{2} \frac{\hat{\rho}}{\rho + \hat{\rho}} \\ \Rightarrow \frac{\partial g(\rho)}{\partial \rho} &= \frac{1}{2} \log \frac{\rho}{\rho + \hat{\rho}} \end{aligned}$$

- (b) Consider the following computation ,



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

by definition :  $\text{sigmoid}(z) = \frac{1}{1+e^{-z}}$  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:**

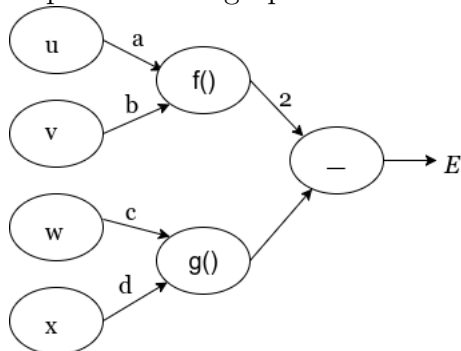
$$\begin{aligned}
 \frac{\partial L}{\partial w} &= -\frac{y}{f(x)} \frac{\partial f(x)}{\partial w} \\
 \Rightarrow \frac{\partial L}{\partial w} &= -\frac{y}{f(x)} \frac{\partial f(x)}{\partial z} \frac{\partial z}{\partial w} \\
 z &= \frac{1 + \frac{e^{wx+b} - e^{-(wx+b)}}{e^{wx+b} + e^{-(wx+b)}}}{2} \\
 \Rightarrow z &= \frac{1}{1 + e^{-2(wx+b)}} \\
 \frac{\partial L}{\partial w} &= -\frac{y}{f(x)} \frac{e^{-z}}{(1 + e^{-z})^2} \frac{\partial \frac{1}{1 + e^{-2(wx+b)}}}{\partial w} \\
 \Rightarrow \frac{\partial L}{\partial w} &= -\frac{ye^{-z}}{(1 + e^{-z}) (e^{wx+b} + e^{-(wx+b)})^2} 2x \\
 \frac{\partial L}{\partial b} &= -\frac{y}{f(x)} \frac{\partial f(x)}{\partial b} \\
 \Rightarrow \frac{\partial L}{\partial b} &= -\frac{y}{f(x)} \frac{\partial f(x)}{\partial z} \frac{\partial z}{\partial b} \\
 \Rightarrow \frac{\partial L}{\partial b} &= -\frac{y}{f(x)} \frac{e^{-z}}{(1 + e^{-z})^2} \frac{\partial \frac{1}{1 + e^{-2(wx+b)}}}{\partial b} \\
 \Rightarrow \frac{\partial L}{\partial b} &= -\frac{ye^{-z}}{(1 + e^{-z}) (e^{wx+b} + e^{-(wx+b)})^2} 2
 \end{aligned}$$

## 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:

$$f(z) = \text{sigm}(z)$$

$$g(z) = \text{tanh}(z)$$

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:**

$$\begin{aligned}
 f(au + bv) &= \frac{1}{1 + e^{-(au+bv)}} \\
 g(cw + dx) &= \frac{e^{cw+dx} - e^{-(cw+dx)}}{e^{cw+dx} + e^{-(cw+dx)}} \\
 \frac{\partial E}{\partial a} &= 2 \frac{\partial f(au + bv)}{\partial a} - \frac{\partial g(cw + dx)}{\partial a} \\
 \Rightarrow \frac{\partial E}{\partial a} &= 2 \frac{\partial f(au + bv)}{\partial(au + bv)} \frac{\partial(au + bv)}{\partial a} + 0 \\
 \Rightarrow \frac{\partial E}{\partial a} &= 2 \frac{e^{-(au+bv)}}{(1 + e^{-(au+bv)})^2} u \\
 \frac{\partial E}{\partial b} &= 2 \frac{\partial f(au + bv)}{\partial b} - \frac{\partial g(cw + dx)}{\partial b} \\
 \Rightarrow \frac{\partial E}{\partial b} &= 2 \frac{\partial f(au + bv)}{\partial(au + bv)} \frac{\partial(au + bv)}{\partial b} + 0 \\
 \Rightarrow \frac{\partial E}{\partial b} &= 2 \frac{e^{-(au+bv)}}{(1 + e^{-(au+bv)})^2} v \\
 \frac{\partial E}{\partial c} &= 2 \frac{\partial f(au + bv)}{\partial c} - \frac{\partial g(cw + dx)}{\partial c} \\
 \Rightarrow \frac{\partial E}{\partial c} &= 0 - \frac{\partial g(cw + dx)}{\partial(cw + dx)} \frac{\partial(cw + dx)}{\partial c} \\
 \Rightarrow \frac{\partial E}{\partial c} &= - \frac{(e^{cw+dx} + e^{-(cw+dx)})^2 - (e^{cw+dx} - e^{-(cw+dx)})^2}{(e^{cw+dx} + e^{-(cw+dx)})^2} w \\
 \Rightarrow \frac{\partial E}{\partial c} &= \frac{-4w}{(e^{cw+dx} + e^{-(cw+dx)})^2} \\
 \frac{\partial E}{\partial d} &= 2 \frac{\partial f(au + bv)}{\partial d} - \frac{\partial g(cw + dx)}{\partial d} \\
 \Rightarrow \frac{\partial E}{\partial d} &= 0 - \frac{\partial g(cw + dx)}{\partial(cw + dx)} \frac{\partial(cw + dx)}{\partial d} \\
 \Rightarrow \frac{\partial E}{\partial d} &= - \frac{(e^{cw+dx} + e^{-(cw+dx)})^2 - (e^{cw+dx} - e^{-(cw+dx)})^2}{(e^{cw+dx} + e^{-(cw+dx)})^2} x \\
 \Rightarrow \frac{\partial E}{\partial d} &= \frac{-4x}{(e^{cw+dx} + e^{-(cw+dx)})^2}
 \end{aligned}$$

- (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:**

$$\begin{aligned}\frac{\partial z}{\partial u} &= \frac{\partial z}{\partial x} \frac{\partial x}{\partial u} + \frac{\partial z}{\partial y} \frac{\partial y}{\partial u} \\ \implies \frac{\partial z}{\partial u} &= f_x v + \frac{f_y}{v} \\ \frac{\partial z}{\partial v} &= \frac{\partial z}{\partial x} \frac{\partial x}{\partial v} + \frac{\partial z}{\partial y} \frac{\partial y}{\partial v} \\ \implies \frac{\partial z}{\partial v} &= f_x u - \frac{u f_y}{v^2}\end{aligned}$$

- (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:** The Jacobian matrix  $J$  is given by :

$$\begin{aligned}J &= \begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{bmatrix} \\ \implies J &= \begin{bmatrix} v & u \\ \frac{1}{v} & -\frac{u}{v^2} \end{bmatrix}\end{aligned}$$

The Jacobian determinant is given by  $\det(J)$  where  $J$  is the Jacobian matrix

$$\begin{aligned}\det(J) &= \frac{-u}{v} - \frac{u}{v} \\ \det(J) &= -2\frac{u}{v}\end{aligned}$$

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

**Solution:** The Jacobian matrix  $J$ , is given by :

$$\begin{aligned}J &= \begin{bmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} & \frac{\partial x}{\partial z} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} & \frac{\partial y}{\partial z} \\ \frac{\partial z}{\partial r} & \frac{\partial z}{\partial \theta} & \frac{\partial z}{\partial z} \end{bmatrix} \\ J &= \begin{bmatrix} \sin(\theta) & r \cos(\theta) & 0 \\ \cos(\theta) & -r \sin(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}\end{aligned}$$

The Jacobian determinant is given by  $\det(J)$  where  $J$  is the Jacobian matrix

$$\det(J) = -r\sin^2(\theta) - r\cos^2(\theta)$$

$$\det(J) = -r$$

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

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

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.

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

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:** Let us approximate the function around the point  $x_0 = e - 5$

$$f(1) = f(e-5+6-e) \approx f(e-5) + \frac{6-e}{e}$$

$$f(1) \approx \frac{6}{e} = 2.207$$

$$f(1.1) = f(e-5+6.1-e) \approx f(e-5) + \frac{6.1-e}{e}$$

$$f(1.1) \approx \frac{6.1}{e} = 2.244$$

$$f(2.5) = f(e-5+7.5-e) \approx f(e-5) + \frac{7.5-e}{e}$$

$$f(2.5) \approx \frac{7.5}{e} = 2.759$$

- (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:** Let the approximation above be  $f'(x)$ .

$$f'(1) - f(1) = 2.207 - 1.792 = 0.415$$

$$f'(1.1) - f(1) = 2.244 - 1.808 = 0.436$$

$$f'(2.5) - f(2.5) = 2.759 - 2.015 = 0.744$$

We can clearly see that the error is increasing as we increase the value of  $h$ . What we are essentially doing with the above approximation is linearizing the function about a point  $x_o$ . However, the log function is not linear. This is a first order approximation and hence there is a mismatch in the values. As we move away from  $x_o$ , the  $h$  value is increasing. But the first order approximation is only valid for infinitesimal  $h$ , hence a discrepancy in the value.

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

**Solution:** Yes, we can get a better approximation by using higher order terms in the Taylor series equation

$$f(x_o + h) = f(x_o) + h \frac{df(x_o)}{dx} + \frac{h^2}{2!} \frac{d^2 f(x_o)}{dx^2} + \dots$$

Or if we still want to use the first order approximation, then we need to choose a point which is infinitesimally close to the function argument and approximate it around that point.

- (d) Consider  $g(x) = a + be^x + c \cos(x)$ . Find  $a, b, c \in \mathbb{R}$  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:**

$$g(x) = a + be^x + c \cos(x)$$

Using Taylor Series approximation, we can approximate the functions around  $x = 0$  as

$$g(x) \approx a + b\left(1 + \frac{x}{1!} + \frac{x^2}{2!} + \dots\right) + c\left(1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \dots\right)$$

An approximation for  $f$  around  $x=0$  using Taylor Series is

$$f(x) \approx \ln(5) + \frac{x}{5} - \frac{x^2}{25 \cdot 2!} + \dots$$

Matching the respective derivative values, we obtain the following equations

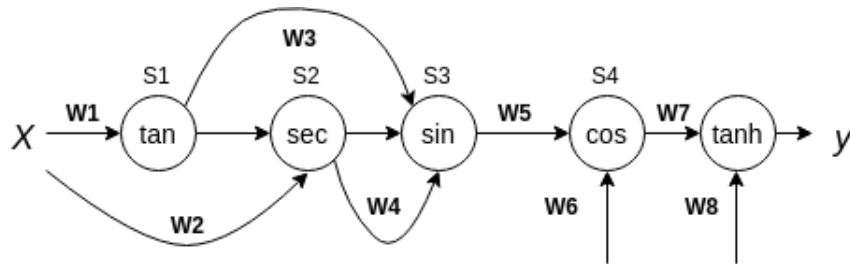
$$a + b + c = \ln(5)$$

$$\begin{aligned}
b &= \frac{1}{5} = 0.2 \\
\frac{b-c}{2} &= \frac{-1}{50} \\
\Rightarrow c &= \frac{6}{25} = 0.24 \\
\Rightarrow a &= \ln(5) - \frac{11}{25} = 1.169
\end{aligned}$$

#### 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:**

$$\begin{aligned}
\frac{dy}{dw_1} &= \frac{\partial y}{\partial (w_7 s_4 + w_8)} \frac{\partial (w_7 s_4 + w_8)}{\partial w_1} \\
\Rightarrow \frac{dy}{dw_1} &= \frac{4}{(e^{w_7 s_4 + w_8} + e^{-(w_7 s_4 + w_8)})^2} w_7 \frac{\partial s_4}{\partial w_1} \\
\Rightarrow \frac{dy}{dw_1} &= \frac{-4w_7 w_5}{(e^{w_7 s_4 + w_8} + e^{-(w_7 s_4 + w_8)})^2} \sin(w_5 s_3 + w_6) \frac{\partial s_3}{\partial w_1} \\
\Rightarrow \frac{dy}{dw_1} &= \frac{-4\sin(w_5 s_3 + w_6)w_7 w_5}{(e^{w_7 s_4 + w_8} + e^{-(w_7 s_4 + w_8)})^2} \cos(w_3 s_1 + w_4 s_2) (w_3 \frac{\partial s_1}{\partial w_1} + w_4 \frac{\partial s_2}{\partial w_1})
\end{aligned}$$

$$\Rightarrow \frac{dy}{dw_1} = \frac{-4\sin(w_5s_3 + w_6)\cos(w_3s_1 + w_4s_2)w_7w_5}{(e^{w_7s_4+w_8} + e^{-(w_7s_4+w_8)})^2} (w_3x\sec^2(w_1x) + w_4\sec(w_2x + s_1)\tan(w_2x + s_1)\frac{\partial(w_2x + s_1)}{\partial w_1})$$

$$\Rightarrow \frac{dy}{dw_1} = \frac{-4\sin(w_5s_3 + w_6)\cos(w_3s_1 + w_4s_2)w_7w_5}{(e^{w_7s_4+w_8} + e^{-(w_7s_4+w_8)})^2} (w_3x\sec^2(w_1x) + w_4\sec(w_2x + s_1)\tan(w_2x + s_1)(x\sec^2(w_1x)))$$

$$\Rightarrow \frac{dy}{dw_1} = \frac{-4w_7w_5x\sin(w_5s_3 + w_6)\cos(w_3s_1 + w_4s_2)\sec^2(w_1x)}{(e^{w_7s_4+w_8} + e^{-(w_7s_4+w_8)})^2} (w_3 + w_4\tan(w_2x + s_1)\sec(w_2x + s_1))$$

$$\frac{dy}{dw_2} = \frac{\partial y}{\partial(w_7s_4 + w_8)} \frac{\partial(w_7s_4 + w_8)}{\partial w_2}$$

$$\Rightarrow \frac{dy}{dw_2} = \frac{4}{(e^{w_7s_4+w_8} + e^{-(w_7s_4+w_8)})^2} w_7 \frac{\partial s_4}{\partial w_2}$$

$$\Rightarrow \frac{dy}{dw_2} = \frac{-4w_7w_5}{(e^{w_7s_4+w_8} + e^{-(w_7s_4+w_8)})^2} \sin(w_5s_3 + w_6) \frac{\partial s_3}{\partial w_2}$$

$$\Rightarrow \frac{dy}{dw_2} = \frac{-4\sin(w_5s_3 + w_6)w_7w_5}{(e^{w_7s_4+w_8} + e^{-(w_7s_4+w_8)})^2} \cos(w_3s_1 + w_4s_2) (w_3 \frac{\partial s_1}{\partial w_2} + w_4 \frac{\partial s_2}{\partial w_2})$$

$$\Rightarrow \frac{dy}{dw_2} = \frac{-4\sin(w_5s_3 + w_6)\cos(w_3s_1 + w_4s_2)w_7w_5}{(e^{w_7s_4+w_8} + e^{-(w_7s_4+w_8)})^2} (0 + (w_4x)\sec(w_2x + s_1)\tan(w_2x + s_1))$$

$$\Rightarrow \frac{dy}{dw_2} = \frac{-4(w_7w_5w_4x)\sin(w_5s_3 + w_6)\cos(w_3s_1 + w_4s_2)\sec(w_2x + s_1)\tan(w_2x + s_1)}{(e^{w_7s_4+w_8} + e^{-(w_7s_4+w_8)})^2}$$

## 5. Differentiation of vectors/matrices

Consider vectors  $\mathbf{u}, \mathbf{b} \in \mathbb{R}^d$ , and matrix  $\mathbf{A} \in \mathbb{R}^{n \times n}$ .



The derivative of a scalar  $f$  w.r.t. a vector  $\mathbf{u}$  is a vector by itself, given by

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

( **Hint:** The derivative of a scalar  $f$  w.r.t. a matrix  $\mathbf{X}$ , 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  $\mathbf{X}$ .)

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

**Solution:** Let  $\mathbf{u}^T = [u_1 \ u_2 \ \dots \ u_n]$ ,  $\mathbf{b}^T = [b_1 \ b_2 \ \dots \ b_n]$  and

$$\mathbf{A} = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \dots & A_{nn} \end{bmatrix}$$

$$\mathbf{u}^T \mathbf{A} \mathbf{u} + \mathbf{b}^T \mathbf{u} = \sum_{i=1}^n b_i u_i + (\mathbf{u}^T \mathbf{A}) \mathbf{u}$$

$$\mathbf{u}^T \mathbf{A} \mathbf{u} + \mathbf{b}^T \mathbf{u} = \sum_{i=1}^n b_i u_i + \left[ \sum_{i=1}^n u_i A_{i1} \quad \sum_{i=1}^n u_i A_{i2} \quad \dots \quad \sum_{i=1}^n u_i A_{in} \right] \mathbf{u}$$

$$\mathbf{u}^T \mathbf{A} \mathbf{u} + \mathbf{b}^T \mathbf{u} = \sum_{i=1}^n b_i u_i + \sum_{k=1}^n u_k \left( \sum_{i=1}^n u_i A_{ik} \right)$$

Let  $f = \mathbf{u}^T \mathbf{A} \mathbf{u} + \mathbf{b}^T \mathbf{u}$ . Hence,

$$\frac{\partial f}{\partial u_i} = b_i + \sum_{k=1, k \neq i}^n u_k A_{ik} + \sum_{k=1, k \neq i}^n u_k A_{ki} + 2u_i A_{ii}$$

$$\frac{\partial f}{\partial u_i} = b_i + \sum_{k=1}^n u_k (A_{ik} + A_{ki})$$

This is one element of the gradient vector. We can obtain the whole gradient vector by the following equation

$$\implies \nabla (\mathbf{u}^T \mathbf{A} \mathbf{u} + \mathbf{b}^T \mathbf{u}) = \mathbf{b} + (\mathbf{A} + \mathbf{A}^T) \mathbf{u}$$

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

**Solution:** In the case of scalar values,  $\mathbf{A} = a$ ,  $\mathbf{b} = b$  and  $\mathbf{u} = u$ . Hence from the previous result obtained,

$$\nabla(\mathbf{u}^T \mathbf{A} \mathbf{u} + \mathbf{b}^T \mathbf{u}) = b + 2au$$

The derivative of the scalar representation w.r.t  $u$  is :

$$\frac{d(au^2 + bu)}{du} = b + 2au$$

Both the derivatives are consistent.

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

**Solution:** Let the  $(i, j)^{th}$  element of the Hessian matrix be  $\frac{\partial^2 f}{\partial u_i \partial u_j}$ . From the previous results obtained

$$\frac{\partial^2 f}{\partial u_i \partial u_j} = \frac{\partial(b_j + \sum_{k=1}^n u_k (A_{jk} + A_{kj}))}{\partial u_i}$$

$$\frac{\partial^2 f}{\partial u_i \partial u_j} = A_{ji} + A_{ij}$$

Hence the Hessian matrix  $H$  will be

$$H = \mathbf{A} + \mathbf{A}^T$$

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:** Let the number of symbols to be represented by a code word =  $N = 11$ .  
 The number of bits required to encode a symbol =  $\lceil \log_2 N \rceil = 4$   
 Let the length of the phrase =  $L = 34$   
 Thus the number of bits required =  $\lceil \log_2 N \rceil * L = 136 \text{ bits}$

- (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:** The weighted average of information content over all symbols  $H(s)$  is

$$H(s) = - \sum_{s \in S} p(s) \log_2 p(s)$$

The  $p(\cdot)$  (probability of occurrence of a symbol), values can be approximated from the table as follows :

$$p(s) = \frac{n}{n_l}$$

where

$n = \text{frequency of symbol}, n_l = \text{length of the phrase}$

$$p('a') = p('c') = p('space') = \frac{5}{34}$$

$$p('d') = p('r') = p('i') = \frac{1}{34}$$

$$p('m') = p('e') = \frac{3}{34}$$

$$p('l') = p('n') = \frac{4}{34}$$

$$p('s') = \frac{2}{34}$$

$$\Rightarrow H(s) = - \left( 3 \frac{5}{34} \log_2 \frac{5}{34} + 3 \frac{1}{34} \log_2 \frac{1}{34} + 2 \frac{3}{34} \log_2 \frac{3}{34} + 2 \frac{4}{34} \log_2 \frac{4}{34} + \frac{2}{34} \log_2 \frac{2}{34} \right)$$

$$\Rightarrow H(s) = 3.254$$

Hence the theoretical number of bits needed to code the entire phrase is  $34 * 3.254 = 110.635 \text{ bits}$

## 7. Plotting Functions for Great Good

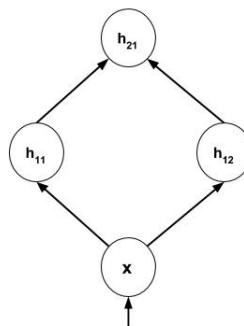
(a) Consider the variable  $x$  and functions  $h_{11}(x)$ ,  $h_{12}(x)$  and  $h_{21}(x)$  such that

$$h_{11}(x) = \frac{1}{1 + e^{-(500x+30)}}$$

$$h_{12}(x) = \frac{1}{1 + e^{-(500x-30)}}$$

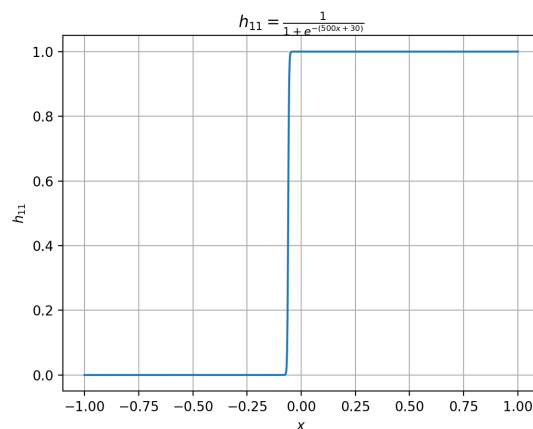
$$h_{21} = h_{11}(x) - h_{12}(x)$$

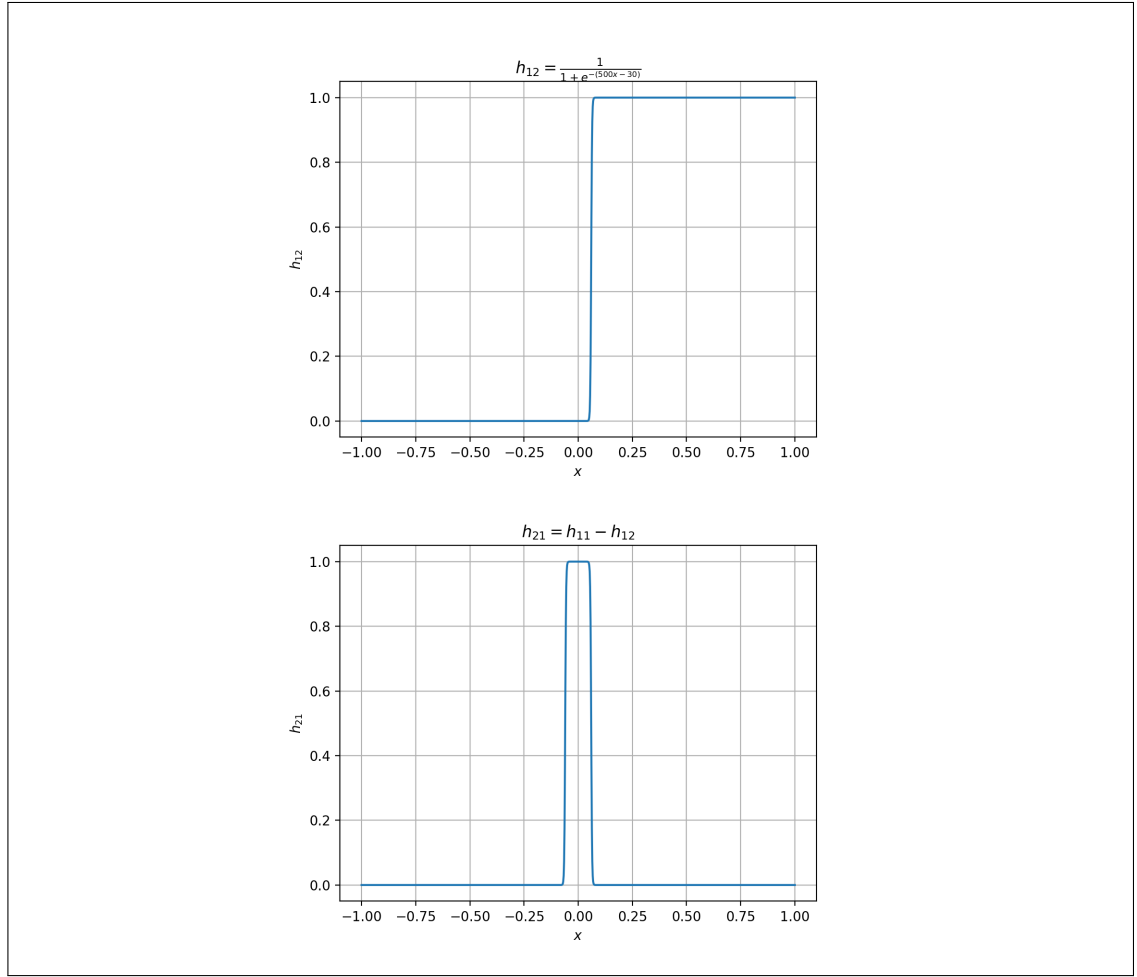
The above set of functions are summarized in the graph below.



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

**Solution:**

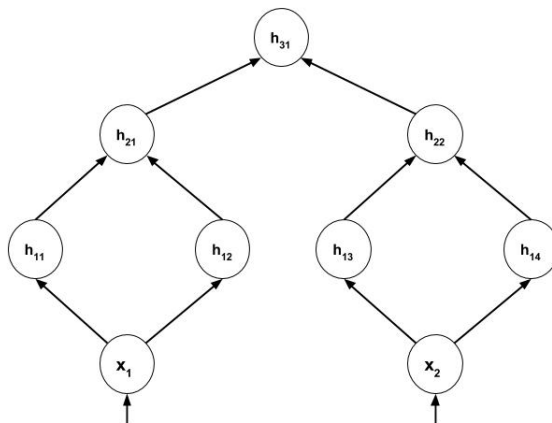




- (b) Now consider the variables  $x_1, x_2$  and the functions  $h_{11}(x_1, x_2), h_{12}(x_1, x_2), h_{13}(x_1, x_2), h_{14}(x_1, x_2), h_{21}(x_1, x_2), h_{22}(x_1, x_2), h_{31}(x_1, x_2)$  and  $f(x_1, x_2)$  such that

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

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:**

