Saarland University, Department of Computer Science
Neural Network Assignment 5

Deborah Dormah Kanubala (7025906) , Irem Begüm Gündüz (7026821), Anh Tuan Tran (7015463)

December 19, 2022

Exercise 5.1 (Newton's Method)

Exercise 5.1.a

Given a function $H : \mathcal{R}^n \Longrightarrow \mathcal{R}^m$
For the Taylor approximations:

$$H(x + \triangle x) = H(x) + \Delta H(x) \triangle x + o(|\triangle x|) = 0$$

In linear approximation, we can neglect $o(|\nabla x|)$ because it is a constant.

$$H(x) + \Delta H(x) \triangle x = 0$$

NB: $\triangle x = x_{k+1} - x_k$

$$\begin{aligned} H(x) + \Delta H(x) \triangle x &= 0 \qquad (1) \\ x_{k+1} - x_k &= -[\Delta H(x)]^{-1} H(x) \\ x_{k+1} &= x_k - [\Delta H(x)]^{-1} H(x) \end{aligned}$$

Exercise 5.1.b

Given $f(x) = x_1 - 3x_1 + x_2 - x_1x_2$
Update Rule for Gradient Descent:

$$\theta' = \theta^0 - \alpha \nabla J(\theta)$$

$$\frac{\partial f(x)}{\partial x_1} = 2x_1 - 3 - x_2; \qquad \frac{\partial f(x)}{\partial x_2} = 2x_2 - x_1$$

$$x^0 = [1,1]$$

$$x_1^0 = 1 - 0.5(2(1) - 3 - 1)$$

$$\nabla x_1' = -2; \qquad x_1^0 = 2$$

$$x_2^0 = 1 - 0.5(2(1) - 1)$$

$$\nabla x_2^0 = 1; \qquad x_2^0 = 0.5 \quad \therefore l_{2-\mathbf{norm}} = \sqrt{(-2)^2 + (1)^2} = 2.236$$

$$x' = [2, 0.5]$$

$$x_1' = 2 - 0.5(2(2) - 3 - 0.5)$$

$$\nabla x_1' = 0.5; \qquad x_1' = 1.75$$

$$x_2' = 0.5 - 0.5(2(0.5) - 2)$$

$$\nabla x_2' = -1; \quad x_2' = 1 \qquad \therefore l_{2-\mathbf{norm}} = \sqrt{(0.5)^2 + (1)^2} = 1.118$$

$$x^2 = [1.75, 1]$$

$$x_1^2 = 1.75 - 0.5(2(1.75) - 3 - 1)$$

$$\nabla x_1^2 = -0.5 \quad ; \quad x_1^2 = 2$$

$$x_2^2 = 1 - 0.5(2(1) - 1.75)$$

$$\nabla x_2^2 = 0.25; x_2^2 = 0.875 \qquad \therefore l_{2-\mathbf{norm}} = \sqrt{(-0.5)^2 + (0.25)^2} = 0.559$$

$$x^3 = [2, 0.875]$$

$$x_1^3 = 2 - 0.5(2(2) - 3 - 0.875)$$

$$\nabla x_1^3 = 0.125; \quad x_1^3 = 1.9375$$

$$x_2^3 = 0.875 - 0.5(2(0.875) - 2)$$

$$\nabla x_2^3 = -0.25; \qquad x_2^3 = 1 \quad \therefore l_{2-\text{ norm}} = \sqrt{(0.125)^2 + (-0.15)^2} = 0.28$$

$$x^4 = [1.9375, 1]$$

$$x_1^4 = 1.9375 - 0.5(2(1.9375) - 3 - 1)$$

$$\nabla x_1^4 = -0.125; \qquad x_1^4 = 2$$

$$x_2^4 = 1 - 0.5(24) - 1.9375) \quad \therefore l_{2-\text{ norm}} = \sqrt{(-0.125)^2 + (0.0625)^2} = 0.14$$

$$\nabla x_2^4 = 0.0625; \qquad 0.96875$$

$$x^5 = [2, 0.96875]$$

Exercise 5.1.c

$f(x) = x_1^2 - 3x_1 + x_2^2 - x_1x_2$ With the Newton method, we need to find the Hessian matrix which is given as:

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 x_2} \\ \frac{\partial^2 f}{\partial x_2 x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix}$$

$$\frac{\partial f}{\partial x_1} = 2x_1 - 3 - x_2; \quad \frac{\partial^2 f}{\partial x_1^2} = 2 \text{ and } \frac{\partial^2 f}{\partial x_2} = 2x_2 - x_1; \quad \frac{\partial^2 f}{\partial x_2^2} = 2$$

$$\frac{\partial^2 f}{\partial x_1 \partial x_2} = -1 \text{ and } \frac{\partial^2 f}{\partial x_2 \partial x_1} = -1$$

$$\therefore \quad H = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

$$\left(H(f)|_{x_t}\right)^{-1} = \frac{1}{3}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

$$\nabla f = \begin{bmatrix} 2x_1 - 3 - x_2 \\ 2x_2 - x_1 \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - 0.5\left[\frac{1}{3}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\right]\begin{bmatrix} 2x_1 - 3 - x_2 \\ 2x_2 - x_1 \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} x_1 - \frac{3}{2} - \frac{1}{2}x_2 \\ -\frac{1}{2} + \frac{1}{2}x_2 \end{bmatrix}$$

$$\Rightarrow 2x_1 + \frac{1}{2}x_2 = -\frac{1}{2}$$

$$\frac{3}{2}x_2 = 3/2$$

$$x_2 = 1, x_1 = -0.5$$

Exercise 5.1.d

No. It is not always applicable as the function may be twice continuously differentiable but be constant throughout. Using the function for instance we have:

$$\begin{aligned} f &= 2x^3 - 5x \\ f^{'}(x) &= 6x^2 - 5 \\ f^{''}(x) &= 12x \\ for \quad x &= 0; f^{''}(x) = 0 \end{aligned} \tag{2}$$

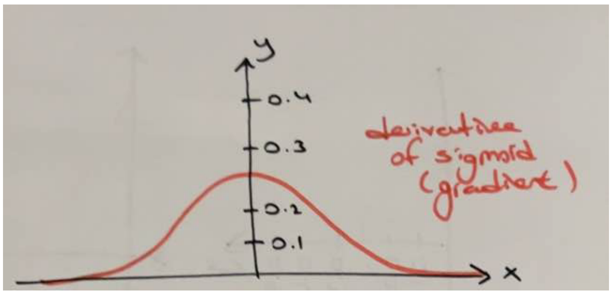
This therefore will mean that there is a saddle point. in machine learning, we often do not want to get stuck in a saddle point.

Exercise 5.2 Sigmoid Function

Exercise 5.2.1 Sigmoid Function Derivation

$$
\begin{aligned}
\frac{d}{dx} \cdot (x) &= \frac{d}{dx}\left[\frac{1}{1+e^{-2}}\right] \\
&= \frac{d}{dx}\left(1+e^{-2}\right)^{-1} \\
&= -\left(1+e^{-x}\right)^{-2}\left(-e^{-x}\right) \\
&= \frac{e^{-x}}{\left(1+e^{-x}\right)^{2}} \\
&= \frac{1}{1+e^{-x}} \cdot \frac{e^{-x}}{1+e^{-x}} \\
&= \frac{1}{1+e^{-x}} \cdot \frac{\left(1+e^{-x}\right)-1}{1+e^{-x}} \\
&= \frac{1}{1+e^{-x}} \cdot \left(\frac{1+e^{-x}}{1+e^{-x}} - \frac{1}{1+e^{-2}}\right) \\
&= \frac{1}{1+e^{-x}} \cdot \left(1 - \frac{1}{1+e^{-x}}\right) \\
&= \sigma(x) \cdot (1-\sigma(x))
\end{aligned}
$$

Exercise 5.2.2



The sigmoid function is differentiable in everywhere, which also indicates the sigmoid function is continous.

Exercise 5.2.3

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

$$\frac{1}{1+e^{-x}} + \frac{1}{1+e^{x}} = 1 \text{ (point symmotiic)}$$

$$\left(1+e^{x}\right)\left(1+e^{-x}\right)$$

$$\frac{1+e^{x}+1+e^{-x}}{\left(1+e^{-x}\right)\left(1+e^{x}\right)} = 1 \rightarrow \frac{2+e^{x}+e^{-x}}{1+e^{x}+e^{-x}+e^{-x^2}} = 1$$

$$2+e^{x}+e^{x} = 1+e^{x}+e^{-k}+e^{-x^2}$$

$$1 = e^{-x^2}$$

$$\log 1 = \log e^{-x^2}$$

$$0 = -x^2 \quad x = 0$$

Exercise 5.2.4

$$\sigma(x) = \frac{1}{1+e^{-z}} - 1$$
$$\sigma(x) = \frac{1}{1+\sum_{k=0}^{\infty}\frac{(-x)^n}{k!}} - 1 = \frac{x}{4} - \frac{x^3}{48} + \frac{x^5}{480} - \frac{17x^7}{80640}$$

where x = 0

Exercise 5.3 (Basics of Forward and Backward passes in computational graphs)

Exercise 5.3.a

Forward pass:

$$\begin{aligned} h_1 =& ReLU(x_1 * 2 + x_2 * 1 + x_3 * -2) \\ =& ReLU(3 * 2 + 1 * 1 + -1 * -2) \\ =& ReLU(9) \\ =& max(0, 9) \\ =& 9 \end{aligned} \tag{3}$$

$$\begin{aligned} h_2 =& ReLU(x_1 * -1 + x_2 * -3 + x_3 * 5) \\ =& ReLU(3 * -1 + 1 * -3 + -1 * 5) \\ =& ReLU(-11) \\ =& max(0, -11) \\ =& 0 \end{aligned} \tag{4}$$

$$\begin{aligned} o_1^{'} =& h_1 * 6 + h_2 * -1 \\ =& 9 * 6 + 0 * -1 \\ =& 54 \end{aligned} \tag{5}$$

$$\begin{aligned} o_2^{'} =& h_1 * 2 + h_2 * 8 \\ =& 9 * 2 + 0 * 8 \\ =& 18 \end{aligned} \tag{6}$$

$$\begin{aligned} o_1 =& \frac{exp(o_1^{'})}{exp(o_1^{'}) + exp(o_2^{'})} \\ =& \frac{exp(54)}{(exp(54) + exp(18))} \\ \approx& 1 \end{aligned} \tag{7}$$

$$\begin{aligned} o_2 =& \frac{exp(o_2^{'})}{exp(o_1^{'}) + exp(o_2^{'})} \\ =& \frac{exp(18)}{exp(54) + exp(18)} \\ \approx& 0 \end{aligned} \tag{8}$$

In a binary classification problem, the output will be the label that's corresponds for o_1. For example, if o_1 corresponds for class 0, then the output is 0.

Exercise 5.3.b

$$\begin{aligned} e =& c * d \\ \Rightarrow & \frac{\partial e}{\partial c} = d \;\; and \;\; \frac{\partial e}{\partial d} = c \end{aligned} \tag{9}$$

$$\begin{aligned} c =& a + b \\ \Rightarrow & \frac{\partial c}{\partial a} = 1 \;\; and \;\; \frac{\partial c}{\partial b} = 1 \end{aligned} \tag{10}$$

$$\begin{aligned} d =& b + 1 \\ \Rightarrow & \frac{\partial d}{\partial b} = 1 \;\; and \;\; \frac{\partial d}{\partial a} = 0 \end{aligned} \tag{11}$$

Therefore, we have

$$\frac{\partial e}{\partial b} = \frac{\partial e}{\partial c} * \frac{\partial c}{\partial b} + \frac{\partial e}{\partial d} * \frac{\partial d}{\partial b} \tag{12}$$

and

$$\frac{\partial e}{\partial a} = \frac{\partial e}{\partial c} * \frac{\partial c}{\partial a} + \frac{\partial e}{\partial d} * \frac{\partial d}{\partial a} = \frac{\partial e}{\partial c} * \frac{\partial c}{\partial a} \tag{13}$$

Exercise 5.4 - Weight Space Symmetry

Exercise 5.4.a

The number of ways to interchanges (both input and output) weights of neurons is equal to the number of permutations of M neurons: $M!$.

The number of ways to choose neurons for change of sign of input weight and output weights are: 2^M.

Therefore, the total number of equivalent transformations is $M! * 2^M$

Exercise 5.5 (Hessian and Optimization: Bonus)

Exercise 5.4.b

With N layers, each layer has $M!*2^M$ equivalent transformations. The total number of equivalent transformations is $(M! * 2^M)^N$.

Exercise 5.5.a

- **Figure a:** Hessian is positive because it is a convex curve.
- **Figure b:** The Hessian has negative eigenvalues because it is a concave curve.
- **Figure c:** The Hessian has one positive/negative and one eigenvalue been 0
- **Figure d:** The Hessian has one eigenvalue been positive and the other negative.

Exercise 5.5.b

- **Figure a:** Hessian is positive-definite of all positive eigenvalues - concave-up or convex.
- **Figure b:** Hessian is negative-definite of all positive eigenvalues - concave-up or convex.
- **Figure c:** Hessian is neither but it could also be a semi-positive definite incase the other eigenvalue is positive.

- **Figure d:** Hessian is neither.