Machine Learning
336546
HW 4

name : Yohan Lascar **I.D :** 800007221
name: Lior Ravia **I.D :** 206913014

Part 1: Theory

1. It is given that X, Y are random variables.

   i. Given : X, Y are statistically independent
   Need to prove: X, Y are not correlated
   Answer:

   To calculate the correlation between X, Y , we can use the following equation:

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_x \sigma_y}$$

$$Cov(X,Y) = E[X \cdot Y] - E[X] \cdot E\,[Y]$$

   Statistically independent random variables X, Y establish:

$$E[X \cdot Y] = E[X] \cdot E[\,Y]$$
$$\Downarrow$$
$$Cov(X,Y) = E[X \cdot Y] - E[X] \cdot E\,[Y] = 0$$
$$\Downarrow$$
$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_x \sigma_y} = 0$$

   Therefore, if X, Y are statistically independent, they are not correlated as well.

   ii. If X, Y are not correlated, we can't deduce they are statistically independent.
   counter example:

$$X \sim Uniform[-1,1] \longrightarrow f(x) = \begin{cases} 0.5 & for \;\; X \in [-1,1] \\ 0 \; otherwise \end{cases}$$

$$Y = X^2 \longrightarrow f(y) = \begin{cases} 1 & for \;\; Y \in [0,1] \\ 0 \; otherwise \end{cases}$$

   Calculations:

$$E[X] = \int_{X=-1}^{1} x \cdot f(x)dx = \int_{X=-1}^{1} x \cdot 0.5\, dx = 0.5 \cdot \left(0.5x^2 | \begin{matrix} 1 \\ -1 \end{matrix} \right) = 0$$

$$E[X \cdot Y] = \int_{X=-1}^{1} x^3 \cdot f(x)dx = 0.5 \cdot \left(0.25x^4 | \begin{matrix} 1 \\ -1 \end{matrix} \right) = 0$$
$$\Downarrow$$

$$Cov(X,Y) = E[X \cdot Y] - E[X] \cdot E\,[Y] = 0$$
$$\Downarrow$$
$$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_x \sigma_y} = 0$$

We can infer that X, Y are not correlated.

Now we would like to show that X, Y are not statistically independent

$$p(\, Y < 0.5) = 0.5$$

$$p(\, Y < 0.5 | X = 0.2) = 1$$
$$\Downarrow$$

$$p(\, Y < 0.5) \neq \, p(\, Y < 0.5 | X = 0.2)$$

We can infer that X, Y are not statistically independent.

2. TRUE/FALSE:

   i. $X \; and \; X^2$ are always independent - $False$
   Counter example, as in the previous question -1.ii :
   $$X \sim Uniform[-1,1] \longrightarrow \; f(x) = \begin{cases} 0.5 & for \;\; X \in [-1,1] \\ & 0 \; otherwise \end{cases}$$
   We showed that
   $$p(\, X^2 < 0.5) \neq \, p(\, X^2 < 0.5 | X = 0.2)$$
   Therefore, in this example X, X^2 are not statistically independent

   ii. $X \; and \; X^2$ are never correlated - $False$
   Counter example:
   $$X \sim Bernouli[0.5] \longrightarrow \; X = \begin{cases} 1 \, , p = 0.5 \\ 0 \, , q = 0.5 \end{cases}$$

   Calculations:
   $$E[X] = 1 \cdot 0.5 + 0 \cdot 0.5 = 0.5$$
   $$E[X^2] = 1^2 \cdot 0.5 + 0^2 \cdot 0.5 = 0.5$$
   $$E[X^3] = 1^3 \cdot 0.5 + 0^3 \cdot 0.5 = 0.5$$
   $$\Downarrow$$
   $$Cov(X, X^2) = E[X \cdot \; X^2] - E[X] \cdot E\,[X^2] = 0.5 - 0.5 \cdot 0.5 = 0.25$$
   $$\Downarrow$$
   $$\rho(X, Y) = \frac{Cov(X, X^2)}{\sigma_x \sigma_{x^2}} \neq 0$$
   Therefore, in this example X, X^2 are correlated. (correlation$\neq$ 0)

iii. $X \; and \; X^2$ are always correlated - $False$

Counter example, as in the previous question -1.ii :

$$X \sim Uniform[-1,1] \longrightarrow \; f(x) = \begin{cases} 0.5 \;\; for \;\; X \in [-1,1] \\ \;\;\; 0 \; otherwise \end{cases}$$

We showed that

$$E[X] = 0, \; E[X \cdot X^2] = 0$$
$$\Downarrow$$
$$Cov(X, X^2) = E[X \cdot X^2] - E[X] \cdot E\,[X^2] = 0$$
$$\Downarrow$$
$$\rho(X,Y) = \frac{Cov(X, X^2)}{\sigma_x \sigma_{x^2}} = 0$$

Therefore, in this example X, X^2 are not correlated.

iv. $X \; and \; X^2$ are never independent – $True$

Proof:

$X = x \; \rightarrow \; X^2 = x^2$ deterministically:

$$p(X^2 = y \, |X = x) = \begin{cases} \;\; 1\,, y = x^2 \\ 0\,, otherwise \end{cases}$$

According to probability rules:

$$p(X^2 = y \, |X = x) = \frac{p(X^2 = y \; and \; X = x)}{p(X = x\,)}$$
$$\Downarrow$$
$$p(X^2 = y \; and \; X = x) = \; p(X^2 = y \, |X = x) \cdot \; p(X = x\,)$$

If $\; y \neq x^2$, then $\; p(X^2 = y \, |X = x) = 0 \; \rightarrow p(X^2 = y \; and \; X = x) = 0$.

1. If $\; y = x^2$, then $\; p(X^2 = y \, |X = x) = 1 \; \rightarrow \; p(X^2 = y \; and \; X = x) = \; p(X = x\,)$

For $X \; and \; X^2$ to be independent, the following equation must be established for all possible values of $X \; and \; X^2$:

$$2. \;\; p(X^2 = y \; and \; X = x) = \; p(X = x\,) \cdot p(X^2 = y\,)$$
$$1 + 2 \Downarrow$$

The following equation must be established:

$$p(X^2 = y\,) = 1$$

The independence condition can only be satisfied if X^2 is constant. For $X^2$to be constant, it depends on the distribution of X itself. Therefore, , $X \; and \; X^2 \;\; are$ never independent.

- We will mention that for $X^2 \; = const$ the independence condition is established.
  Therefore, if you consider this case as independency though X^2 is determined according to X so then the statement is false. We will add a **counter example:**

$$X = \begin{cases} 1\,, p = 0.5 \\ -1\,, q = 0.5 \end{cases}$$

$$\Downarrow$$

$$X^2 = 1$$

We will notice that $p(\,X^2 = a) = \,p(\,X^2 = a|X = b)$ for every value of a $(b = -1\ or\ 1\ \ )$. This happens because $X^2 = const$. Therefore, in this manner $X\ and\ X^2$ are independent

3. Given : $X\ is\ a\ nxn\ matrix\ , C = X^T X$
   Need to prove: $C\ is\ symmetric\ (C = C^T)\ and\ semi\ positive\ definite(\,z^T Cz \geq 0\ for\ every\ z \in R_n)$

   Answer:

   - First, we would like to prove that $C\ is\ symmetric\ (C = C^T)$:

   $$C^T = (X^T X)^T = \,X^T (X^T)^T = X^T X = C$$

   - Second, we would like to prove that $C\ is\ semi\ positive\ definite(\,z^T Cz \geq 0\ for\ every\ z \in R_n)$:

   $$z^T Cz = z^T (X^T X)z = (Xz)^T (Xz) = \|Xz\|^2 \geq 0$$

   I. we used $C = X^T X$
   II. we used the transposed rule: $(AB)^T = B^T A^T$
   III. $Xz \in nx1$
   IV. $\|Xz\|^2 \geq 0,\ \ since\ \|Xz\| \geq 0\,, second\ norms\ are\ always\ non\ negative$

4. Given: $Matrix\ X =$

$$\begin{array}{cc} 0 & 1 \\ 1 & 4 \\ 1 & 5 \\ 2 & 4 \\ 3 & 6 \\ 5 & 7 \end{array}$$

i. We would compute C, the covariance matrix:
   a. Calcaulation of the mean vector μ :

$$\mu_{j=1} = \frac{0+1+1+2+3+5}{6} = 2, \qquad \mu_{j=1} = \frac{1+4+5+4+6+7}{6} = 4.5$$

$$\mu = [2{,}4.5]$$

b. Calculating the centered data matrix, $X-\mu$:

$$\begin{matrix} 0-2 & 1-4.5 \\ 1-2 & 4-4.5 \\ 1-2 & 5-4.5 \\ 2-2 & 4-4.5 \\ 3-2 & 6-4.5 \\ 5-2 & 7-4.5 \end{matrix} \quad \rightarrow X-\mu = \quad \begin{matrix} -2 & -3.5 \\ -1 & -0.5 \\ -1 & +0.5 \\ 0 & -0.5 \\ +1 & +1.5 \\ +3 & +2.5 \end{matrix}$$

c. Computing the covariance matrix C:

$$C=\frac{1}{n-1}\cdot(X-\mu)^T(X-\mu)$$

$$(X-\mu)^T = \begin{matrix} -2 & -1 & -1 & 0 & +1 & +3 \\ -3.5 & -0.5 & +0.5 & -0.5 & +1.5 & +2.5 \end{matrix}$$

$$X-\mu = \begin{matrix} -2 & -3.5 \\ -1 & -0.5 \\ -1 & +0.5 \\ 0 & -0.5 \\ +1 & +1.5 \\ +3 & +2.5 \end{matrix}$$

$n=6$

$$\Downarrow$$

$$\boldsymbol{C}=\frac{\mathbf{1}}{\mathbf{5}}\begin{pmatrix} \mathbf{16} & \mathbf{16} \\ \mathbf{16} & \mathbf{21.5} \end{pmatrix}=\begin{pmatrix} \mathbf{3.2} & \mathbf{3.2} \\ \mathbf{3.2} & \mathbf{4.3} \end{pmatrix}$$

ii. According to the spectral decomposition theorem, every real symmetric matrix has a spectrum. Therefore, the covariance matrix C can be written as:

$$C=U\Lambda U^T$$

- U is an orthonormal matrix, containing the eigenvectors of C
- Λ is a diagonal matrix with the eigenvalues of C on the diagonal

We will find U,Λ and verify $C=U\Lambda U^T$

a. First, we will find the eigenvalues:

$$\det(C-\lambda I)=0$$

$$\begin{vmatrix} 3.2-\lambda & 3.2 \\ 3.2 & 4.3-\lambda \end{vmatrix} = 0$$

$$(3.2-\lambda)(4.3-\lambda) - 3.2^2 = 0$$

$$\lambda^2 - 7.5\lambda + 3.52 = 0$$

$$\Downarrow$$

$$\boldsymbol{\lambda_1 = 6.997, \lambda_2 = 0.503}$$

$$\Lambda = \begin{pmatrix} \mathbf{6.997} & \mathbf{0} \\ \mathbf{0} & \mathbf{0.503} \end{pmatrix}$$

b. Next, we will compute the eigenvectors:
For each λ, we will solve:

$$(C - \lambda I) \cdot v = 0$$

For $\boldsymbol{\lambda_1 = 6.997}$:

$$\begin{pmatrix} 3.2-6.997 & 3.2 \\ 3.2 & 4.3-6.997 \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

1. $-3.797 \cdot v_1 + 3.2 \cdot v_2 = 0$
2. $3.2 \cdot v_1 - 2.697 \cdot v_2 = 0$

These two equations are proportional, so we need to choose one of the values to find the other. Choosing $v_2 = 1$, we get $v_1 = 0.843$
Therefore, the corresponding eigenvector to $\boldsymbol{\lambda_1 = 6.997}$ is: $\begin{pmatrix} \mathbf{0.843} \\ \mathbf{1} \end{pmatrix}$, and after normalization: $\begin{pmatrix} \mathbf{0.644} \\ \mathbf{0.765} \end{pmatrix}$

For $\boldsymbol{\lambda_2 = 0.503}$:

$$\begin{pmatrix} 3.2-0.503 & 3.2 \\ 3.2 & 4.3-0.503 \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

1. $2.697 \cdot v_1 + 3.2 \cdot v_2 = 0$
2. $3.2 \cdot v_1 + 3.797 \cdot v_2 = 0$

These two equations are proportional, so we need to choose one of the values to find the other. Choosing $v_2 = 1$, we get $v_1 = -1.186$
Therefore, the corresponding eigenvector to $\boldsymbol{\lambda_2 = 0.503}$ is: $\begin{pmatrix} \mathbf{-1.186} \\ \mathbf{1} \end{pmatrix}$, and after normalization: $\begin{pmatrix} \mathbf{-0.765} \\ \mathbf{0.644} \end{pmatrix}$

$$\Downarrow$$

$$\mathrm{U} = \begin{pmatrix} \mathbf{0.644} & \mathbf{-0.765} \\ \mathbf{0.765} & \mathbf{0.644} \end{pmatrix}$$

We can notice that U consists of orthonormal eigenvectors, so $U^T U = I$

c. Finally, we will verify $C = U\Lambda U^T$

$$\boldsymbol{U\Lambda U^T} = \begin{pmatrix} 0.644 & -0.765 \\ 0.765 & 0.644 \end{pmatrix} \begin{pmatrix} 6.997 & 0 \\ 0 & 0.503 \end{pmatrix} \begin{pmatrix} 0.644 & 0.765 \\ -0.765 & 0.644 \end{pmatrix} = \begin{pmatrix} \mathbf{3.2} & \mathbf{3.2} \\ \mathbf{3.2} & \mathbf{4.3} \end{pmatrix} = \boldsymbol{C}$$

iii. The main component is the direction in which the data varies the most. This corresponds to the eigenvector associated with the largest eigenvalue of the matrix C.
The largest eigenvalue of C is $\lambda_1 = 6.997$ ($> \lambda_2 = 0.503$), and the corresponding eigenvector is $\begin{pmatrix} 0.644 \\ 0.765 \end{pmatrix}$.

Therefore, $\begin{pmatrix} \mathbf{0.644} \\ \mathbf{0.765} \end{pmatrix}$ is the main component.

iv. The main eigenvalue is the largest eigenvalue of the matrix C. It represents the amount of variance captured by the main principal component.
The largest eigenvalue of C is $\lambda_1 = 6.997$.

Therefore, $\boldsymbol{\lambda_1 = 6.997}$ is the main eigenvalue.

v. The projection of the new data point $x_7 = (3,2)$ on the main component $v_{\lambda_1} = \begin{pmatrix} 0.644 \\ 0.765 \end{pmatrix}$:

$$projection = (x_7 \cdot v_{\lambda_1}) v_{\lambda_1}$$

$$projection = (3 \cdot 0.644 + 2 \cdot 0.765) \begin{pmatrix} 0.644 \\ 0.765 \end{pmatrix} = 3.462 \cdot \begin{pmatrix} 0.644 \\ 0.765 \end{pmatrix} = \begin{pmatrix} 2.23 \\ 2.65 \end{pmatrix}$$

The full projection - the actual vector representing the shadow of $\boldsymbol{x_7}$ in the direction of $\boldsymbol{v_{\lambda_1}}$:

$$\mathbf{3.462} \cdot \begin{pmatrix} \mathbf{0.644} \\ \mathbf{0.765} \end{pmatrix} = \begin{pmatrix} \mathbf{2.23} \\ \mathbf{2.65} \end{pmatrix}$$

5.

a. The matrices $W^{(1)}, W^{(2)}$ according to the convention that has been seen in the tutorial:

$$W^{(1)} = \begin{pmatrix} w_{11} & w_{21} \\ w_{12} & w_{22} \end{pmatrix}$$

$$W^{(2)} = \begin{pmatrix} w_{31} \\ w_{32} \end{pmatrix}$$

- Each row -same source
- Each column - same destination

b. We will write explicitly $y^{\wedge}$ using $W^{(1)}, W^{(2)}$

$$y^{\wedge} = O_1 \cdot w_{31} + O_2 \cdot w_{32}$$
$$\Downarrow$$
$$y^{\wedge} = \varphi_1(v_1) \cdot w_{31} + \varphi_2(v_2) \cdot w_{32}$$
$$\Downarrow$$
$$\boldsymbol{y^{\wedge} = \varphi_1(1 \cdot w_{11} + x \cdot w_{12}) \cdot w_{31} + \varphi_2(1 \cdot w_{21} + x \cdot w_{22}) \cdot w_{32}}$$
$$\Downarrow$$
$$y^{\wedge} = \varphi\left(\begin{bmatrix} 1 & x \end{bmatrix} \begin{pmatrix} w_{11} & w_{21} \\ w_{12} & w_{22} \end{pmatrix}\right) \begin{pmatrix} w_{31} \\ w_{32} \end{pmatrix}$$
$$\Downarrow$$
$$\boldsymbol{y^{\wedge} = (ReLU(\begin{bmatrix} 1 & x \end{bmatrix} W^{(1)}))W^{(2)}}$$

c. We will write $y(x)$ explicitly using $\varphi(x) = max\{0, x\}$

$$y(x) = \begin{cases} x + 1\,, & -2 \le x \le -1 \\ 0\,, & -1 \le x \le 1 \\ -0.5x + 0.5\,, & 1 \le x \le 3 \end{cases}$$
$$\Downarrow$$
$$\boldsymbol{y(x) = -\varphi(-1 - x) - \varphi(-0.5 + 0.5x)}$$

We will find 6 weights that will make the neural output y hat to be equal to the given function $y(x)$**.**

$$y^{\wedge} = \varphi_1(1 \cdot w_{11} + x \cdot w_{12}) \cdot w_{31} + \varphi_2(1 \cdot w_{21} + x \cdot w_{22}) \cdot w_{32}$$

Without loss of generality, we will choose the first neuron to be activated for $x \le -1$, therefore:

$$\boldsymbol{w_{11} = -1},\ \boldsymbol{w_{12} = -1}$$

We will choose the second neuron to be activated for $x \ge 1$, therefore:

$$\boldsymbol{w_{21} = -0.5},\ \boldsymbol{w_{22} = 0.5}$$

In addition:

$$\boldsymbol{w_{31} = -1},\ \boldsymbol{w_{32} = -1}$$

d. Assuming $L = (y - y^{\wedge}(x))^2$, we will calculate ∇L:

$$\nabla L = (\frac{\vartheta L}{\vartheta w_{11}}, \frac{\vartheta L}{\vartheta w_{12}}, \frac{\vartheta L}{\vartheta w_{21}}, \frac{\vartheta L}{\vartheta w_{22}}, \frac{\vartheta L}{\vartheta w_{31}}, \frac{\vartheta L}{\vartheta w_{32}})$$

(It has 6 elements according to the number of weights).

In the loss function: $L = (y - y^{\wedge}(x))^2$, we will substitute:

$$y^{\wedge} = \varphi_1(1 \cdot w_{11} + x \cdot w_{12}) \cdot w_{31} + \varphi_2(1 \cdot w_{21} + x \cdot w_{22}) \cdot w_{32}$$

$$\varphi_1 = \varphi_2 = \varphi$$

$$\Downarrow$$

$$L = (y - (\varphi(1 \cdot w_{11} + x \cdot w_{12}) \cdot w_{31} + \varphi(1 \cdot w_{21} + x \cdot w_{22}) \cdot w_{32}))^2$$

Next, we will calculate the partial derivatives:

$$\frac{\vartheta \mathrm{L}}{\vartheta \mathrm{w}_{11}} = 2 \cdot (\mathrm{y} - \mathrm{y}^{\wedge}(\mathrm{x})) \cdot (-w_{31}(\varphi'(1 \cdot w_{11} + x \cdot w_{12})))$$

$$\frac{\vartheta \mathrm{L}}{\vartheta \mathrm{w}_{12}} = 2 \cdot (\mathrm{y} - \mathrm{y}^{\wedge}(\mathrm{x})) \cdot (-w_{31} \cdot x \cdot (\varphi'(1 \cdot w_{11} + x \cdot w_{12})))$$

$$\frac{\vartheta \mathrm{L}}{\vartheta \mathrm{w}_{21}} = 2 \cdot (\mathrm{y} - y^{\wedge}(\mathrm{x})) \cdot (-w_{32}(\varphi'(1 \cdot w_{21} + x \cdot w_{22})))$$

$$\frac{\vartheta \mathrm{L}}{\vartheta \mathrm{w}_{22}} = 2 \cdot (\mathrm{y} - y^{\wedge}(\mathrm{x})) \cdot (-w_{32} \cdot x \cdot (\varphi'(1 \cdot w_{21} + x \cdot w_{22})))$$

$$\frac{\vartheta \mathrm{L}}{\vartheta \mathrm{w}_{31}} = 2 \cdot \left(\mathrm{y} - y^{\wedge}(\mathrm{x})\right) \cdot (-\upphi(1 \cdot \mathrm{w}_{11} + \mathrm{x} \cdot \mathrm{w}_{12}))$$

$$\frac{\vartheta \mathrm{L}}{\vartheta \mathrm{w}_{32}} = 2 \cdot \left(\mathrm{y} - y^{\wedge}(\mathrm{x})\right) \cdot (-\upphi(1 \cdot \mathrm{w}_{21} + \mathrm{x} \cdot \mathrm{w}_{22}))$$

Since the term $-2 \cdot (\mathrm{y} - \mathrm{y}^{\wedge}(\mathrm{x}))$ repeats in all the partial derivatives, we will define a new variable: **r =** $\mathbf{-2 \cdot (y - y^{\wedge}(x))}$

Therefore, we will get:

$$\boldsymbol{\nabla L = (r \cdot w_{31} \cdot \varphi'(1 \cdot w_{11} + x \cdot w_{12}), r \cdot w_{31} \cdot x \cdot \varphi'(1 \cdot w_{11} + x \cdot w_{12}), r \cdot w_{32} \cdot \varphi'(1 \cdot w_{21} + x \cdot w_{22}), r \cdot w_{32} \cdot x \cdot \varphi'(1 \cdot w_{21} + x \cdot w_{22}), r \cdot \upphi(1 \cdot \mathrm{w}_{11} + \mathrm{x} \cdot \mathrm{w}_{12}), r \cdot \upphi(1 \cdot \mathrm{w}_{21} + \mathrm{x} \cdot \mathrm{w}_{22}))}$$

e. Using the previous section, we will write the update policy of every weight given learning rate η :

vectorized representation:

$$\mathbf{W_{new} = W_{old} - \eta \cdot \nabla} \boldsymbol{L}$$

- $\mathrm{W} = (w_{11}, w_{12}, w_{21}, \mathrm{w}_{22}, w_{31}, w_{32})$

component-wise representation:

$$\mathbf{W_{i,j,new} = W_{i,j,old} - \eta \cdot \frac{\vartheta L}{\vartheta w_{\mathit{ij}}}}$$

f. <u>Given</u>:
$(w_{11}, w_{12}) = (w_{31}, \mathrm{w}_{32}) = (1{,}1)\ and\ (w_{21}, w_{22}) = (-1, -1)$
The first example is $(x, y) = (0{,}0)$

We will calculate the first value of the loss :

$$L = [y - (\varphi(1 \cdot w_{11} + x \cdot w_{12}) \cdot w_{31} + \varphi(1 \cdot w_{21} + x \cdot w_{22}) \cdot w_{32})]^2$$

$$L = [0 - \{\varphi(1 \cdot 1 + 0 \cdot 1) \cdot 1 + \varphi(1 \cdot (-1) + 0 \cdot (-1)) \cdot (-1)\}]^2$$

$$L = [0 - \{\varphi(1) \cdot 1 + \varphi(-1) \cdot (-1)\}]^2$$

$$L = [0 - \{1 \cdot 1 + 0 \cdot (-1)\}]^2$$

$$L = (-1)^2$$

$$\boldsymbol{L = 1}$$

g. To determine whether a weight value will increase or decrease, we should check if the term $\frac{\vartheta \mathrm{L}}{\vartheta \mathrm{w}_{ij}}$ is positive or negative .
According to the equation : $\mathrm{W_{i,j,new}} = \mathrm{W_{i,j,old}} - \eta \cdot \frac{\vartheta \mathrm{L}}{\vartheta \mathrm{w}_{ij}}$
if $\frac{\vartheta \mathrm{L}}{\vartheta \mathrm{w_{ij}}} > 0$ it means that the value decreased, else if $\frac{\vartheta \mathrm{L}}{\vartheta \mathrm{w_{ij}}} < 0$ it means that the value increased

$\frac{\boldsymbol{\vartheta L}}{\boldsymbol{\vartheta w_{11}}} = 2 \cdot \left(\mathrm{y} - y\hat{}(\mathrm{x})\right) \cdot \left(-w_{31}(\varphi'(1 \cdot w_{11} + x \cdot w_{12}))\right) = 2 \cdot (-1) \cdot \left(-1(\varphi'(1+0))\right) = \mathbf{2 > 0} \rightarrow \boldsymbol{decrased}$

$\frac{\boldsymbol{\vartheta L}}{\boldsymbol{\vartheta w_{12}}} = 2 \cdot \left(\mathrm{y} - y\hat{}(\mathrm{x})\right) \cdot \left(-w_{31} \cdot x \cdot (\varphi'(1 \cdot w_{11} + x \cdot w_{12}))\right) = 2 \cdot (-1) \cdot \left(-1 \cdot 0 \cdot (\varphi'(1+0))\right) = \mathbf{0} \rightarrow$
$\boldsymbol{no\ update}$

$\frac{\boldsymbol{\vartheta L}}{\boldsymbol{\vartheta w_{21}}} = 2 \cdot \left(y - y\hat{}(x)\right) \cdot (-w_{32}(\varphi'(1 \cdot w_{21} + x \cdot w_{22}))) = 2 \cdot (-1) \cdot \left(-1 \cdot (\varphi'(1 \cdot (-1) + 0))\right) = \mathbf{0} \rightarrow$
$\boldsymbol{no\ update}$

$\frac{\boldsymbol{\vartheta L}}{\boldsymbol{\vartheta w_{22}}} = 2 \cdot \left(y - y\hat{}(x)\right) \cdot (-w_{32} \cdot x \cdot (\varphi'(1 \cdot w_{21} + x \cdot w_{22}))) = 2 \cdot (-1) \cdot 0 \cdot \left(-1 \cdot (\varphi'(-1+0))\right) = \mathbf{0} \rightarrow$
$\boldsymbol{no\ update}$

$$\frac{\boldsymbol{\vartheta}\mathbf{L}}{\boldsymbol{\vartheta}\mathbf{w}_{\mathbf{31}}} = 2\cdot\left(\,\mathrm{y} - y^{\wedge}(\mathrm{x})\right)\cdot(-\varphi(1\cdot\mathrm{w}_{11} + \mathrm{x}\cdot\mathrm{w}_{12})) = 2\cdot(\,-1)\cdot(-\varphi(1\cdot 1 + 0\cdot 1)) = \ \mathbf{2} > \mathbf{0}\ \rightarrow \boldsymbol{decrased}$$

$$\frac{\boldsymbol{\vartheta}\mathbf{L}}{\boldsymbol{\vartheta}\mathbf{w}_{\mathbf{31}}} = 2\cdot\left(\,\mathrm{y} - y^{\wedge}(\mathrm{x})\right)\cdot(-\varphi(1\cdot\mathrm{w}_{21} + \mathrm{x}\cdot\mathrm{w}_{22})) = 2\cdot(\,-1)\cdot(-\varphi(1\cdot(-1) + 0)) = \ \mathbf{0}\ \rightarrow \boldsymbol{no\ update}$$