Automated Learning and Data Analysis

Team G40

TEAM MEMBER 1 Sujith Tumma,stumma2

TEAM MEMBER 2 Vishnu Challa,vchalla2

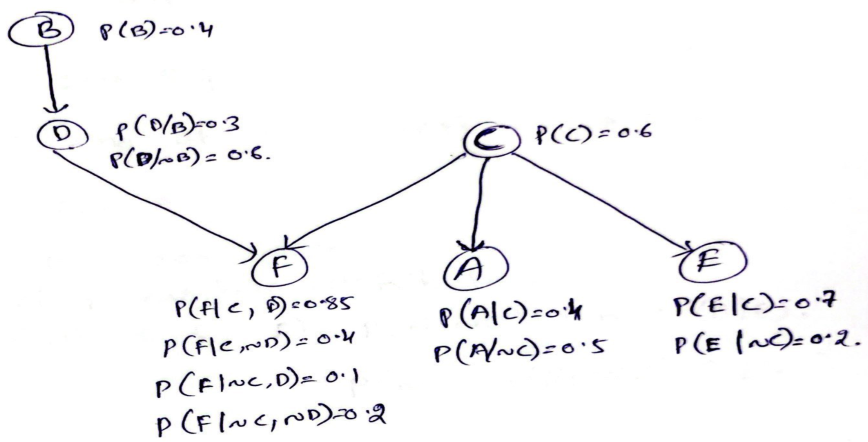
TEAM MEMBER 3 Indu Chenchala,ichench

April 5, 2022

1 BN Inference

1. 1a) (2 points) Compute P(A). Show your work.

1 BN Inference.



1a) To compute P(A),

we know that,

$$P(A) = P(A/C) \cdot P(C) + P(A|\sim C) \cdot P(\sim C).$$

$$= (0.4)(0.6) + (0.5)(0.4)$$

$$= 0.44$$

$$\boxed{\therefore P(A) = 0.44}$$

2. 1b) (2 points) Compute P(D—B, A). Show your work.

(b) To compute $P(D/B, \sim A)$

we can expand $P(D/B, \sim A)$ as,

$$P(D|B, \sim A) = P(D/B) \times P(D|\sim A).$$

AS, we know that 'A' and 'D' are marginally independent. It is safe to ignore the second term on left hand side i.e; $P(D|\sim A)$.

$$\therefore P(D|B, \sim A) = P(D/B).$$

given, that, $P(D|B) = 0.3$

we can say that,

$$\boxed{P(D|B, \sim A) = 0.3}$$

3. 1c) (2 points) Compute P(A, B, ~C, D, E, F). Show your work.

4. 1d) (2 points) Under the Causal Markov Assumption, are E and F conditionally independent given C? Justify your answer in 1 sentence.

(1c)

To, compute $P(A, B, \sim C, D, E, F)$

we can expand the given probability by looking at the parent's of each and every node,

so, we can write
By, Joint probability distribution,

$$P(A, B, \sim C, D, E, F) = P(A|\sim C) \cdot P(B) \cdot P(D|B) \cdot P(\sim C) \cdot P(F|\sim C, D) \cdot P(E|\sim C)$$

$$= (0.5) \cdot (0.4) \cdot (0.3) \cdot (0.4) \cdot (0.1) \cdot (0.2)$$

$$= 0.00048.$$

(1d) Yes,
If 'C' is given:
The occurences of 'E' and 'F' are independent to each other. Because, 'C' is common parent of 'E' and 'F'.
If 'C' is not given:
The occurences of 'E' and 'F' are not independent to each other.

By, above statement's we can say, 'E' and 'F' are conditionally independent given 'C'.

5. 1e) (2 points) Under the Causal Markov Assumption, are A and B marginally independent? Justify your answer in 1 sentence.

6. 1f) (2 points) Given evidence that A = true, C = true D = false, and E = true, use the Bayes Net to predict whether F is more likely to be true or false, or whether both are equally likely.

(1 e) Yes,

Knowing ~~either~~ value of 'A' is either True/False does not effect the value of 'B'. Because they not have a parent in common.

Similarly, vice versa, knowing value of 'B' does not effect the value of 'A'.

∴ We can say that, 'A' and 'B' are marginally independent

(1 f)

Given,

when given evidence A = true, C = true, D = false and E = true

need to find, whether 'F' ~~whether~~ is more likely to be true or false.

i.e; $P(F \mid A, C, \sim D, E)$ and $P(\sim F \mid A, C, \sim D, E)$

(1f)

(i) probability of 'F' to be true,

$\Rightarrow P(A, C, \overline{D}, E, F) = P(A/C) \times P(C) \times [P(\overline{D}/B) \times P(B) + P(\overline{D}|\overline{B}) \times P(\overline{B})] \times P(E|C) \times P(F|C, \overline{D})$.

$= (0.4) \times (0.6) \times (0.7 \times 0.4 + 0.4 \times 0.6) \times 0.7 \times 0.4$

$= 0.4 \times 0.6 \times 0.52 \times 0.7 \times 0.4$

$= 0.034944.$

$\simeq 0.035$

(ii) $P(A, C, \overline{D}, E, \overline{F})$

$\Rightarrow P(A|C) \times P(C) \times [P(\overline{D}|B) \times P(B) + P(\overline{D}|\overline{B}) \times P(\overline{B})] \times P(E/C) \times P(\overline{F}|C, \overline{D})$

$= 0.4 \times 0.6 \times [0.7 \times 0.4 + 0.4 \times 0.6] \times 0.7 \times 0.6$

$= 0.4 \times 0.6 \times 0.52 \times 0.7 \times 0.6$

$= 0.052416.$

$\simeq 0.052$

Now,

$P(A, C, \bar{D}, E, F) = 0.035$

$P(A, C, D, E, F) = 0.052$

So,

$0.052 > 0.035.$

The Bayes Net more likely predict 'F' to be "False".

2 Linear Regression

1. Determine the values of w1 and w0 and show each step of your work.

2 Linear Regression

(2a) Given,

Training Data points and linear Regression is of the form $y = w_1 x^2 + w_0$

x	y	$\phi(x) = x^2$	$x_i y_i$
-2	0	4	0
0	-1	0	0
2	-4	4	-8
1	2	1	2
$\Sigma x_i = 1$	$\Sigma y_i = -3$	$\Sigma x_i^2 = 9$	$\Sigma x_i y_i = -6$

$y = w_1 x^2 + w_0$, number of Data points $= n = 4$

$$\boxed{w_0 = \frac{\Sigma_i (y_i - w_1 x_i^2)}{n}} \rightarrow (1)$$

Taking the Derivative of Error with respect to w_1 and set to 0:

$$\frac{\partial}{\partial w} \sum_i (y_i - (w_1 x_i^2 + w_0))^2$$

$$= 2 \sum_i -x_i^2 (y_i - w_1 x_i^2 - w_0)$$

$$\Rightarrow 2 \sum_i -x_i^2 (y_i - w_1 x_i^2 - w_0) = 0$$

$$\Rightarrow \sum_i x_i^2 (y_i - w_0) = \sum_i w_1 x_i^4$$

$$\boxed{w_1 = \frac{\sum_i x_i^2 (y_i - w_0)}{\Sigma x_i^4}} \rightarrow (2)$$

Figure 1: 2(i)Derivation for w1 and w0

On Substituting w_0 in Equation (2)

$$w_1 = \frac{\sum_i x_i^2 \left(y_i - \frac{\sum_i (y_i - w_1 x_i^2)}{n}\right)}{\sum_i x_i^4}$$

$$w_1 = \frac{\sum_i x_i^2 \left(\frac{n y_i - \sum_i (y_i - w_1 x_i^2)}{n}\right)}{\sum_i x_i^4}$$

$$w_1 = \frac{n \sum_i x_i^2 y_i - \sum_i y_i \sum_i x_i^2 + \sum_i x_i^2 \times \sum_i w_1 x_i^2}{n \sum_i x_i^4}$$

$$w_1 \sum_i x_i^4 n = n \sum_i x_i^2 y_i - \sum_i y_i \sum_i x_i^2 + \left(\sum_i x_i^2\right)^2 w_1$$

$$w_1 \left[\left(n \sum_i x_i^4\right) - \left(\sum_i x_i^2\right)^2\right] = n \sum_i x_i^2 y_i - \sum_i y_i \sum_i x_i^2$$

$$\boxed{w_1 = \frac{n \sum_i x_i^2 y_i - \sum_i y_i \sum_i x_i^2}{n \sum_i x_i^4 - \left(\sum_i x_i^2\right)^2}}$$

Substituting the Datapoints

$$w_1 = \frac{4(4(0) + 0(-1) + 4(-4) + 2(1)) - (-3(9))}{4(16 + 0 + 16 + 1) - (4 + 0 + 4 + 1)^2}$$

$$w_1 = \frac{-56 + 27}{132 - 81}$$

$$\boxed{w_1 = -0.56}$$

Figure 2: 2(i) calculation for w1

$$W_0 = \frac{\sum_i (y_i - w_1 x_i^2)}{n}$$

Substituting Datapoints and w_1 :

$$w_0 = \frac{\sum_i y_i - w_1 \sum_i x_i^2}{n}$$

$$w_0 \Rightarrow \frac{-3 + 0.56(9)}{4}$$

$$\boxed{w_0 = 0.51}$$

Figure 3: 2(ii) calculation for w0

2. Calculate the training RMSE for the fitted linear regression. Show your work.

(ii) Training RMSE:-

x	y	$f(x)$	Error	Squared Error
-2	0	-1.73	1.73	2.99
0	-1	0.51	-1.51	2.28
2	4	-1.73	-2.27	5.15
1	2	-0.05	2.05	4.20

$$\text{RMSE} = \sqrt{\frac{\sum_i^n (y_i - f(x_i))^2}{n}}$$

$$= \sqrt{\frac{2.99 + 2.28 + 5.15 + 4.20}{4}}$$

$$= \sqrt{\frac{14.62}{4}}$$

$$= \sqrt{3.655}$$

$$\boxed{\text{RMSE} = 1.91}$$

Figure 4: 2(ii) calculation for RMSE

3 ANN + Backpropagation

3) ANN + Backpropagation

(a) Given output of each units for $X1 = 1$ $X2 = (-1)$, and $Y = 1$

activation of $H1 = a_{H1}$

$$a_{H1} \Rightarrow f(w1X1 + w3X2)$$
$$\Rightarrow f(0.3 \times 1 + 0.4 \times (-1))$$
$$\Rightarrow f(-0.1) = M \times (-0.1)$$

Since activation function

$f(z) = Mz$ where $\boxed{M = 1}$

$$\Rightarrow f(-0.1) = 1 \times (-0.1)$$

$$\therefore \boxed{a_{H1} = -(0.1)}$$

$$a_{H2} \Rightarrow f(w2X1 + w4X2)$$
$$\Rightarrow f((-0.1) \times 1 + 0.7 \times (-1))$$
$$\Rightarrow f(-0.1 - 0.7)$$
$$\Rightarrow f(-0.8)$$
$$\Rightarrow 1 \times (-0.8)$$

$$\therefore \boxed{a_{H2} = -(0.8)}$$

$$a_0 \Rightarrow f(w5 a_{H1} + w6 a_{H2})$$
$$\Rightarrow f((-0.2) \times -(0.1) + 0.6 \times -(0.8))$$
$$\Rightarrow f(0.02 - 0.48)$$
$$\Rightarrow f(-0.46)$$

Figure 5: `3(a) Calculation for Forward Pass`

$\Rightarrow 1 \times -0.46$

$$\boxed{a_0 = -(0.46)}$$

3(b)(i)

(i) Cost function $(C) = \dfrac{(Y_i - a_0)^2}{n}$

where $n = 1$ since we have only one instance

$$\boxed{C = (Y_i - a_0)^2}$$

(ii) Activation function $(a_0) = w5\, a_{H1} + w6\, a_{H2}$

$$\therefore \boxed{a_0 = w5\, a_{H1} + w6\, a_{H2}}$$

(iii) Activation Function $(a_{H1}) = w1 X1 + w3 X2$

$$\therefore \boxed{a_{H1} = w1 X1_i + w3 X2_i}$$

(ii) For the second layer weights w5 and w6

$$\frac{\partial C}{\partial a_0} = -2(Y_i - a_0)$$

$$\Rightarrow -2(1 - (-0.46))$$

$$\boxed{\frac{\partial C}{\partial a_0} = -(2.92)}$$

CamScanner

Figure 6: `3(b)(i) Formula for backward pass`

Since $C = (Y_i - a_o)^2$

$$a_o = f(z) \Rightarrow Mz \quad (\text{where } M = 1)$$

$$z = X1_1 w1 w5 + X2_1 w3 w5 + X1_1 w2 w6 + X2_1 w4 w6$$

$$\Rightarrow \frac{\partial z}{\partial w5} = (X1_1 w1 + X2_1 w3)$$

$$\Rightarrow \frac{\partial z}{\partial w6} = (X1_1 w2 + X2_1 w4)$$

$$\Rightarrow \frac{\partial a_o}{\partial z} = 1$$

$$\therefore \frac{\partial a_o}{\partial w5} = \frac{\partial a_o}{\cancel{\partial z}} \times \frac{\cancel{\partial z}}{\partial w5}$$

$$\Rightarrow 1 \times (X1_1 w1 + X2_1 w3)$$

$$\Rightarrow a_{H1}$$

$$\therefore \boxed{\frac{\partial a_o}{\partial w5} = -(0.1)}$$

$$\therefore \frac{\partial a_o}{\partial w6} = \frac{\partial a_o}{\partial z} \times \frac{\partial z}{\partial w5}$$

$$\Rightarrow 1 \times (X1_1 w2 + X2_1 w4)$$

$$\Rightarrow a_{H2}$$

$$\therefore \boxed{\frac{\partial a_o}{\partial w6} = -(0.8)}$$

Figure 7: 3(b)(ii) Calculation for Layer-2 weights

(iii) Thus the gradients for $w5$ and $w6$ are:

$$\frac{\partial C}{\partial w5} \Rightarrow \frac{\partial C}{\cancel{\partial a_0}} \times \frac{\cancel{\partial a_0}}{\partial w5}$$ } from the equations in subquestion (ii)

$$\Rightarrow -2.92 \times -0.1$$

$$\therefore \boxed{\frac{\partial C}{\partial w5} \Rightarrow 0.292}$$

$$\frac{\partial C}{\partial w6} = \frac{\partial C}{\cancel{\partial a_0}} \times \frac{\cancel{\partial a_0}}{\partial w6}$$ } from the equations in subquestion (ii)

$$\Rightarrow -2.92 \times -0.8$$

$$\therefore \boxed{\frac{\partial C}{\partial w6} \Rightarrow 2.336}$$

(iv) for weight $w1$ equations are

$$a_0 = a_{H1} w5 + a_{H2} w6$$

$$a_{H1} = X1_1 w1 + X2_1 w3$$

$$\therefore \boxed{\frac{\partial a_0}{\partial a_{H1}} \Rightarrow w5 = -(0.2)}$$

Figure 8: 3(b)(iii) & (iv) Calculation for Layer-2 weights gradient descent

$$\therefore \boxed{\frac{\partial a_{H1}}{\partial w1} = x1_1 \Rightarrow 1}$$

$$\text{and } \boxed{\frac{\partial C}{\partial a_o} = -(2.92)}$$

from the equation in subquestion (ii)

(V) Thus the gradient for w1 is:

$$\frac{\partial C}{\partial w1} = \frac{\partial C}{\cancel{\partial \alpha_o}} \times \frac{\cancel{\partial \alpha_o}}{\cancel{\partial \alpha_{H1}}} \times \frac{\partial \alpha_{H1}}{\partial w1}$$

$$\Rightarrow -2.92 \times -0.2 \times 1$$

$$\therefore \boxed{\frac{\partial C}{\partial w1} = 0.584}$$

(vi) Updating the weights w1 & w; from the gradients in subquestion (V)

<u>for w1.</u>

$$w1' = w1 - \alpha \frac{\partial C}{\partial w1}$$

$$\text{Since } \boxed{\alpha = 0.1}$$

$$w1' \Rightarrow 0.3 - 0.1 \times 0.584$$

$$\boxed{w1^1 = 0.2416}$$

Figure 9: 3(b)(v) Calculation for w1 updating and gradient descent

for w5

$$w5' = w5 - \alpha \frac{\partial C}{\partial w5}$$

$$\Rightarrow -0.2 - 0.1 \times 0.292$$

$$\boxed{w5' = -0.2292}$$

for w6

$$w6' = w6 - \alpha \frac{\partial C}{\partial w6}$$

$$= 0.6 - 0.1 \times 2.336$$

$$\boxed{w6' = 0.3664}$$

for w2

for weight w2 equations are

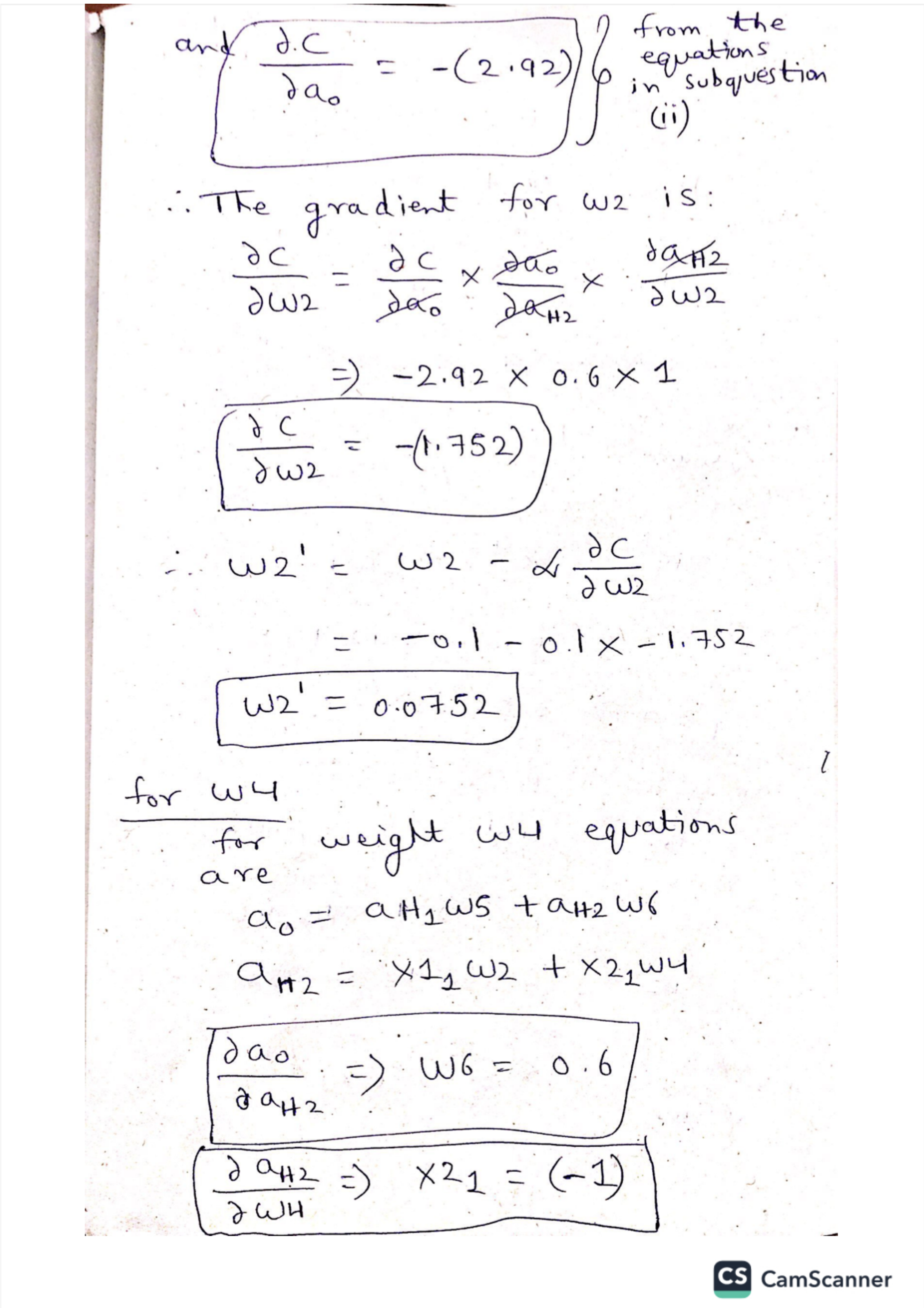
$$a_o = a_{H_1} w5 + a_{H_2} w6$$

$$a_{H2} = x1_1 w2 + x2_1 w4$$

$$\therefore \boxed{\frac{\partial a_o}{\partial a_{H2}} \Rightarrow w6 = 0.6}$$

$$\therefore \boxed{\frac{\partial a_{H2}}{\partial w2} = x1_1 = 1}$$

Figure 10: 3(b)(vi) Weights updating calculations



and $\boxed{\frac{\partial C}{\partial a_o} = -(2.92)}$ from the equations in subquestion (ii)

$\therefore$ The gradient for w2 is:

$$\frac{\partial C}{\partial w2} = \frac{\partial C}{\cancel{\partial a_o}} \times \frac{\cancel{\partial a_o}}{\cancel{\partial a_{H2}}} \times \frac{\cancel{\partial a_{H2}}}{\partial w2}$$

$$\Rightarrow -2.92 \times 0.6 \times 1$$

$$\boxed{\frac{\partial C}{\partial w2} = -(1.752)}$$

$$\therefore w2' = w2 - \alpha \frac{\partial C}{\partial w2}$$

$$= -0.1 - 0.1 \times -1.752$$

$$\boxed{w2' = 0.0752}$$

for w4

for weight w4 equations are

$$a_o = a_{H_1} w5 + a_{H2} w6$$

$$a_{H2} = x1_1 w2 + x2_1 w4$$

$$\boxed{\frac{\partial a_o}{\partial a_{H2}} \Rightarrow w6 = 0.6}$$

$$\boxed{\frac{\partial a_{H2}}{\partial w4} \Rightarrow x2_1 = (-1)}$$

Figure 11: 3(b)(vi) Weights updating calculations

and $\boxed{\dfrac{\partial C}{\partial a_o} = -(2.92)}$ } from the equations in subquestion (ii)

$\therefore$ The gradient for w4 is

$$\frac{\partial C}{\partial w4} = \frac{\partial C}{\partial a_o} \times \frac{\partial a_o}{\partial a_{H2}} \times \frac{\partial a_{H2}}{\partial w4}$$

$$\Rightarrow -2.92 \times 0.6 \times -1$$

$$\boxed{\frac{\partial C}{\partial w4} \Rightarrow 1.752}$$

$$\therefore w_4' = w4 - \alpha \frac{\partial C}{\partial w4}$$

$$\Rightarrow 0.7 - 0.1 \times 1.752$$

$$\boxed{w_4' \Rightarrow 0.5248}$$

<u>for w3</u>

for weight w3 equations are

$$a_o = a_{H1} w5 + a_{H2} w6$$

$$a_{H1} = X1_1 W1 + X2_1 W3$$

$$\therefore \boxed{\frac{\partial a_o}{\partial a_{H1}} = w5 = -(0.2)}$$

Figure 12: `3(b)(vi) Weights updating calculations`

$$\therefore \boxed{\frac{\partial a_{H1}}{\partial w3} = X2_1 = -(1)}$$

and $$\boxed{\frac{\partial C}{\partial a_o} = -(2.92)}$$ from the equations in subquestion (ii)

$\therefore$ The gradient for $w3$ is

$$\frac{\partial C}{\partial w3} = \frac{\partial C}{\partial a_o} \times \frac{\partial a_o}{\partial a_{H1}} \times \frac{\partial a_{H1}}{\partial w3}$$

$$\Rightarrow -2.92 \times -0.2 \times -1$$

$$\boxed{\frac{\partial C}{\partial w3} = -(0.584)}$$

$$\therefore w3' = w3 - \alpha \times \frac{\partial C}{\partial w_3}$$

$$\Rightarrow 0.4 - 0.1 \times -0.584$$

$$\therefore \boxed{w_3' = 0.4584}$$

Figure 13: `3(b)(vi) Weights updating calculations`