Linear Regression

Definition: Linear regression is a statistical method used in machine learning and statistics to model the relationship b/w a dependent variable (also known as response variable) and one or more independent variables (predictors or features).
The relationship is modeled as a linear equation, hence the term "linear regression".

The general form of a linear regression equation for a single independent variable is:

$$\boxed{Y = b_0 + b_1 X + \varepsilon}$$

where:
- Y is the dependent variable.
- X is the independent variable.
- b_0 is the y-intercept (value of Y when X is 0)
- b_1 is the slope of the line (the
- ε represents the error term, accounting for the variability in Y that cannot be explained by the linear relationship with X.

Aim: The goal of linear regression is to find the values of b_0 and b_1 that minimize the sum of squared differences between the observed values of Y, and the values predicted by the linear equation.
This process is known as "least squares" fitting.

- Extensions of linear regression to multiple independent variables involve the equation:

$$\boxed{Y = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_K X_K + \varepsilon}$$

Here, $X_1, X_2, \dots, X_K$ are the indep. variables,
& $b_0, b_1, \dots, b_K$ are the coeff. to be determined

① Assumptions

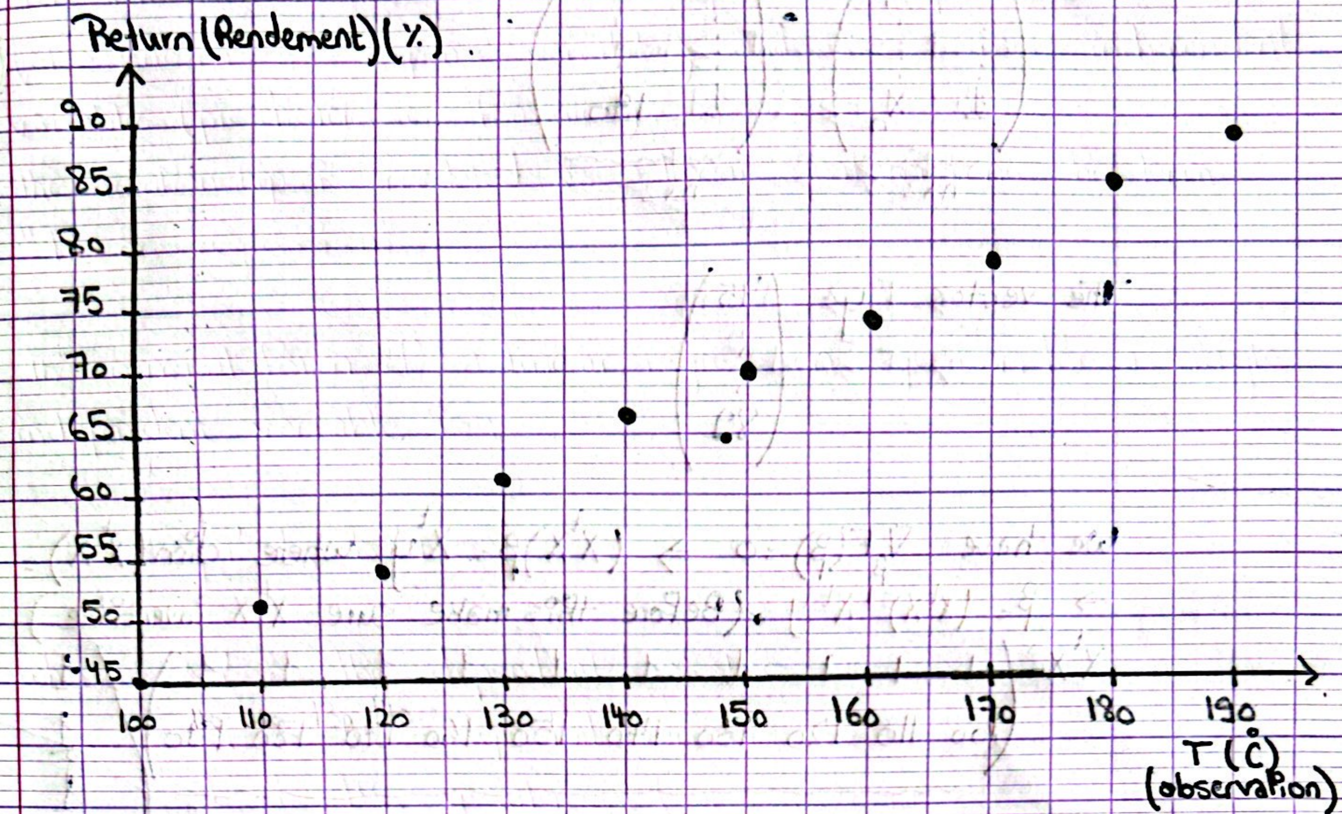
- Linearity: the relationship b/w. variables is assumed to be linear
- Independence: observations are assumed to be indep. of each other.
- Homoscedasticity: diff. b/w observed and predicted values should have constant variance
- Normality: Residuals should be approximately normally distributed

② Interpretation

The coefficients $b_0, b_1, \dots, b_K$ represent the strength and direction of the relationship b/w the indep. and dep. variables.

Exercises- Linear Regression.

Exercise 1.



⇒ The graphic representation is almost linear.

We can determine the predictive model using linear regression.

$\boxed{R = \beta_0 + \beta_1 T}$ $\quad \beta_0 = ? \quad \beta_1 = ? \Rightarrow$ Vector $\beta = (\beta_0 \ \beta_1)^T = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$.

R: Function of linear regression.

Least square sum difference $= (y - X\beta)^T (y - X\beta) = F(\beta)$ Scalar Function

$(-X^T)(y - X\beta) + (y - X\beta)^T(-X)$

[R = 45 51 54 61 66 70 74 78 85 89]

We have $n=10$ and $p=1$.

The matrix X is given by:

$$X = \underbrace{\begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}}_{n\times 2} = \underbrace{\begin{pmatrix} 1 & 100 \\ \vdots & \vdots \\ 1 & 190 \end{pmatrix}}_{n\times 2}.$$

The vector $y = \begin{pmatrix} 45 \\ \vdots \\ 89 \end{pmatrix}$

We have $\nabla_\beta F(\beta) = 0 \Rightarrow (X^t X)\beta = X^t y$ where $\dim(X^t X) = 2\times 2$

$\rightarrow \beta = (X^t X)^{-1} X^t y$. (Before this, make sure $X^t X$ invertible)

$$X^t X = \underbrace{\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 100 & 110 & 120 & 130 & 140 & 150 & 160 & 170 & 180 & 190 \end{pmatrix}}_{2\times n} \underbrace{\begin{pmatrix} 1 & 100 \\ 1 & 110 \\ 1 & 120 \\ 1 & 130 \\ 1 & 140 \\ 1 & 150 \\ 1 & 160 \\ 1 & 170 \\ 1 & 180 \\ 1 & 190 \end{pmatrix}}_{n\times 2}$$

$$\Rightarrow X^t X = \begin{pmatrix} 10 & 1450 \\ 1450 & 218500 \end{pmatrix}$$

$\det(X^tX) = 2185000 - (1450)^2 = 82500 \neq 0$

$\Rightarrow (X^tX)^{-1}$ exists.

Let $H = (X^tX)^{-1}$

We want to find $(X^tX)^{-1}$.

$$\boxed{A^{-1} = \frac{1}{\det A}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}}$$

$$(X^tX)^{-1} = \frac{1}{\det(X^tX)}\begin{pmatrix} 218500 & -1450 \\ -1450 & 10 \end{pmatrix}$$

$$= \frac{1}{82500}\begin{pmatrix} 218500 & -1450 \\ -1450 & 10 \end{pmatrix}$$

Now, let us find X^ty

$$X^ty = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 100 & 110 & 120 & 130 & 140 & 150 & 160 & 170 & 180 & 190 \end{pmatrix}\begin{pmatrix} 45 \\ \vdots \\ 89 \end{pmatrix} = \begin{pmatrix} 673 \\ 101570 \end{pmatrix}.$$

Hence, $\beta = \frac{1}{82500}\begin{pmatrix} 218500 & -1450 \\ -1450 & 10 \end{pmatrix}\begin{pmatrix} 673 \\ 101570 \end{pmatrix}$

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \frac{1}{82500}\begin{pmatrix} 218500 \times 673 - 1450 \times 101570 \\ -1450 \times 673 + 10 \times 101570 \end{pmatrix}$$

$\Rightarrow \beta_0 = \frac{-226000}{82500} = -2.7394$ & $\beta_1 = \frac{39850}{82500} = 0.483$

$\Rightarrow \beta = \begin{pmatrix} -2.7394 \\ 0.483 \end{pmatrix}$

$\Rightarrow \boxed{P(T) = -2.7394 + 0.483T}$

Papillon

2) If $T=80 \Rightarrow P(T) = 0.483\times 80 - 2.7394 = 35.9006 \approx 36\%$

3) Returns $\geqslant 100$?

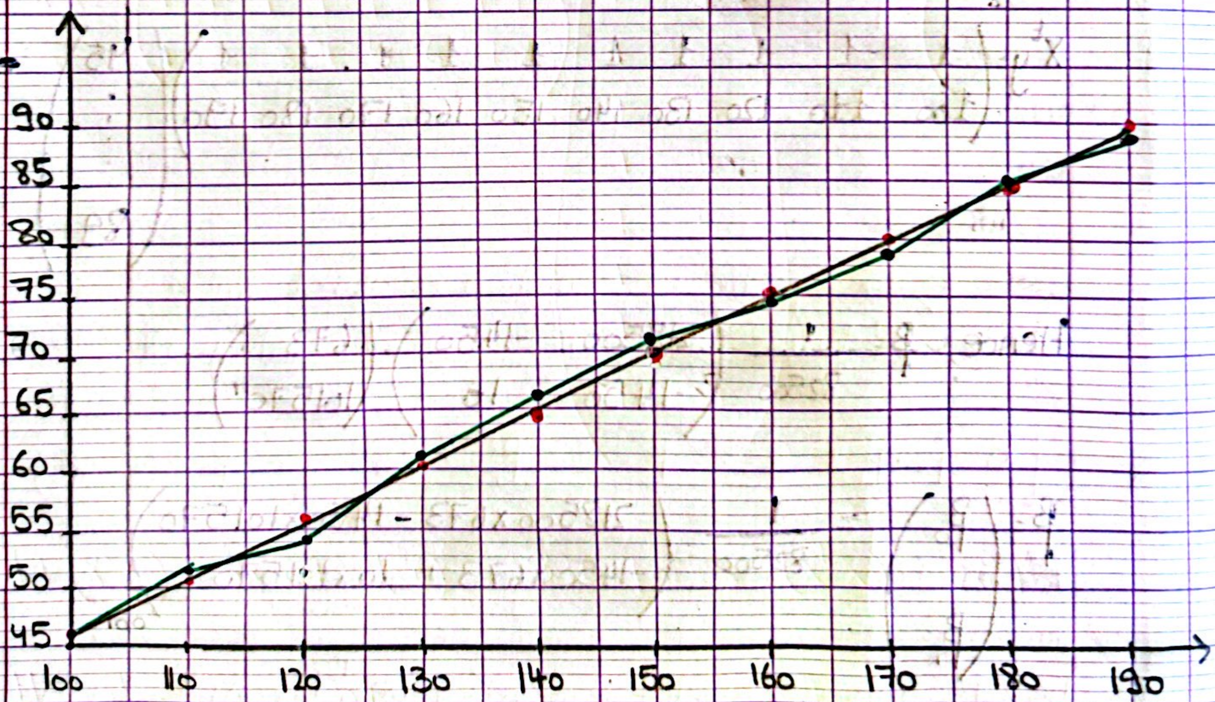
$$-2.7394 + 0.483T \geqslant 100$$

$$0.438T \geqslant 102.7394$$

$$T \geqslant \frac{102.7394}{0.438T}$$

$$\Rightarrow T \geqslant 212.711$$

$$\Rightarrow T \geqslant 212^{\circ}$$



Restime: 45.5606 / 50.3906 / 55.2206 / 60.0506 / 64.8806 / 69.7106 / 74.5406
79.3706 / 84.2006 / 89.0306

Error vector: 0.5606 / −0.6094 / 1.2206 / −0.9494 / −1.1194 / −0.2894 / 0.5406 /
1.3706 / −0.7994 / 0.0306

norm error vector: 7.2244

Papillon

Artificial Neural Networks.

_Artificial Neural Networks (ANNs) are a class of machine learning models inspired by the structure and functioning of the human brain. They consist of interconnected nodes, also known as artificial neurons or perceptrons, organized into layers.
The three main types of layers in a typical neural network are: "the input layer, hidden layers, and output layer.

1) Neurons/ Nodes/ Perceptrons:

_These are the basic units of the neural network.
_Each node receives one or more input, processes them using weights that are adjusted during training, and produces an output using an activation function.

2) Layers:

_Input layer: this layer receives the initial input data.
_Hidden Layers: these layers come between the input and output layers and are responsible for learning complex patterns in the data.
_Output Layer: this layer produces the final output of the network.

3) Weights and Bias:

_Weights: each connection between nodes has an associated weight that determines the strength of the connection. These weights are adjusted during training to optimize the network's performance.
_Bias: each node has an associated bias that allows it to have some level of activation even when the inputs are all zero.

4) Activation Functions:

_The activation function determines the output of a node given its inputs and weights. It introduces non-linearity to the network, allowing it to learn complex patterns.

5) FeedForward and Backpropagation:

_FeedForward is the process of passing input data through the network to get the output.

_Backpropagation is the training algorithm that adjusts the weights and biases based on the error between the predicted output and actual output.

6) Training:

During training, the network learns from a labled dataset by adjusting its weights and biases to minimize the difference between predicted and actual outputs.

7) Deep Learning:

_ANNs with multiple hidden layers are referred to as deep neural networks. The use of deep learning has been particularly sucessful in tasks like image and speech recognition.

Scanned with CamScanner

→ Mathematical Description

1) Representation:

- The input to an artificial neural network can be represented as a vector, and the entire dataset can be organized as a matrix.

2) Neurons and Activation Functions:

- Each neuron in the network performs a weighted sum of its inputs, which can be expressed mathematically as a dot product between the input vector and a weight vector.

Mathematically, for a single neuron j in a layer, the weighted sum (z_j) is given by:

$$\boxed{z_j = \sum_{i=1}^{n} (w_{ij} \cdot x_i) + b_j}$$

- Here, w_{ij} is the weight connecting the i-th input to the j-th neuron, x_i is the i-th input / b_j is the bias of the j-th neuron / n nb of input

- The output of the neuron is then passed through an activation function, denoted as $a_j = f(z_j)$ where f is the activation function.

3) Feedforward Operation:

- The output of one layer serves as the input to the next layer. The feedforward operation can be expressed as matrix multiplication for efficiency.

$$\boxed{a^{(l)} = f(W^{(l)} \cdot a^{(l-1)} + b^{(l)})}$$

Here:
- $W^{(l)}$ is the weight matrix.
- $a^{(l-1)}$ is the output of the previous layer.
- $b^{(l)}$ is the bias vector
- f is the activation function.

4) Loss Function:

The performace of the network is measured by a loss function, which quantifies the difference between the predicted output and the actual output. The goal during training is to minimize the loss.

5) Backpropagation:

- Backpropagation involves computing the gradient of the loss with respect to the weights and biases of the network. This is done using chain rule. $\frac{\partial Loss}{\partial W}$ and $\frac{\partial Loss}{\partial b}$

- The weights and biases are then updated in the opp direction of the gradient to minimize the loss.

6) Optimization:

Optimization algorithms, such as gradient descent, are used to update the weights and biases iteratively during training.

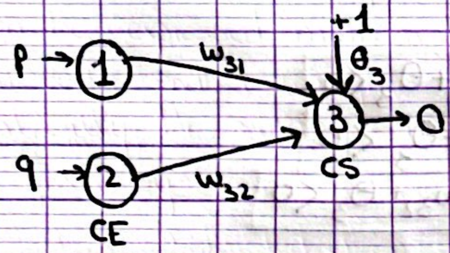
Exercises - ANN.

Exercise 1

Figure 1 shows the structure of the neural network that is used to perform the two logical Fcts (NAND) and (NOR) $O=\overline{pq}$ and $O=\overline{p+q}$

The activation Fct of the output neuron is: $P(A)=\begin{cases}1 & \text{if } A\geq 0\\ 0 & \text{if } A<0\end{cases}$

Find the parameters w_{31}, w_{32}, and θ_3 of each neural network



① **NAND:** $O=\overline{pq}$

p	q	O_d
0	0	1
0	1	1
1	0	1
1	1	0

$A_3 = w_{31}p + w_{32}q + \theta_3$

Final output

step Fct

$O = P(A_3)$

$p=q=0$, $O_d=1 \Rightarrow A_3 \geq 0 \Rightarrow \boxed{\theta_3 \geq 0}$

$p=0$, $q=1$, $O_d=1 \Rightarrow \boxed{w_{32}+\theta_3 \geq 0}$

$p=1$, $q=0$, $O_d=1 \Rightarrow \boxed{w_{31}+\theta_3 \geq 0}$

$p=1=q$, $O_d=0 \Rightarrow \boxed{w_{31}+w_{32}+\theta_3 < 0}$

$\Rightarrow w_{31}=-0.5$, $w_{32}=-0.5$ and $\theta_3=0.8$

② **NOR:** $O = \overline{p+q}$

p	q	O_d
0	0	1
0	1	0
1	0	0
1	1	0

$A_3 = w_{31} p + w_{32} q + \theta_3$

$O = f(A_3)$

$p = q = 0$, $\boxed{\theta_3 \geq 0}$

$p = 0, q = 1, O_d = 0$, $\boxed{w_{32} + \theta_3 < 0}$

$p = 1, q = 0, O_d = 0$, $\boxed{w_{31} + \theta_3 < 0}$

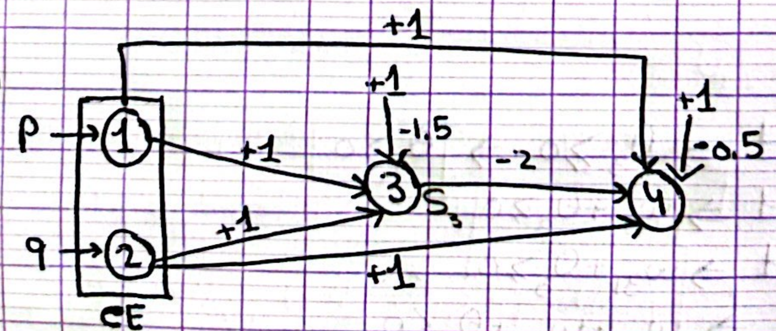
$p = q = 1, O_d = 0$, $\boxed{w_{31} + w_{32} + \theta_3 < 0}$

We can choose, $w_{31} = -0.8$, $w_{32} = -0.8$ and $\theta_3 = 0.5$

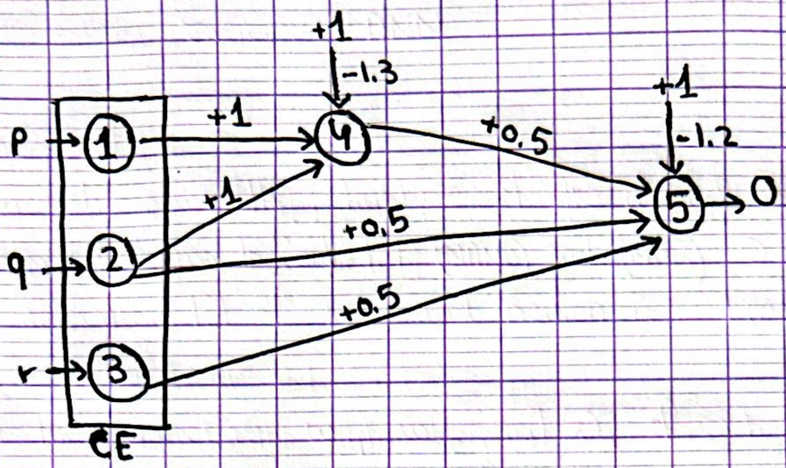
Exercise 2

Consider the two neural networks shown in Fig. 2. The activation function of the hidden neurons and the output neuron is the step fct. Determine the logic fcts performed by these two neural networks

(a)



(b)



(a) $A_3 = p + q - 1.5$

$A_4 = -2S_3 - 0.5 + p + q$

p	q	A_3	S_3	A_4	O
0	0	-1.5	0	-0.5	0
0	1	-0.5	0	0.5	1
1	0	-0.5	0	0.5	1
1	1	0.5	1	-0.5	0

$\Rightarrow O = \bar{p}q + p\bar{q} \Rightarrow O = p (XOR) q$

(b) $A_4 = p + q - 1.3$

$A_5 = 0.5S_4 + 0.5q + 0.5r - 1.2$

p	q	r	A_4	S_4	A_5	O
0	0	0	-1.3	0	-1.2	0
0	0	1	-1.3	0	-0.7	0
0	1	0	-0.3	0	-0.7	0
0	1	1	-0.3	0	-0.2	0
1	0	0	-0.3	0	-1.2	0
1	0	1	-0.3	0	-0.7	0
1	1	0	0.7	1	-0.7	0
1	1	1	0.7	1	0.3	1

$O = pqr$

Exercise 5.

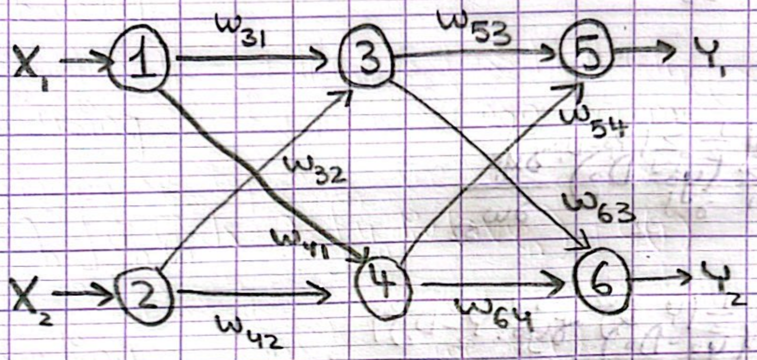
Let's consider the MLP network given below. The hidden neurons have a sigmoidal activation function, while the output neurons have a linear activation function.

a) Calculate MLP outputs Y_1 and Y_2 as a fct of external inputs X_1 and X_2

b) Calculate $\frac{\partial E}{\partial w_{53}}$, $\frac{\partial E}{\partial w_{32}}$, $\frac{\partial E}{\partial w_{41}}$ and $\frac{\partial Y_2}{\partial X_1}$ (w_{40} is the threshold of neuron

where $E = \frac{1}{2}\sum_{i=1}^{2}(Y_i - D_i)^2$ is the calculated error at the network

D_1 and D_2 are the desired MLP outputs.



solutions

a) $Y_1 = A_5$; $A_5 = w_{53}S_3 + w_{54}S_4 + w_{50}$;

$S_3 = f(A_3)$; $A_3 = w_{31}X_1 + w_{32}X_2 + w_{30}$;

$S_4 = f(A_4)$; $A_4 = w_{41}X_1 + w_{42}X_2 + w_{40}$;

So, $Y_1 = w_{53}f(w_{31}X_1 + w_{32}X_2 + w_{30}) + w_{54}f(w_{41}X_1 + w_{42}X_2 + w_{40}) + w_{50}$

Similarly, $Y_2 = A_6$; $A_{63} = w_{63}S_3 + w_{64}S_4 + w_{60}$;

$Y_2 = w_{63}f(w_{31}X_1 + w_{32}X_2 + w_{30}) + w_{64}f(w_{41}X_1 + w_{42}X_2 + w_{40}) + w_{60}$

b) $\frac{\partial E}{\partial w_{53}} = ?$

$$\frac{\partial E}{\partial w_{53}} = \frac{\partial}{\partial w_{53}}\left[\frac{1}{2}\sum_{i=1}^{2}(y_i - D_i)^2\right]$$

$$= \frac{\partial}{\partial w_{53}}\left[\frac{1}{2}(y_1 - D_1)^2\right] = (y_1 - D_1)\frac{\partial y_1}{\partial w_{53}}$$

$$y_1 = A_5 \Rightarrow \frac{\partial y_1}{\partial w_{53}} = \frac{\partial A_5}{\partial w_{53}} = S_3 = F(A_3)$$

$$\Rightarrow \frac{\partial E}{\partial w_{53}} = (y_1 - D_1)F(A_3)$$

$\frac{\partial E}{\partial w_{32}} = ?$

$$\frac{\partial E}{\partial w_{32}} = \frac{\partial\left(\frac{1}{2}\sum_{i=1}^{2}(Y_i - D_i)^2\right)}{\partial w_{32}}$$

$$= \frac{\partial\left(\frac{1}{2}(Y_1 - D_1)^2 + \frac{1}{2}(Y_2 - D_2)^2\right)}{\partial w_{32}}$$

$$= (Y_1 - D_1)\frac{\partial Y_1}{\partial w_{32}} + (Y_2 - D_2)\frac{\partial Y_2}{\partial w_{32}}$$

$$= (Y_1 - D_1)\frac{\partial A_5}{\partial w_{32}} + (Y_2 - D_2)\frac{\partial A_6}{\partial w_{32}}$$

$$= (Y_1 - D_1)\frac{\partial(w_{53}S_3 + w_{54}S_4 + w_{50})}{\partial w_{32}} + (Y_2 - D_2)\frac{\partial(w_{63}S_3 + w_{64}S_4 + w_{60})}{\partial w_{32}}$$

$$= (Y_1 - D_1)w_{53}\frac{\partial S_3}{\partial w_{32}} + (Y_2 - D_2)w_{63}\frac{\partial S_3}{\partial w_{32}}$$

$$= (Y_1 - D_1)w_{53}\frac{\partial(F(A_3))}{\partial A_3}\frac{\partial A_3}{\partial w_{32}} + (Y_2 - D_2)w_{63}\frac{\partial(F(A_3))}{\partial A_3}\frac{\partial A_3}{\partial w_{32}}$$

$$= (Y_1 - D_1)w_{53}F'(A_3)X_2 + (Y_2 - D_2)w_{63}F'(A_3)X_2$$

Papillon

$$\Rightarrow \frac{\partial E}{\partial w_{32}} = \sum_{i=1}^{2} (y_i - D_i)\, w_{(4+i)3}\, f'(A_3)\, X_2$$

$\odot$ $$\frac{\partial E}{\partial w_{41}} = \sum_{i=1}^{2} (y_i - D_i) \frac{\partial y_i}{\partial w_{41}}$$

$$= \sum_{i=1}^{2} (y_i - D_i) \frac{\partial A_{4+i}}{\partial S_4} \cdot \frac{\partial S_4}{\partial A_4} \cdot \frac{\partial A_4}{\partial w_{41}}$$

$$= \sum_{i=1}^{2} (y_i - D_i) \cdot w_{(4+i)4} \cdot f'(A_4) \cdot X_1$$

$\odot$ $$\frac{\partial y_2}{\partial x_1} = \frac{\partial A_6}{\partial x_1} = \frac{\partial}{\partial x_1} (w_{63} S_3 + w_{64} S_4 + w_{60})$$

$$= w_{63} \frac{\partial S_3}{\partial x_1} + w_{64} \frac{\partial S_4}{\partial x_1}$$

$$= w_{63} \cdot f'(A_3) \cdot \frac{\partial A_3}{\partial x_1} + w_{64} \cdot f'(A_4) \cdot \frac{\partial A_4}{\partial x_1}$$

$$= w_{63} \cdot f'(A_3) \cdot w_{31} + w_{64} \cdot f'(A_4) \cdot w_{41}$$

$$\Rightarrow \frac{\partial y_2}{\partial x_1} = \sum_{K=3}^{4} w_{6K} \cdot f'(A_K)\, w_{K1}$$