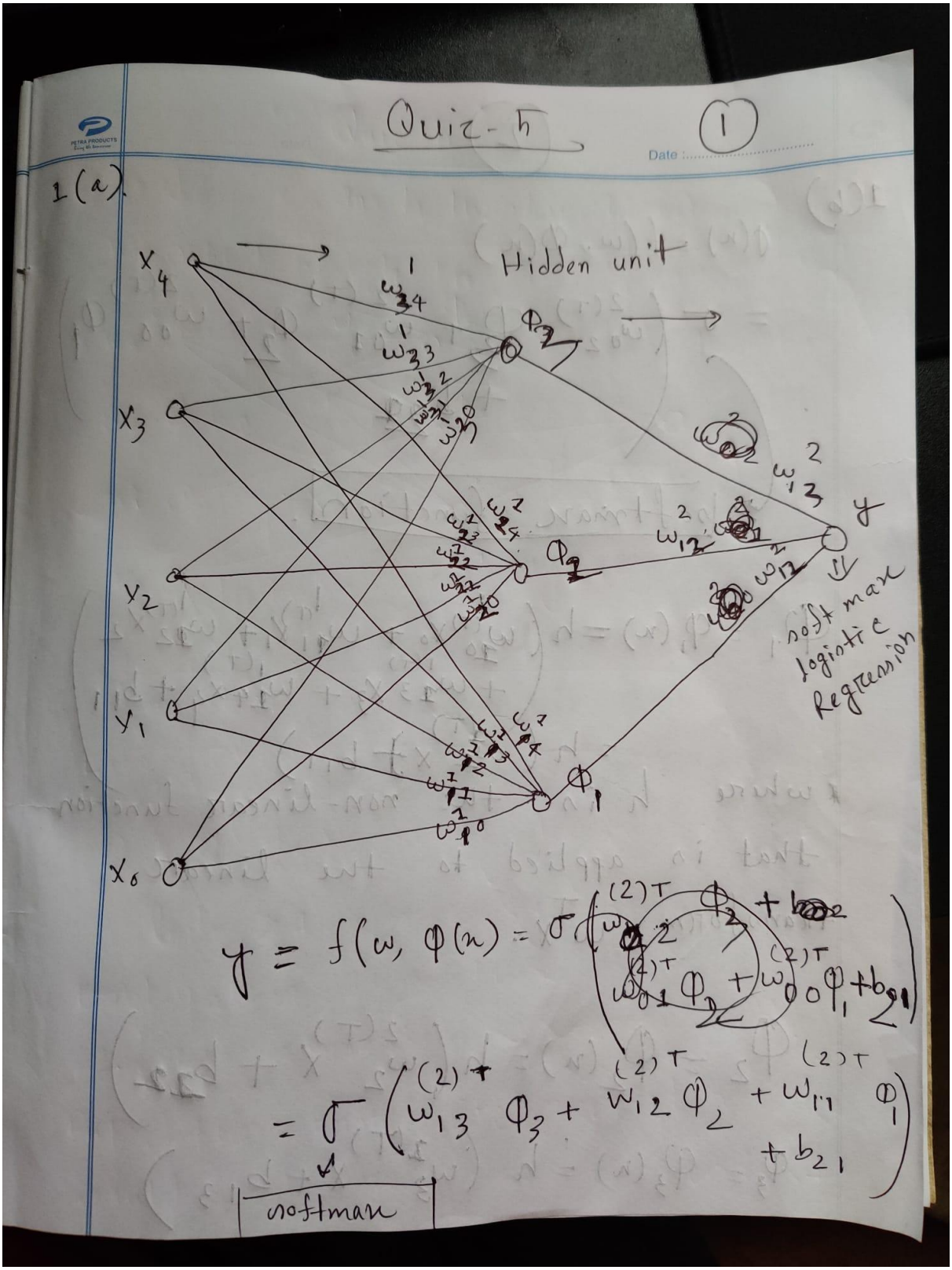
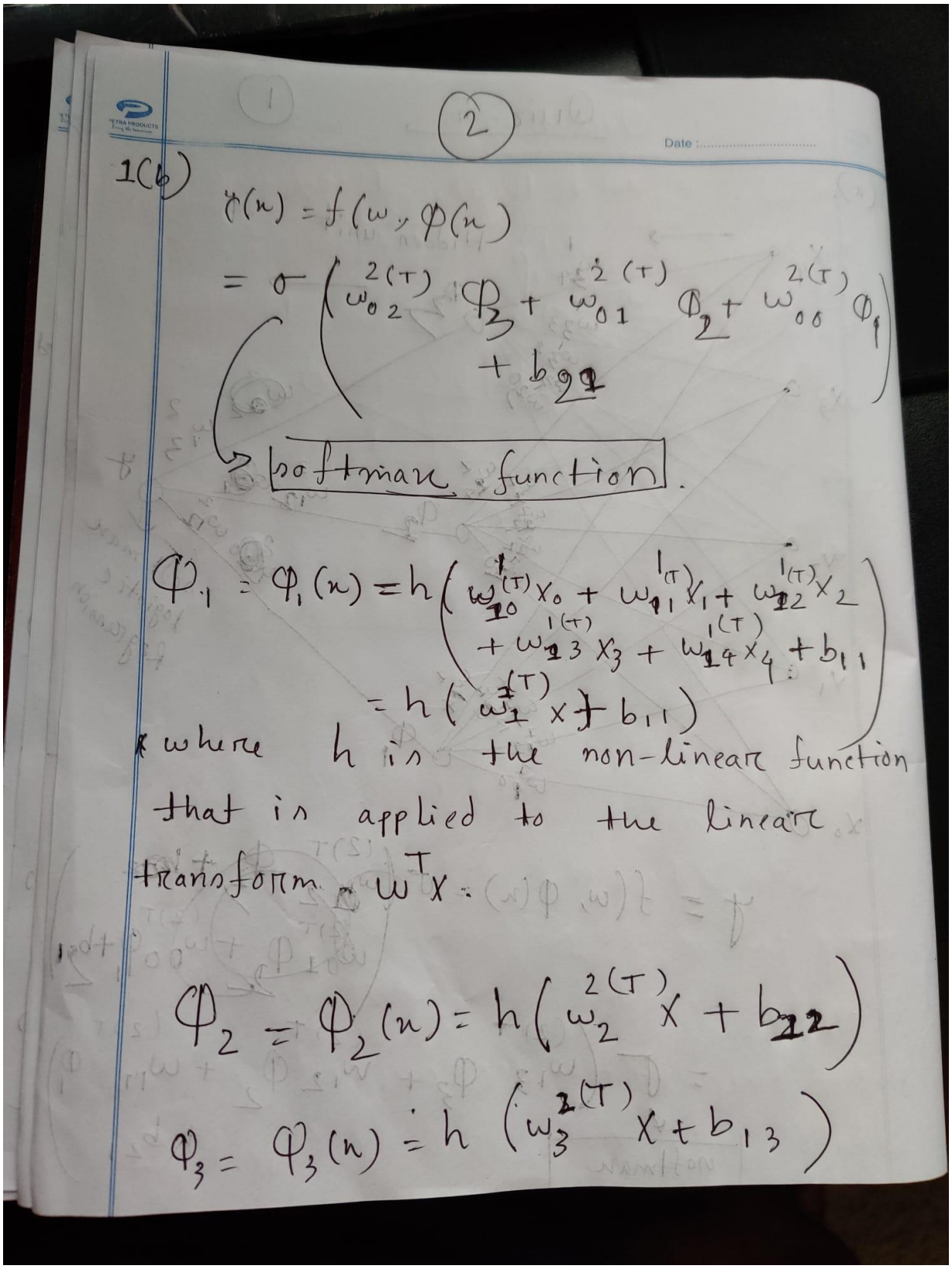
Ans to the Question 1 (a)



Ans to the Question 1 (b)



1(b)

$$y(n) = f(w, \varphi(n))$$

$$= \sigma\left(w_{02}^{2(T)} \varphi_3 + w_{01}^{2(T)} \varphi_2 + w_{00}^{2(T)} \varphi_1 + b_{21} \right)$$

→ softmax function.

$$\varphi_1 = \varphi_1(n) = h\left(w_{10}^{1(T)} x_0 + w_{11}^{1(T)} x_1 + w_{12}^{1(T)} x_2 + w_{13}^{1(T)} x_3 + w_{14}^{1(T)} x_4 + b_{11} \right)$$

$$= h\left(w_1^{1(T)} x + b_{11} \right)$$

* where h is the non-linear function that is applied to the linear transform $w^T x$.

$$\varphi_2 = \varphi_2(n) = h\left(w_2^{2(T)} x + b_{12} \right)$$

$$\varphi_3 = \varphi_3(n) = h\left(w_3^{2(T)} x + b_{13} \right)$$

Date

* The non-linearity / non-linear function is applied to learn the non-linear relations in the data. Different functions; Relu, Crelu, Sigmoid etc.

* Roles of Φ; These are called hidden neurons of units. Each neuron learns a different feature or relation in the data. At the end the final output is the ~~different~~ scaled/weighted function of the output of different neurons where the input to each of the neuron is the scaled / weighted ~~function~~ version of each input.