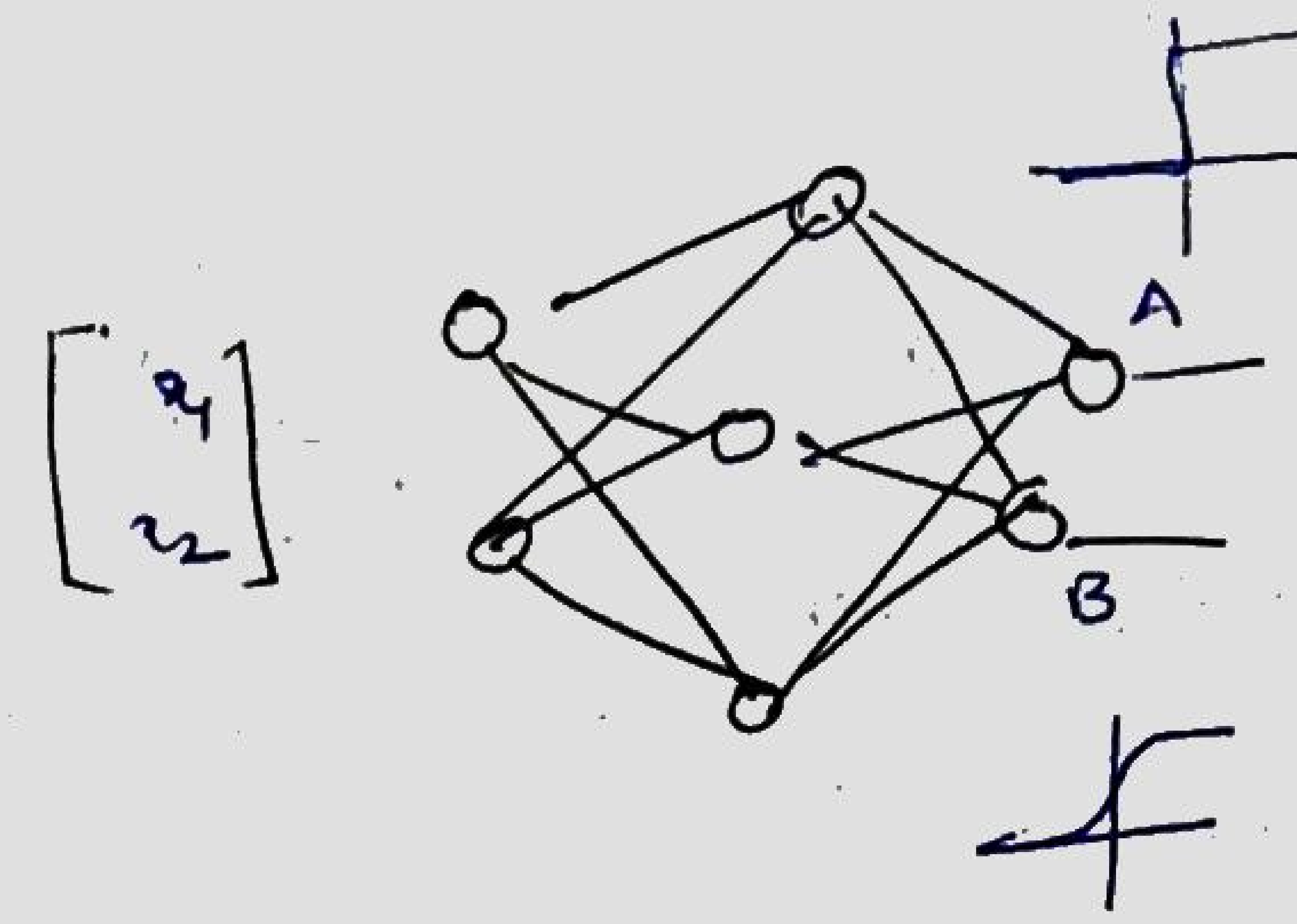
11/10/18

Lec-18

An activation.



- If sigmoid : [0, 1) → but suppose the net is supposed to predict 10, so no activation layer

- If A, B are ~~firing~~ both not firing both
  Good and Bad

→ for "wrong" sample need not fire | → ~~when~~ But sometimes, we would
•
and Both can ~~simplify~~ fire

— Softmax: can handle

[Find]

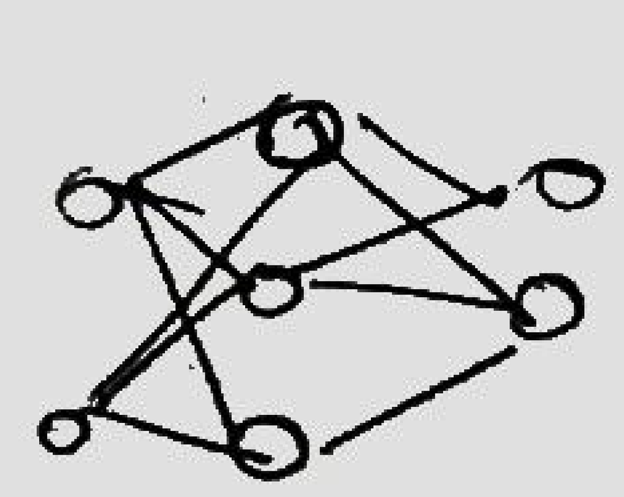
- +ve is defined, only samples u have
  -ve is not defined. Not +ve = -ve

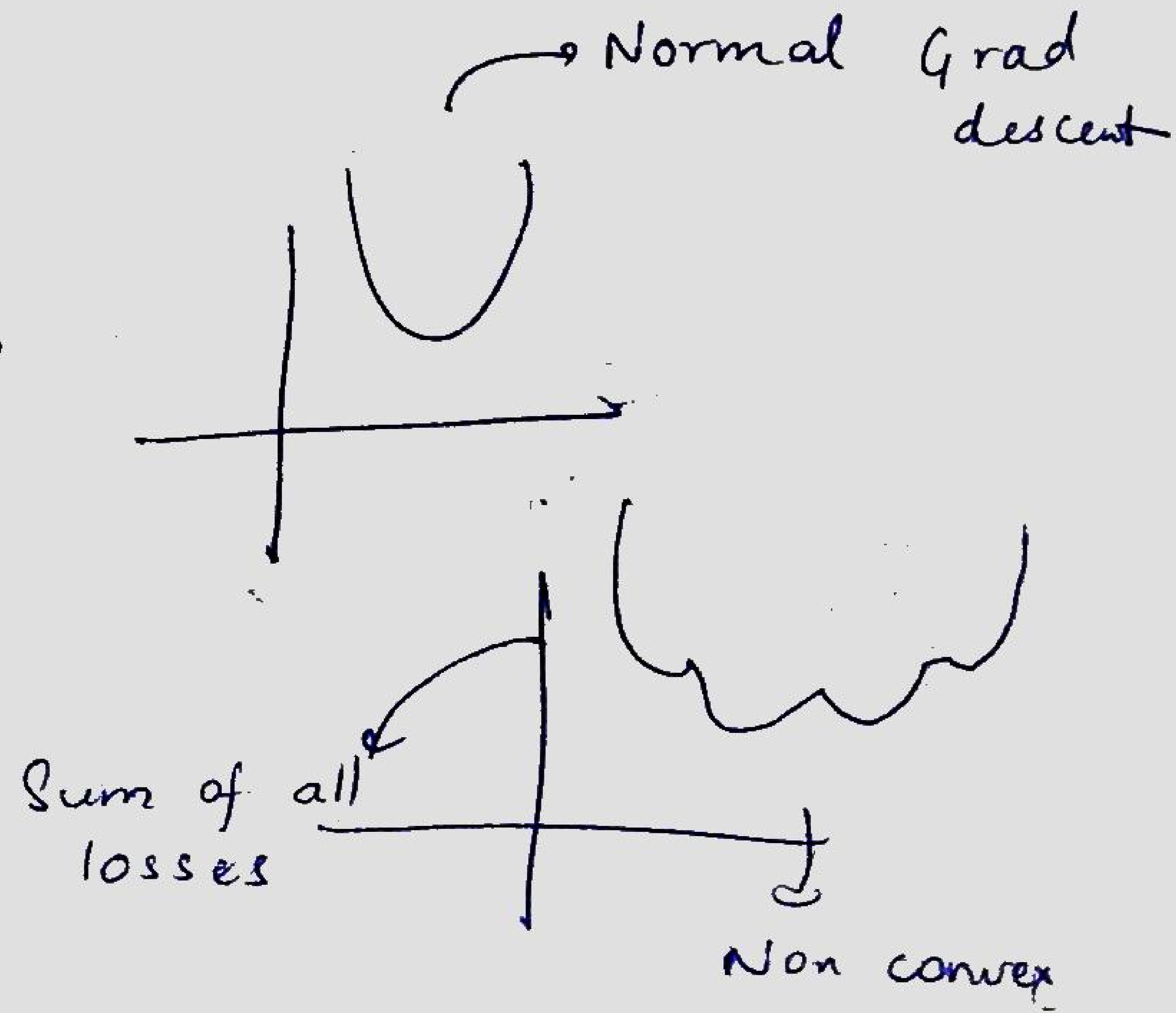
- If 2 classes v neither

— Learning w:-

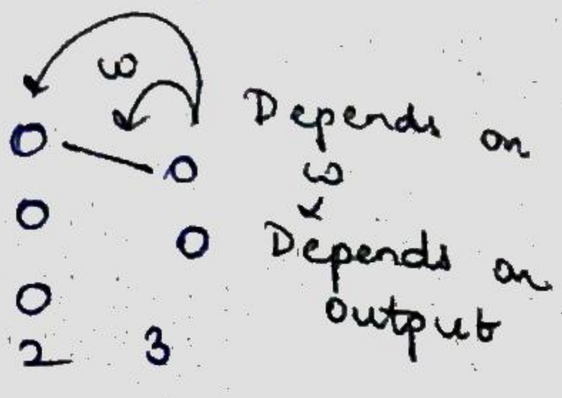
Prob (x_i, y_i) $i = 1$ to N

find optimum ω





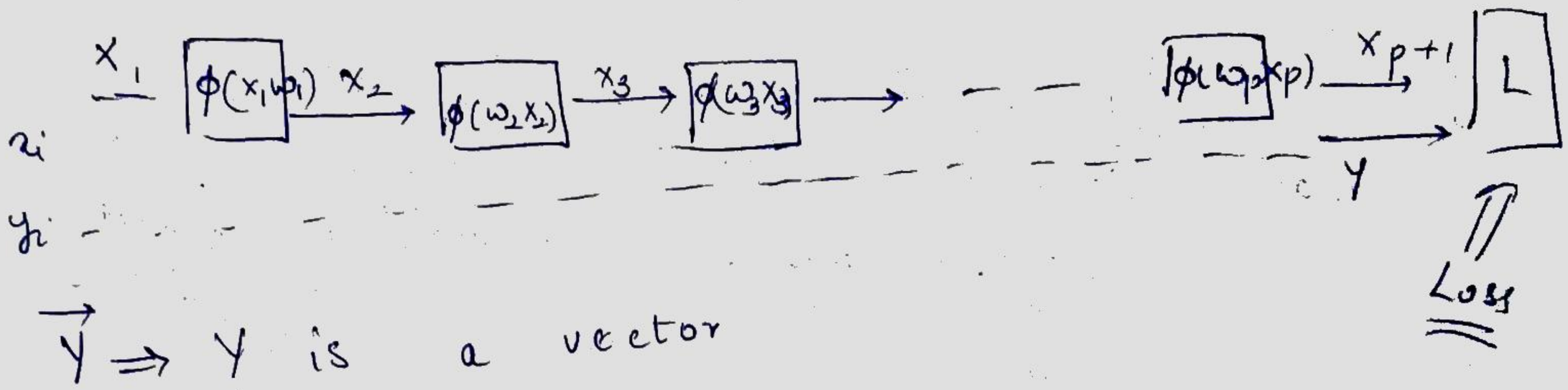
If we assume ~~only~~ last layer is ~~being~~ trained



- If we ~~assume~~ say ~~of 2nd layer~~ only 2 to 3 wts ~~fixed outputs~~ can be changed

  ⇒ we assume outputs of Lay-2 are constant ⇒ No learning

⇒ So, we need more



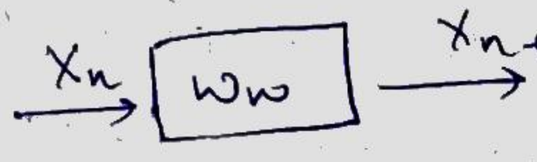
$\vec{Y}$ ⇒ Y is a vector

$$x_{p+1} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ ! \end{bmatrix}$$

Only one ele is 1

$$w^{(k+1)} = w^{(k)} - \eta \frac{\partial L}{\partial w}$$

Take one block



$$x_{n+1} = \phi(x_n w_n)$$

$$L = \sum_{i=1}^{N} [x^i_{p+1} - Y]^T [x^i_{p+1} - Y]$$

$$\frac{\partial L}{\partial w_2} = \frac{\partial L}{\partial x_2} \times \boxed{\frac{\partial x_2}{\partial w_1}} \rightarrow \text{we have it}$$

$$\frac{\partial L}{\partial x_{p+1}} \times \frac{\partial x_{p+1}}{\partial x_p} \times \frac{\partial x_p}{\partial x_{p-1}} \cdots \frac{\partial x_3}{\partial x_2}$$

$$w_p^{k+1} \leftarrow w_p - \eta \frac{\partial L}{\partial w_p}$$

$$\frac{\partial L}{\partial w_p} = \frac{\partial L}{\partial x_{p+1}} \times \frac{\partial x_{p+1}}{\partial x_p}$$

If $\frac{\partial L}{\partial w_p}$ is computed

$$\Rightarrow \frac{\partial L}{\partial w_{p-1}} = \frac{\partial L}{\partial x_{p-1}} \times \partial x$$

Initialization :- Initialize $w_1 - - w_p$ randomly

FP $\rightarrow$ Compute loss for all N samples

Update w_p

$$w_i^{K+1} \leftarrow w_i^k - \eta \frac{\partial L}{\partial w_i} \Big| \begin{matrix} i = p \\ \text{to } 1 \end{matrix} \quad \text{or} \quad \begin{matrix} i = 1 \\ \text{to } p \end{matrix}$$

So,

Generally skips are there