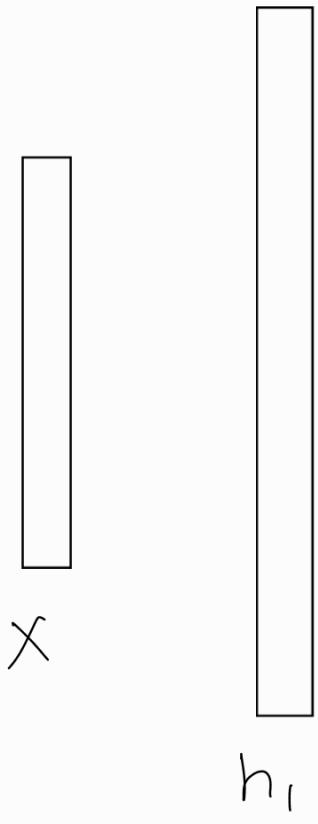
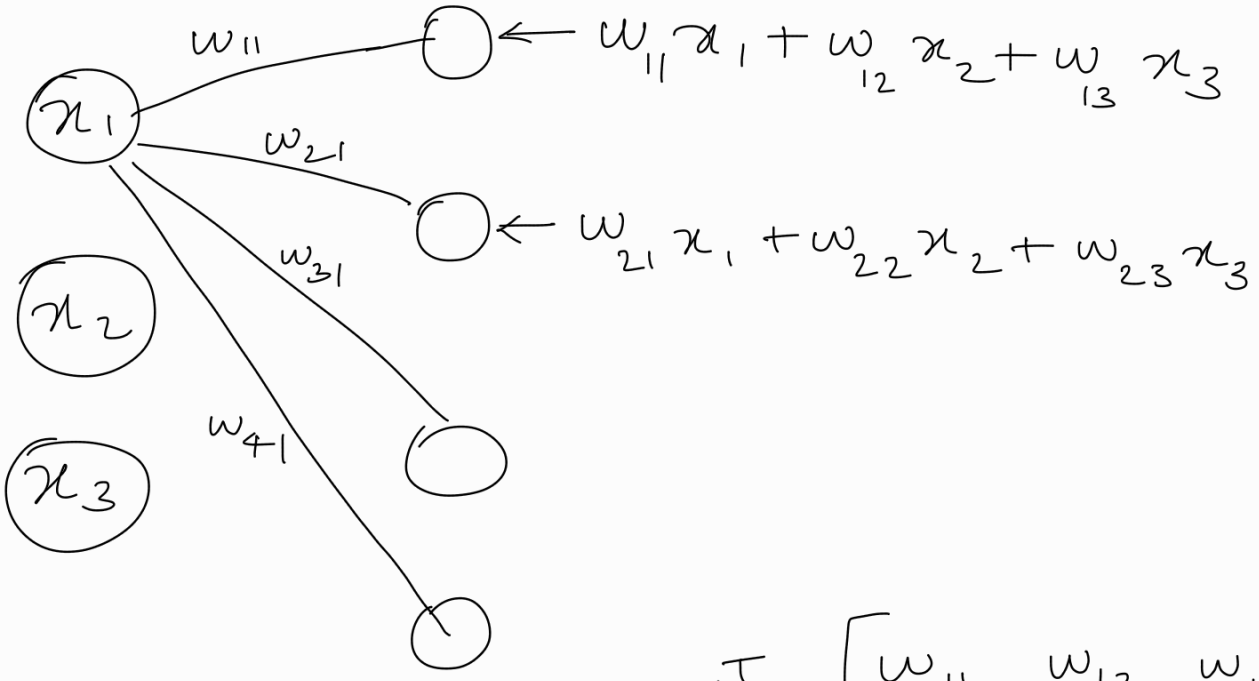
13/2/24



$$h_1 = W^T X$$

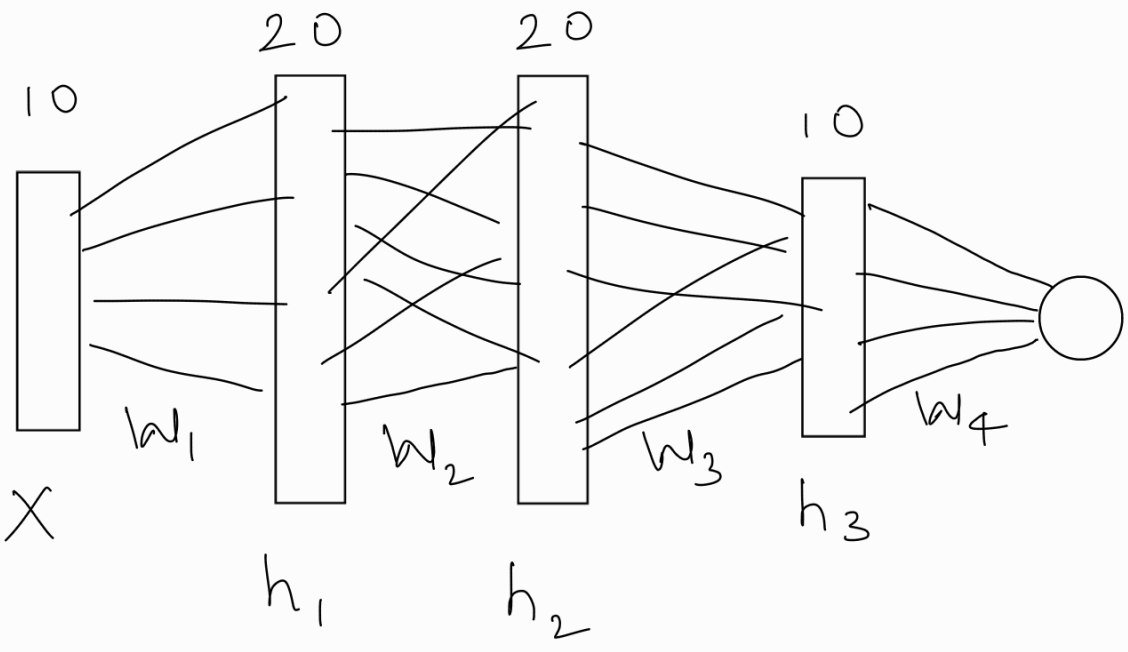


$$W^T = \begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & & \\ w_{31} & & \\ w_{41} & & w_{43} \end{bmatrix}$$

$$h_1 = \phi(W^T x)$$

ϕ is a non-linear function

$$h_n = \phi(W^T h_{n-1})$$



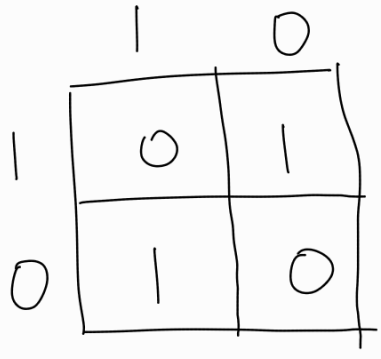
$$L = \sum (p_i - y_i)^2$$

$$\frac{\partial L}{\partial W_1} \quad \frac{\partial L}{\partial W_2} \quad \frac{\partial L}{\partial W_3} \quad \frac{\partial L}{\partial W_4}$$

Linear Classification will only act as one layer → Perceptron

This is why the non-linear function (ϕ) is important. Otherwise, all layers will act as a single layer.

Eg: Drawback of linear classification



→ You can't classify this using linear classifier

sign ($W^T X$)

↳ +ve one class

-ve other class