1.

$$\begin{aligned} y &= \mathrm{f}[x, \boldsymbol{\phi}] \\ &= \phi_0 + \phi_1 \mathrm{a}[\theta_{10} + \theta_{11}x] + \phi_2 \mathrm{a}[\theta_{20} + \theta_{21}x] + \phi_3 \mathrm{a}[\theta_{30} + \theta_{31}x]. \end{aligned} \quad (3.1)$$

If $a(z) = \psi_0 + \psi_1 z$

$$\phi_1 h_1 = \phi_1 a(\theta_{10} + \theta_{11} x)$$

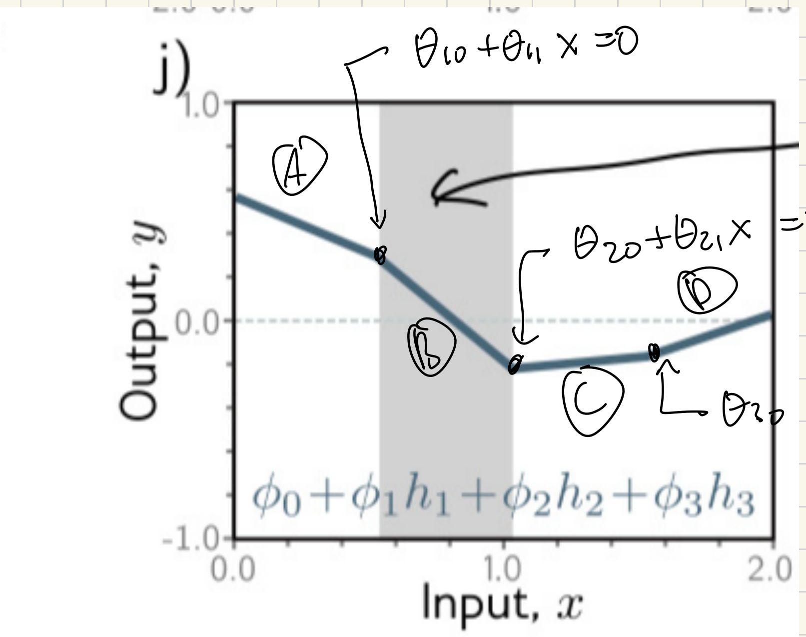
$$= \phi_1 (\psi_0 + \psi_1 \theta_{10} + \theta_{11} \psi_1 x)$$

$$= \underbrace{\phi_1 \psi_0 + \psi_1 \theta_{10} \phi_1}_{b} + \underbrace{\phi_1 \theta_{11} \psi_1}_{\theta} x = b + \theta x$$

This is just a linear function.

If $a(z) = z$ → also just linear.

3.



Gradients: A: $\theta_{31}\,\phi_3$

B: $\theta_{11}\,\phi_1 + \theta_{31}\,\phi_3$

C: $\theta_{11}\,\phi_1 + \theta_{21}\,\phi_2 + \theta_{31}\,\phi_3$

D: $\theta_{11}\,\phi_1 + \theta_{21}\,\phi_2$

5. $$\text{ReLU}(z) = \begin{cases} z & z > 0 \\ 0 & z \leq 0 \end{cases}$$

$$\text{ReLU}(\alpha z) = \begin{cases} \alpha z & z > 0 \\ 0 & z \leq 0 \end{cases}$$

$$\alpha\,\text{ReLU}(z) = \begin{cases} \alpha z & z > 0 \\ 0 & z \leq 0 \end{cases}$$

$\therefore$ $\text{ReLU}(\alpha z) = \alpha\,\text{ReLU}(z)$ for $\alpha \in \mathbb{R}^+$

7.

$$\begin{aligned} y &= \mathrm{f}[x, \boldsymbol{\phi}] \\ &= \phi_0 + \phi_1 \mathrm{a}[\theta_{10} + \theta_{11}x] + \phi_2 \mathrm{a}[\theta_{20} + \theta_{21}x] + \phi_3 \mathrm{a}[\theta_{30} + \theta_{31}x]. \end{aligned} \tag{3.1}$$

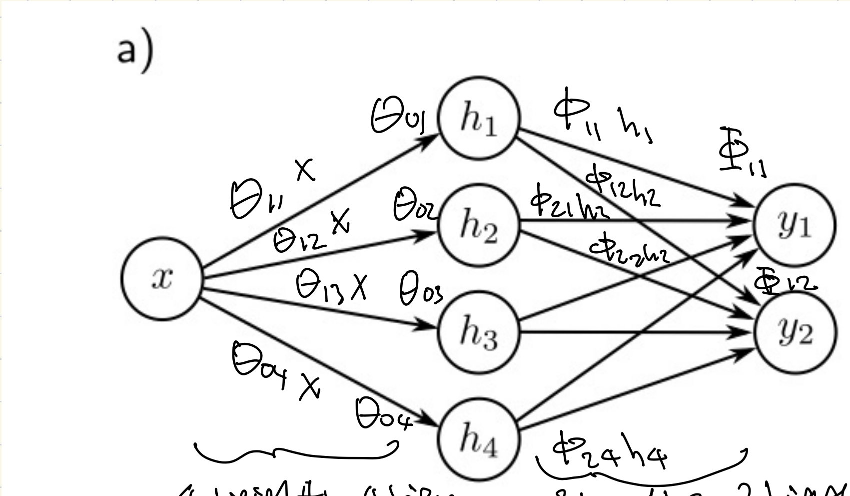
Assume $a = \text{ReLU}$. If $\theta_{10} + \theta_{11}x < 0$ and $\theta_{20} + \theta_{21}x < 0$ for example, then the loss function.

$$L(\underline{\phi}) = \sum_{i=1}^{I} (y_i - f(x_i, \underline{\phi}))^2$$

may take the same value for many different $\theta_{10}, \theta_{11}, \theta_{20}, \theta_{21}$. The non-linearity of a makes the loss function non-convex. There may be equivalent optima, hence we cannot say that it has one single global minima.

9. Shown already in Q3 (ie. slope of C is sum of A & D)

11.

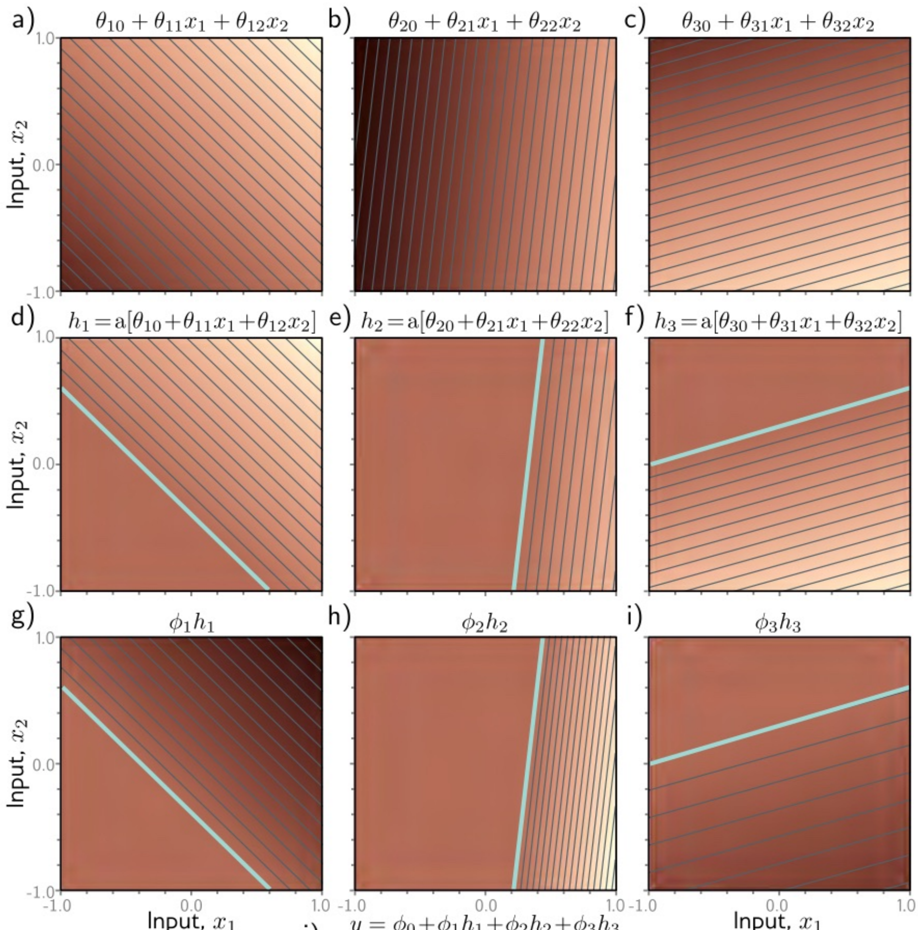


4 weights, 4 biases

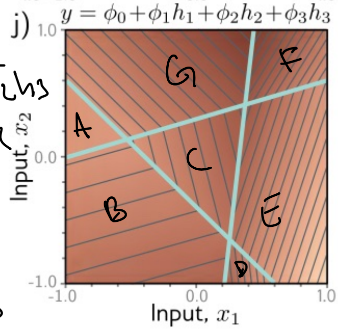
8 weights, 2 biases.

$\therefore$ 18 parameters.

13.



Region	Active	Not
A		$h_1 h_2 h_3$
B	h_3	$h_1 h_2$
C	$h_1 h_3$	h_2
D	$h_2 h_3$	h_1
E	$h_1 h_2 h_3$	
F	$h_1 h_2$	h_3
G	h_1	$h_2 h_3$



15. Input dimension is 3 (D_i)
No hyperplanes is 3 (D)

Zaslavsky (1975)

$$\sum_{j=0}^{D_i} \binom{D}{j}$$

$$= \binom{3}{0} + \binom{3}{1} + \binom{3}{2} + \binom{3}{3}$$

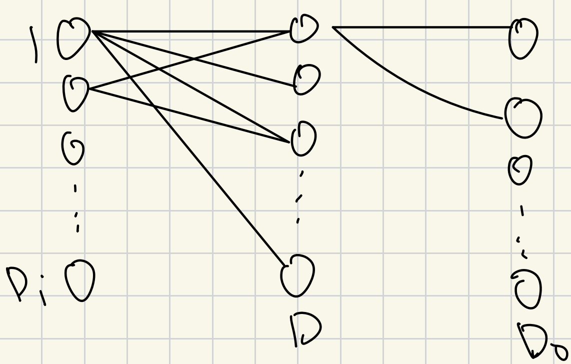
$$= 1 + 3 + 3 + 1 = 8$$

17. 3.11

$$h_d = \mathrm{a}\left[\theta_{d0} + \sum_{i=1}^{D_i} \theta_{di} x_i\right],$$

3.12

$$y_j = \phi_{j0} + \sum_{d=1}^{D} \phi_{jd} h_d,$$



$D_i D$ weights $\quad$ $D_o D$ weights

D biases $\quad$ D_o biases

Total no params:

$$D_i D + D + D_o D + D_o$$

$$D(D_i + 1) + D_o(D + 1)$$