1. $W = \begin{bmatrix} w_{00} & w_{01} & w_{02} \\ w_{10} & w_{11} & w_{12} \\ w_{20} & w_{21} & w_{22} \end{bmatrix}$ $\qquad X = \begin{bmatrix} x_{00} & x_{01} & x_{02} \\ x_{10} & x_{11} & x_{12} \\ x_{20} & x_{21} & x_{22} \end{bmatrix}$

Stride $= 4$ $\qquad$ padding $= 2$

$$\text{output size} = \frac{3-3+2(2)}{4} + 1 = 2 \Rightarrow 2\times 2 \text{ matrix}$$

$$X_{\text{padded}} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & x_{00} & x_{01} & x_{02} & 0 & 0 \\ 0 & 0 & x_{10} & x_{11} & x_{12} & 0 & 0 \\ 0 & 0 & x_{20} & x_{21} & x_{22} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$w_1 = W_{22}X_{00}$ $\qquad$ $w_2 = W_{20}X_{02}$

$w_3 = W_{02}X_{20}$ $\qquad$ $w_4 = W_{00}X_{22}$

$$WX = \begin{bmatrix} W_{22}X_{00} & W_{20}X_{02} \\ W_{02}X_{20} & W_{00}X_{22} \end{bmatrix}$$

$$Y = \begin{bmatrix} W_{22}X_{00} & W_{20}X_{02} & W_{02}X_{20} & W_{00}X_{22} \end{bmatrix}^T$$

$$Y = \underbrace{\begin{bmatrix} W_{22} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & W_{20} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & W_{02} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & W_{00} \end{bmatrix}}_{A} \begin{bmatrix} x_{00} \\ x_{01} \\ x_{02} \\ x_{10} \\ x_{11} \\ x_{12} \\ x_{20} \\ x_{21} \\ x_{22} \end{bmatrix} = Ax$$

2. $W = \begin{bmatrix} w_{00} & w_{01} \\ w_{10} & w_{11} \end{bmatrix} \qquad X = \begin{bmatrix} x_{00} & x_{01} \\ x_{10} & x_{11} \end{bmatrix}$

stride $= 2$

define $w$ as

$x_{00}$ $\qquad$ $x_{01}$

$$w = \begin{bmatrix} w_{00} & w_{01} & w_{00} & w_{01} \\ w_{10} & w_{11} & w_{10} & w_{11} \\ w_{00} & w_{01} & w_{00} & w_{01} \\ w_{10} & w_{11} & w_{10} & w_{11} \end{bmatrix} = \begin{bmatrix} x_{00}w_{00} & x_{00}w_{01} & x_{01}w_{00} & x_{01}w_{01} \\ x_{00}w_{10} & x_{00}w_{11} & x_{01}w_{10} & x_{01}w_{11} \\ x_{10}w_{00} & x_{10}w_{01} & x_{11}w_{00} & x_{11}w_{01} \\ x_{10}w_{10} & x_{10}w_{11} & x_{11}w_{10} & x_{11}w_{11} \end{bmatrix}$$

$x_{10}$ $\qquad$ $x_{11}$

row-wise $w = [\, x_{00}w_{00} \;\; x_{00}w_{01} \;\; x_{01}w_{00} \;\; x_{01}w_{01} \;\; x_{00}w_{10} \;\; x_{00}w_{11} \;\; x_{01}w_{10} \;\; x_{01}w_{11}$
$\quad \cdots \; x_{11}w_{11} \,]$

$$Y = \underbrace{\begin{bmatrix} w_{00} & 0 & 0 & 0 \\ w_{01} & 0 & 0 & 0 \\ 0 & w_{00} & 0 & 0 \\ 0 & w_{01} & 0 & 0 \\ w_{10} & 0 & 0 & 0 \\ w_{11} & 0 & 0 & 0 \\ 0 & w_{10} & 0 & 0 \\ 0 & w_{11} & 0 & 0 \\ 0 & 0 & w_{00} & 0 \\ 0 & 0 & w_{01} & 0 \\ 0 & 0 & 0 & w_{00} \\ 0 & 0 & 0 & w_{01} \\ 0 & 0 & w_{10} & 0 \\ 0 & 0 & w_{11} & 0 \\ 0 & 0 & 0 & w_{10} \\ 0 & 0 & 0 & w_{11} \end{bmatrix}}_{A} \begin{bmatrix} x_{00} \\ x_{01} \\ x_{10} \\ x_{11} \end{bmatrix} = Ax$$

3. Kerenel $(o \times r^2, i, k, k) \equiv$ Kernel $(o, i, k \times r, k \times r)$

$o = 1 \quad r = 2 \quad k = 1$

$\Rightarrow (4,1,1,1) \equiv (1,1,2,2)$

For $w$ of size $(4,1,1,1)$

$$W \Rightarrow \begin{matrix} W_1 = [w_1] \\ W_2 = [w_2] \\ W_3 = [w_3] \\ W_4 = [w_4] \end{matrix} \qquad X = \begin{bmatrix} X_{00} & X_{01} \\ X_{10} & X_{11} \end{bmatrix}$$

$$Y = \begin{bmatrix} W_1 & 0 & 0 & 0 \\ 0 & W_1 & 0 & 0 \\ 0 & 0 & W_1 & 0 \\ 0 & 0 & 0 & W_1 \\ W_2 & 0 & 0 & 0 \\ 0 & W_2 & 0 & 0 \\ 0 & 0 & W_2 & 0 \\ 0 & 0 & 0 & W_2 \\ W_3 & 0 & 0 & 0 \\ 0 & W_3 & 0 & 0 \\ 0 & 0 & W_3 & 0 \\ 0 & 0 & 0 & W_3 \\ W_4 & 0 & 0 & 0 \\ 0 & W_4 & 0 & 0 \\ 0 & 0 & W_4 & 0 \\ 0 & 0 & 0 & W_4 \end{bmatrix} \begin{bmatrix} X_{00} \\ X_{01} \\ X_{10} \\ X_{11} \end{bmatrix} = \begin{bmatrix} W_1 X_{00} \\ W_1 X_{01} \\ W_1 X_{10} \\ W_1 X_{11} \\ W_2 X_{00} \\ W_2 X_{01} \\ W_2 X_{10} \\ W_2 X_{11} \\ W_3 X_{00} \\ W_3 X_{01} \\ W_3 X_{10} \\ W_3 X_{11} \\ W_4 X_{00} \\ W_4 X_{01} \\ W_4 X_{10} \\ W_4 X_{11} \end{bmatrix}$$

For $w$ of size $(1,1,2,2)$

$$w = \begin{bmatrix} w_1 & w_2 \\ w_3 & w_4 \end{bmatrix} \qquad X = \begin{bmatrix} X_{00} & X_{01} \\ X_{10} & X_{11} \end{bmatrix}$$

From $q^2$ we can write:

$$Y = \begin{bmatrix} W_1 & 0 & 0 & 0 \\ W_2 & 0 & 0 & 0 \\ 0 & W_1 & 0 & 0 \\ 0 & W_2 & 0 & 0 \\ W_3 & 0 & 0 & 0 \\ W_4 & 0 & 0 & 0 \\ 0 & W_3 & 0 & 0 \\ 0 & W_4 & 0 & 0 \\ 0 & 0 & W_1 & 0 \\ 0 & 0 & W_2 & 0 \\ 0 & 0 & 0 & W_1 \\ 0 & 0 & 0 & W_2 \\ 0 & 0 & W_3 & 0 \\ 0 & 0 & W_4 & 0 \\ 0 & 0 & 0 & W_3 \\ 0 & 0 & 0 & W_4 \end{bmatrix} \begin{bmatrix} X_{00} \\ X_{01} \\ X_{10} \\ X_{11} \end{bmatrix} = \begin{bmatrix} W_1 X_{00} \\ W_2 X_{00} \\ W_1 X_{01} \\ W_2 X_{01} \\ W_3 X_{00} \\ W_4 X_{00} \\ W_3 X_{01} \\ W_4 X_{01} \\ W_1 X_{10} \\ W_2 X_{10} \\ W_1 X_{11} \\ W_2 X_{11} \\ W_3 X_{10} \\ W_4 X_{10} \\ W_3 X_{11} \\ W_4 X_{11} \end{bmatrix}$$

We can see that both Y's have same elements, just in different order

4. $f(x) = \begin{cases} 1 & \text{if } w^T x + b \geq 0 \\ 0 & \text{if } w^T x + b < 0 \end{cases}$

| $x_1$ | $x_2$ | $f_{AND}(x)$ |
|---|---|---|
| 0 | 0 | 0 |
| 0 | 1 | 0 |
| 1 | 0 | 0 |
| 1 | 1 | 1 |

$$\boxed{W_{AND} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad b_{AND} = -2}$$

$$f\left(W_{AND}^T \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + b_{AND}\right) = f_{AND}(x)$$

$x_1 = 0$, $x_2 = 1$ $\Rightarrow [1 \;\; 1] \begin{bmatrix} 0 \\ 1 \end{bmatrix} - 2 = -1 \Rightarrow 0$

$x_1 = 1$, $x_2 = 0$ $\Rightarrow [1 \;\; 1] \begin{bmatrix} 1 \\ 0 \end{bmatrix} - 2 = -1 \Rightarrow 0$

$x_1 = 1$, $x_2 = 1$ $\Rightarrow [1 \;\; 1] \begin{bmatrix} 1 \\ 1 \end{bmatrix} - 2 = 0 \Rightarrow 1$

$x_1 = 0$, $x_2 = 0$ $\Rightarrow [1 \;\; 1] \begin{bmatrix} 0 \\ 0 \end{bmatrix} - 2 = -2 \Rightarrow 0$

$$\boxed{W_{OR} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad b_{OR} = -0.1}$$

$x_1 = 0$, $x_2 = 1$ $\Rightarrow [1 \;\; 1] \begin{bmatrix} 0 \\ 1 \end{bmatrix} - 0.1 = 0.9 \Rightarrow 1$

$x_1 = 1$, $x_2 = 0$ $\Rightarrow [1 \;\; 1] \begin{bmatrix} 1 \\ 0 \end{bmatrix} - 0.1 = 0.9 \Rightarrow 1$

$x_1 = 1$, $x_2 = 1$ $\Rightarrow [1 \;\; 1] \begin{bmatrix} 1 \\ 1 \end{bmatrix} - 0.1 = 1.9 \Rightarrow 1$

$x_1 = 0$, $x_2 = 0$ $\Rightarrow [1 \;\; 1] \begin{bmatrix} 0 \\ 0 \end{bmatrix} - 0.1 = -0.1 \Rightarrow 0$

5.

| $x_1$ | $x_2$ | $f_{XOR}(x)$ |
|---|---|---|
| 0 | 0 | 0 |
| 0 | 1 | 1 |
| 1 | 0 | 1 |
| 1 | 1 | 0 |

proof by contradiction:

If $f(x) = w^T x + b$ can output a linear model function

$\downarrow$ Let $x_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ $x_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

Then $f(x_1) < f(x_2)$

$$f\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) < f\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right)$$

$$w \begin{bmatrix} 1 \\ 1 \end{bmatrix} + b < w \begin{bmatrix} 0 \\ 1 \end{bmatrix} + b$$

Let $w = [w_1 \ w_2]$

$$[w_1 \ w_2] \begin{bmatrix} 1 \\ 1 \end{bmatrix} + b < [w_1 \ w_2] \begin{bmatrix} 0 \\ 1 \end{bmatrix} + b$$

$$w_1 + w_2 + b < w_2 + b$$

$$w_1 < 0$$

Now let's look at for $x_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ $x_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$:

$$f(x_1) < f(x_2)$$

$$w \begin{bmatrix} 0 \\ 0 \end{bmatrix} + b < w \begin{bmatrix} 1 \\ 0 \end{bmatrix} + b$$

$$b < w_1 + b$$

$$0 < w_1$$

$\Rightarrow$ we have a contradiction, $\therefore$ XOR can not be represented using Linear Model ∎

6. $h(x) = W^{(3)} \max\{0, W^{(2)}\{0, W^{(1)}x + b^{(1)}\} + b^{(2)}\} + b^{(3)}$

$W^{(1)} = \begin{bmatrix} 1.5 \\ 0.5 \end{bmatrix} \quad W^{(2)} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \quad W^{(3)} = [1 \ \ 1]$

$b^{(1)} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad b^{(2)} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad b^{(3)} = -1$

$\underline{X_0 = 2}$

$[1 \ \ 1] \max\left\{0, \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \max\left\{0, \begin{bmatrix} 1.5 \\ 0.5 \end{bmatrix}[2] + \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right\} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right\} - 1$

$= [1 \ \ 1] \max\left\{0, \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 3 \\ 2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right\} - 1$

$= [1 \ \ 1] \max\left\{0, \begin{bmatrix} 7 \\ 8 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right\} - 1$

$= [1 \ \ 1]\begin{bmatrix} 7 \\ 9 \end{bmatrix} - 1 = 7 + 9 - 1 = 15$

$\left.\begin{matrix} x = 2 \\ w = 7 \\ b = 1 \end{matrix}\right\} \Rightarrow y = 7x + 1 = 7(2) + 1 = 15$

$\frac{dh}{dx} = w = 7$

$\underline{X_0 = -1}$

$[1 \ 1] \max\left\{0, \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \max\left\{0, \begin{bmatrix} 1.5 \\ 0.5 \end{bmatrix}[-1] + \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right\} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right\} - 1$

$= [1 \ 1] \max\left\{0, \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \max\left\{0, \begin{bmatrix} -1.5 \\ 0.5 \end{bmatrix}\right\} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right\} - 1$

$= [1 \ 1] \max\left\{0, \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 0 \\ 0.5 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right\} - 1$

$= [1 \ 1] \max\left\{0, \begin{bmatrix} 1 \\ 0.5 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right\} - 1$

$= [1 \ 1]\begin{bmatrix} 1 \\ 1.5 \end{bmatrix} - 1 = 1 + 1.5 - 1 = 1.5$

$\left.\begin{matrix} x = -1 \\ w = -1 \\ b = 0.5 \end{matrix}\right\} \ y = -1x + 0.5 = (-1)(-1) + 0.5 = 1.5$

$\frac{dh}{dx} = w = -1$

$\underline{X_0 = 1}$

$$[1 \quad 1] \max\left\{0, \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \max\left\{0, \begin{bmatrix} 1.5 \\ 0.5 \end{bmatrix} [1] + \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right\} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right\} - 1$$

$$= [1 \quad 1] \max\left\{0, \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1.5 \\ 1.5 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right\} - 1$$

$$= [1 \quad 1] \max\left\{0, \begin{bmatrix} 4.5 \\ 4.5 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right\} - 1$$

$$= [1 \quad 1] \begin{bmatrix} 4.5 \\ 5.5 \end{bmatrix} - 1 = 4.5 + 5.5 - 1 = 9$$

$$\left.\begin{matrix} x = 1 \\ w = 8 \\ b = 1 \end{matrix}\right\} \; y = 8x + 1 = 8(1) + 1 = 9$$

$$\frac{dh}{dx} = w = 8$$

7. Given : $f(x) = |x|$ divides input space into 2 regions for value $y$

For each linear region, the region of the input is a bijection

$$W_{ij}^{(1)} = \begin{cases} 2 & \text{if } i=j \\ 0 & \text{if } i \neq j \end{cases} \leftarrow W \text{ is a diagonal matrix}$$

$$b_i^{(1)} = -1$$

$$f_1(x) = |W^{(1)} x + b|$$

Proof : Take hypercube and partition it into halves with hyperplanes in each dimension then for each region $R_i$, there is a bijection

$$x_i = \begin{cases} \frac{-y_i+1}{2} \\ \frac{+y_i+1}{2} \end{cases} \text{if } y_i = 1, \quad x_i = \begin{Bmatrix} -1 \\ +1 \end{Bmatrix}$$

$$f_i(x_i) = |W^{(1)} x + b^{(1)}| = \sum_j W_{ij} x_i + b_i$$

$$= |2x_i + b_i|$$

$$= \left|2\left(\frac{\pm y_i + 1}{2}\right) - 1\right|$$

$$= |\pm y_i + 1 - 1| = y_i$$

assume $x_1 \neq x_2$

$\Rightarrow 2x_1 \neq 2x_2$

$\Rightarrow |f(x_1)| \neq |f(x_2)|$

$\Rightarrow$ we have a bijection and $|R| = |(0,1)^d| = 2^d$ possible choices of input regions $R_i$

Alternatively, we can see that for $d=1$

$O = (0,1)^1$

input space $R = \{(-1,0), (0,1)\} \Rightarrow 2$ input regions

then it is intuitive to see that

$O = (0,1)^d \Rightarrow 2^d$ input regions ∎

8. $g$ has $n_g$ regions of $(0,1)^d$

$f$ has $n_f$ regions of $(0,1)^d$

$f \circ g(\cdot) = f(g(\cdot))$

Let $G$ be input region of $g()$

Let $F$ be input region of $f()$

① Each subregion in $G$ bijects onto $f$ because $g(G)$ bijects onto $(0,1)^d$

② $f$ bijects between $F$ and $O$

① + ② $\Rightarrow$ $f \circ g$ bijects between $G$ and $O$

$g(G) \Rightarrow |n_g|$

$f(g(G)) \Rightarrow |n_f \, n_g|$ regions

9. $h_1 = |W_1 x + b_1|$

$h_2 = |W_2 h_1 + b_2|$

$\vdots$

$h_L = |W_L h_{L-1} + b_L|$

$x \in (0,1)^d \qquad f(x) = h_L$

Each $h_i$ is implicitly a function of $x$:

$h_1(x) = |W_1 x + b_1|$

$h_2(x) = |W_2(W_1 x + b_1) + b_2|$

$\vdots$

$h_L \circ h_{L-1} \circ \cdots \circ h_1(x) = |W_L h_L \cdots h_1(x) + b_L|$

Proof by induction:

Base case: $L = 1$ we know from q4 that $h_1(x)$ identifies $2^d$ input region space

$2^{Ld} = 2^{1d} = 2^d$

Inductive Step:

Give true for $L = \ell - 1 \Rightarrow 2^{(\ell-1)d}$ input region space, prove true for $L + 1 = \ell$

Let $g(x) = h_{\ell-1} h_{\ell-2} \cdots h_1(x) + b_{\ell-1}$ identify $2^{(\ell-1)d}$ input region space

Let $f(x) = h_\ell(g(x))$

By q8 we know $f \circ g(\cdot) = |n_f n_g|$ input region space

$h_\ell$ identifies $2^d$ input region space

$\Rightarrow f(x) = h_\ell(g(x)) = h_\ell \circ g \Rightarrow |n_{h_\ell} n_g| = |2^d \cdot 2^{(\ell-1)d}| = 2^{\ell d}$ input region space

∎

10. $$\min_{w \in \mathbb{R}^d} f(w) = \sum_{i=1}^{n} (y_i - w^T x_i)^2 \quad ①$$

$$w^{(t+1)} = w^{(t)} - \eta \nabla f(w^{(t)})$$

$$\vdots$$

$$w^{gd} = w^{(gd-1)} - \eta \nabla f(w^{(gd-1)})$$

$$|||$$

$$w^{gd} = \underset{w \in \mathbb{R}^d}{\operatorname{argmin}} \|w\|_2^2 \quad \text{s.t. } Xw = y \quad ②$$

$$\|w\|_2^2 = \sum_i^d w_i^2$$

For ②:

Introduce Lagrange Multipliers

$$L(x, \lambda) = \|w\|_2^2 + x^T (wX - y)$$

$$\text{s.t. } \nabla_x L = 2X + w^T \lambda = 0$$

$$\nabla_y L = wX - y = 0$$

$$\Rightarrow \lambda = -2(ww^T)^{-1} y$$

$$\Rightarrow X = w^T (ww^T)^{-1} y \quad ③$$

For ①: $$w^{gd} = (wX - y)^2 + 2(y - wX)$$

$$X = (w^T w + I)^{-1} w^T y$$

$$= w^T (ww^T)^{-1} y \quad ④$$

③ = ④ $\Rightarrow$ solutions found by both are same

11. (25) means sol'n will have smallest norm of any sol'n

(24) is the optimization problem

$\Rightarrow$ gradient descent will find the optimal sol'n for the loss function

12. Least squares regression can always find global min of linear system if it has a soln.