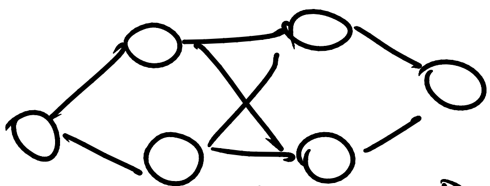
sábado, 26 de abril de 2025 02:49

4.6) $D_i = 1$, $D_o = 1$, $K = 10$ LAYERS, $D = 10$ hidden units



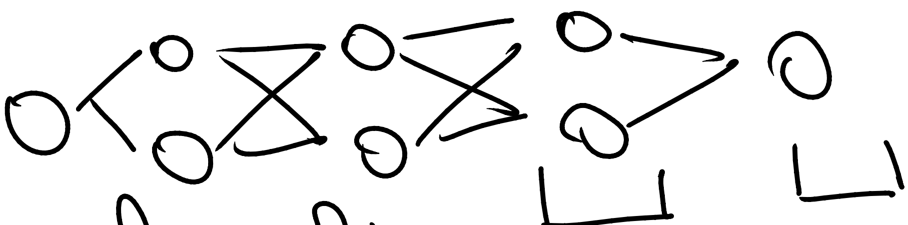
$D_i = 1$ $D_o = 1$

$D_1 = 2$ $D_2 = 2$

$\Omega_0 \in R^{D_1 \times D_i}$ $\Omega_1 \in R^{D_2 \times D_1}$ $\Omega_2 \in R^{D_o \times D_2}$

$B_0 \in D^1$ $B_1 \in D^2$ $B_2 \in R^{D_o}$.

• AUMENTO EL WIDTH.



Ω_0 Ω_1 NUEVA $\Omega_3 \in R^{D_o \times D_3}$

B_0 B_1 Ω_2 $B_3 \in R^{D_o}$.

B_2

Cambian las dimensiones!

$\Omega_2 \in R^{D_3 \times D_2}$

$B_2 \in D^3$

• Se suman

D_K params (corresponden a los B de la NUEVA CAPA K.)

y

$D_K \times (D_{K-1})$ (matriz Ω_K nueva)

$D_K \times (D_{K-1})$ (matriz nueva Ω_{K-1})

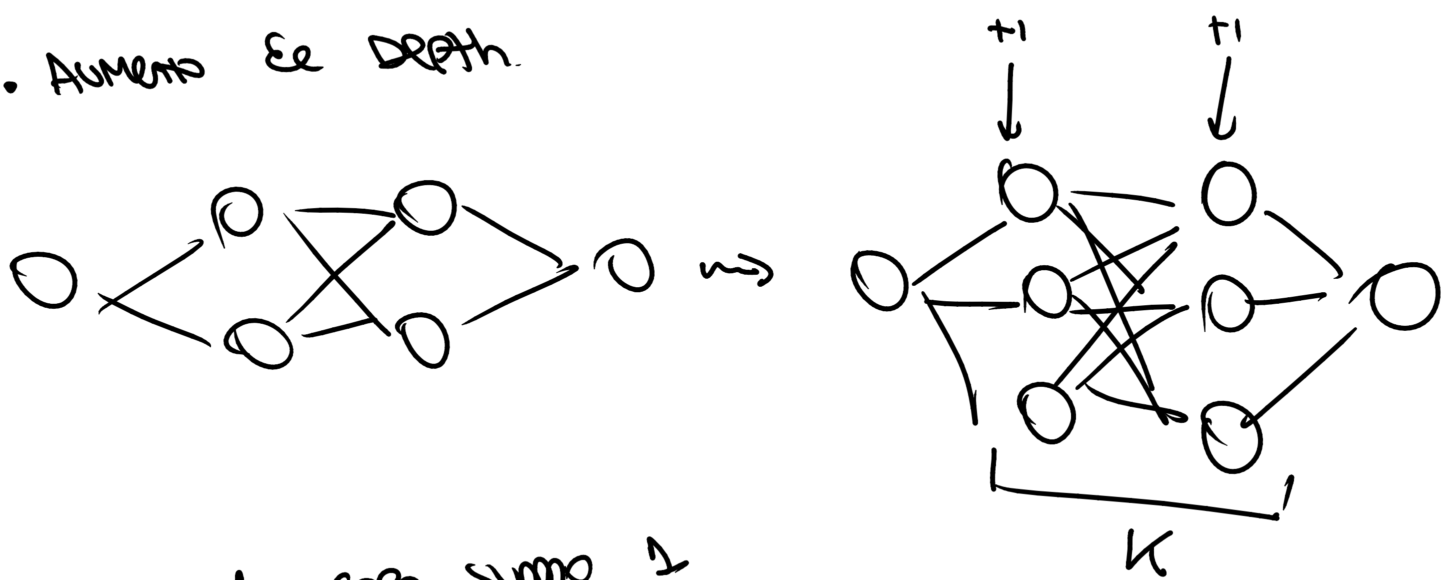
$D_K + D_K D_{K-1} = D_K (1 + D_{K-1})$

$= D(1+D) = D + D^2$

$\downarrow$ (H)

$\boxed{= 10 + 100 = 110}$

- Aumento de depth.



* por cada capa sumo 1 neurona → 1 bias ⇝ K Biases nuevos

* En el input tengo 1 peso. (D) ⇝ 1

* por cada hidden layer los D neuronas viejas tienen un peso más. (solo ult capa)

$D(K-1)$

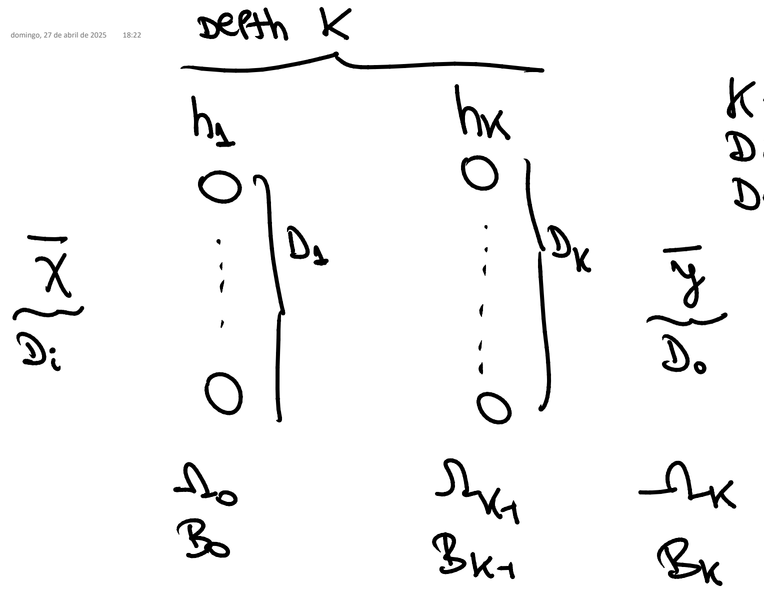
* En ult capa tengo 1 conexión más

$$K + 1 + D(K-1) + 1 =$$

$$K + D(K-1) + 2.$$

$$= 10 + 10(9) + 2 =$$

$$\boxed{10 + 90 + 2 = 102}$$



K = # HIDDEN LAYERS

D_i = Input Dim

D_o = Output Dimension

$D_n, n \in \{1, \dots, K\}$ = WIDTH OF n^{th} hidden Layer.

$\Omega_{i \in \{0, \dots, K\}} \in \mathbb{R}^{D_{i+1} \times D_i}$

$B_i \in \mathbb{R}^{D_{i+1}}$

$\left.\begin{matrix} x \in \mathbb{R}^{D_i \times 1} \\ y \in \mathbb{R}^{D_o \times 1} \end{matrix}\right\}$ Column vector

$\phi = \{B_k, \Omega_k\}_{k=0}^{K}$

↳ ALL the NN weights.

¿ $|\phi|$? how many?

$\Omega_0 + B_0 + \Omega_1 + B_1$
$+ \dots + \Omega_{K-1} + B_{K-1}$
$+ \Omega_K + B_K =$

$$\sum_{k=0}^{K} |\Omega_k| + |B_k| = \overbrace{|\Omega_0| + |B_0|}^{\text{IN}} + \sum_{k=1}^{K-1} |\Omega_k| + |B_k| + \underbrace{|\Omega_K| + |B_K|}_{\text{OUT}}.$$

$$= \underbrace{(D_1 \times D_i) + (D_1)}_{\text{IN}} + \underbrace{\sum_{k=1}^{K-1} |\Omega_k| + |B_k|}_{\text{HIDDEN UNITS}} + \underbrace{(D_o \times D_K) + (D_o)}_{\text{OUT}}$$

$D_1 (D_i + 1)$

$$\sum_{k=1}^{K-1} (D_{k+1} \times D_k) + (D_{k+1})$$

$$\sum_{k=1}^{K-1} D_{k+1} (D_k + 1)$$

$D_o (D_K + 1)$

¿ $|\phi| =$

$K-1$

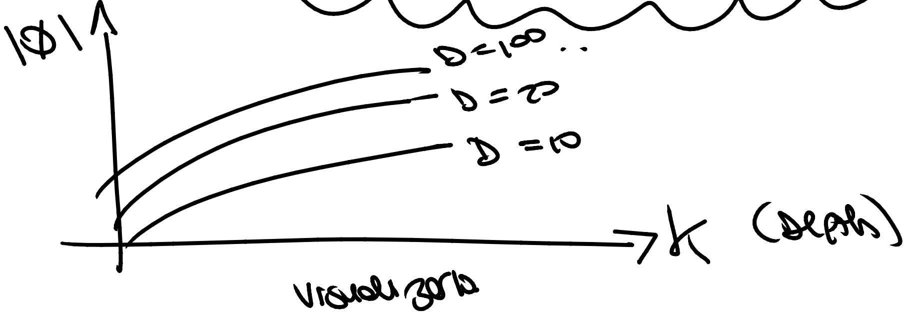
$$\sum_{K=1}$$

So… $|\phi| =$

$$D_1(D_i+1) + D_o(D_K+1) + \sum_{K=1}^{K-1} D_{K+1}(D_K+1)$$

Podría fijar distintos D = # hidden

$D_i, D_o = 1$



Computando el Resultado:

$$h_1 = a(\beta_0 + \Omega_0 x).$$

$$h_2 = a(\beta_1 + \Omega_1 h_1)$$

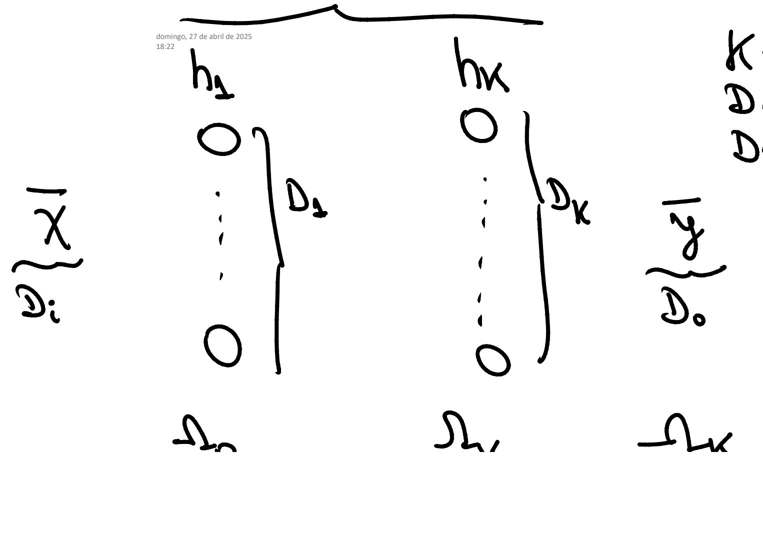
$$\vdots$$

$$h_K = a(\beta_{K-1} + \Omega_{K-1} h_{K-1})$$

$$\boxed{y = \underbrace{\beta_K}_{\text{Beta}} + \underbrace{\Omega_K}_{\text{Omega.}} h_K\,.}$$

Forward pass / inference

domingo, 27 de abril de 2025 18:22



K = # Hidden layers

D_i = Input Dim

D_o = Output Dimension

$D_n, n \in \{1, \dots, K\}$ = Width of n^{th} hidden layer.

$\Omega_{i \in \{0, \dots, K\}} \in \mathbb{R}^{D_{i+1} \times D_i}$

$\beta_i \in \mathbb{R}^{D_{i+1}}$

$\Omega_0 \qquad \Omega_{K-1} \qquad \Omega_K$

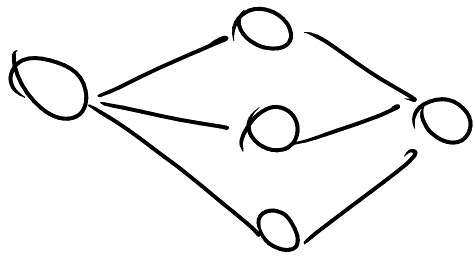
$\beta_0 \qquad \beta_{K-1} \qquad \beta_K$

$\beta_i \in \mathbb{R}^{D_{i+1}}$

$x \in \mathbb{R}^{D_i \times 1}$

$y \in \mathbb{R}^{D_o \times 1}$) Column vector

Ex FOR 1 - 1 - 3 NN



$$X = \begin{pmatrix} | & | & & | \\ x^{(1)} & x^{(2)} & \cdots & x^{(n\_data)} \\ | & | & & | \end{pmatrix}$$

$$\Omega_0 X = \begin{pmatrix} \phi_{11} & \phi_{12} & \cdots \\ \phi_{21} & \phi_{22} & \\ & \cdots & \end{pmatrix} \begin{pmatrix} | & | & \\ x^{(1)} & x^{(2)} & \cdots \\ | & | & \end{pmatrix}$$

$$\beta_1 = \begin{pmatrix} \phi_{10} \\ \phi_{20} \end{pmatrix}$$