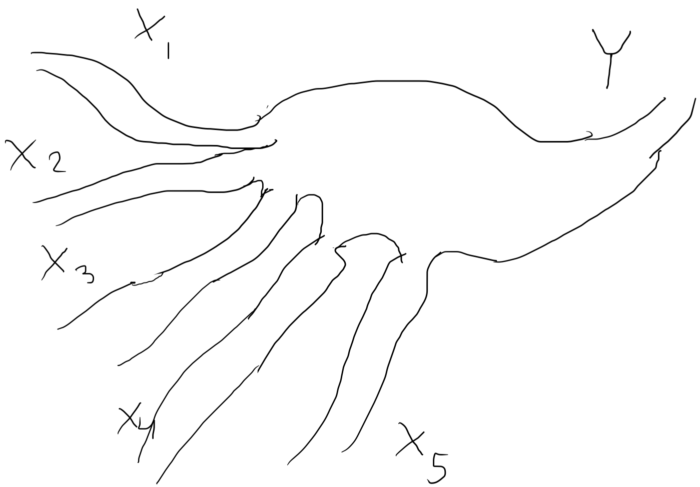
Antecedentes: 1943 McCulloch-Pitts

Se considera la neurona



$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_5 \end{pmatrix} = \text{Señal de entrada}$$

y = señal de salida

la señal X se procesa en la neurona.

El procesamiento se describe como una asignación de pesos a cada señal en términos de una transformación lineal

$$X \longrightarrow \sum w_i x_i \in \mathbb{R}$$

Con $w_i \in \mathbb{R}$, esos se nombran pesos

La señal puede activar o no a la neurona

Este fenómeno se describe de la sig. forma:

Suponemos que existe un umbral de activación que denotamos con $\theta \in \mathbb{R}$

Entonces, si

$$\sum w_i x_i \geq \theta,$$

la neurona se activa

si $\sum w_i x_i < \theta$, la neurona no se activa.

De manera matemática se modela con la función característica o indicadora:

Dado un conjunto A, definimos

$$\chi_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \notin A. \end{cases} \quad \left(\mathbb{1}_A(x), \text{ en proba.} \right)$$

Denotamos por $\mathbb{R}^+ = [0, \infty)$

Entonces el proceso de activación se describe de la sig manera:

$$\chi_{\mathbb{R}^+}\left(\sum w_i x_i - \theta \right),$$

de forma que si x es tal que

$\chi_{\mathbb{R}}\left(\sum w_i x_i - \theta \right) = 1$, entonces la neurona se activa y si

$\chi_{\mathbb{R}}\left(\sum w_i x_i - \theta \right) = 0$, entonces la neurona no se activa.

La señal de salida y describe el proceso de activación

$$y(x) = \chi_{\mathbb{R}^+}\left(\sum w_i x_i - \theta\right)$$

y la interpretación es

$$y(x) = \begin{cases} 1, & \text{la neurona se activa} \\ 0, & \text{la neurona no se activa} \end{cases}$$

En el caso anterior se tienen los sig. objetos matemáticos

$x \in \mathbb{R}^5$ señal de entrada

$T : \mathbb{R}^5 \longrightarrow \mathbb{R}$, $T(x) = \sum w_i x_i - \theta$, procesamiento de la señal y umbral

$\chi_{\mathbb{R}^+}$ función de activación

$y(x) = \chi_{\mathbb{R}^+}(T(x))$ señal de salida

De forma general, se considera el sig modelo:

1. $T: \mathbb{R}^m \longrightarrow \mathbb{R}^n$ una transformación afín.

Recordar que una transformación afin es una transformación lineal compuesta con una traslación:

$T(x) = Ax + \theta$

con A una trasformación lineal y $\theta \in \mathbb{R}^n$. A se puede representar como una matriz. Normalmente se identifica a A con la matriz que la representa en las bases canónicas de $\mathbb{R}^m$ y $\mathbb{R}^n$ y las componentes de A, Ae_j, se nombran pesos.

2. $\rho : \mathbb{R} \to \mathbb{R}$ es una función de activación.
(ρ es cualquier función)

Usando ρ, se pueden definir funciones que denotamos temporalmente por

$$\rho^n : \mathbb{R}^n \longrightarrow \mathbb{R}^n,$$

definida por

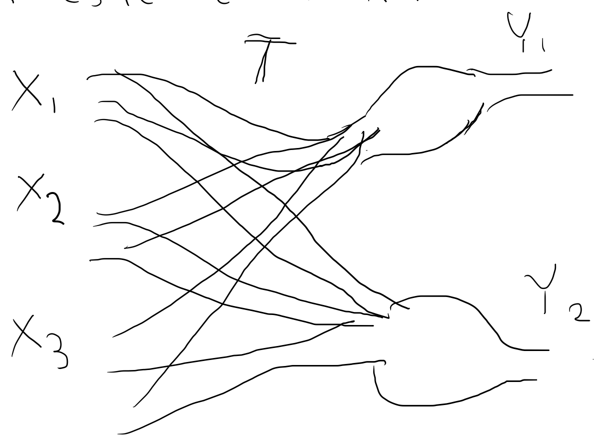
$$\rho^n \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix} := \begin{pmatrix} \rho(z_1) \\ \rho(z_2) \\ \vdots \\ \rho(z_n) \end{pmatrix}, \text{ con } z_i \in \mathbb{R}, \ \forall i.$$

ρ^n y ρ son funciones distintas, sin embargo normalmente se identifican y se omite el super índice n. Es decir

$$\rho^n \equiv \rho$$

de manera que se genera un abuso de notación qu se usa todo el tiempo en el contexto de redes neuronales.

En este contexto tenemos la sig. figura



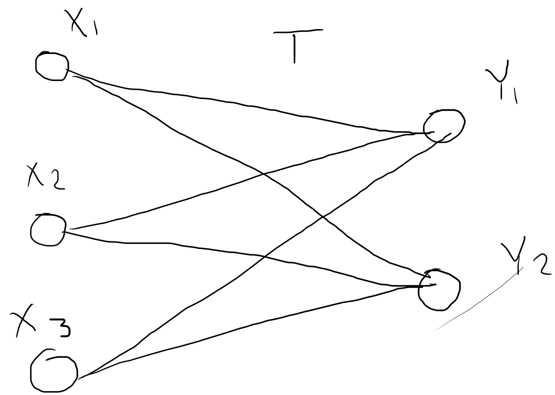
$T: \mathbb{R}^3 \longrightarrow \mathbb{R}^2$

transformacion afin

$$X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \text{señal de entrada}$$

$$f \circ T(x) = f(T(X)) = Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \text{Señal de salida}$$

De forma esquemática se usa la siguiente figura



Las bolas nombran neuronas. Las líneas representan señales. $X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$ es la capa de entrada. $Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$ es la capa de salida.

De forma general, se pueden considerar capas intermedias (capas ocultas) entre X y Y

de manera que la señal de entrada pasa por varias neuronas agrupadas en capas antes de llegar a la señal de salida

Este proceso de señales trasmitidas y análisis se perceptrón multi-capa y es el objeto fundamental de las redes neuronales.

Def. Perceptrón Multi-Capa.

Un perceptrón multi-capa "Multilayer Perceptron (MLP)" de dimensión d, L niveles y función de activación ρ es una función

$$f : \mathbb{R}^d \longrightarrow \mathbb{R}^n$$

Que tiene las sig. propiedades

Existe un arreglo de transformaciones afines

$$\phi = (T_1, T_2, \ldots, T_L)$$

en donde T_ℓ es afín, para toda ℓ, tales que

$$f = R_\rho(\phi) \equiv R(\phi),$$

en donde

$$R(\phi) = T_L \circ \rho \circ T_{L-1} \circ \rho \circ T_{L-2} \circ \rho \circ \cdots \circ \rho \circ T_1$$

Los dominios y rangos de las transformaciones afines tienen que ser compatibles para que f tenga sentido.
Ademas se tiene que cumplir
$$\mathrm{dom}(T_1) = \mathbb{R}^d.$$

Def. (Red Neuronal).

Al arreglo de transformaciones afines que se describe en la def. anterior se le nombra red neuronal

$\phi = (T_1, T_2, \ldots, T_L)$ red neuronal

Def (Realización de una Red Neuronal)

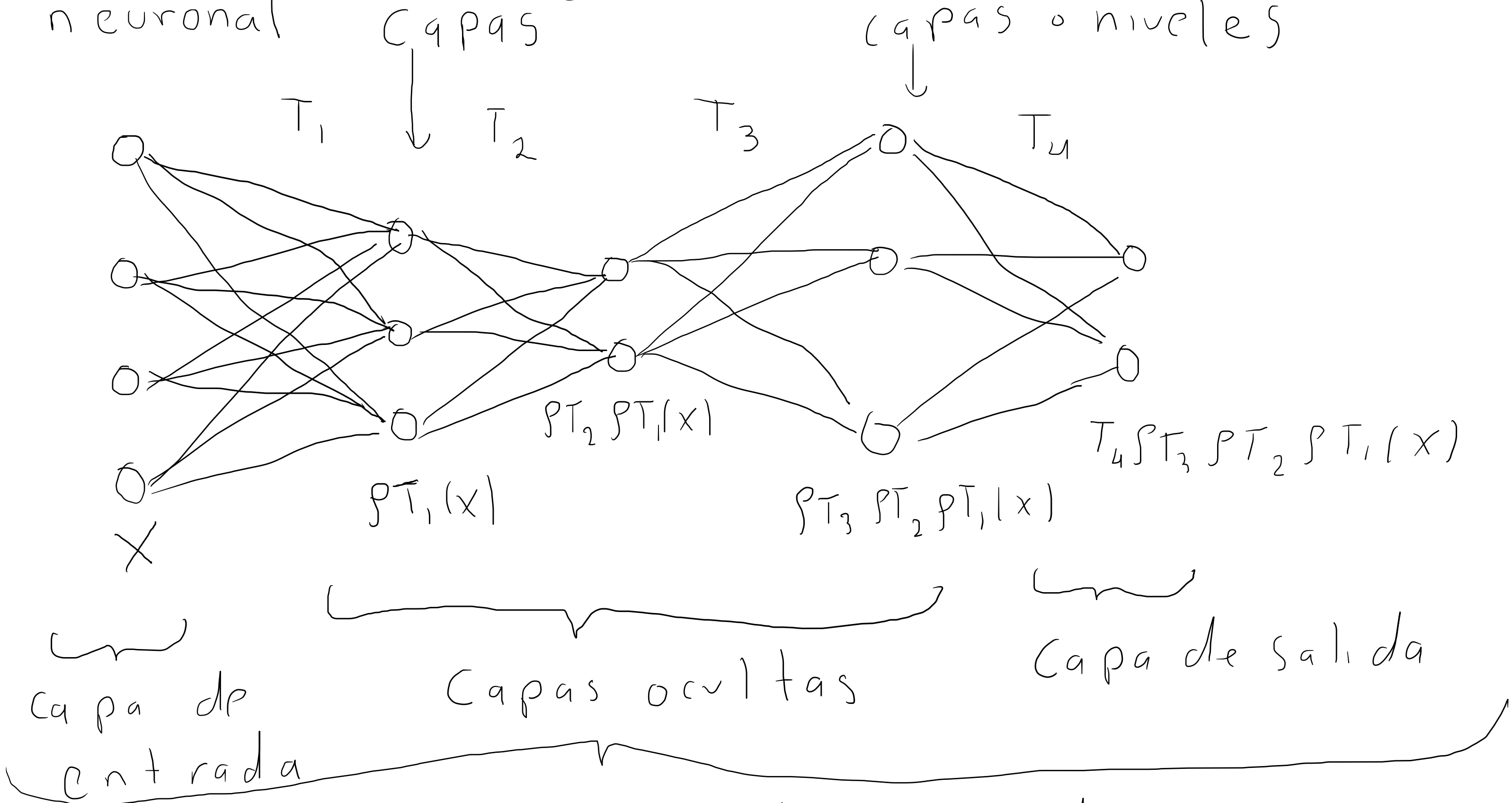
La función

R : Redes Neuronales $\longrightarrow$ Perceptrones Multi-capa

$$R(\phi) = T_L \circ \rho \circ T_{L-1} \circ \rho \circ \cdots \circ \rho \circ T_1$$

es nombrada realización de la red neuronal $\phi = (T_1, T_2, \ldots, T_L)$.

Representación gráfica de una red neuronal



Arquitectura de la red

Podemos considerar la sucesión de señales

$X_0 = X$ = señal de entrada

$X_1 = \rho T_1(X)$

$X_2 = \rho T_2 \, \rho T_1(X) = \rho T_2(X_1)$

$X_3 = \rho T_3(X_2)$

$X_4 = \rho T_4(X_3)$ señal de salida

Entonces X_ℓ se puede ver como la señal de entrada de la red neuronal $(T_{\ell+1}, \dots, T_4)$.

$y = X_4$ es la señal de salida.

Operaciones con Redes Neuronales

Dada una transformación afín, existe una matriz A y un vector b tales que

$T(x) = Ax + b$ (A y b expresadas en bases canónicas).

Identificamos

$T \equiv (A, b)$

Dado $x \in \mathbb{R}^d$ usamos como antes la notación

$x_0 = x,\quad x_1 = \rho T_1 x_0 \;\cdots,\; x_\ell = \rho T_\ell x_{\ell-1}, \cdots, x_L = T_L x_{L-1}$

En donde $\phi = (T_1, T_2, \ldots, T_L)$

Entonces $R(\phi)$ manda á x en x_L

Supongamos que $T_\ell : \mathbb{R}^{N_{\ell-1}} \longrightarrow \mathbb{R}^{N_\ell}$, $N_0 = d$

Definimos

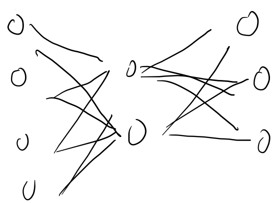
1. Dimensión de entrada $= d$
2. Número de niveles o capas $= L \equiv L(\phi)$
3. Número de neuronas $:= d + \sum_{j=1}^{L} N_j$
4. Dada una matriz A y un vector b denotamos por
   $\|A\|_0$ = entradas no cero de A.
   $\|b\|_0$ = entradas no cero de b
5. Dada una transformación afín $T \equiv (A, b)$, denotamos
   $$\|T\|_0 = \|A\|_0 + \|b\|_0$$
6. Denotamos por
   $M_j(\phi) = \|T_j\|_0$,

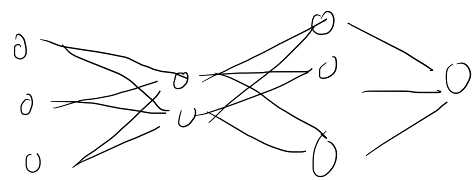
7.- Denotamos por $M(\phi) = \sum_{j=1}^{L} M_j(\phi)$

8.- Dimensión de salida $:= N_L$

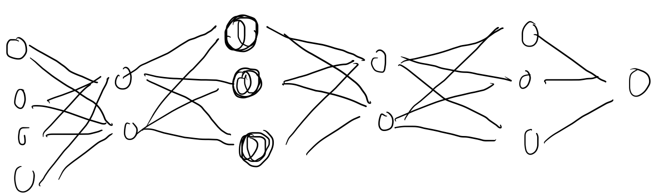
Def (concatenación)

Explicación intuitiva. Supongamos que tenemos las sig. redes





concatenar significa "pegar" las redes de forma horizontal, formando la sig. red



Sean $\phi^{(1)}, \phi^{(2)}, \ldots, \phi^{(k)}$ redes neuronales con $\phi^{(j)} = \left(T_1^{(j)}, \ldots, T_{\theta_j}^{(j)}\right)$

Definimos

$$\phi^{(1)} \bullet \phi^{(2)} \bullet \cdots \bullet \phi^{(k)} :=$$

$$\Big(T_1^{(k)}, T_2^{(k)}, \ldots T_{\theta_{k-1}}^{(k)}, T_1^{(k-1)} \circ T_{\theta_k}^{(k)}, T_2^{(k-1)}, T_3^{(k-1)}, \ldots$$

$$T_{\theta_{k-1}-1}^{(k-1)}, T_1^{(k-2)} \circ T_{\theta_{k-1}}^{(k-1)}, T_2^{(k-2)}, \ldots,$$

$$T_1^{(2)} \circ T_{\theta_3}^{(3)}, T_2^{(2)}, \ldots, T_{\theta_2-1}^{(2)}, T_1^{(1)} \circ T_{\theta_2}^{(2)}, T_2^{(1)}, \ldots, T_{\theta_1}^{(1)}\Big)$$

Esta es una red neuronal de

$$\sum_{j=1}^{k} \theta_j - (k-1)$$

Esto se puede hacer siempre y cuando la dimensión de salida de $\phi^{(j)}$ coincida con la dimensión de entrada de $\phi^{(j-1)}$, para toda j.

<u>Prop.</u>

$$R(\phi^{(1)} \circ \phi^{(2)} \circ \cdots \circ \phi^{(k)}) = R(\phi^{(1)}) \circ R(\phi^{(2)}) \circ \cdots \circ R(\phi^{(k)})$$

<u>Dem</u>

$$R(\phi^{(1)} \circ \phi^{(2)} \circ \cdots \circ \phi^{(k)}) =$$

$$\underline{T^{(1)}_{\theta_1} \circ \rho \circ \cdots \circ \rho \circ T^{(1)}_2 \circ \rho \circ T^{(1)}_1} \circ \underline{T^{(2)}_{\theta_2} \circ \rho \circ T^{(2)}_{\theta_2 - 1}}$$

$$\underline{\circ \circ \cdots \circ \rho \circ T^{(2)}_2 \circ \rho \circ T^{(2)}_1} \circ T^{(3)}_{\theta_3} \circ \rho \circ \cdots \circ \rho \circ T^{(k-2)}_2 \circ \rho \circ$$

$$T^{(k-2)}_1 \circ T^{(k-1)}_{\theta_{k-1}} \circ \rho \circ T^{(k-1)}_{\theta_{k-1}} \circ \rho \circ \cdots \circ T^{(k-1)}_2 \circ \rho \circ T^{(k-1)}_1 \circ T^{k}_{\theta_k} \circ \rho$$

$$\circ T^{(k)}_{\theta_k - 1} \circ \rho \circ \cdots \circ T^{(k)}_2 \circ \rho \circ T^{(k)}_1 = R(\phi^{(1)}) \circ R(\phi^{(2)}) \circ \cdots \circ R(\phi^{(k)})$$

$\square$

Def. Dadas $T^{(1)}, T^{(2)}, \dots, T^{(k)}$ transformaciones afines, definimos

1.- (En el caso de que $T^{(1)}, T^{(2)}, \dots, T^{(k)}$ actúen en el mismo espacio)

$$\begin{pmatrix} T^{(1)} \\ T^{(2)} \\ \vdots \\ T^{(k)} \end{pmatrix}(X) = \begin{pmatrix} T^{(1)}(X) \\ T^{(2)}(X) \\ \vdots \\ T^{(k)}(X) \end{pmatrix}$$

2.- $T_1 \oplus T_2 \oplus \cdots \oplus T_K \begin{pmatrix} X^{(1)} \\ X^{(2)} \\ X^{(3)} \\ \vdots \\ X^{(K)} \end{pmatrix} = \begin{pmatrix} T_1(X^{(1)}) \\ T_2(X^{(2)}) \\ \vdots \\ T_K(X^{(K)}) \end{pmatrix}$

Def. Dadas redes neuronales

$$\phi^{(j)} = \left(T_1^{(j)}, T_2^{(j)}, \dots, T_L^{(j)}\right), \quad j \in \{1, 2, \dots, K\}$$

Definimos

1. $P(\phi^{(1)}, \phi^{(2)}, \dots, \phi^{(k)}) =$

$$\left(\begin{pmatrix} T_1^{(1)} \\ T_1^{(2)} \\ \vdots \\ T_1^{(k)} \end{pmatrix}, T_2^{(1)} \oplus \dots \oplus T_2^{(k)}, T_3^{(1)} \oplus \dots \oplus T_3^{(k)}, \dots, T_L^{(1)} \oplus \dots \oplus T_L^{(k)} \right)$$

si las dimensiones de entrada de $\phi^{(1)}, \phi^{(2)}, \dots, \phi^{(k)}$ son las mismas

2. $FP(\phi^{(1)}, \phi^{(2)}, \dots, \phi^{(k)}) =$

$$\left(T_1^{(1)} \oplus \dots \oplus T_1^{(k)}, T_2^{(1)} \oplus \dots \oplus T_2^{(k)}, \dots, T_L^{(1)} \oplus \dots \oplus T_L^{(k)} \right)$$

Prop

1. $M(P(\phi^{(1)}, \phi^{(2)}, \ldots, \phi^{(n)})) = \sum_j M(\phi^{(j)})$

$= M(FP(\phi^{(1)}, \phi^{(2)}, \ldots, \phi^{(n)}))$

2. $R(P(\phi^{(1)}, \phi^{(2)}, \ldots, \phi^{(n)}))(x) = \begin{pmatrix} R(\phi^{(1)})(x) \\ R(\phi^{(2)})(x) \\ \vdots \\ R(\phi^{(n)})(x) \end{pmatrix}$

Tarea

$$\left\{ R(FP(\phi^{(1)}, \phi^{(2)}, \ldots, \phi^{(n)}))\begin{pmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(n)} \end{pmatrix} = \begin{pmatrix} R(\phi^{(1)})(x^{(1)}) \\ \vdots \\ R(\phi^{(n)})(x^{(n)}) \end{pmatrix} ?\right.$$

Dem de 2.

$$R(P(\phi^{(1)}, \phi^{(2)}, \dots, \phi^{(k)}))(x)$$

$$= R\left(\begin{pmatrix} T_1^{(1)} \\ \vdots \\ T_1^{(k)} \end{pmatrix}, T_2^{(1)} \oplus \dots \oplus T_2^{(k)}, \dots, T_L^{(1)} \oplus \dots \oplus T_L^{(k)} \right)(x)$$

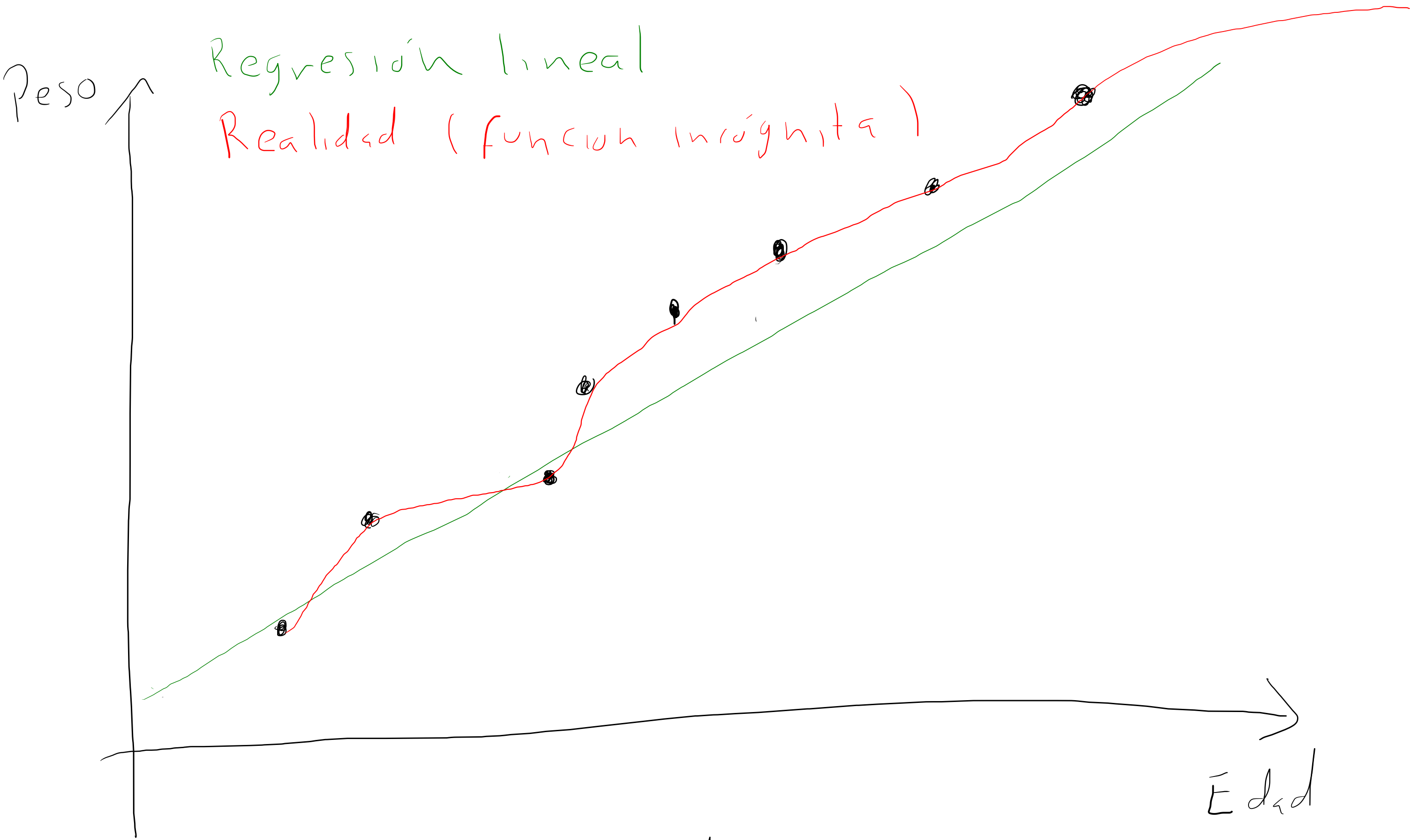
$$= T_L^{(1)} \oplus \dots \oplus T_L^{(k)} \circ \rho \circ \dots T_2^{(1)} \oplus \dots \oplus T_2^{(k)} \begin{pmatrix} \rho(T_1^{(1)}(x)) \\ \vdots \\ \rho(T_1^{(k)}(x)) \end{pmatrix}$$

$$= T_L^{(1)} \oplus \dots \oplus T_L^{(k)} \circ \rho \dots \circ \rho T_3^{(1)} \oplus \dots \oplus T_3^{(k)} \circ \rho \begin{pmatrix} T_2^{(1)}(\rho(T_1^{(1)}(x))) \\ \vdots \\ T_2^{(k)}(\rho(T_1^{(k)}(x))) \end{pmatrix}$$

$$= \begin{pmatrix} T_L^{(1)} \circ \rho \circ T_{L-1}^{(1)} \circ \rho \circ \cdots \circ T_3^{(1)} \circ \rho \circ T_2^{(1)} \circ \rho \circ T_1^{(1)}(X) \\ T_L^{(2)} \circ \rho \circ T_{L-1}^{(2)} \circ \rho \circ \cdots \circ T_3^{(2)} \circ \rho \circ T_2^{(2)} \circ \rho \circ T_1^{(2)}(X) \\ \vdots \\ T_L^{(K)} \circ \rho \circ T_{L-1}^{(K)} \circ \rho \circ \cdots \circ T_3^{(K)} \circ \rho \circ T_2^{(K)} \circ \rho \circ T_1^{(K)}(X) \end{pmatrix}$$

$$= \begin{pmatrix} R\phi^{(1)}(X) \\ R\phi^{(2)}(X) \\ \vdots \\ R\phi^{(K)}(X) \end{pmatrix}$$

$\square$



Medimos pesos y edades de 8 personas
Deseo: Que la línea recta aproxime a la función de la realidad.

Respuesta al deseo:

No todas las funciones se pueden aproximar con líneas rectas

Alternativa: Considerar otras familias de funciones que si puedan aproximar a cualquier función. Por ejemplo

1. Polinomios (aproximan a las funciones continuas definidas en intervalos cerrados y acotados).

2. Series de Fourier.

3. Realizaciones de redes neuronales

Prop Sean $f_1, f_2, \dots, f_K$

$f_i : \mathbb{R}^d \longrightarrow \mathbb{R}^{m_i}$.

Supongamos que dada $\varepsilon > 0$ existe una red neuronal $\phi_\varepsilon^{(i)}$ tal que

$\| R\phi_\varepsilon^{(i)}(x) - f_i(x) \| \leq \varepsilon, \quad \forall x \in \mathbb{R}^d.$

Entonces, para toda $\varepsilon > 0$ existe una red neuronal ϕ_ε tal que

$$\left\| R\phi_\varepsilon(x) - \begin{pmatrix} f_1(x) \\ \vdots \\ f_K(x) \end{pmatrix} \right\| \leq \varepsilon, \quad \forall x \in \mathbb{R}^d.$$