



REDES NEURONALES 2024

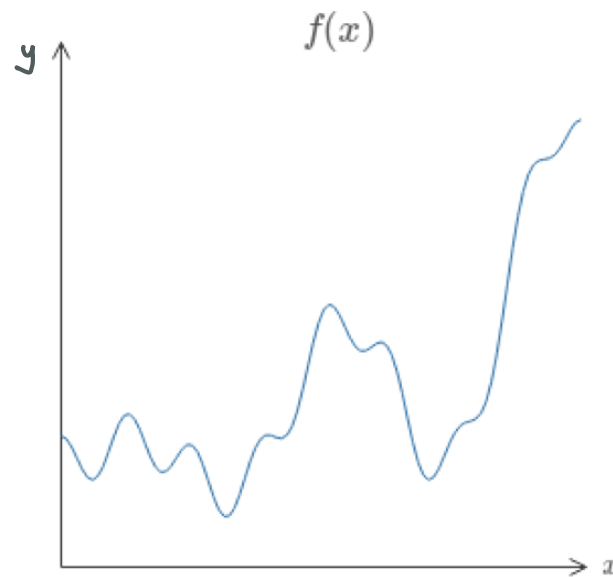
Clase 20

Jueves 31 de octubre 2024

FAMAF, UNIVERSIDAD NACIONAL DE CÓRDOBA

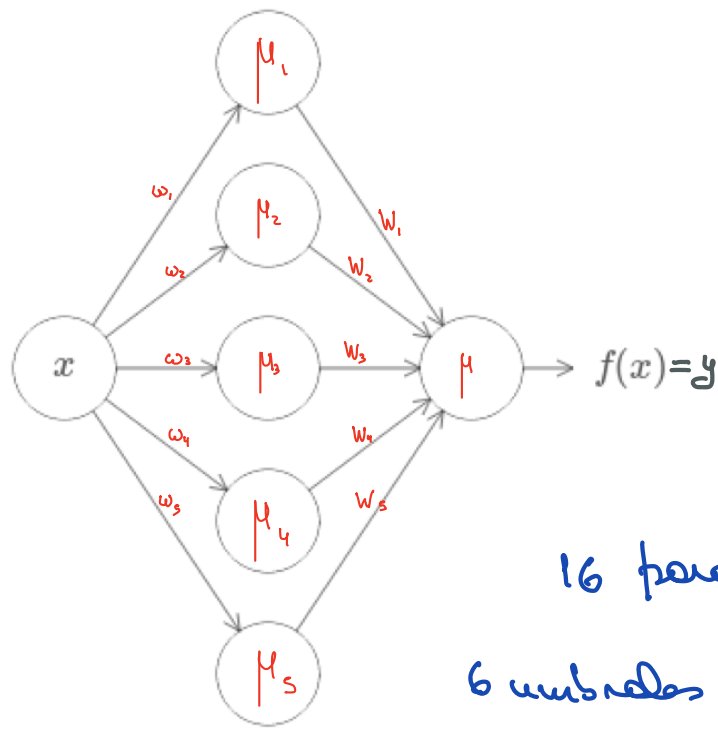
INSTITUTO DE FÍSICA ENRIQUE GAVIOLA (UNC-CONICET)

Un ejemplo muy interesante y didáctico de aprendizaje automático supervisado usando redes neuronales feed-forward es aprender una función arbitraria $f(x)$. Supongamos que alguien nos dice que la relación entre una variable independiente x y una variable dependiente y viene dada por la siguiente gráfica:



En lugar de aproximarla mediante alguna función predeterminada, como podría ser un polinomio de alto grado, a partir de ciertos puntos medidos, trataremos de **APRENDERLA**. Como x e y son escalares (números reales) nos construimos una red muy simple:

$$\begin{aligned}
 y = \mathcal{O}_1 &= g_2 \left(\sum_{j=1}^5 w_{1j} V_j - \mu \right) \\
 &= g_2 \left(\sum_{j=1}^N w_{1j} g_1(w_{1j} x_j - \mu_j) - \mu \right)
 \end{aligned}$$



16 parameters
6 neurons & 10 synapses

Asumamos que tenemos un conjunto de entrenamiento:

$$(x_i, \hat{y}_i) \quad i=1, 2, \dots, p$$

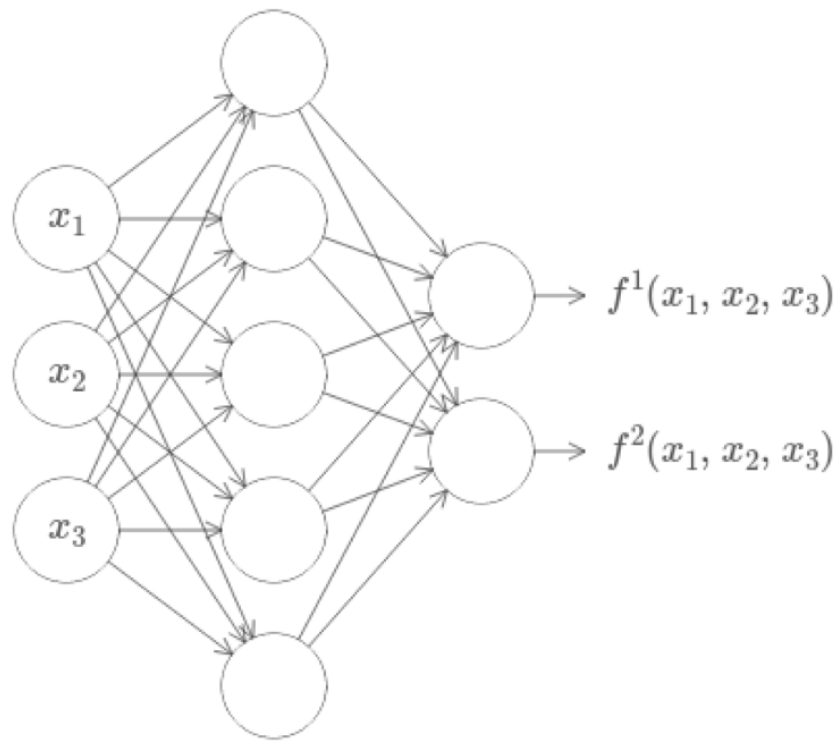
NOTA: esto se puede generalizar a una función arbitraria de \mathbb{R}^n a \mathbb{R}^m .

Pensemos por ejemplo en una función con tres entradas reales y dos salidas reales:

$$f : \mathbb{R}^3 \rightarrow \mathbb{R}^2$$

$$u = f_1(x, y, z)$$

$$v = f_2(x, y, z)$$



Por suerte disponemos de un teorema que nos garantiza el éxito en este proceso y a la vez explica qué estamos haciendo.

***Approximation** by superpositions of a sigmoidal function, by George Cybenko (1989). The result was very much in the air at the time, and several groups proved closely related results. Cybenko's paper contains a useful discussion of much of that work. Another important early paper is **Multilayer feedforward networks are universal approximators**, by Kurt Hornik, Maxwell Stinchcombe, and Halbert White (1989). This paper uses the Stone-Weierstrass theorem to arrive at similar results.

Todo problema de aprendizaje automático supervisado es en realidad un problema de aproximación por medio de combinaciones de composiciones de funciones no lineales $g(x)$. Lejos de buscar una función simple y de pocos parámetros, aproximamos con muchos parámetros, que no son más que los acoplamientos sinápticos.

Una vez definida la arquitectura de la red, tenemos definida la función que usaremos para aproximar. Luego variaremos los parámetros y elegiremos aquellos que minimiza el error fuera de la muestra E_{out} .

ANALOGÍA CON LAS CIENCIAS NATURALES

La idea del aprendizaje automático supervisado acerca a las ciencias de la computación, vista usualmente como una ciencia formal (como la matemática o la lógica) con las ciencias naturales. Las ideas de error, realización, distribución, entre otras, no son frecuentes en las ciencias exactas, salvo en el estudio de la teoría de probabilidades y la estadística. De hecho, podemos decir que estamos trabajando en extensiones de la ESTADÍSTICA, la gran ciencia formal de datos. Sin estadística no habría métodos cuantitativos en las ciencias naturales, sociales, de la vida e ingenierías. En este sentido, el aprendizaje automático se puede ver como una parte disruptiva de la estadística, y ésta siempre nos guía en la generación de métodos, en la validación de resultados y la previa curación de los datos. Sin este auxilio, no podríamos considerar al aprendizaje automático como el conjunto de

¿Que es un modelo cuantitativo?

Como ustedes seguramente están descubriendo poco a poco, el aprendizaje de máquina o aprendizaje automático es una su disciplina esencialmente multidisciplinaria, donde saberes muy diversos deben dialogar.

El primer pre-requisito para dialogar es tener voluntad de hacerlo. El segundo es compartir una lengua. Y el tercero es estar de acuerdo en el vocabulario.

Esta tercera condición representa un serio problema para quienes trabajamos en aprendizaje automático neuronal, pues hay términos que cada disciplina usa, muchas veces al mismo tiempo que otras disciplinas usan otras. Un ejemplo claro es el de las conexiones neuronales. Para un biólogo estas se denominan sinapsis. Sin embargo, para un informático esta denominación no tiene sentido y lo mejor es llamarlos parámetros. Ambos tiene razón pues en diferentes contextos cada denominación tiene sus ventajas. De la misma manera unos usan la palabra sesgo, otros umbrales y los informáticos los llaman también parámetros. Un científico natural sin duda prefiere llamar función error a lo que un informático llama función perdida y en economista llama función costo. No hay duda de la validez de las tres denominaciones. Cuando estas cosas pasan los humanos generamos consensos muy útiles que nos permiten ponernos rápidamente de acuerdo.

Sin embargo, hay un problema más difícil que nos cuesta sobrellevar, y tiene que ver con lo que cada disciplina entiende por modelo.

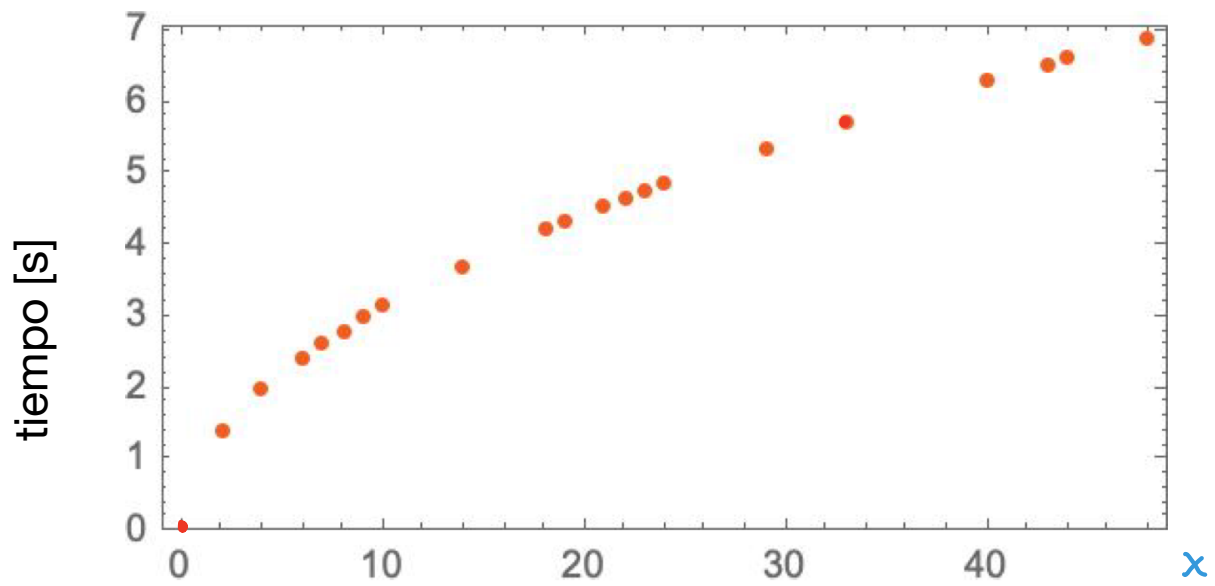
¿A qué nos referimos cuando hablamos de modelar un fenómeno o un proceso?

Hay modelos meramente cualitativos, en los cuales las hipótesis y las tesis se plasman solamente a través del lenguaje natural. Esta forma sigue siendo muy usada y muy útil en las ciencias naturales y en la tecnología. Sin embargo, en estas disciplinas hemos aprendido con el paso de los siglos, a generar modelos cuantitativos, en los cuales el fenómeno o el proceso es descrito por un modelo matemático que expresa una relación funcional entre variables que podemos controlar, de entrada podemos decir, y las variables que podemos medir, o salidas.

El padre del modelado cuantitativo sin duda fue Isaac Newton, quien a través de su segunda ley nos permitió plantear una ecuación diferencial de segundo orden para la relación entre el tiempo en que medimos y la posición en que se encuentra un objeto dado cuyo movimiento queremos describir. Tener un modelo matemático tiene la enorme ventaja de permitirnos hacer predicciones sobre el comportamiento futuro del sistema en consideración, un futuro en el cual por cierto no podemos medir.

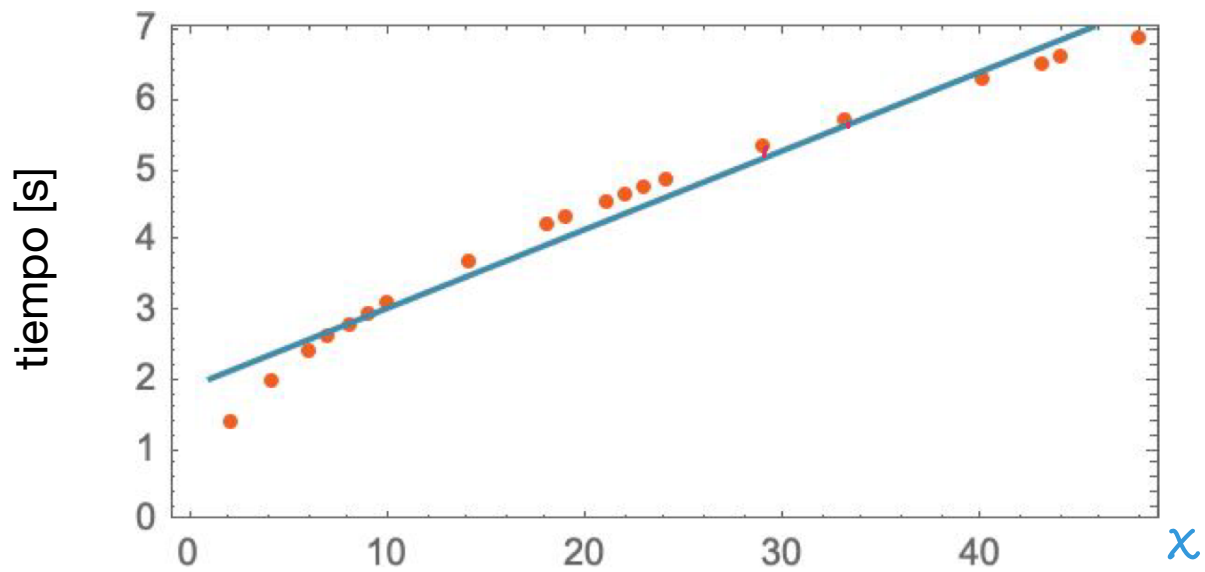
En la ingeniería, en la física o en la astronomía, un modelo cuantitativo viene dado en forma de una o varias ecuaciones

UN EJEMPLO SIMPLE DE LA FISICA



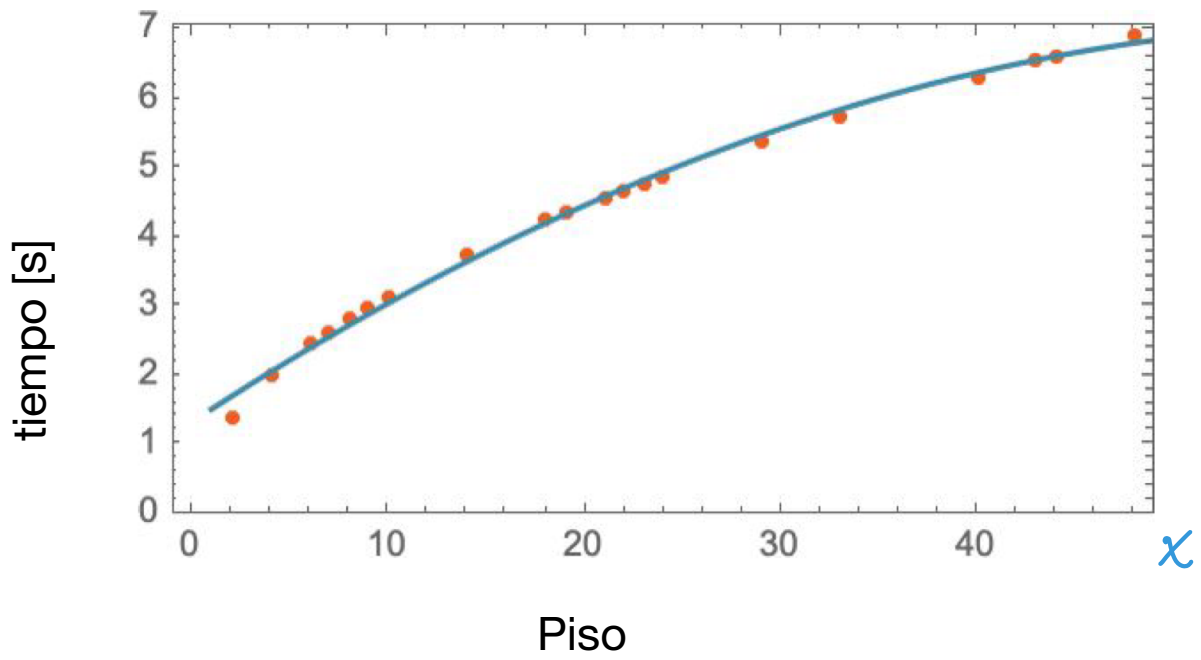
Piso

$$t = f(x) + \eta$$

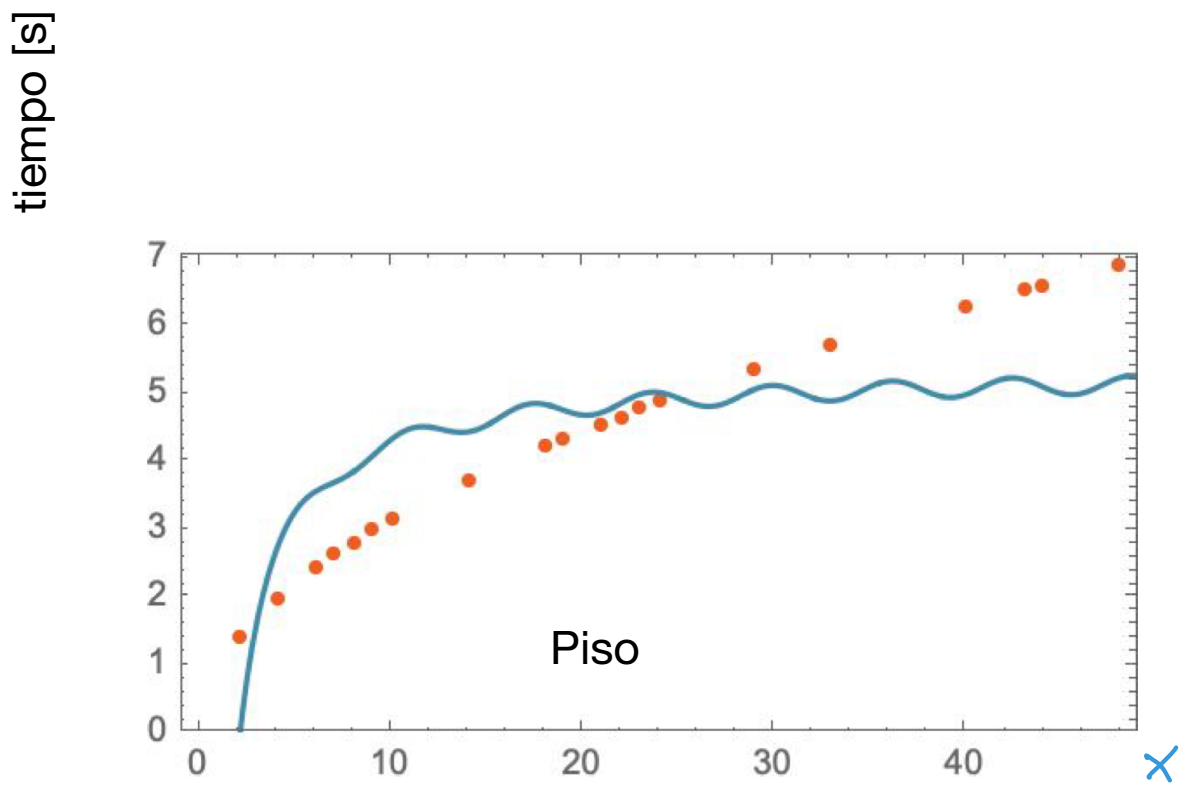


Piso

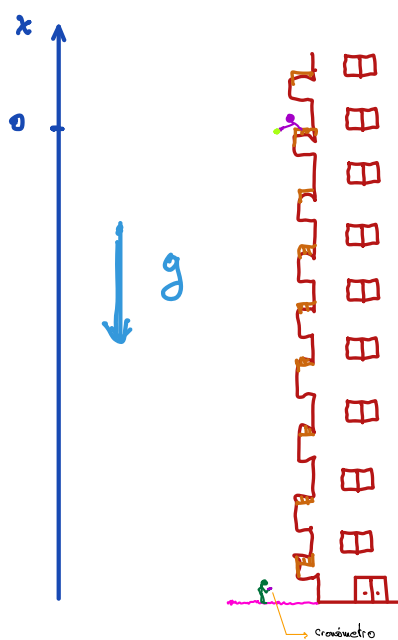
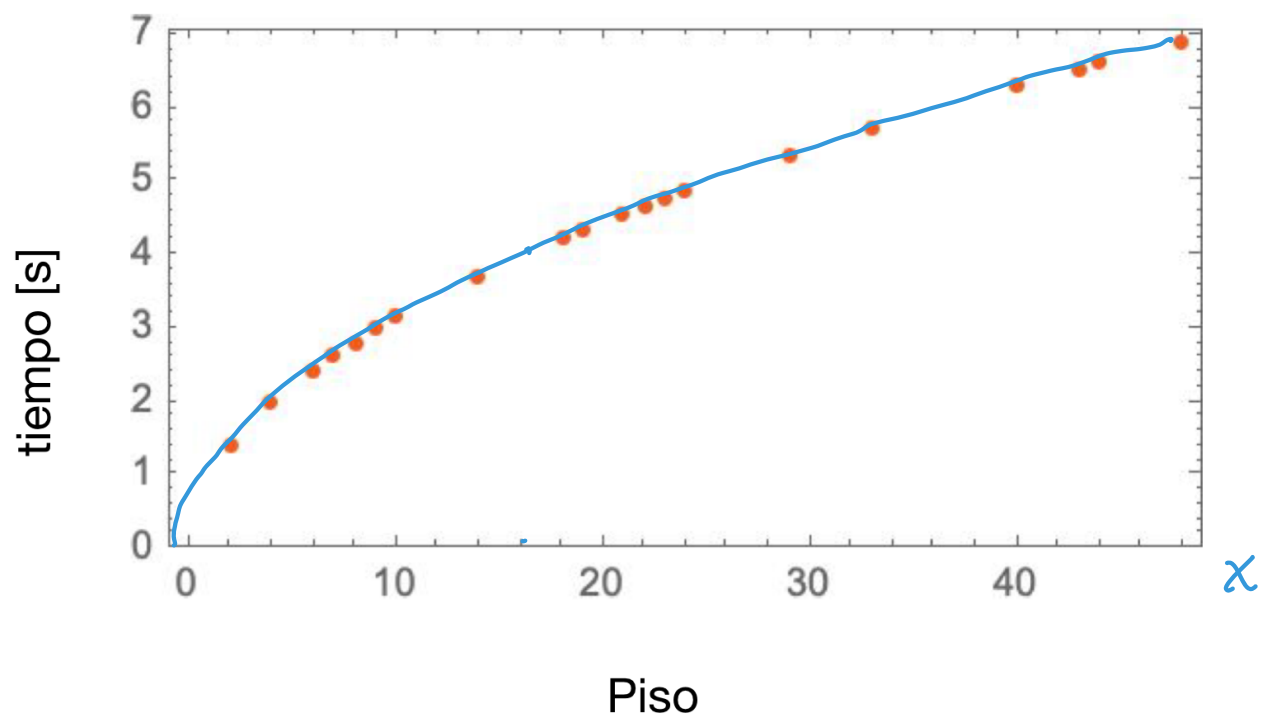
$$t = a + bx$$



$$t = a + bx + cx^2$$



$$t = a + \frac{b}{x} + c \sin(x)$$



$$a(t) = -g$$

$$v(t) = -gt + v_0$$

$$x(t) = -\frac{1}{2}gt^2 + v_0 t + x_0$$

$$x(t=0) = x_0 = 0$$

$$v(t=0) = v_0 = 0$$

$$x(t) = \frac{1}{2}gt^2$$

$$t = \sqrt{\frac{2x}{g}}$$

El modelo es aproximado, pues desprecia el rozamiento que produce el aire, desprecia las corrientes de aire transversales, el error en la determinación de la posición y velocidad inicial, la forma de la pelota, entre muchos otros factores.

Nuestro modelo, basado en las leyes de la cinemática, tiene un único parámetro g . ¿Cómo determinamos g ? Con cuadrados mínimos.

Lo elegimos como el valor de g que minimiza el error cuadrático medio entre el valor medido y el valor que predice el modelo.

$$E_{\text{error}}(g) = \frac{1}{20} \sum_{i=1}^{20} (\gamma_i - \hat{\gamma}_i)^2$$

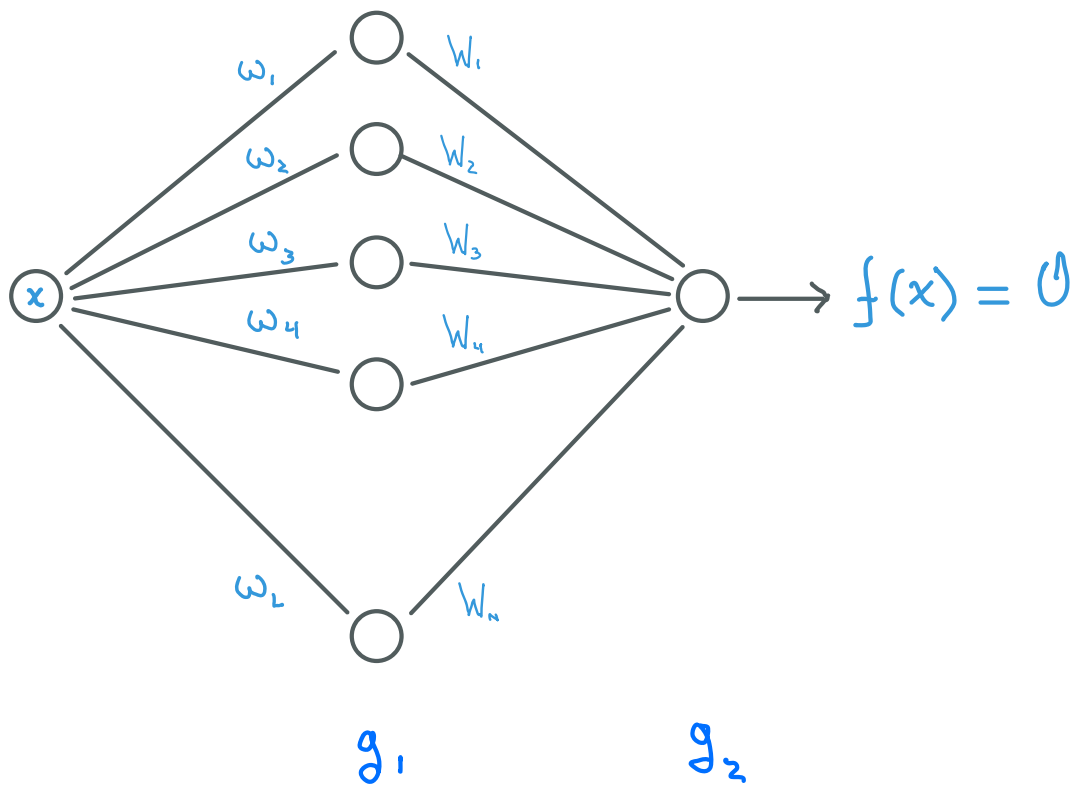
$\hat{\gamma}_i$: valor del modelo

γ_i : valor de la medición

En machine learning el problema es parecido. Tenemos datos y queremos predecir un número o varios. Imaginemos que tenemos una variable dependiente y una independiente. Medimos e intuimos qué hay una relación

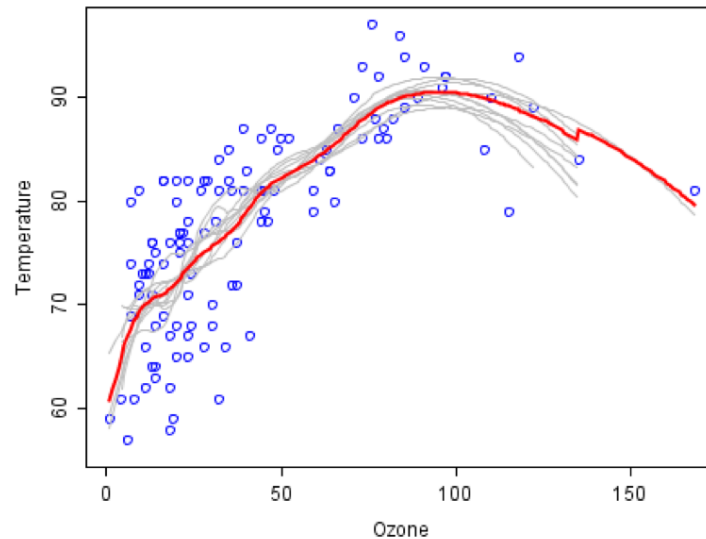
$$y = f(x)$$

Hacemos una red neuronal de una capa oculta para aprender lo mejor posible la función objetivo $f(x)$



$$\mathcal{O} = g_2 \left(\sum_j W_j g_1 \left(\sum_k \omega_k x \right) \right)$$

Consideremos un caso particular en el cual no podemos suponer una relación analítica a priori entre la entrada (el tamaño de la capa de ozono) y la temperatura en la tierra en cierto lugar específico que no nos interesa ahora (y que desconozco).



$$T = f(x) + \eta \quad \text{--- ruido}$$

Los círculos azules son mediciones. Cada una responde a un modelo más un ruido sin sesgo que no sabemos controlar. La red neuronal aprende la función objetivo f sin necesidad de suponer su forma matemática. En este caso la función f es un aproximador universal. Depende de la única variable independiente x y tiene $2L$ parámetros, donde L puede ser muy grande. Como cuando hacemos mínimos cuadrados, minimizamos el error cuadrático mínimo pero en forma algorítmica. Hacerlo algebraicamente sería muy costoso.

