

**Diplomatura en Ciencia de Datos, Inteligencia Artificial y
sus aplicaciones a la economía y los negocios
FCE y FAMAF**

MÓDULO 3

UNIDAD 2

Clase 1 Parte 3

FCE y FAMAF

**Universidad Nacional de Córdoba
Sábado 23 de octubre 2021**

EL APRENDIZAJE SUPERVIZADO

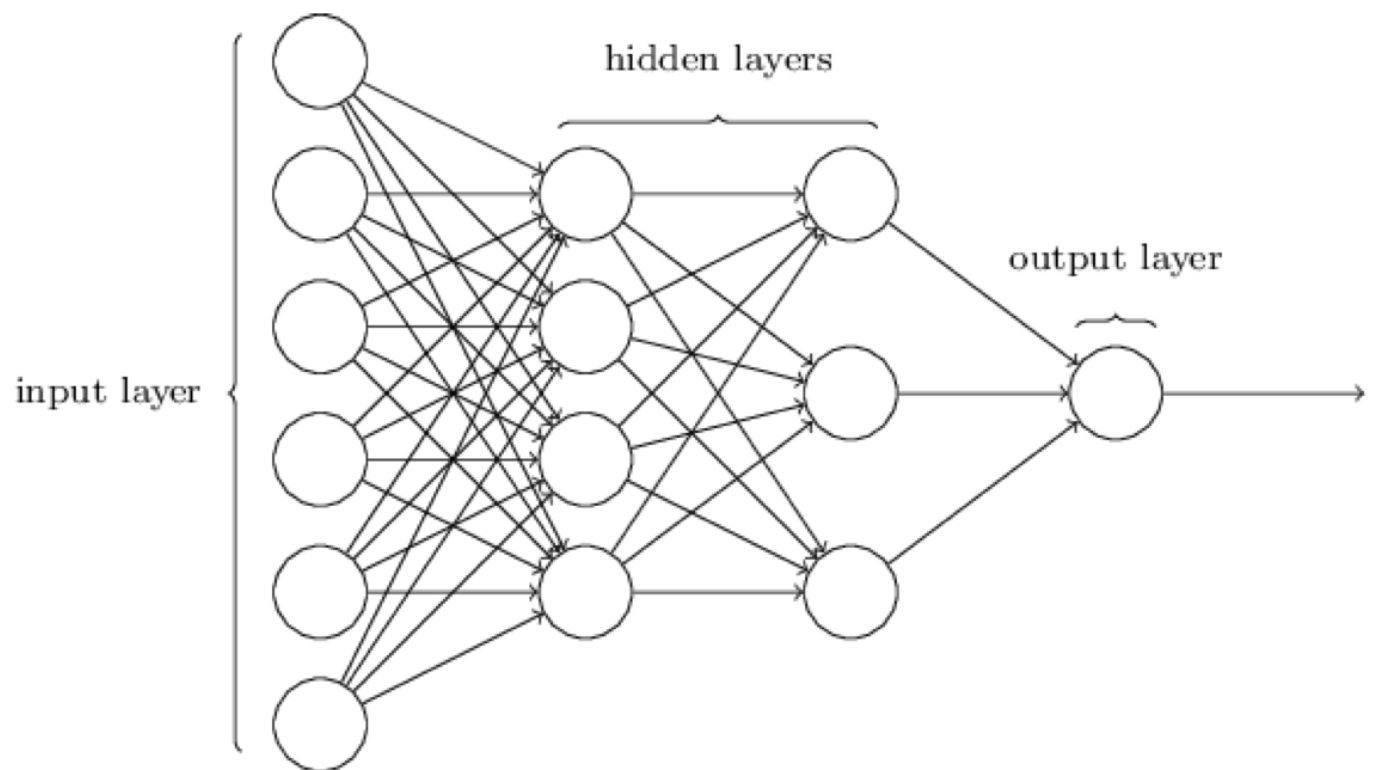
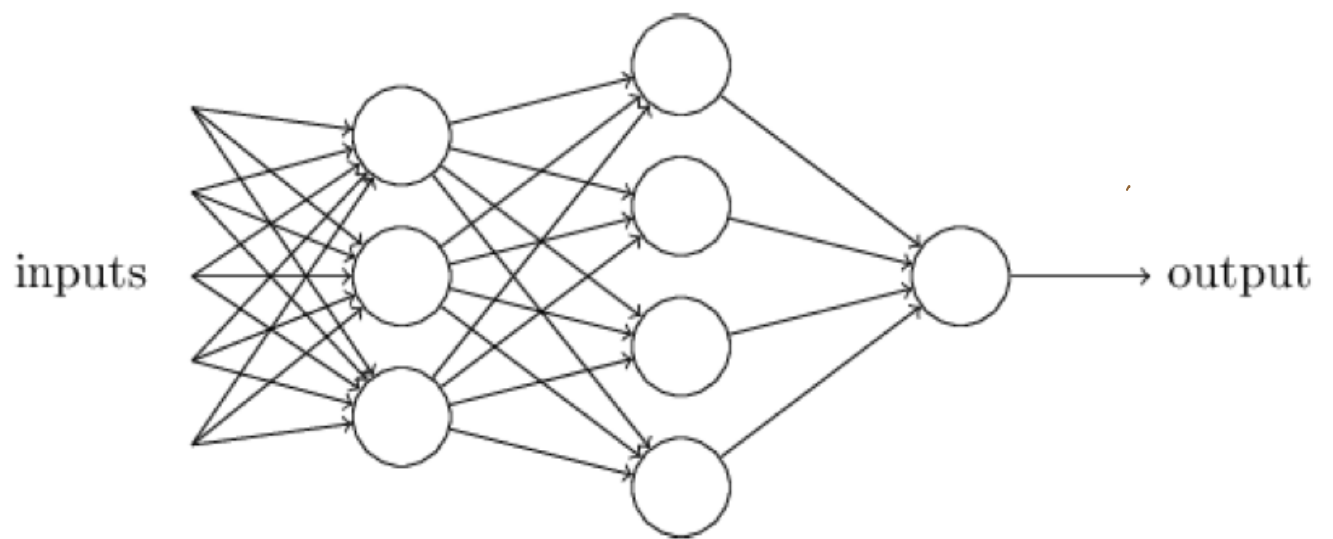
El aprendizaje automático es una técnica del aprendizaje de máquina por medio de la cual podemos estimar el valor de una función arbitraria sin conocer su expresión funcional y a partir de un conjunto de valores conocidos de la función llamado conjunto de entrenamiento.

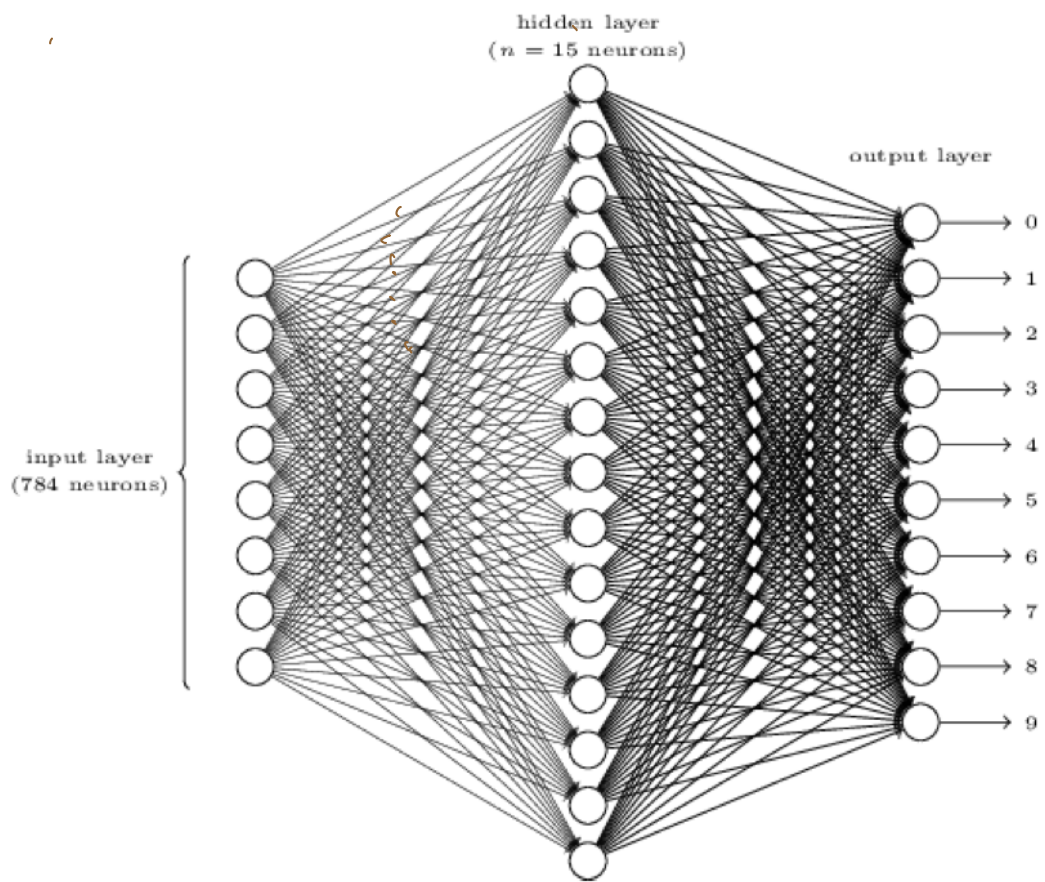
REDES FEED-FORWARD

En una red neuronal feed forward las neuronas se organizan en capas sucesivas. La información entra en la llamada capa de entrada, que no es una capa de neuronas. De ahí viaja a la primera capa. Una vez procesada la información por las neuronas de la primera capa, estas se la envían a las neuronas de la segunda capa, hasta llegar a la última capa o capa de neuronas de salida.

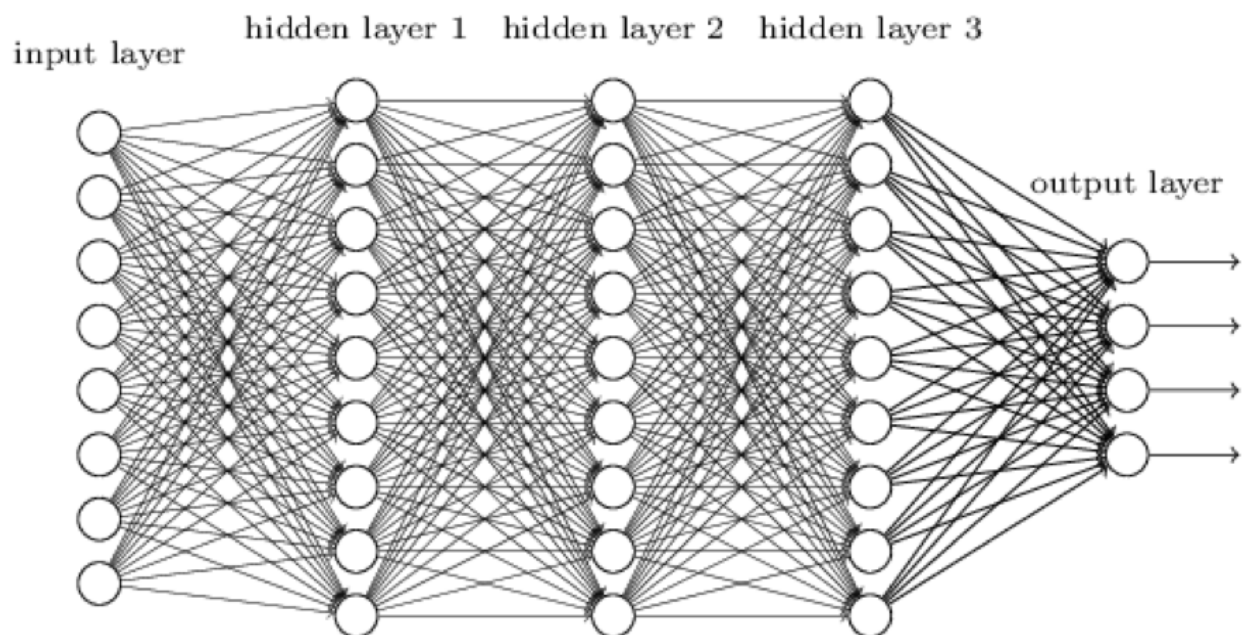
No puede haber intercambio de información dentro de una capa, ni hacia atrás ni saltarse capas hacia adelante. Esta arquitectura tan ordenada facilita la elaboración y la programación de los algoritmos.

Por fuera de las redes feed-forward, la otra arquitectura popular es la llamada arquitectura de redes recurrentes.

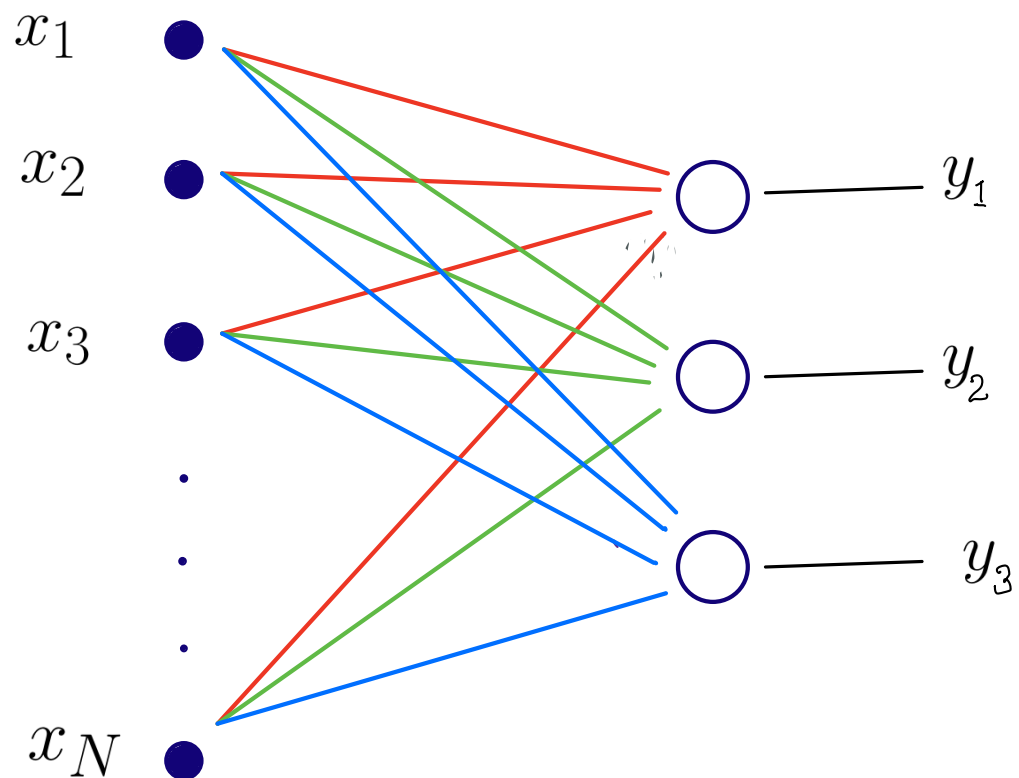




5 0 4 1 9 2



Consideremos una red muy simple como la de la figura siguiente, con cinco entradas y tres neuronas en su única capa, que es a la vez la capa de salida. Noten que la capa de entrada no es una capa de neuronas pues no procesan nada, solo presentan.



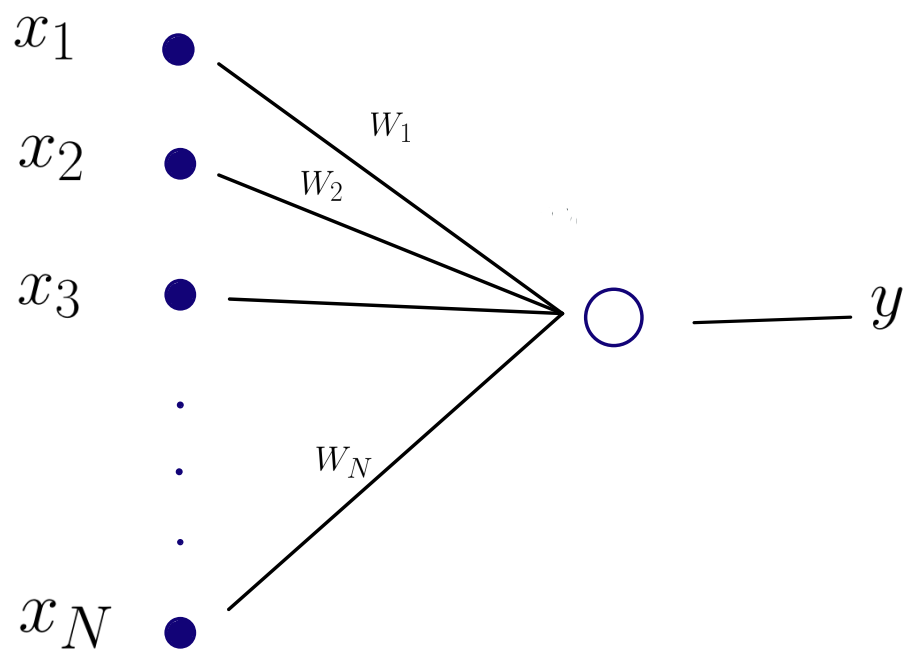
Llamaremos siempre x_k a la entradas (features)

Llamaremos siempre y_i a los estados de las neuronas de salida

Tenemos tres redes, separadas por colores, cada una independiente de las otras, y cada una con una única salida.

Esto nos permite desacoplar toda red feed-forward de una única capa en tantos PERCEPTRONES simples como neuronas en la capa de salida tengamos.

En nuestro caso lo desacoplamos en tres perceptrones simple.



El conjunto de variables de entrada forman un vector en

$$\vec{X} = (x_1, x_2, \dots, x_N) \in R^N$$

Las variables de entrada $\bar{\mathbf{X}}$ pueden tomar valores discretos o continuos, acotados o no.

El rango de valores que puede tomar y depende exclusivamente de la función de activación $g(\bar{\mathbf{X}})$.

Si $g(\bar{\mathbf{X}})$ es discreta nuestra red aprenderá a clasificar. Si es continua aprenderá a hacer una regresión. Volveremos sobre este punto muchas veces.

$$y_i = g(h_i(t) - \mu_i) = g \left(\sum_k^N W_{ik} y_k - \mu_i \right)$$

Aprendizaje supervisado en un perceptron simple


Supongamos que deseamos modelar cierto proceso que asigna a vectores en \mathbb{R}^N un único valor real y . No tenemos ninguna información sobre la expresión matemática de dicha asignación, pero podemos imaginarla como en función real que come un vector de N componentes reales y devuelve un número real.

Llamemos f a esa relación; entonces

$$f(\vec{X}) = y \qquad f : \mathbb{R}^N \rightarrow \mathbb{R}$$

Supongamos que la única información de la cual disponemos es el valor que esta función asigna a cierto conjunto de valores de entrada al cual denominamos *conjunto de entrenamiento*. Supongamos que son p los datos de entrada cuya solución conozco y los enumeramos con la letra griega α .

$$\vec{X}^\alpha = (x_1^\alpha, x_2^\alpha, \dots, x_N^\alpha) \rightarrow \hat{y}^\alpha$$


Experto

$\alpha = 1, 2, \dots, p$

$$\{\vec{X}^\alpha, \hat{y}^\alpha\}_{\alpha=1}^p \quad \text{—} \quad \text{Conjunto de entrenamiento}$$

y^{α} representa el valor que debería tomar f cuando le presentamos \bar{x}^{α} .

Pero la neurona de salida tomará el valor

$$y = g(h(t) - \mu) = g \left(\sum_k^N W_k x_k - \mu \right)$$

Nuestro desafío es ahora encontrar un conjunto posible de eficacias sinápticas \bar{W} y un valor del umbral μ que sean capaces de hacer que nuestra red asigne correctamente las entradas y las respectivas salidas para el conjunto de entrenamiento.

Si tenemos suerte y si el conjunto de entrenamiento es suficientemente grande y variado, esta red, que aprendió el conjunto de entrenamiento, será capaz de resolver bien incluso los casos que nunca vió. O sea, podrá asignar una salida para una entrada arbitraria.

Esto es equivalente al problema de ajustar los parámetros de una función modelo mediante cuadrados mínimos.

Observemos que podemos pensar a las eficacias sinápticas, que recordamos pueden tomar cualquier valor real, como formando un vector en \mathbb{R}^N .

$$\vec{W} = (W_1, W_2, \dots, W_N)$$

Ambos vectores, \vec{X} y \vec{W} viven en el mismo espacio.

¡Son vecinos!

Si tenemos N entradas o features y M neuronas de salida, podemos aprender una función arbitraria

$$f : \mathbb{R}^N \rightarrow \mathbb{R}^M$$

Ustedes ya vieron tres ejemplos en la unidad anterior

- Si la función de activación es discreta podemos aprender a clasificar.
- Si la función de activación es lineal pueden hacer una regresión lineal.
- Si la función de activación es sigmoide, pueden hacer una regresión logística.

