

Mecanismo de Aprendizaje de una Red Neuronal (RN)

El mecanismo de aprendizaje de una RN se basa en la configuración simultánea de todos los parámetros internos a partir del procesamiento de un gran conjunto de datos, con su resultado correctamente etiquetado. Este mecanismo crea conexiones entre neuronas, destruye conexiones entre neuronas o modifica las conexiones preexistentes para optimizar los resultados y que sean más correctos en cada ciclo de prueba con cada dato disponible. Este mecanismo se llama retropropagación (backpropagation).

Estructura básica de una RN:

Una red neuronal consta de capas de neuronas interconectadas como se muestra en la figura 1. Estas capas se dividen en tres tipos principales: la capa de entrada, capas ocultas y la capa de salida. Cada neurona en una capa está conectada a todas las neuronas de la capa siguiente y anterior mediante conexiones ponderadas por pesos. La información fluye desde la capa de entrada, a través de las capas ocultas, hasta la capa de salida, donde se obtiene la predicción final.

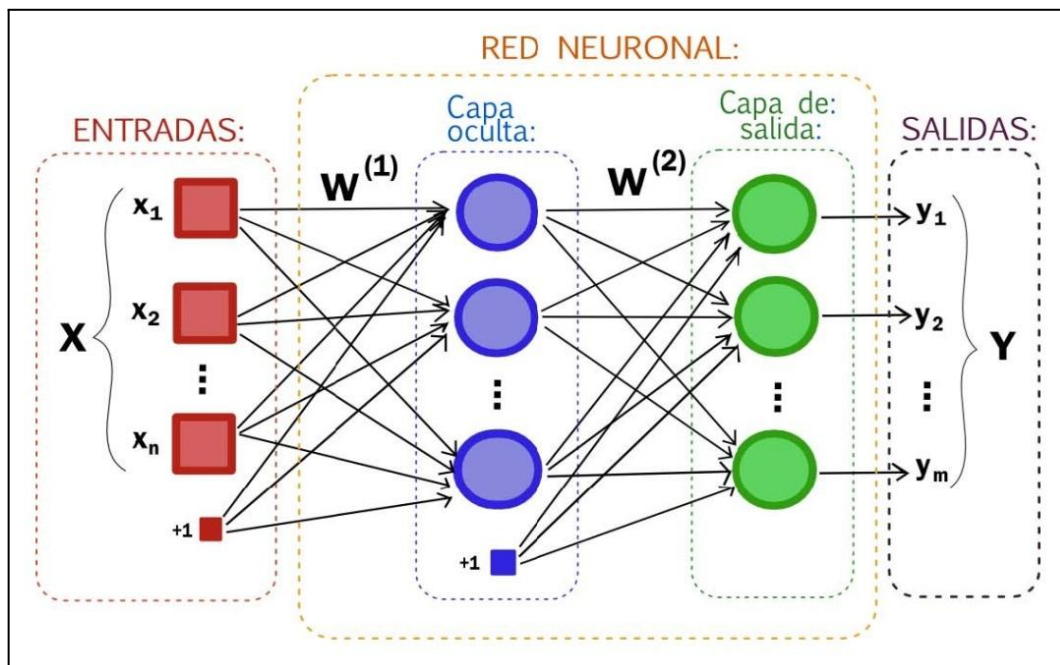


Figura 1: Esquema de una red neuronal

Funcionamiento de una Neurona:

Cada neurona realiza dos operaciones fundamentales: la suma ponderada de las entradas y la aplicación de una función de activación. La suma ponderada se calcula multiplicando cada entrada (S_n) por su peso correspondiente (α_n) y sumando los resultados. La función de activación (típicamente la función sigmoide, que tiene bajo costo computacional) introduce no linealidades en el modelo, permitiendo que la red aprenda patrones más complejos.

La salida de cada neurona respeta la ecuación $S_t = \text{sigmoid}(\sum_{n=0}^{N, si n \neq t} \alpha_n \times S_n)$ y en proceso de backpropagation se ajustan los α_n de a un paso a la vez.

Backpropagation:

La retropropagación es el proceso mediante el cual se ajustan los pesos de las conexiones para minimizar el error. El algoritmo utiliza el gradiente descendente para encontrar el mínimo de la función de pérdida. El gradiente de la función de pérdida con respecto a los pesos se calcula utilizando la regla de la cadena, y se utiliza para actualizar los pesos de las conexiones.

El proceso de aprendizaje en una red neuronal se lleva a cabo en dos fases: propagación hacia adelante y retropropagación. Durante la propagación hacia adelante, la entrada se pasa a través de la red, y se calculan las salidas de cada neurona. Estas salidas se comparan con las salidas deseadas para calcular el error, esta comparación que se hace ciclo a ciclo aplica la diferenciación discreta de los valores obtenidos contra los esperados en la última capa, y retropropaga la corrección del error responsabilizando a cada neurona conectada desde la capa final, hasta la capa de entrada.

Desde el punto de vista matemático, la retropropagación es calcular la derivada de la función error, el error es una función de N dimensiones de entrada, dimensión de salida 1, con N correspondiente a la cantidad de neuronas de cada capa, para la Figura 2 se graficó solo 2 dimensiones de los α_n y en el eje Z se gráfico la altura del "error". Al inyectar un dato a la entrada del sistema con todos los α_n seteados en algún valor aleatorio, se consigue aparecer en algún punto del plano error (coloreado en la Figura 2) para minimizar el error, lo que se hace es derivar en N-dimensiones (la operación gradiente) devuelve la dirección del máximo crecimiento del error, y configurar los α_n que vayan en la dirección contraria del crecimiento del error (para que el error en cada iteración disminuya, ya que se considera que la función error es una función continua). Como se puede deducir la retropropagación es un proceso de

ajustar la salida hacia un mínimo local, pero no asegura encontrar el mínimo global de la función error, para se inicializa los α_n en muchos valores aleatorios para encontrar el que mejor se ajuste en algunas iteraciones. además es probable que en algún eje caiga en un punto silla, que es un máximo en un eje, y un mínimo en otro eje (entonces el gradiente da 0, pero no es ni un mínimo ni un máximo). Estos puntos son fáciles de detectar porque solo se puede llegar si en la primera iteración se inicializa con un eje en un máximo, (donde la dirección de máximo crecimiento también da 0)

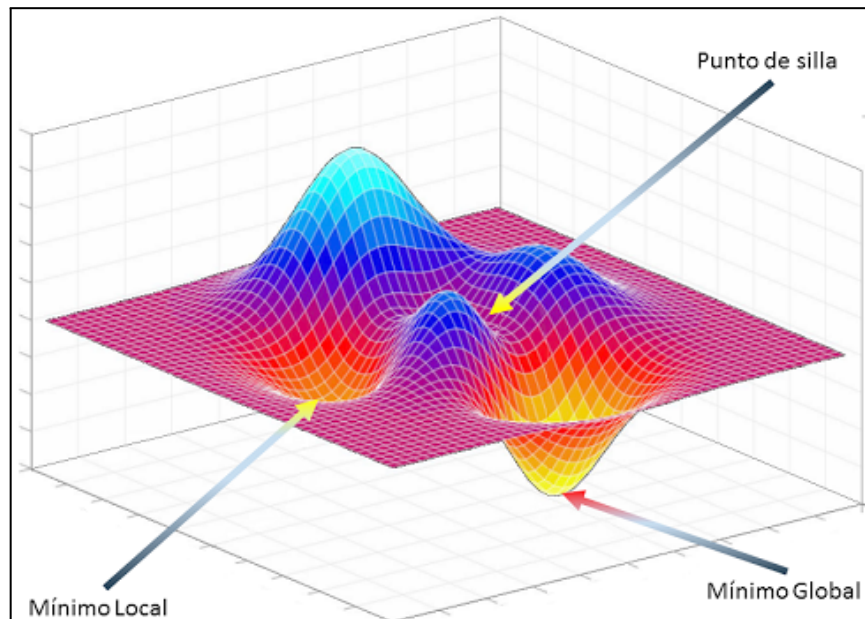


Figura 2: puntos importantes en la función error de una capa IA

El mecanismo aprendizaje

Los α_n son ajustados cíclicamente de manera que la red aproxime la ecuación de la solución intrínseca en los datos de entrenamiento lo mejor posible, ya que la ecuación que representa la RN está codificada en los α_n que se ajustan para que la red resuelva con el menor error posible los datos de entrenamiento, y por interpolación resuelve para cualquier dato de entrada. En la Figura 3 finalmente se plasma este mecanismo de aprendizaje, donde los datos de entrada se procesan desde la capa de entrada hacia la capa de salida, los resultados de la capa de salida se comparan con los resultados correctos, y se calcula la corrección del error para cada peso α_n de cada capa.

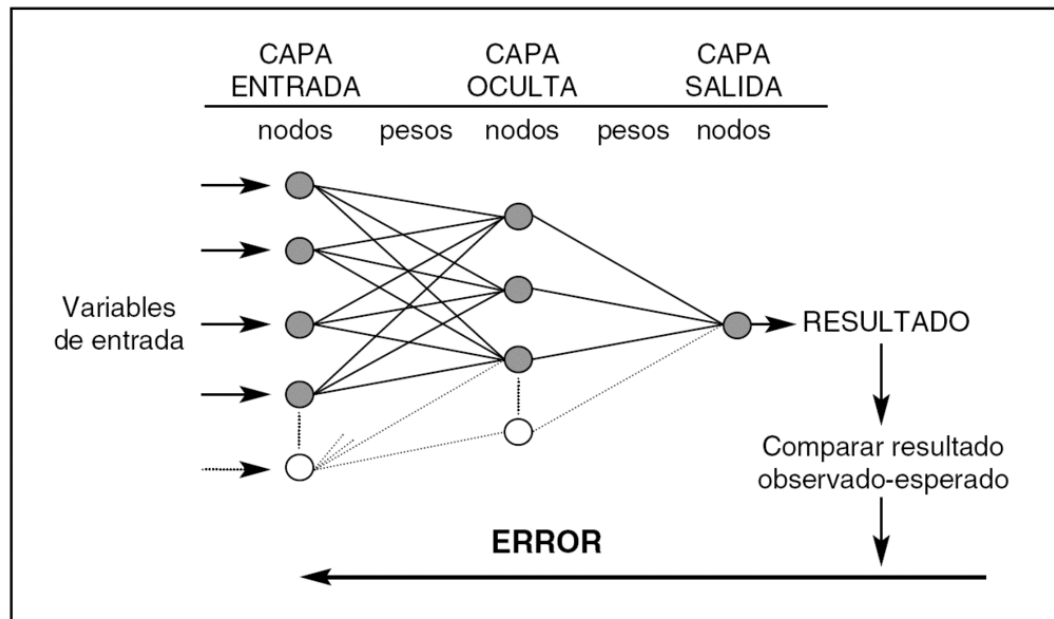


Figura 3: retropropagación ajuste del error en una RN

VOZ ACTIVA marcada con marcada con verde

Vos PASIVA marcada con amarillo

Glosario:

Red Neuronal (RN): Un modelo computacional compuesto por capas de neuronas interconectadas que se utiliza para realizar tareas de aprendizaje automático y reconocimiento de patrones.

Retropropagación (Backpropagation): Un algoritmo utilizado en el entrenamiento de redes neuronales para ajustar los pesos de las conexiones entre neuronas y minimizar la diferencia entre las predicciones del modelo y las salidas deseadas.

Capa de Entrada: La primera capa de una red neuronal que recibe las entradas del sistema.

Capas Ocultas: Capas intermedias entre la capa de entrada y la capa de salida que procesan la información de manera no lineal, permitiendo a la red aprender patrones más complejos.

Capa de Salida: La capa final de una red neuronal que produce la predicción o salida del modelo.

Descenso del Gradiente: Un algoritmo de optimización utilizado para minimizar la función de pérdida ajustando los pesos de las conexiones en la dirección opuesta al gradiente.

Neurona: Un componente fundamental de una red neuronal que realiza operaciones de suma ponderada y aplica una función de activación a sus entradas.

Función de Activación: Una función aplicada a la salida de una neurona que introduce no linealidades en el modelo, permitiendo a la red aprender patrones más complejos.

Suma Ponderada: La operación en la que se multiplican las entradas de una neurona por sus pesos correspondientes y se suman los resultados.

Función de Pérdida: Una medida que cuantifica la discrepancia entre las predicciones de la red y las salidas deseadas.

Regla de la Cadena: Una regla matemática utilizada en la retropropagación para calcular el gradiente de la función de pérdida con respecto a los pesos de las conexiones.

Derivada: La tasa de cambio instantánea de una función, utilizada en el cálculo del gradiente durante la retropropagación.

Punto Silla: Un punto en la función de error donde el gradiente es cero, pero no es ni un mínimo ni un máximo. Se deben tener en cuenta en el proceso de optimización.

Biografía:

- Algoritmo del descenso del Gradiente :
<https://es.khanacademy.org/math/multivariable-calculus/applications-of-multivariable-derivatives/optimizing-multivariable-functions/a/what-is-gradient-descent>