

Perceptrones multicapa

Un perceptrón multicapa es un tipo de red neuronal feedforward que consta de neuronas completamente conectadas con un tipo de función de activación no lineal. Se utiliza ampliamente para distinguir datos que no son separables linealmente.

Los MLP se han utilizado ampliamente en diversos campos, incluido el reconocimiento de imágenes, el procesamiento del lenguaje natural y el reconocimiento de voz, entre otros. Su flexibilidad en la arquitectura y su capacidad para aproximarse a cualquier función bajo ciertas condiciones los convierten en un componente fundamental en el aprendizaje profundo y la investigación de redes neuronales. Profundicemos en algunos de sus conceptos clave.

Capa de entrada

La capa de entrada consta de nodos o neuronas que reciben los datos de entrada iniciales. Cada neurona representa una característica o dimensión de los datos de entrada. El número de neuronas en la capa de entrada está determinado por la dimensionalidad de los datos de entrada.

capa oculta

Entre las capas de entrada y salida, puede haber una o más capas de neuronas. Cada neurona en una capa oculta recibe entradas de todas las neuronas de la capa anterior (ya sea la capa de entrada u otra capa oculta) y produce una salida que se pasa a la siguiente capa. La cantidad de capas ocultas y la cantidad de neuronas en cada capa oculta son hiperparámetros que deben determinarse durante la fase de diseño del modelo.

Capa de salida

Esta capa consta de neuronas que producen el resultado final de la red. La cantidad de neuronas en la capa de salida depende de la naturaleza de la tarea. En la clasificación binaria, puede haber una o dos neuronas según la función de activación y que represente la probabilidad de pertenecer a una clase; mientras que en las tareas de clasificación de clases múltiples, puede haber múltiples neuronas en la capa de salida.

Pesos

Las neuronas de las capas adyacentes están completamente conectadas entre sí. Cada conexión tiene un peso asociado, que determina la fuerza de la conexión. Estos pesos se aprenden durante el proceso de entrenamiento.

Neuronas de sesgo

Además de las neuronas de entrada y ocultas, cada capa (excepto la capa de entrada) generalmente incluye una neurona de polarización que proporciona una entrada constante a las neuronas de la siguiente capa. La neurona de polarización tiene su propio peso asociado a cada conexión, que también se aprende durante el entrenamiento.

La neurona de sesgo cambia efectivamente la función de activación de las neuronas en la capa siguiente, lo que permite a la red aprender un desplazamiento o sesgo en el límite de decisión. Al ajustar los pesos conectados a la neurona de polarización, el MLP puede aprender a controlar el umbral de activación y adaptarse mejor a los datos de entrenamiento.

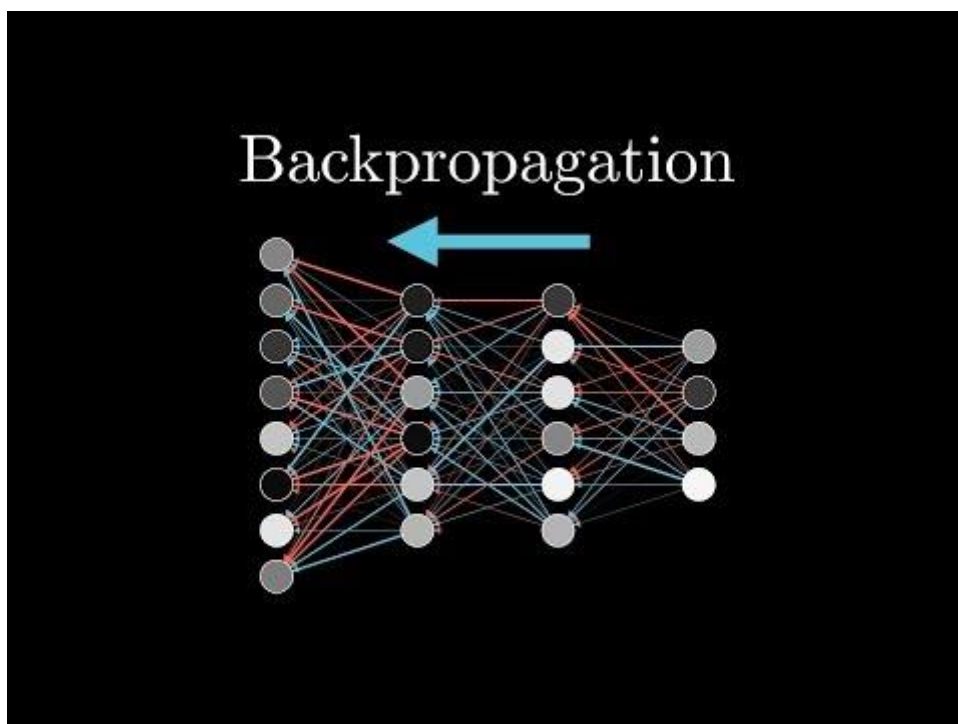
Nota: Es importante señalar que en el contexto de las MLP, *bias* puede referirse a dos conceptos relacionados pero distintos: sesgo como término general en aprendizaje automático y neurona de sesgo (definida anteriormente). En el aprendizaje automático general, el sesgo se refiere al error introducido al aproximar un problema del mundo real con un modelo simplificado. El sesgo mide qué tan bien el modelo puede capturar los patrones subyacentes en los datos. Un sesgo alto indica que el modelo es demasiado simplista y puede no ajustarse a los datos, mientras que un sesgo bajo sugiere que el modelo está captando bien los patrones subyacentes.

Función de activación

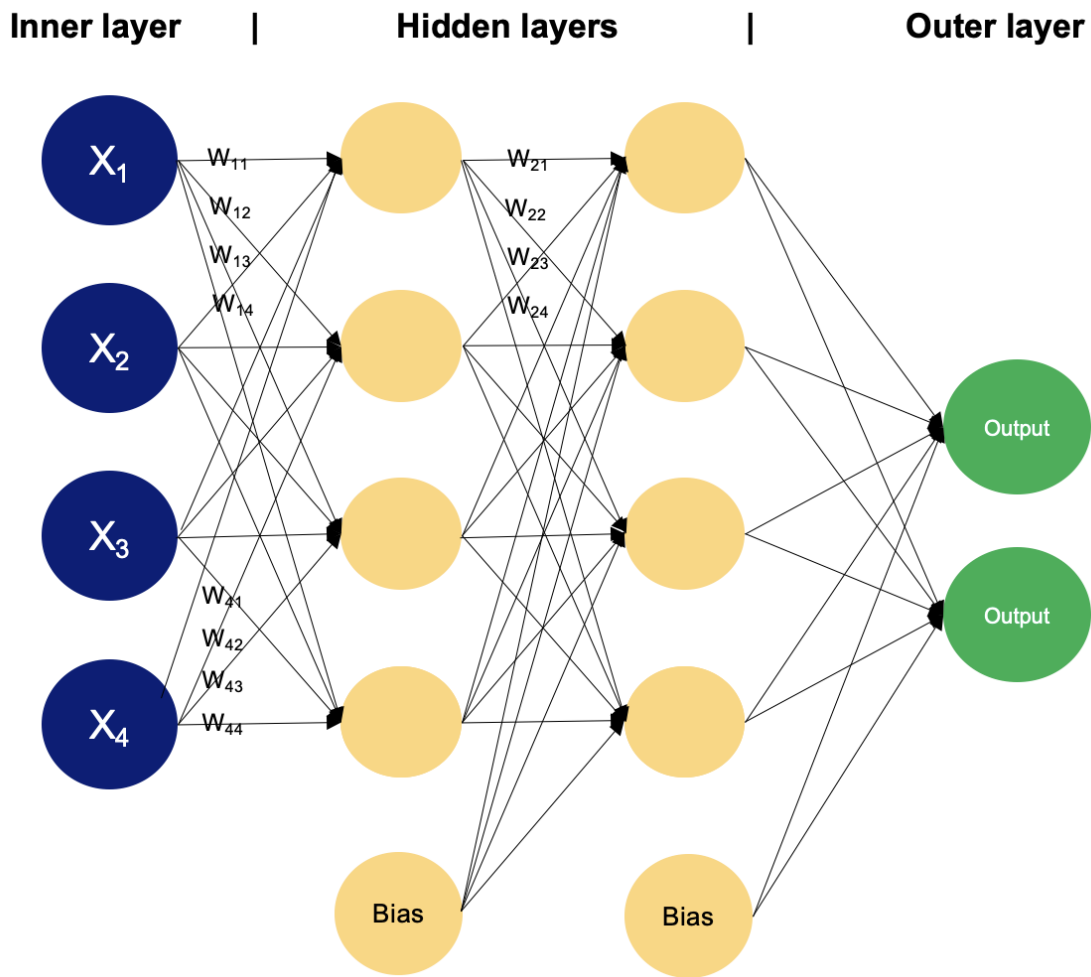
Normalmente, cada neurona en las capas ocultas y en la capa de salida aplica una función de activación a su suma ponderada de entradas. Las funciones de activación comunes incluyen sigmoide, tanh, ReLU (Unidad lineal rectificada) y softmax. Estas funciones introducen no linealidad en la red, lo que le permite aprender patrones complejos en los datos.

Entrenamiento con retropropagación

Los MLP se entrenan utilizando el algoritmo de retropropagación, que calcula los gradientes de una función de pérdida con respecto a los parámetros del modelo y actualiza los parámetros de forma iterativa para minimizar la pérdida.



Funcionamiento de un perceptrón multicapa: capa por capa



Ejemplo de un MLP que tiene dos capas ocultas

En un perceptrón multicapa, las neuronas procesan información paso a paso, realizando cálculos que implican sumas ponderadas y transformaciones no lineales. Caminemos capa por capa para ver la magia que hay en su interior.

Capa de entrada

- La capa de entrada de un MLP recibe datos de entrada, que podrían ser características extraídas de las muestras de entrada en un conjunto de datos. Cada neurona en la capa de entrada representa una característica.
- Las neuronas de la capa de entrada no realizan ningún cálculo; simplemente pasan los valores de entrada a las neuronas de la primera capa oculta.

Capas ocultas

- Las capas ocultas de un MLP consisten en neuronas interconectadas que realizan cálculos sobre los datos de entrada.
- Cada neurona de una capa oculta recibe información de todas las neuronas de la capa anterior. Las entradas se multiplican por los pesos correspondientes, denotados como w . Los pesos determinan cuánta influencia tiene la entrada de una neurona sobre la salida de otra.
- Además de los pesos, cada neurona de la capa oculta tiene un sesgo asociado, denominado b . El sesgo proporciona una entrada adicional a la neurona, permitiéndole ajustar su umbral de salida. Al igual que las pesas, los prejuicios se aprenden durante el entrenamiento.
- Para cada neurona en una capa oculta o en la capa de salida, se calcula la suma ponderada de sus entradas. Esto implica multiplicar cada insumo por su peso correspondiente, sumar estos productos y sumar el sesgo:

$$\text{Weighted Sum} = \sum_{i=1}^n (w_i * x_i) + b$$

Dónde n es el número total de conexiones de entrada, w_i es el peso de la i -ésima entrada, y x_i es el i -ésimo valor de entrada.

- Luego, la suma ponderada pasa a través de una función de activación, denotada como f . La función de activación introduce no linealidad en la red, lo que le permite aprender y representar relaciones complejas en los datos. La función de activación determina el rango de salida de la neurona y su comportamiento en respuesta a diferentes valores de entrada. La elección de la función de activación depende de la naturaleza de la tarea y de las propiedades deseadas de la red.

Capa de salida

- La capa de salida de un MLP produce las predicciones o resultados finales de la red. La cantidad de neuronas en la capa de salida depende de la tarea que se realiza (por ejemplo, clasificación binaria, clasificación multiclase, regresión).
- Cada neurona de la capa de salida recibe información de las neuronas de la última capa oculta y aplica una función de activación. Esta función de activación suele ser diferente de las utilizadas en las capas ocultas y produce el valor de salida final o predicción.

Durante el proceso de entrenamiento, la red aprende a ajustar los pesos asociados con las entradas de cada neurona para minimizar la discrepancia entre las salidas predichas y los valores objetivo verdaderos en los datos de entrenamiento. Al ajustar los pesos y aprender las funciones de activación apropiadas, la red aprende a aproximar patrones y relaciones complejos en los datos, lo que le permite hacer predicciones precisas sobre muestras nuevas e invisibles.