

Redes Neuronales en Clasificación

Daniel Stiven Martinez

2023-06-05

Sección 6: Arquitectura, aplicación y avances recientes de las redes neuronales

Arquitectura

La arquitectura de las redes neuronales se compone de una capa de entrada, salida y oculta. Las propias redes neuronales, o redes neuronales artificiales (ANN), son un subconjunto de aprendizaje automático diseñado para imitar la potencia de procesamiento del cerebro humano. Las redes neuronales funcionan pasando datos a través de las capas de una neurona artificial los cuales se transforman internamente mediante una función de activación y como ultimo nos entregan una salida con la cual se toma una decisión.

Gráficamente una red neuronal tiene una estructura como sigue

Componentes principales de una red neuronal

Como acabamos de mencionar una red neuronal se componen de una capa entrada , una capa oculta y una de salida, sin embargo internamente hay una manipulación de los datos mediante una transformación de los mismos.

Para poder describir las componentes de la red neuronal, recordemos su expresión matemática

$$\psi(x) = c_0 + \sum_{i=1}^d c_i \sigma(\psi_i(x)),$$

a partir de esta tenemos que las componentes que conforman una red neuronal son las siguientes

- Entrada : Variable independiente x , que a fines prácticos son los datos que se le dan a modelo con fines de aprendizaje y entrenamiento
- Pesos: Corresponde a las constantes c_i las cuales nos ayudan a organizar las variables según su importancia y contribución al modelo que se este estudiando.
- Función de transferencia: Corresponde al procedimiento de sumar todas las entradas transformadas con el fin de combinarlas en una sola variable de salida, específicamente es la suma que aparece en la expresión matemática.
- Función sigmoide (σ) : En la literatura y contextos de redes neuronales conocida como función de activación, es la que nos ayuda a decidir cuando si o no una neurona debe ser activada, en base a la importancia de las entradas que reciba la neurona. Como recordatorio estas funciones tienen la característica de que $\lim_{x \rightarrow \infty} \sigma(x) = 1$ y $\lim_{x \rightarrow -\infty} \sigma(x) = -1$

Tipos de arquitecturas en redes neuronales

En el contexto de aprendizaje de maquina, las redes neuronales son una forma eficiente para solucionar problemas de la vida diaria, como puede ser en clasificación. Las redes neuronales nos ofrecen respuestas muy bastante precisas en lo que a solución de problemas se refiere, para poder obtener esta precisión se cuenta con distintos tipos de redes que son especificas para problemas concretos, ya que seleccionando la red adecuada

podemos incrementar la eficiencia en los resultados que estemos buscando. Entre las redes neuronales mas comunes tenemos

Redes neuronales estándar:

Estas son las redes neuronales mas básicas entre ellas tenemos:

- **Perceptron simple:**

El discriminante lineal o perceptron simple es la red neuronal mas sencilla en donde no tenemos capas ocultas, en el perceptron los datos(input) son transformados de manera lineal mediante una ponderación de manera que obtenemos una respuesta(output), brevemente la idea detrás de esto es que tomamos una decisión

$$\phi(x) = \begin{cases} 0 & \text{si } \psi(x) \geq 1/2 \\ 1 & \text{e.o.c} \end{cases}$$

basados en la combinación lineal

$$\psi(x) = c_0 + \sum_{i=1}^d c_i x_i,$$

donde como mencionamos en un inicio los c_i s son los pesos, c_0 es el sesgo y $x = (x_1, \dots, x_d)$.

- **Redes neuronales prealimentadas (FNN):**

Este tipo de redes es de las primeras en donde consideramos capas ocultas son perceptron con capa oculta, brevemente la idea detrás de este tipo de redes es que ahora la decisión que se toma depende de

$$\psi(x) = c_0 + \sum_{i=1}^d c_i \sigma(\psi_i(x))$$

donde los c_i son igual a como se mencionaron antes y cada $\psi_i(x)$ es de la forma

$$\psi_i(x) = b_i + \sum_{j=1}^d a_{ij} x_j$$

En este caso de perceptron con una capa oculta decimos que hay k neuronas ocultas y la salida de la i -ésima neurona oculta es $u_i = \sigma(\psi_i(x))$. De modo que podemos reescribir a $\psi(x)$ como

$$\psi(x) = c_0 + \sum_{i=1}^k c_i u_i$$

De igual forma a como consideramos una capa oculta las redes neuronales prealimentadas, se puede considerar mas capas ocultas en la red neuronal, veamos el caso con dos capas ocultas y casos generales con mas capas siguen la misma idea.

Para un perceptron con dos capas ocultas tenemos

$$\psi_i(x) = c_0 + \sum_{j=1}^l c_j z_j,$$

donde

$$z_i = \sigma \left(d_{i0} + \sum_{j=1}^k d_{ij} u_j \right), \quad 1 \leq i \leq l$$

y

$$u_j = \sigma \left(b_j + \sum_{i=1}^k a_{ij} x_i \right), \quad 1 \leq j \leq k$$

donde los d_{ij} , b_j y a_{ij} son constantes. En este caso concreto contamos con k neuronas ocultas en la primera capa y l neuronas ocultas para la segunda capa.

Redes neuronales recurrentes (RNN)

Las redes neuronales recurrentes (RNN) son un tipo de redes muy útiles, las cuales recuerdan las predicciones aprendidas previamente para ayudar a hacer predicciones futuras con precisión (tienen memoria) el esquema de este tipo de redes se puede observar en la imagen anterior. Este tipo de redes son muy usados en modelos de Deep learning de problemas ordinales o temporales, como pueden ser la traducción idiomas o el reconocimiento de voz (asistente como Alexa)

La como acabamos de mencionar las RNN recuerdan las predicciones que han hecho para mejorar la precisión un ejemplo de aplicación de este tipo de redes es el de predicción de texto, como por ejemplo cuando escribes una oración para enviárselo a un amigo, el celular aprende de las palabras que has usado y en el orden en que lo has hecho para hacerte sugerencias y escribir mas rápido, como puede ser en la oración **¿Como te ha ido?**, si empiezas con la palabra **como**, tu teclado te ira sugiriendo **te**, **ha** y **ido**, en incluso como va aprendiendo puede predecir que la palabra siguiente sera **hoy**.

- **Redes con memoria a corto y largo plazo (LSTM)**

Las redes neuronales con memoria a corto y largo plazo (LSTM por sus siglas en ingles) son un tipo especial de redes recurrentes. La característica principal de las redes recurrentes es que la información puede persistir introduciendo ciclos en la red, por lo que, básicamente, pueden “recordar” estados previos y utilizar esta información para predecir la salida siguiente. Esta característica las hace muy adecuadas para manejar datos que provienen de series de tiempo. Mientras las redes recurrentes estándar pueden modelar dependencias a corto plazo (es decir, relaciones cercanas en la serie cronológica), las LSTM pueden aprender dependencias largas, por lo que se podría decir que tienen una memoria a más largo plazo

- **Redes neuronales convolucionales**
 - AlexNet
 - Redes neuronales de capsula (CNN)
- **Redes generativas antagónicas (GAN)**
- **Redes neuronales transformer**

Implementación y ejemplos