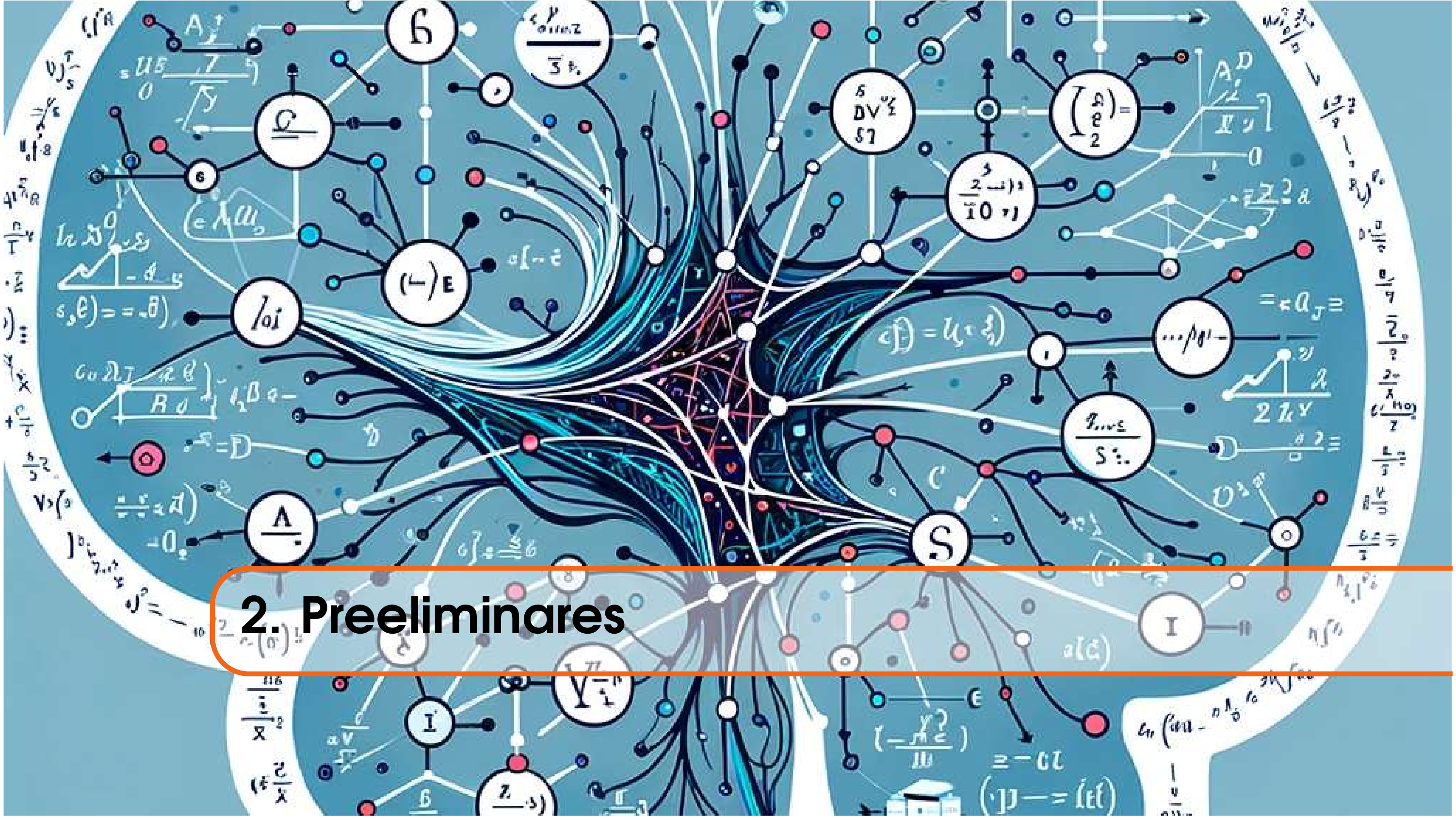
Parte Uno: Fundamentos



2. Preeliminares

2.1 Introducción

En este capítulo, nos adentraremos en los cimientos matemáticos que son esenciales para comprender y trabajar efectivamente con redes neuronales. A menudo, el poder y la eficacia de las redes neuronales en tareas de aprendizaje automático pueden parecer casi mágicos. Sin embargo, este "truco de magia" se basa firmemente en principios matemáticos bien establecidos. Aquí, exploraremos estos principios con el fin de desmitificar cómo las redes neuronales aprenden a partir de datos y realizan predicciones o clasificaciones asombrosamente precisas.

Los temas incluidos en este capítulo son fundamentales en la ciencia de datos y la inteligencia artificial: álgebra lineal, cálculo diferencial, probabilidad y estadística. Cada uno de estos temas contribuye de manera única al funcionamiento de las redes neuronales.

La álgebra Lineal es el lenguaje en el que se expresan las redes neuronales. Entender conceptos como vectores, matrices, y operaciones como la multiplicación de matrices, es esencial para comprender cómo las redes procesan y transforman los datos.

Por otra parte, las redes neuronales aprenden ajustando sus parámetros para mejorar su rendimiento. Este proceso de aprendizaje es guiado por el cálculo diferencial, en particular por conceptos como derivadas y gradientes. Comprender estos conceptos es crucial para entender cómo se optimizan las redes durante el entrenamiento.

En otro canal igual de importante, podemos ver que las redes neuronales, en su núcleo, realizan tareas de inferencia estadística. Entender la teoría de probabilidad y las técnicas estadísticas nos permite comprender cómo las redes neuronales modelan la incertidumbre y hacen predicciones basadas en datos. Pero sobre todo nos dan las herramientas para desmitificar las cajas negras y comprender a las redes como procesos estocásticos que modelan problemas del mundo real.

2.2 Algebra lineal

El álgebra lineal es una rama de las matemáticas que estudia conceptos tales como vectores, matrices, espacio dual, sistemas de ecuaciones lineales y en su enfoque de manera más formal, espacios vectoriales y sus transformaciones lineales. Esta rama de las matemáticas es esencial en una amplia gama de campos, incluyendo la ciencia, la ingeniería, la economía, las finanzas y en especial la inteligencia artificial.

2.2.1 Vectores y espacios vectoriales

En matemáticas, un vector es un objeto que tiene magnitud y dirección. Se puede representar gráficamente como un segmento de recta orientado, o de forma algebraica como una tupla de números.

$$\vec{v} = \begin{bmatrix} 1 \\ 2 \\ \vdots \\ n \end{bmatrix}$$

La notación para describir en que espacio se encuentra un vector, v, puede ser algún conjunto de números como los reales, $\mathbb{R}$, y este puede tener un número n de dimensiones, teniendo así:

$$\vec{v} \in \mathbb{R}^n$$

La magnitud de un vector es su longitud, y se puede medir en unidades de longitud, como metros, kilómetros o pies. La dirección de un vector es la dirección en la que está apuntando, y se puede especificar por un ángulo o por un punto cardinal.

Los vectores se pueden sumar, restar, multiplicar por escalares.

- La suma de dos vectores es otro vector que apunta en la misma dirección que los dos vectores originales, pero con una magnitud que es la suma de las magnitudes de los dos vectores originales.
- La resta de dos vectores es otro vector que apunta en la dirección opuesta a la de los dos vectores originales, pero con una magnitud que es la diferencia de las magnitudes de los dos vectores originales.
- La multiplicación de un vector por un escalar es otro vector que tiene la misma dirección que el vector original, pero con una magnitud que es el producto del escalar por la magnitud del vector original.
- El producto escalar de dos vectores es un número real que representa el producto de las magnitudes de los vectores originales y el coseno del ángulo entre ellos.

Suma de vectores

La suma de dos vectores se puede calcular algebraicamente utilizando las componentes de los vectores. Si los vectores tienen componentes (x_1, y_1) y (x_2, y_2), entonces su suma es el vector $(x_1 + x_2, y_1 + y_2)$.

Resta de vectores

La resta de dos vectores se puede calcular algebraicamente utilizando las componentes de los vectores. Si los vectores tienen componentes (x_1, y_1) y (x_2, y_2), entonces su resta es el vector $(x_1 - x_2, y_1 - y_2)$.

Multiplicación de un vector por un escalar

La multiplicación de un vector por un escalar, a, también se puede calcular algebraicamente utilizando las componentes del vector. Si el vector tiene componentes (x, y) y el escalar es a, entonces el producto del vector por el escalar es el vector (ax, ay).

Producto escalar de dos vectores

Si los vectores tienen componentes (x_1,y_1) y (x_2,y_2), entonces su producto escalar es:

$$(x_1,y_1)\cdot(x_2,y_2) = x_1\cdot x_2 + y_1\cdot y_2$$

2.2.2 Matrices

Las matrices son arreglos de filas y columnas, útiles para hacer operaciones numéricas en aplicaciones como en las matemáticas. Es común que este arreglo sea de m-filas y n-columnas, creando así un arreglo con tamaño $m\times n$:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \tag{2.1}$$

misma que es posible reescribir como $A = [a_{ij}]$.

Suma de matrices

Si tenemos dos matrices A y B de $m\times n$, el resultado de la suma de ambas será una matriz $m\times n$, $A+B$:

$$A+B = (a_{ij}+b_{ij}) = \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \cdots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \cdots & a_{2n}+b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & \cdots & a_{mn}+b_{mn} \end{bmatrix} \tag{2.2}$$

es importante recordar que para realizar la operación de suma ambas matrices deben tener las mismas dimensiones de lo contrario será imposible realizar esta operación ya que como se observa la suma se realiza elemento a elemento.

Producto

El producto entre dos matrices solo es posible si satisface una condición y esta requiere que la primera matriz tenga un tamaño $m\times n$ y la segunda $n\times p$ (nótese que ambas comparten en sus dimensiones el valor n), dando como resultado a una nueva matriz con dimensiones $m\times p$; es decir, si se tienen dos matrices AyB con dimensiones $2\times 3y3\times 4$ respectivamente, el resultado será una nueva matriz cuyas dimensiones serán 2×4.

$$AB = (a_{ij}b_{ij}) = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mp} \end{bmatrix} \tag{2.3}$$

y para calcular los elementos de cada una de las filas desplazaremos de acuerdo al elemento que estemos calculado de la matriz resultante, dado por el subíndice de c_{ij} donde i será la fila correspondiente de A y j la columna correspondiente de B, para un primer elemento tendremos:

$$c_{11} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \end{bmatrix} \begin{bmatrix} b_{11} \\ b_{21} \\ \vdots \\ b_{n1} \end{bmatrix} = a_{11}b_{11} + a_{12}b_{21} + \cdots + a_{1n}b_{n1} \tag{2.4}$$

y de esta misma forma se calcula cada uno de los elementos correspondientes de la matriz resultante de la operación de producto.

Ejercicios

Dadas las siguientes matrices:

$$A = \begin{bmatrix} -8 & -5 & -7 \\ -3 & 9 & 5 \end{bmatrix}, B = \begin{bmatrix} 4 & -5 \\ -8 & -3 \end{bmatrix}, C = \begin{bmatrix} 7 & 5 & 8 \\ -6 & 5 & -4 \end{bmatrix},$$

$$D = \begin{bmatrix} -2 \\ 4 \end{bmatrix}, E = \begin{bmatrix} 9 & 0 & 8 \\ -3 & 8 & 4 \\ -8 & -9 & 5 \end{bmatrix}, F = \begin{bmatrix} 10 & -5 \\ 6 & -4 \\ 0 & -15 \end{bmatrix},$$

$$G = \begin{bmatrix} 1 & 6 \\ 9 & 2 \end{bmatrix}, H = \begin{bmatrix} 2 & -8 & -1 \\ -8 & 4 & 11 \\ 6 & 5 & 4 \end{bmatrix}$$

1.- Calcular:

1. A+C
2. C-A
3. EH
4. H+E
5. GD
6. DG
7. GB
8. AE
9. CF
10. FD

2.- Hallar la transpuesta de todas las matrices (A-H).

Matriz inversa

La inversa de una matriz cuadrada, A, es otra matriz cuadrada, A^{-1}, que satisface la siguiente ecuación $AA^{-1} = I$, siendo I la matriz identidad; es decir una matriz inversa es aquella que multiplicada por la matriz original da como resultado a la matriz identidad. Para que una matriz tenga inversa el determinante de esta debe ser distinto de cero.

Algunos métodos para calcular la matriz inversa son:

- Método del adjunto
- Método de Gauss-Jordad
- Método de Cramer

Matriz identidad

La matriz identidad, I, es una matriz de dimensiones $n \times n$, es decir es una matriz cuadrada en la que sus elementos de la diagonal principal tienen como valor unos y el resto tienen valor cero.

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

La inversa de la matriz identidad siempre será la misma matriz identidad.

Determinante

El determinante de una matriz cuadrada, A, es un numero real y este es de utilidad para calcular la distancia entre dos vectores, determinar si es que una matriz es invertible, resolver sistemas de ecuaciones lineales. Y lo podemos encontrar representado como:

$$det(A) = |A|$$

Algunos métodos para calcular el determinante son: la regla de Sarrus, Teorema de Laplace o método de cofactores

2.3 Cálculo diferencial

Si $y = f(x)$ entonces la derivada esta definida como $f'(x) = \lim_{h \to 0} \dfrac{f(x+h) - f(x)}{h}$.

Si $y = f(x)$ entonces las siguientes notaciones para la derivada son válidas.

$$f'(x) = y' = \frac{df}{dx} = \frac{dy}{dx} = \frac{d}{dx}(f(x))$$

Si $y = f(x)$ entonces, podemos interpretar la derivada como,

2.3.1 Ejercicios

1. Determine la derivada de las siguientes funciones.

2.3.2 Lecturas recomendadas

- Capítulos 2 y 3. Stewart, James. Cálculo De Una Variable: Trascendentes Tempranas. 4a. ed. México D.F.: Thomson, 2001.

2.4 Probabilidad y estadística

2.4.1 Variable aleatoria

Una variable aleatoria es aquella variable que puede tomar diferentes valores aleatoriamente y esta es una función que asigna un valor numérico a cada resultado de un experimento aleatorio. Las variables aleatorias se pueden clasificar en dos tipos:

- Variables aleatorias continuas: Son variables aleatorias que pueden tomar cualquier valor dentro de un intervalo.
- Variables aleatorias discretas: Son variables que pueden tomar valores de un numero finito o infinito numerable.

2.4.2 Distribución de probabilidad

La distribución de probabilidad es una función que asigna una probabilidad a cada valor posible de una variable aleatoria; es decir, describe como se distribuyen los valores de una variable aleatoria.

Distribución de probabilidad discreta

A la distribución de probabilidad sobre variables aleatorias discretas se le puede describir haciendo uso de una función de masa de probabilidad (PMF, por sus siglas en ingles). Esta función se representa con la letra P y es la encargada de mapear de un estado de la variable aleatoria a la probabilidad de que esa variable toma en ese estado, la probabilidad de una variable x es denotada por $P(x)$. Una PMF debe satisfacer las siguientes condiciones:

- El dominio de P debe ser el conjunto de todos los posible estados de x.
- Un evento imposible tiene probabilidad cero y ningun estado puede ser menos probable que eso.
- La suma de las probabilidades de todos los estados es igual a uno. Algunos se le refieren como ser normalizada.

Distribución de probabilidad continua

A la distribución de probabilidad sobre variables aleatorias continuas se le puede describir haciendo uso de una función de densidad de probabilidad, p,(PDF, por sus siglas en ingles). Esta funcion no describe la probabilidad de un estado directamente, sino la probabilidad de que esta caiga en una region infinitesimal con volumen δx es dada por $p(x)\delta x$. Una PDF debe satisfacer las siguientes condiciones:

- El dominio de p debe ser el conjunto de todos los posible estados de x.
- Todos los estados tienen una probabilidad mayor o igual a cero y no en necesario que $p(x)$ sea menor o igual a uno.
- Su integral en todo el rango de valores posibles de la variable aleatoria debe ser igual a uno

Distribución uniforme: esta distribución hace que cada uno de los estados tenga una probabilidad igual.

2.4.3 Probabilidad marginal

Algunas veces se conoce la distribución de probabilidad sobre todo el conjunto de estados y necesitamos conocer la distribución de probabilidad solo sobre un subconjunto de estados. A la probabilidad sobre un subconjunto se le conoce como distribución de probabilidad marginal. En otras palabras, la probabilidad marginal es la probabilidad de que ocurra un evento específico, sin tener en cuenta la probabilidad de que ocurran otros eventos en conjunto con él. La probabilidad marginal se puede calcular sumando, o integrando el volumen (según sea el caso), las probabilidades conjuntas de todos los eventos que pueden ocurrir en conjunto con el evento específico.