Introducción al álgebra y el cálculo

Conceptos básicos algebraicos para el Deep Learning

Los conceptos de Álgebra Lineal son cruciales para entender la teoría detrás del Aprendizaje Automático (Machine Learning), especialmente para el Aprendizaje Profundo (Deep Learning). Ofrecen una mejor información para entender cómo los algoritmos realmente funcionan bajo el entramado de códigos de Python o R, lo que te permite tomar mejores decisiones como experto en ciencia de datos. Entonces, si realmente quieres ser un profesional en este campo, no puedes dejar de dominar algunos de sus conceptos. Esta parte del módulo os dará una introducción a los conceptos más importantes de Álgebra Lineal que se utilizan en el Machine Learning.

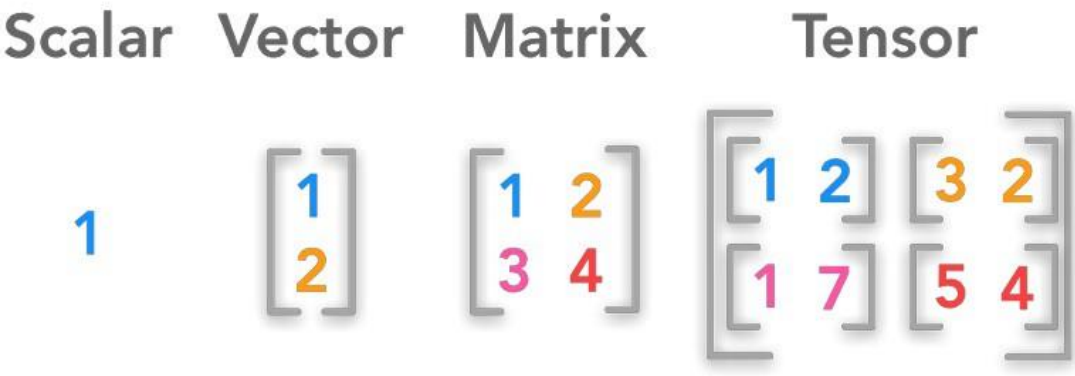
Introducción

El álgebra lineal permite modelar fenómenos naturales y calcularlos de manera eficiente. Sin embargo, muchos informáticos no tienen mucha experiencia con ella. El álgebra lineal también es fundamental para casi todas las áreas de las matemáticas, como la geometría, el cálculo, la estadística y el análisis funcional. Sus conceptos son un requisito previo fundamental para comprender la teoría detrás del Machine Learning, especialmente si está trabajando con algoritmos de Deep Learning. Quizás no es estrictamente necesario que entiendas el Álgebra Lineal antes de comenzar con el Aprendizaje Automático, pero en algún momento, es posible que desees comprender mejor cómo los diferentes algoritmos de Aprendizaje Automático realmente funcionan bajo los códigos prediseñados. Esto le ayudará a tomar mejores decisiones durante el desarrollo de un sistema de aprendizaje automático. Entonces, si realmente quieres ser un profesional en este campo, tendrás que dominar las partes del Álgebra Lineal que son importantes para el Aprendizaje Automático. En el Álgebra Lineal, los datos se representan mediante ecuaciones lineales, que se presentan en forma de matrices y vectores. Por lo tanto, se trata principalmente de matrices y vectores en lugar de escalar (cubriremos estos términos en la siguiente sección). Cuando se tienen las bibliotecas adecuadas de R, se pueden realizar operaciones con vectores y matrices complejas muy fácilmente con solo unas pocas líneas de código.

Elementos algebraicos

Los elementos algebraicos básicos son:

- Escalar
- Vector
- Matriz
- Tensor



Fuente: Medium

Prácticamente todo en matemáticas se puede describir de forma algebráica:

$$\begin{cases} 3x + 5y = 11 \\ x + 4y = 6 \end{cases}$$

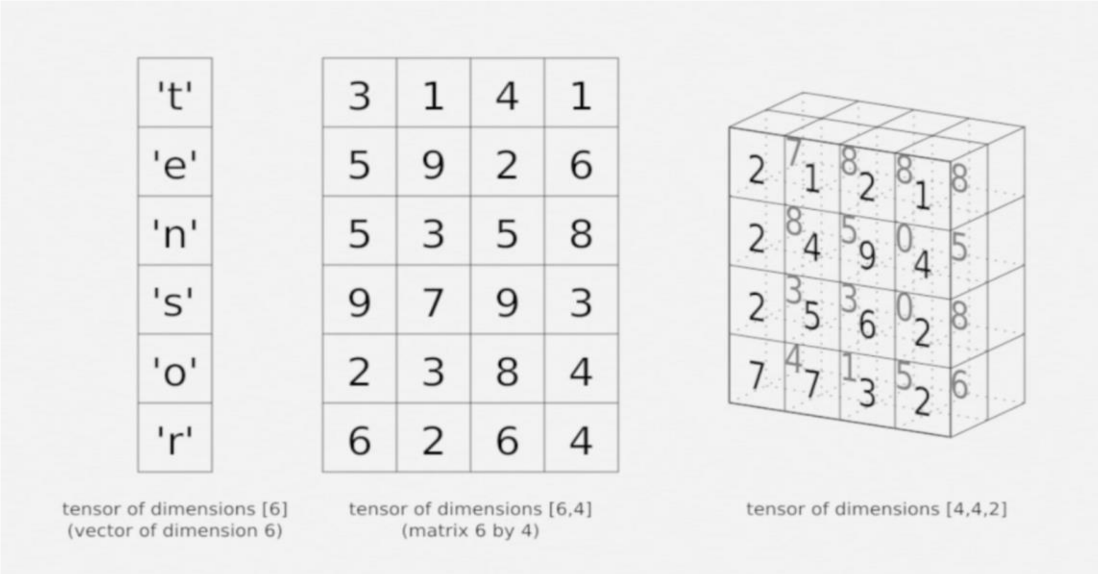
$$\downarrow$$

$$\begin{bmatrix} 3 & 5 \\ 1 & 4 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 11 \\ 6 \end{bmatrix}$$

Fuente: Medium

Parémonos un poco más en el concepto de **tensor**.

Se puede pensar en un Tensor como una matriz de números, dispuestos en una cuadrícula regular, con un número variable de ejes. Un tensor tiene tres índices, donde el primero apunta a la fila, el segundo a la columna y el tercero al eje. Por ejemplo, T232 apunta a la segunda fila, la tercera columna y el segundo eje. Esto se refiere al valor 0 en el Tensor derecho en el siguiente gráfico:

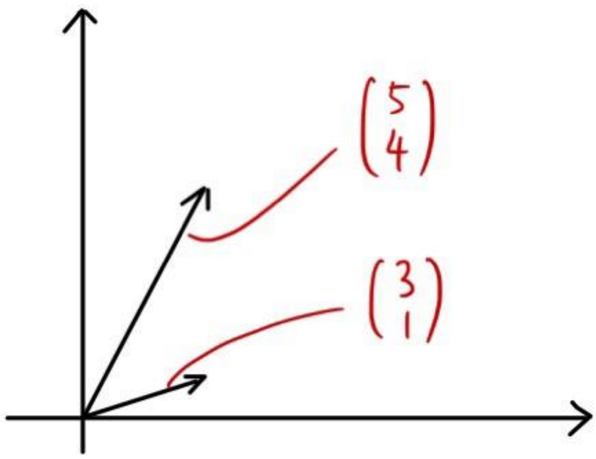


Así, tensor es el término más general para todos estos conceptos anteriores porque un Tensor es una matriz multidimensional y puede ser un Vector y una Matriz, dependiendo del número de índices que tiene. Por ejemplo, un Tensor de primer orden sería un Vector (1 índice). Un Tensor de segundo orden es una Matriz (2 índices) y los Tensores de tercer orden (3 índices) y superiores se denominan Tensores de orden superior (3 o más índices).

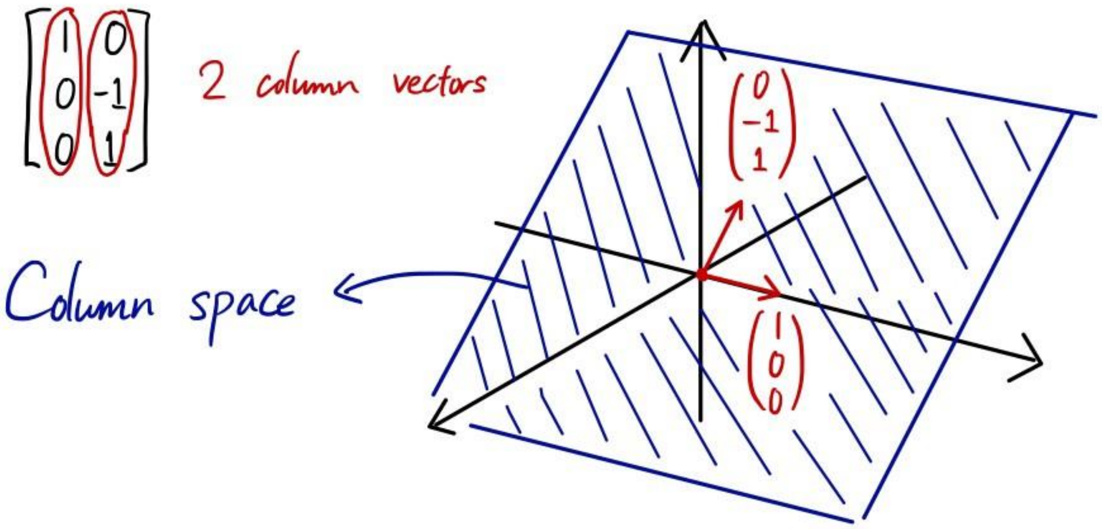
Consideraciones sobre las dimensiones de vectores y matrices

Operaciones con vectores

¡No perdamos de vista el concepto geométrico de vector!



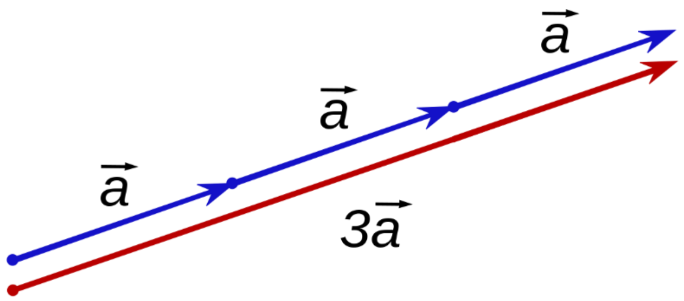
Fuente: Medium



Fuente: Medium

Multiplicación escalar-vector

Las nuevas coordenadas resultan de multiplicar las originales por el escalar



Fuente: Wikipedia

Consideremos el vector **a**= (3, 5). Si lo multiplicamos por los escalares 2, -4 y 1/3, tendríamos:

2**a**= (6, 10)

-4**a**= (-12, -20)

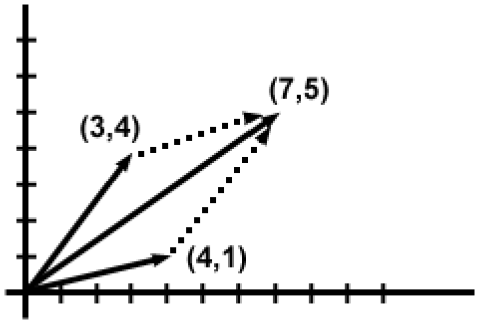
(1/3)**a**= (1/5, 5/3)

Suma y sustracción de vectores

Las nuevas coordenadas se obtienen de la suma (resta) entre las coordenadas correspondientes a ambos vectores.

Si queremos sumar el vector **a**=(3, 5) y **b**=(4, 1), el resultado será **a**+**b**=(3+4, 5+1)=(7,6).

Por su lado, la resta será **a**-**b**=(3-4, 5-1)=(-1,4).



Fuente: SparkNotes

Producto vectorial

Este concepto es un poco más elaborado. Gráficamente, de forma geométrica está relacionado con el coseno del ángulo que forman los vectores que se multiplican. Sin entrar en estos detalles, podemos decir que el producto vectorial, o producto cruzado entre dos vectores es igual a la multiplicación de las coordenadas y su posterior adición.

De esta forma, si tenemos dos vectores de iguales dimensiones (tres en este ejemplo):

a=(a_1, a_2, a_3); **b**=(b_1, b_2, b_3)

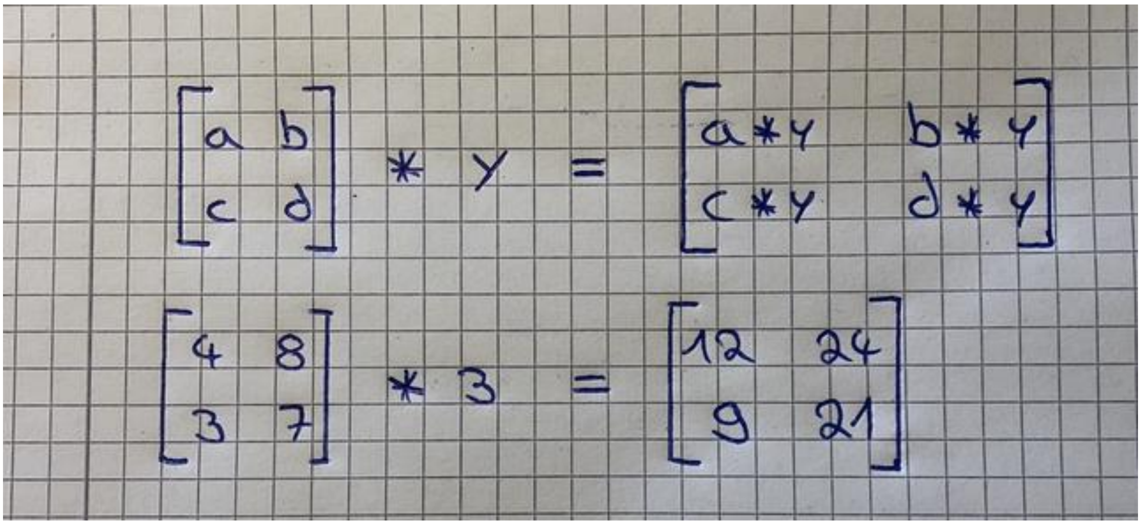
El producto vectorial **a** x **b** = (a_1 b_1) + (a_2 b_2) + (a_3 b_3)

Que es un número escalar

Operaciones con matrices

Operaciones con escalar: Multiplicación

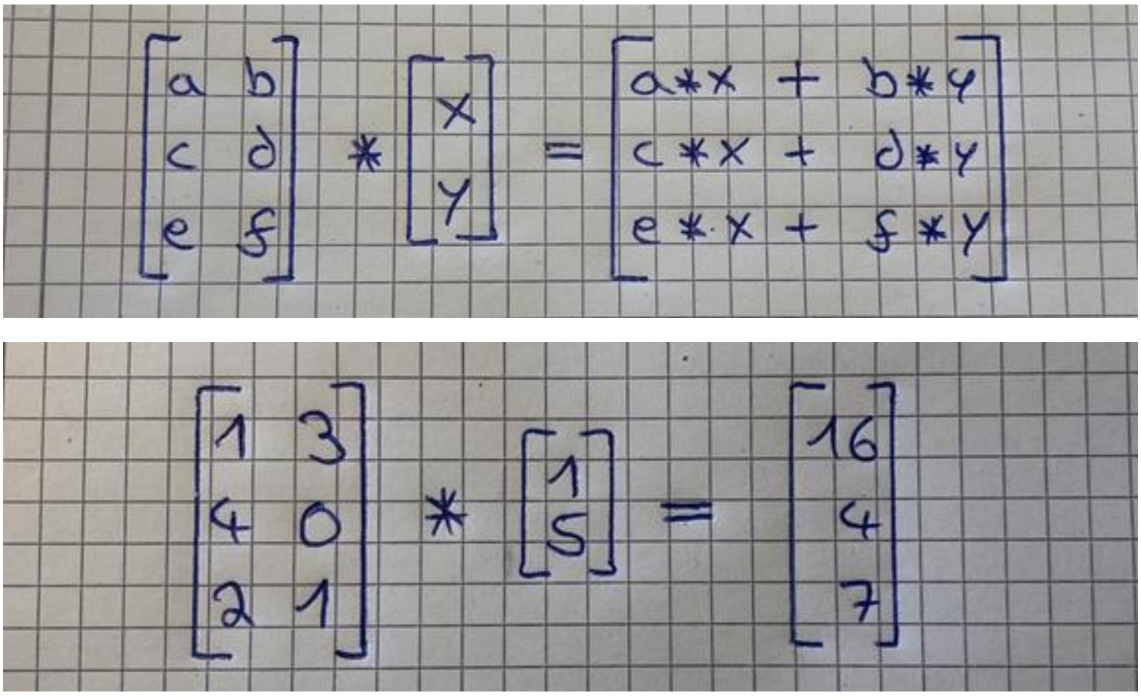
Sólo debemos multiplicar el número por cada uno de los componentes de la matriz. El resultado que obtenemos es una única matriz, producto de la operación realizada.



Fuente: Medium

Operaciones con vectores: Multiplicación

La multiplicación de una matriz por un vector se puede considerar como la multiplicación de cada fila de la matriz por la columna del vector. La salida será un Vector que tiene el mismo número de filas que la Matriz.

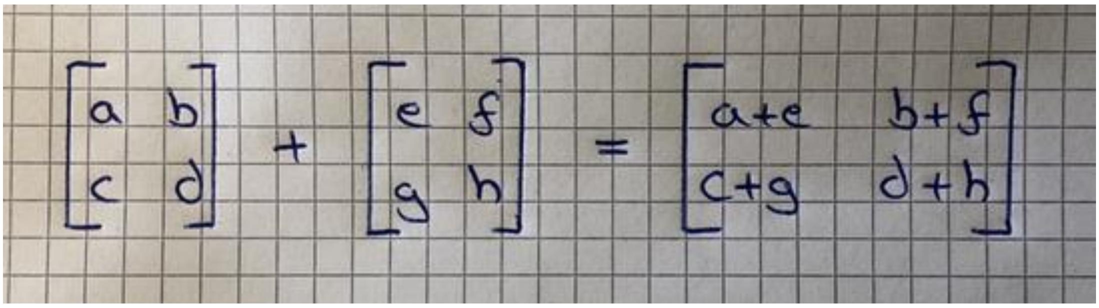


Fuente: Medium

Vemos que el resultante siempre será un vector vertical con el mismo número de filas que la matriz.

Suma y resta de matrices

El requisito es que las matrices tengan las mismas dimensiones y el resultado y el resultado es una matriz que también tiene las mismas dimensiones. Simplemente se agrega o resta cada valor de la primera matriz con su valor correspondiente en la segunda matriz:

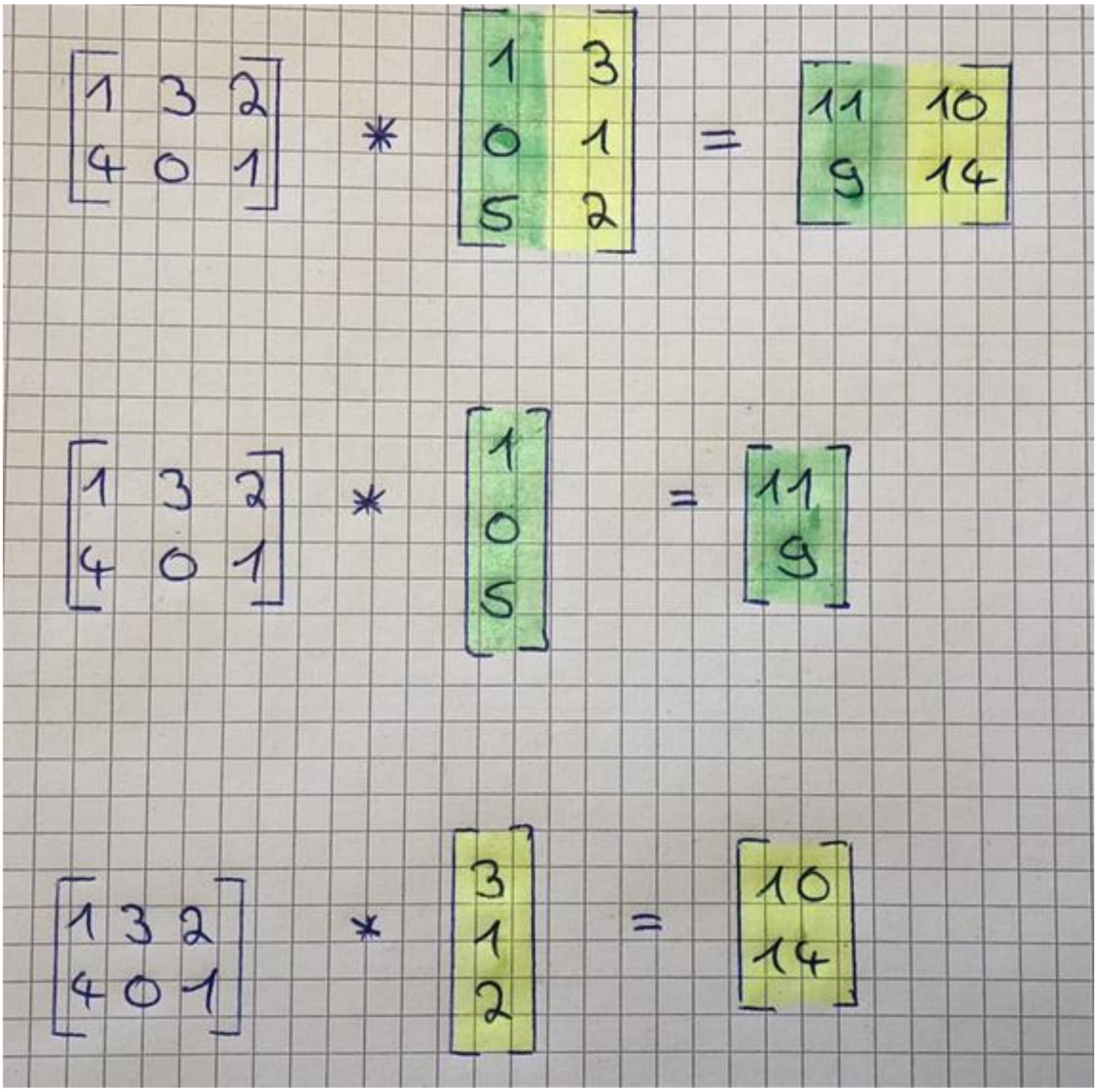


Fuente: Medium

Multiplicación de matrices

Multiplicar dos Matrices juntas tampoco es tan difícil si sabes cómo multiplicar una matriz por un vector. Ten en cuenta que solo se pueden multiplicar matrices si el número de columnas de la primera matriz coincide con el número de filas de la segunda matriz. El resultado será una matriz con el mismo número de filas que la primera y el mismo número de columnas que la segunda. Funciona de la siguiente manera:

Simplemente divide la segunda matriz en vectores de columna y multiplica la primera por separado por cada uno de estos vectores. Luego pones los resultados en una nueva matriz sin sumarlos:



Fuente: Medium

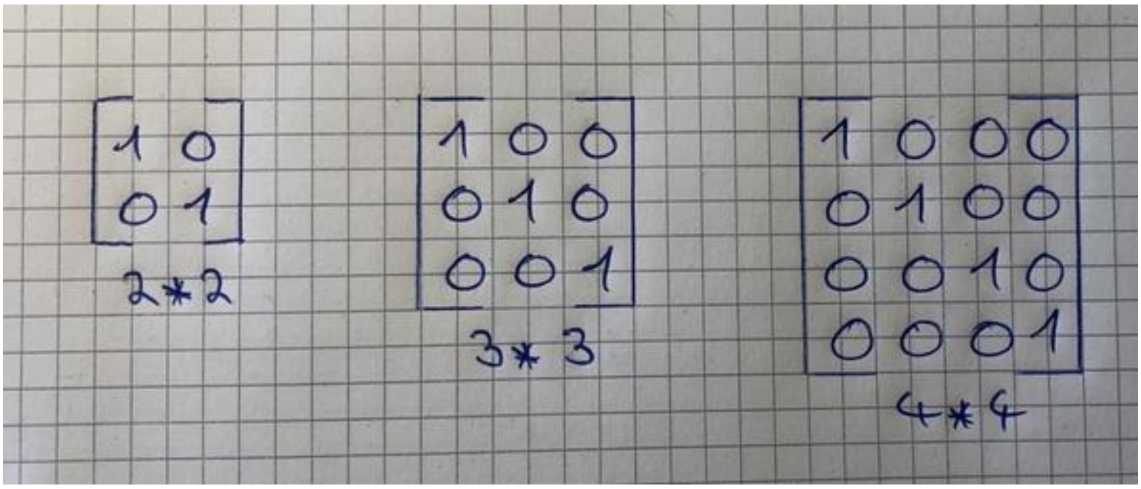
La multiplicación de matrices es:

- No conmutativa: A * B no es lo mismo que B * A.
- Asociativa: A (B * C) = (A * B) C

- Distributiva: A (B + C) es lo mismo que A * B + A * C.

Matriz identidad

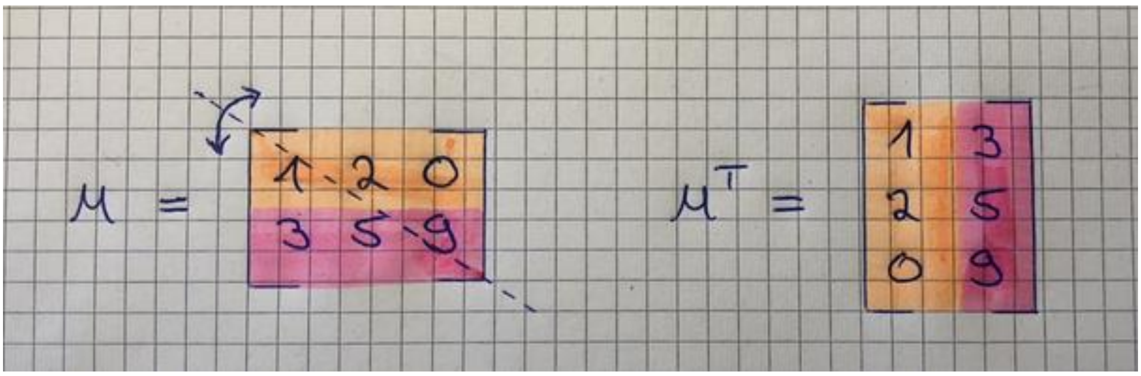
¿Qué es una identidad? Una identidad es un elemento matemático que, multiplicado por otro elemento A, da como resultado dicho elemento A. Así, el número 1 es una Identidad porque todo lo que multiplicas con 1 es igual a sí mismo. La matriz identidad es un tipo especial de. Por lo tanto, cada matriz que se multiplica por una matriz identidades es igual a sí misma. las matrices identidades son matrices cuadradas con los elementos de la diagonal igual a 1 y el resto igual a 0:



Fuente: Medium

Transpuesta de una matriz

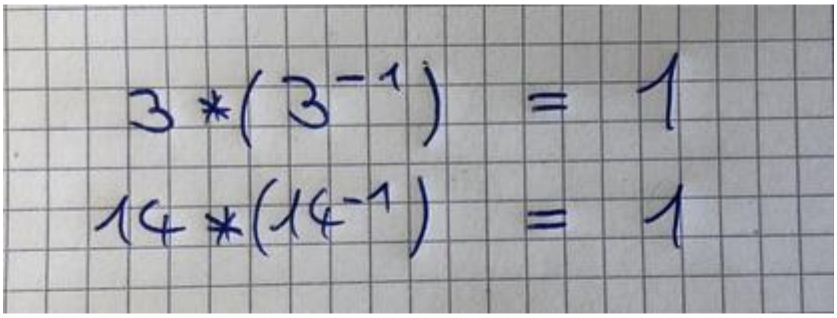
Esta es básicamente la imagen reflejada de una matriz, a lo largo de un eje de 45 grados. Es bastante simple obtener la transpuesta de una matriz. Básicamente, sus filas pasan a ser sus columnas y viceversa. Su primera columna es la primera fila de la matriz de transposición y la segunda columna es la segunda fila de la matriz de transposición. Una matriz m x n se transforma en una Matriz n x m. Además, el elemento a_{ij} de A es igual al elemento A_{ji} de la transpuesta.



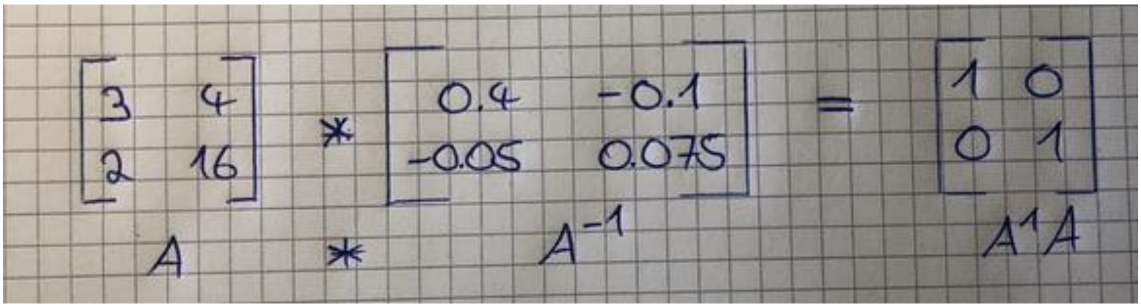
Fuente: Medium

Inversa de una matriz

¿Qué es una inversa? Matemáticamente, un elemento matemático es el inverso de otro si, al multiplicarlos, se obtiene una identidad. Con un escalar:



Con una matriz, la matriz inversa vendrá dada por:



Fuente: Medium

Ojo, no todas las matrices tienen una inversa. Como requisito necesario, la matriz debe ser cuadrada (mismo número de filas que columnas). Aparte, es necesario que tenga determinante diferente de cero.

Más información sobre el cálculo de determinantes e inversas:

Determinantes de orden 1 y 2:

https://www.sangakoo.com/es/temas/calculo-de-determinantes-de-orden-1-y-2

Determinantes de orden 3:

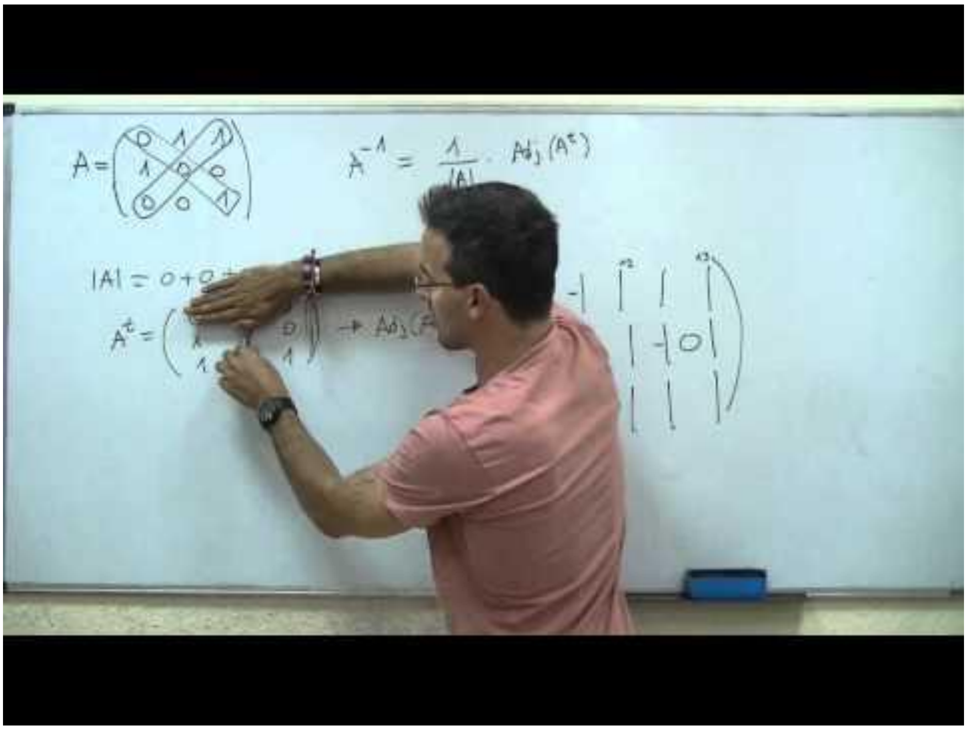
https://www.sangakoo.com/es/temas/regla-de-sarrus-calculo-de-determinantes-de-orden-3

Determinantes de órdenes superiores:

https://www.sangakoo.com/es/temas/metodo-general-para-el-calculo-de-determinantes

Cálculo de inversas:

https://www.youtube.com/watch?v=3BpGef99HEs



Otro método sin determinantes ni adjuntos

https://www.youtube.com/watch?v=Q2EUa_wRgO0

¿Por qué necesitamos un inverso? Porque no podemos dividir matrices. No hay un concepto de división por una matriz, pero podemos multiplicar una matriz por su inversa, lo que resulta esencialmente en la misma cosa.

Autovector y autovalor

Definición. Sea A una matriz cuadrada de orden m. Diremos que un escalar $\lambda \in \mathbb{K}$ ($= \mathbb{R}$ o $\mathbb{C}$) es un autovalor de A si existe un vector $v \in \mathbb{K}^m$, $v \neq 0$ tal que $Av = \lambda v$, en cuyo caso se dice que v es un autovector de A asociado al autovalor λ.
Proposición. Sea λ un autovalor de A y v un autovector asociado, entonces:

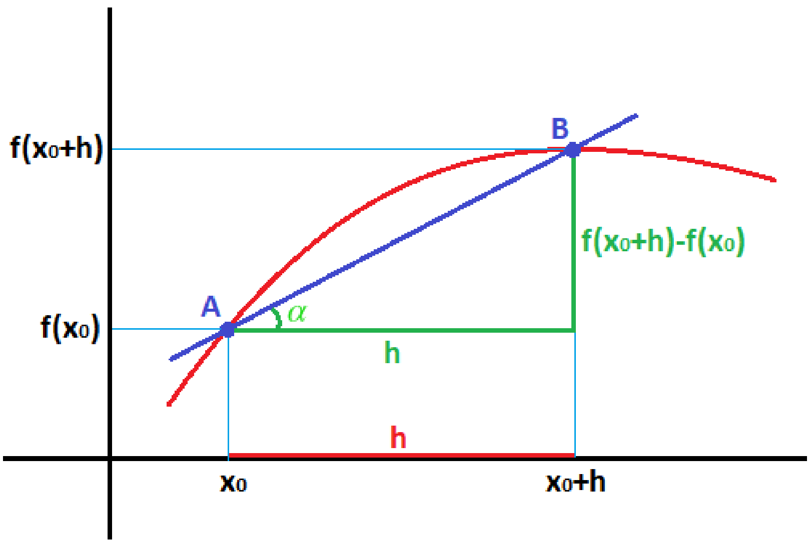
1. $\alpha\lambda$ es un autovalor de αA con autovector v.
2. $(\lambda - \mu)$ es un autovalor de $A - \mu I$ con autovector v.
3. λ^k es un autovalor de A^k con autovector v.
4. Si $q(\cdot)$ es un polinomio, entonces $q(\lambda)$ es un autovalor de $q(A)$ con autovector v. (Ejemplo: $3\lambda^3 + 5\lambda^2 - 7\lambda + 2$ es un autovalor de la matriz $3A^3 + 5A^2 - 7A + 2I$).
5. Si A tiene inversa, entonces $\lambda \neq 0$ y λ^{-1} es un autovalor de A^{-1} con autovector v.

Límite de una función

El concepto de límite es sencillo de entender. El límite de una función *f* en un punto determinado “a” es el valor al que se acerca la función cuando la variable de esta se acerca al punto “a”. Existen muchos métodos para calcular el límite de una función en un punto dependiendo de las características de la función, su rango y su recorrido

Derivada

Es la tasa de cambio instantánea o la pendiente de la tangente a una curva.



Fuente: Eukatio

En cuanto a instantáneo, nos referimos a que la derivada significa el cambio en el valor de la función relativizado con el aumento de la variable que lo ha provocado, cuando este aumento de la variable es infinitamente pequeño.

Cuando tomamos la derivada de una función f (x), obtenemos otra función f '(x). Para obtener la pendiente de la tangente en un punto particular, sustituya los valores en la función f '(x).

Si nuestra variable es el tiempo, la formulación vendrá dada por:

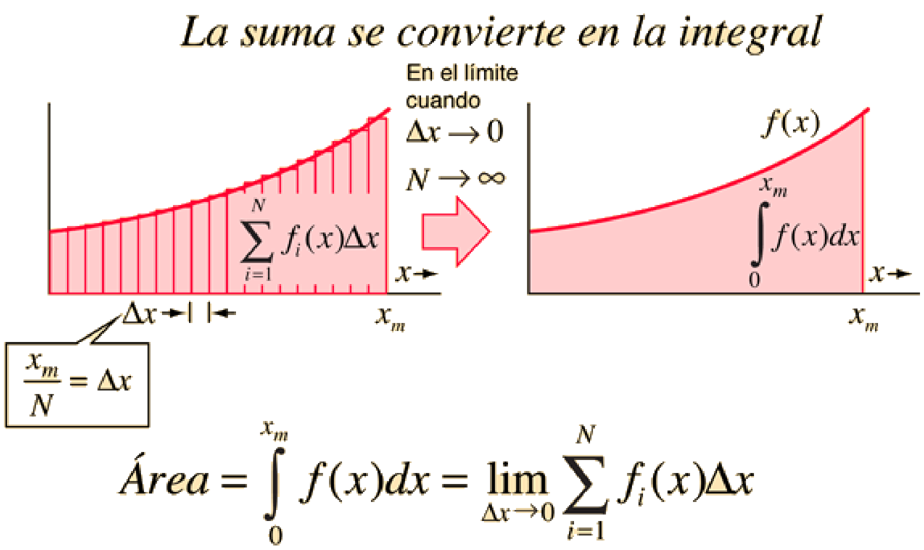
$$\frac{df}{dt} \equiv \lim_{\Delta t \to 0} \frac{f(t + \Delta t) - f(t)}{\Delta t}$$

Fuente: Fuga de celebros

Integral

La forma más sencilla de definir la integral es como el proceso inverso a la derivación. Es decir, para una función *f(x)*, una **integral** de dicha función es otra función, *F(x)*, denominada **primitiva**, que cumpla que *F´(x)=f(x)*. La primera característica evidente de este proceso es que la primitiva de una función no es única.

Una integral definida representa el área que queda bajo la función entre dos puntos determinados:



Fuente: Hyperphysics

Optimización Matemática

La optimización matemática es una rama de las matemáticas que usa el álgebra y el cálculo para encontrar óptimos de funciones. Es un campo bastante complejo y aquí sólo se darán pinceladas a qué forma tienen los diferentes problemas de optimización. Es importante entender que esta rama lleva a encontrar soluciones óptimas de forma matemática a problemas de toma de decisiones bajo ciertas circunstancias. Esta rama también recibe a menudo el nombre de **Investigación Operativa**.

Actualmente la Investigación Operativa incluye gran cantidad de ramas como la Programación Lineal, Programación No Lineal, Programación Dinámica, Simulación, Teoría de Colas, Teoría de Inventarios, Teoría de Grafos, etc..

Se trata de organizar un modelo matemático que relacione las variables que influyen en la decisión, las restricciones y el objetivo. La solución consiste en encontrar los valores de las variables que optimizan (maximizan o minimizan) el valor de la función objetivo y a la vez, satisfaciendo las condiciones expuestas a través de las restricciones. Resumiendo, se puede representar de la siguiente manera:

$$\begin{array}{ll} \textbf{Maximizar o Minimizar} & \text{Función Objetivo} \\ \textbf{Sujeto a:} & \text{Restricciones} \end{array}$$

Problema de programación lineal

Tiene el siguiente planteamiento, donde *c* y *x* son vectores, el primero representando la ponderación de cada variable, y el segundo, las variables involucradas.

$$\begin{array}{ll} \text{Min} & z = c^t x \\ \text{sa:} & x \in S \end{array}$$

Método de resolución habitual: Método del simplex

https://www.gestiondeoperaciones.net/programacion_lineal/metodo-simplex-ejemplo/

Problema de programación no lineal

Usa metodología de cálculo de derivadas, basándose en la convexidad y concavidad de la función objetivo y el conjunto factible. Generalmente tiene la forma:

$$\begin{cases} max f(x_1, \ldots, x_n) \\ g_1(x_1, \ldots, x_n) \leq 0 \\ \vdots \\ g_m(x_1, \ldots, x_n) \leq 0 \\ x_1, \ldots, x_n \geq 0 \end{cases} \quad \text{y} \quad \begin{cases} min f(x_1, \ldots, x_n) \\ g_1(x_1, \ldots, x_n) \geq 0 \\ \vdots \\ g_m(x_1, \ldots, x_n) \geq 0 \\ x_1, \ldots, x_n \geq 0 \end{cases}$$

Como norma general se usa el lagrangiano

Dualidad

Todo problema de programación lineal (que llamamos *primal (P)*) lleva asociado un problema *dual (D)*. la solución al problema dual (de maximzación) provee de un límite mínimo para la solución del problema primal de minimizar.