Contenido

[1. Vectores. 2](#_Toc136508797)

[1.1 Proyección de un vector. 2](#_Toc136508798)

[1.2 Producto punto de vectores. 2](#_Toc136508799)

[2. Matrices. 2](#_Toc136508800)

[2.1 El determinante de una matriz. 2](#_Toc136508801)

[2.2 Tipos de matrices. 3](#_Toc136508802)

[2.2.1 Matriz cuadrada. 3](#_Toc136508803)

[2.2.2 Matriz diagonal. 3](#_Toc136508804)

[2.3 Transposición de una matriz. 3](#_Toc136508805)

[2.4 La convención de suma de Eistein aplicada al producto matricial. 3](#_Toc136508806)

[2.5 Matriz ortogonal: Características. 4](#_Toc136508807)

[2.6 Cambio de base de matrices. 5](#_Toc136508808)

[2.6.1 ¿Qué es el cambio de base? 5](#_Toc136508809)

[2.6.2 El marco de referencia canónico 5](#_Toc136508810)

[2.6.3 Otros marcos de referencia 6](#_Toc136508811)

[2.6.4 Marcos de referencia ortogonales. 6](#_Toc136508812)

[2.7 El proceso de Gram-Schmidt. 6](#_Toc136508813)

[2.8 Reflejando en un plano. 7](#_Toc136508814)

[3. Vectores característicos (característicos). 7](#_Toc136508815)

[3.1 ¿Qué son los vectores eigen? 7](#_Toc136508816)

[3.2 Casos especiales. 8](#_Toc136508817)

[3.3 Descripción matemática de los eigen vectors y eigen values. 8](#_Toc136508818)

[3.4 Cambio de base de vectores propios o característicos. 9](#_Toc136508819)

[3.5 Un ejemplo práctico: El page rank. 9](#_Toc136508820)

[4. Funciones. 11](#_Toc136508821)

[4.1 ¿Qué son las funciones? 11](#_Toc136508822)

[5. Derviada o gradiente 13](#_Toc136508823)

[5.1 ¿Qué es una derivada? 13](#_Toc136508824)

[5.2 Reglas de las derivadas. 13](#_Toc136508825)

[5.3 Derivadas inmediatas: 15](#_Toc136508826)

[5.1 ¿Qué es un gradiente? 15](#_Toc136508827)

[5.2 Derivadas parciales. 15](#_Toc136508828)

[5.3 Derivada total 15](#_Toc136508829)

[5.4 Regla de la cadena en funciones de varias variables. 16](#_Toc136508830)

[5.5 Vector Jacobiano. 16](#_Toc136508831)

[5.6 Matriz jacobiana. 16](#_Toc136508832)

[5.7 La matriz hessiana. 17](#_Toc136508833)

[6. Fundamentos matemáticos de redes neuronales. 18](#_Toc136508834)

[6.1 Modelos. 18](#_Toc136508835)

[6.2 Regresiones lineales. 18](#_Toc136508836)

[6.3 Definición de una neurona a nivel matemático. 18](#_Toc136508837)

[6.3.1 La neurona como función 18](#_Toc136508838)

[6.3.2 El papel del Bias 18](#_Toc136508839)

[6.3.3 La función de activación. 19](#_Toc136508840)

[6.4 Definición matemática formal de una red neuronal. 19](#_Toc136508841)

[6.5 Definición de una red neuronal a nivel matemático. 19](#_Toc136508842)

[7. Python y las matemáticas. 21](#_Toc136508843)

[7.1 Multiplicación de producto punto de matrices en numpy. 21](#_Toc136508844)

[7.2 Vectores y valores propios en Python. 21](#_Toc136508845)

# Vectores.

## Proyección de un vector.

Esta técnica consiste en descomponer un vector dado en términos de sus componentes paralelos y perpendiculares a cada uno de los vectores base ortogonales del marco de referencia.

## Producto punto de vectores.

El producto punto, también conocido como producto escalar o producto interno, es una operación matemática que se aplica entre dos vectores y se define como la suma de los productos de los elementos correspondientes de los vectores. En otras palabras, si tenemos dos vectores A y B de igual longitud, el producto punto se define como:

A · B = A1 \* B1 + A2 \* B2 + ... + An \* Bn

donde A1, A2, ..., An y B1, B2, ..., Bn son los elementos correspondientes de los vectores A y B.

Claves:

* El producto punto es simétrico.

# Matrices.

## El determinante de una matriz.

El determinante de una matriz nos puede decir si los vectores que componen una matriz son o no son linealmente independientes. Si el determinante de una matriz es distinto de cero quiere decir que son linealmente independientes.

## Tipos de matrices.

### Matriz cuadrada.

Es aquella matriz que tiene el mismo número de filas y de columnas, de estas matrices pueden calcularse el determinante

### Matriz diagonal.

Es aquella matriz cuadrada, es decir de dimensiones “n x n” tal que todos los elementos de fuera de la diagonal principal de la matriz son ceros (esto no significa que algunos elementos de la diagonal no puedan ser ceros). Una propiedad interesante de las matrices diagonales es que:

En el calculo de los valores característicos o propios dada una matriz de transformación que es diagonal, podremos afirmar que cada uno de los elementos de la diagonal son “valores propios” o “eigenvalues”.

## Transposición de una matriz.

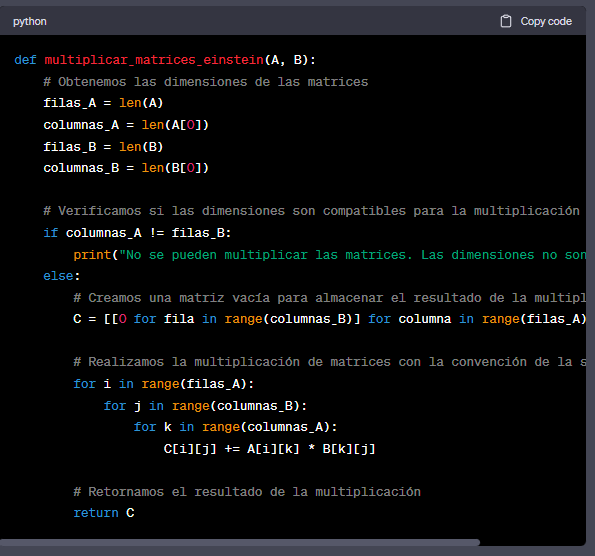
La transposición de matrices es una operación matricial que consiste en intercambiar las filas por las columnas de una matriz. Dada una matriz A de dimensiones m x n, su matriz transpuesta, denotada como , es una matriz de dimensiones n x m, donde el elemento en la fila i y columna j de es el mismo que el elemento en la fila j y columna i de A.

|  |  |
| --- | --- |
| PROPIEDADES DE LA TRANSPOSICIÓN | |
| Doble transposición |  |
| Transposición de la suma |  |
| Transposición de escalar |  |
| Transposición de producto |  |
| Matriz identidad (Solo si A es ortogonal y ortonormal) |  |

## La convención de suma de Eistein aplicada al producto matricial.

En el contexto de la multiplicación de matrices, lo que viene a decir está convención es que dada dos matrices A y B, las cuales se multiplican, en el calculo de una determinada posición “ij” en la matriz resultante viene dada por:

Ayudándonos de está formula, podemos construir de manera sistemática una función que calcule la multiplicación de dos matrices:



Claves:

* Nos permite lidiar con matrices no cuadradas.
* Busca simplificar la formula compleja de: . Por una más sencilla.
* Ya con la librería numpy podemos calcular fácilmente el producto punto de dos matrices.

## Matriz ortogonal: Características.

Las columnas (y también las filas) de una matriz ortogonal forman un conjunto ortonormal. Esto significa que todos los vectores columna (o fila) son perpendiculares entre sí (ortogonales) y tienen magnitud 1 (norma unitaria).

|  |  |
| --- | --- |
| **Propiedad** | **Descripción** |
| **Inversa y transposición** | La inversa de una matriz ortogonal es igual a su matriz transpuesta, es decir, |
| **Determinante** | El determinante de una matriz ortogonal es siempre 1 o -1. |
| **Columnas (y filas) ortonormales** | Las columnas (y filas) de una matriz ortogonal forman un conjunto ortonormal. |
| **Preservación de longitud y ángulos** | Las matrices ortogonales preservan la longitud y el ángulo entre los vectores al aplicar la transformación lineal correspondiente. |
| **Representación de rotaciones y reflexiones** | Las matrices ortogonales pueden representar rotaciones y reflexiones en el espacio euclidiano. |
| **Producto con su transpuesta** | Una matriz ortogonal cumple con la propiedad A^T \* A = A \* A^T = I (matriz identidad). |

Además, debemos considerar que toda matriz diagonal es una matriz ortogonal, es decir que los vectores que la conforman forman una base ortonormal.

**Importante:** En ciencia de datos siempre se están buscando matrices de este tipo.

## Cambio de base de matrices.

### ¿Qué es el cambio de base?

En matemáticas y ciencias aplicadas, a menudo es útil trabajar con diferentes bases para describir un mismo objeto en un espacio vectorial. El cambio de base de matrices es una técnica matemática que nos permite transformar una matriz de coordenadas de un sistema de referencia a otro, es decir, nos permite expresar las mismas coordenadas de un objeto en una base diferente.

Existe una matriz base que define la base del sistema de coordenadas, recordemos que una base es un conjunto de vectores que se utilizan como punto de referencia para describir la posición de un objeto en ese sistema de coordenadas (es como un “marco de referencia”). Cada uno de los vectores que componen la base deben ser linealmente independientes (esto no es sinónimo de que sean perpendiculares entre sí, ya que hay vectores que son linealmente independientes y no son perpendiculares entre sí).

Ejemplo práctico para entenderlo mejor:

Imaginemos que estamos en un avión volando en el aire y de repente recibimos una llamada de la torre de control del aeropuerto más cercano informándonos sobre la posición de una bandada de pájaros que debemos evitar. La torre de control nos proporciona las coordenadas de la posición de los pájaros utilizando un sistema de coordenadas global basado en la Tierra.

En este escenario, podemos pensar en el cambio de base como una herramienta que nos permite convertir las coordenadas de la posición de los pájaros desde el sistema de coordenadas global a un sistema de coordenadas local que esté alineado con nuestro avión. Para realizar esta conversión, necesitamos conocer la matriz de transformación que lleva de un sistema de coordenadas al otro.

Una vez que hemos convertido las coordenadas de los pájaros a nuestro sistema de coordenadas local, podemos determinar su posición relativa con respecto a nuestro avión. Si los pájaros están muy cerca de nuestro avión, podemos utilizar esta información para maniobrar y evitar la colisión.

Aquí vemos la aplicación del campo base para el mundo real.

### El marco de referencia canónico

El marco de referencia canónico es un sistema de coordenadas en el que se utilizan los vectores canónicos como base. Un ejemplo de base cartesiana está definido por los siguientes vectores:

Si, además de esto, se cumple que los ejes del marco los denominamos ejes “x”, “y” y “z” entonces hablamos del marco de referencia “cartesiano”.

### Otros marcos de referencia

Cuando tenemos una base diferente a la canónica, uno de los requisitos que se deben cumplir para que sea una base correcta, es que todos los vectores que la compongan sean linealmente independientes.

Ahora bien, supongamos lo siguiente:

* Tenemos una matriz base A que define un marco de referencia Marco(A) cuyos vectores están definido con respecto del eje de coordenadas canónico.
* Si hacemos la inversa de ese vector A, es decir, calculamos entonces tendremos la base A con respecto del marco de referencia Marco(A)
* Además, tenemos un vector “v” definido en el marco de referencia (A)
* Finalmente, tenemos un vector “w” definido en el marco de referencia (Canónico).

Con estas premisas tenemos que:

* Para transformar el vector “v” referenciado con respecto del marco(A) a un nuevo vector “ v’ ” que tiene como referencia el marco canónico tenemos la siguiente formula:
* Para transformar el vector “w” referenciado con respecto del marco (Canónico) a un nuevo vector w’ con respecto del marco(A) hacemos lo siguiente:

### Marcos de referencia ortogonales.

Vimos en el punto anterior ciertas transformaciones de vectores de un marco de referencia A hacia el marco de referencia canónico, ahora bien, dado un vector v en un marco de referencia B con vectores base ortogonales , y un marco de referencia A con vectores base ortogonales puedes calcular las coordenadas del vector v' en el marco de referencia A mediante proyecciones.

Recordemos que la proyección ortogonal consiste en descomponer un vector dado en términos de sus componentes paralelos y perpendiculares a cada uno de los vectores base ortogonales del marco de referencia

En este sentido, si se tiene un vector v en el marco de referencia X y se quiere transformar a un vector v' en el marco de referencia Y, se puede multiplicar el vector v por cada uno de los vectores base de Y utilizando el producto punto, cada resultado será el valor de una de las coordenadas del vector v’ en el marco de referencia Y.

## El proceso de Gram-Schmidt.

Como **lo ideal para ciencia de datos es que un conjunto de vectores sea ortonormal** formando una matriz ortogonal, dado un conjunto de vectores linealmente independientes no ortonormales podemos generar mediante el proceso de Gram-Schmidt un conjunto de vectores ortonormales.

Es decir, **el proceso de Gram-schmidt** es un método de álgebra lineal para ortogonalizar un conjunto de vectores (que deben ser linealmente independientes) en un espacio vectorial con un producto escalar, generalmente el espacio euclidiano.

De manera genérica seguimos los siguientes pasos:

1. El primer vector de la base se divide entre su modulo. Dando como resultado el primer vector de nuestra base ortonormal .
2. Tras esto para los siguientes “n” vectores hay que seguir dos procedimientos:
   1. El primero es hacer como una especie de operación que nos permite hacer el vector “n” perpendicular a los “n-1” vectores tratados anteriormente, esto se consigue con fórmulas propias de la proyección de un vector con respecto de otros. Por ejemplo, para el caso del segundo vector:

Tal que es el segundo vector de la base origen que se debe ortogonalizar con respecto de . Además, para el caso de tenemos que hacer la ecuación anterior no solo con respecto de sino también con respecto de que dando como sigue:

así, podemos concluir que para el vector tendremos que hacer la formula expresada con respecto de los n-1 vectores normalizados anteriormente.

* 1. El segundo es un proceso de normalización del vector (conseguir que su modulo sea 1), esto se consigue dividiendo el vector resultando del proceso “a” entre su modulo.

## Reflejando en un plano.

# Vectores característicos (característicos).

## ¿Qué son los vectores eigen?

Cuando multiplicamos un vector o conjunto de vectores por una matriz de transformación, el resultado es un nuevo vector o conjunto de vectores, si lo único que varía en el vector o en ciertos vectores tras realizar la operación, es el sentido y la magnitud manteniéndose la dirección igual, entonces hablamos de vectores característicos.

Es decir, podemos decir que los vectores eigen son aquellos vectores que tras multiplicarse por una matriz transformadora mantienen su dirección. Además, también hablamos de “Eigen value” o “valor característico” como el valor en el que se ha proporcionado el vector característicos, es decir, ¿el vector característico es el doble de grande, o tal vez, es un medio más pequeño, o incluso, puede hacer cambiado su sentido, pero no su dirección teniendo un valor característico igual a un número negativo. Para aclarar, el valor característico λ no representa directamente la magnitud del eigenvector v, sino la escala por la cual se transforma el eigenvector cuando se multiplica por la matriz A

Cuando hablamos de un problema Eigen (problema característicos) se esta hablando de encontrar las características de algún tema.

## Casos especiales.

* Dado un conjunto de vectores que sufren una transformación lineal de manera proporcional en todas las direcciones, podemos afirmar que todos los vectores que sufren este tipo de transformación son vectores característicos.
* La rotación de un conjunto de vectores de 180 grados implica que todos los vectores siguen siendo vectores característicos.
* Cuando el eigen value es negativo significa que el vector característico ha cambiado de sentido pero que mantiene su dirección.
* En matrices de transformación diagonales (todo menos la diagonal a cero) los eigen valores son siempre los elementos de la diagonal.

## Descripción matemática de los eigen vectors y eigen values.

En ciertos casos no es fácil detectar los vectores característicos y se deben usar las matemáticas formales para poder encontrarlos. La formula básica queda como sigue:

Tal y como se aprecia, la multiplicación de un vector característicos por una matriz tiene que implicar que el vector mantenga su dirección y que se redimensione su magnitud en una proporción denotada por su “valor característico”. Tras realizar ciertas operaciones matriciales básicas de la anterior formula podemos llegar a la siguiente formula:

En este caso, o bien, el vector es cero, o lo que hay entre paréntesis es cero, para el estudio de los eigen vectors es mucho más interesante estudiar la formula:

De esta ultima ecuación podemos obtener los “valores característicos”, Teniendo en cuenta que a, b, c y d en los problemas serán valores conocidos. Con esto vamos a obtener una ecuación de segundo grado que nos permitirá conocer los valores característicos posibles, es decir, los posibles valores de tras resolver la ecuación, una vez obtenido este valor, volvemos a la formula:

Sustituyendo los valores de lamba en la ecuación podremos obtener los vectores característicos del sistema. Si en el polinomio característico no lograrmos obtener eigen values, entonces podemos afirmar que para esa transformación no existen eigen values.

## Cambio de base de vectores propios o característicos.

Cuando necesitamos multiplicar “n” veces una matriz de transformación sobre un vector podemos apreciar que a medida que crece “n” el problema a nivel de calculo computacional se torna complejo y costoso, por lo tanto, para facilitar estos cálculos, se utiliza la diagonalización. Una matriz diagonal es aquella que en la que todos los elementos fuera de la diagonal principal de la matriz son cero.

Matemáticamente hablando, si tenemos una matriz de transformación y la aplicamos sobre un vector , podremos afirmar que siendo el vector el nuevo vector transformado, si bien, podemos aplicar una segunda transformación al vector nuevamente con la matriz quedándonos que:

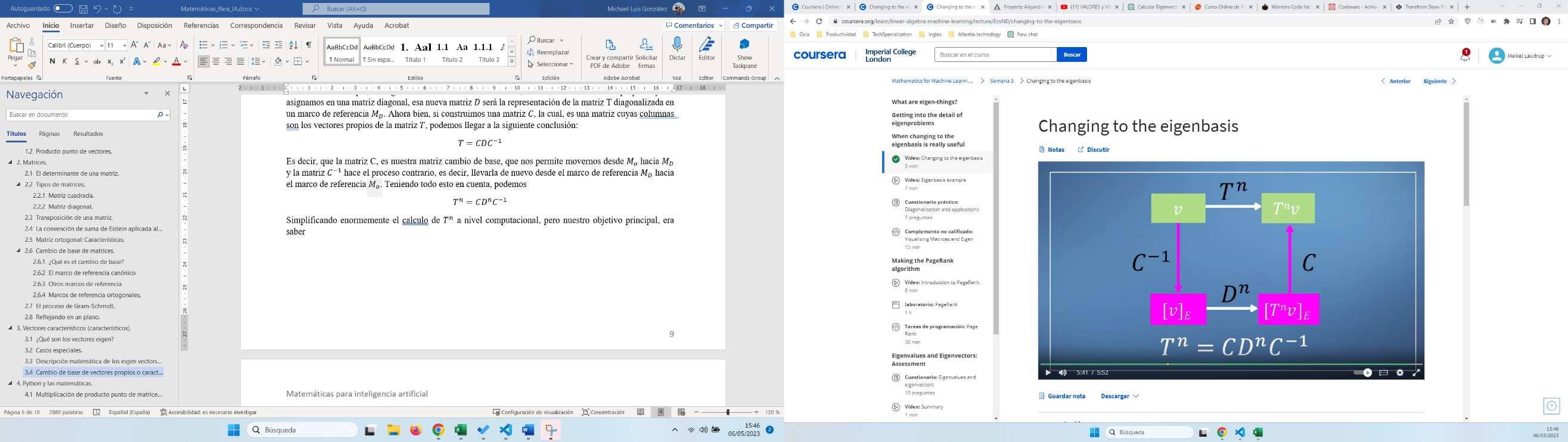
Si continuamos haciendo transformaciones de vectores nos quedaría, de manera genérica que:

Ahora bien, como habíamos comentado al principio de este epígrafe cuando “n” tiende a ser muy grande a nivel computacional el coste de realizar se puede tornar altamente costoso a medida que crece “n”, por todo esto, para optimizar el calculo se busca siempre que la matriz transformadora “T” sea una matriz diagonal.

Cuando la matriz Transformadora “” no sea diagonal, entonces buscaremos llevárnosla a otro marco de referencia cambiándola de base donde si sea una matriz diagonal, efectuar los cálculos propicios y traerla de nuevo a la base o marco de referencia origen para obtener el resultado final, simplificando enormemente el proceso de cálculo de la operación .

Pongamos dos marcos de referencia, el primero, será el **M**arco de referencia **O**rigen donde no es una matriz Diagonal, de ahora en adelante y pongamos el segundo **M**arco de referencia donde T sí que es una matriz **D**iagonal, denotado a partir de ahora como El marco de referencia tiene denotada la base y el marco de referencia tendrá la base .

Podemos afirmar que, si cogemos la matriz transformadora T, calculamos sus “valores propios” y los asignamos en una matriz diagonal, esa nueva matriz será la representación de la matriz T diagonalizada en un marco de referencia . Ahora bien, si construimos una matriz , la cual, es una matriz cuyas columnas son los vectores propios de la matriz , podemos llegar a la siguiente conclusión:

Es decir, que la matriz C, es nuestra matriz cambio de base, que nos permite movernos desde hacia y la matriz hace el proceso contrario, es decir, llevarla de nuevo desde el marco de referencia hacia el marco de referencia . Teniendo todo esto en cuenta, podemos

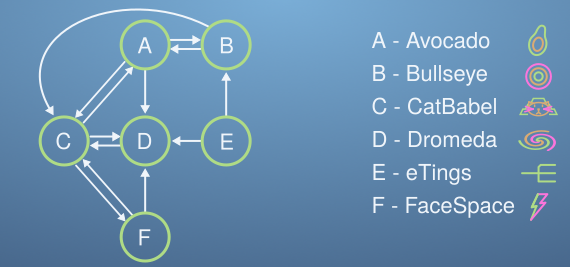
***CAMINO DIFICIL***

Simplificando enormemente el calculo de a nivel computacional.

## Un ejemplo práctico: El page rank.

El PageRank es un algoritmo desarrollado por Larry Page y Sergey Brin, los fundadores de Google, para determinar la importancia de las páginas web en función de la cantidad y la calidad de los enlaces que apuntan hacia ellas. Cuando el usuario realiza una búsqueda recibirá enlaces web ordenados en función de lo s resultados de este algoritmo La idea detrás del PageRank es que una página es más importante si otras páginas importantes enlazan a ella. En términos sencillos, el PageRank intenta determinar la probabilidad de que un usuario que navega aleatoriamente por la web termine en una página específica.

Imaginemos una micro-red de páginas webs, las cuales, son nodos de un grafo y las flechas que unen los nodos son links entre estas paginas web, por ejemplo:



Claramente se aprecia visualmente que la página web más importante será la página C, porque es las que más nodos tienen un enlace hacia a él, en este caso cuatro links externos pueden acceder hacia ella. Ahora, si imaginamos a una persona, que tienen mucho tiempo libre y se ponen a navegar en la web de manera un poco aleatoria. Mapeando todos los posibles links, se puede construir un modelo para estimar la cantidad de tiempo que gastara esta persona en cada pagina web. Podemos describir los links que hay en cada página web como un vector, en la cada posición se asignará un valor que será 0 o bien tal que “n” es el número de links externos que hay en la página web, en el caso de A, hay 3 links externos, por lo que hay una probabilidad de de que uno de los enlaces sea pulsado.

Si aplicamos esto para todos los nodos del grafo, teniendo en cuenta, que cada columna será un vector de cada página web tendremos que:

Al aplicar el algoritmo de PageRank a la matriz de enlaces L, buscamos encontrar un vector de importancia (PageRank) r que sea estable bajo la transformación dada por la matriz de enlaces. Esto significa que cuando multiplicamos la matriz L por el vector r, el vector resultante sigue siendo proporcional al vector r original. En otras palabras, queremos que:

Aquí, λ es el eigenvalor y r es el eigenvector correspondiente. En el caso del algoritmo de PageRank, buscamos el eigenvector correspondiente al eigenvalor dominante (más grande), que generalmente es 1. Este eigenvector r es el vector de PageRank que buscamos.

Entonces, el problema de encontrar el vector de PageRank se reduce a encontrar el eigenvector asociado con el eigenvalor dominante de la matriz de enlaces L. Utilizamos técnicas como la iteración de potencias para encontrar este eigenvector de manera eficiente, especialmente cuando trabajamos con matrices grandes.

Imagina que el Internet es una gran red de páginas web conectadas por enlaces. Un "surfer" (navegante) en esta red comienza en una página y sigue los enlaces de una página a otra de manera aleatoria. El PageRank representa la probabilidad de que un "surfer" que navega aleatoriamente en esta red termine en una página específica después de un número suficientemente largo de pasos.

Las páginas con un PageRank más alto se consideran más importantes y, por lo tanto, tienen más probabilidades de aparecer en los primeros resultados de una búsqueda en Google. El algoritmo de PageRank originalmente fue una parte clave del sistema de clasificación de Google, aunque desde entonces se ha vuelto más complejo e incluye muchos otros factores además del PageRank.

Cuando se dice que el "PageRank es el orden clasificado de las páginas desde la más probable hasta la menos probable que el navegante esté viendo", significa que las páginas se ordenan en función de su importancia en la red (según el algoritmo de PageRank), desde la página con la probabilidad más alta de ser visitada por un "surfer" que navega aleatoriamente hasta la página con la probabilidad más baja de ser visitada. Este ranking es una medida de la importancia y relevancia de cada página en la web.

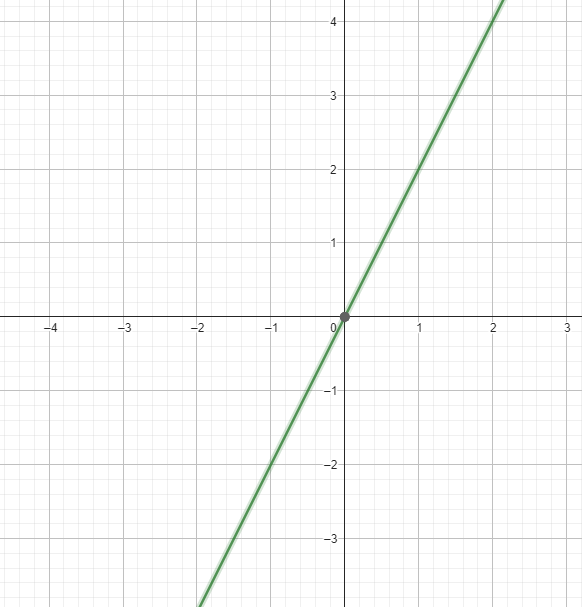
# Funciones.

## ¿Qué son las funciones?

De manera informal una función recibe es como una caja negra que recibe unos inputs y a partir de ellos genera una salida. Una manera semiformal de definirla es como:

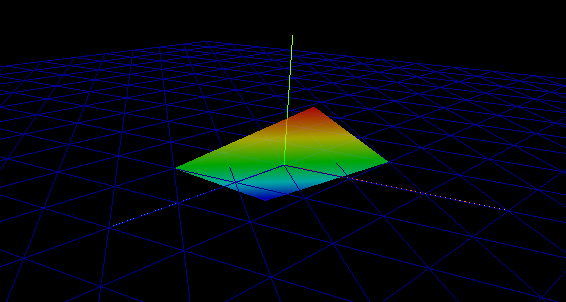
Tal que inputs puede ser un arreglo de variables y un output suele ser un calculo a partir de los inputs. Las funciones más simples son aquellas que solo tienen una variable, por ejemplo, una función que dada un valor cualquiera como entrada, te devolverá su doble:

Normalmente las funciones matemáticas tienen una representación gráfica asociada, esta se consigue dibujar asignado puntos en una gráfica cuya distancia del punto (0,0) a nivel horizontal es el valor que se le da a “x” y cuya distancia con respecto del punto (0,0) a nivel vertical es el resultado de la función, uniendo los puntos conseguimos dibujar la línea que representa la gráfica, por ejemplo, para el caso anterior tenemos que su grafica asociada es:

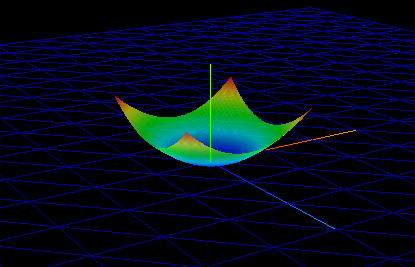


De aquí podemos concluir que las funciones de una variable son una línea (recta o curva) en un gráfico, y esta puede ser continua o discontinua, además hay otras funciones más complejas compuestas por varias variables, por ejemplo, una función que represente la suma, dado dos inputs devuelve su suma:

Las funciones de dos variables suelen representarse como superficies como planos o planos curvos, por ejemplo, para la anterior función tenemos que:



Otro ejemplo podría ser la gráfica para la función , ya finalmente, si subimos un nivel de complejidad más alta, donde tenemos funciones de tras variables, podemos hablar de figuras geométricas en el espacio.



Algunos enlaces de interés son:

* Ver funciones más conocidas: https://rstudio-pubs-static.s3.amazonaws.com/824875\_494efd5800f64deeaee802c3c4dbad5d.html
* Diagramar funciones: <https://www.geogebra.org/graphing?lang=es>
* Diagramar funciones de varias variables: https://www.mathstools.com/section/main/3D\_Functions\_Plotter?lang=es#.ZFqCBXZBxD8

# Derviada o gradiente

## ¿Qué es una derivada?

Las derivadas, gráficamente hablando, pueden considerarse como la pendiente de la recta tangente de una gráfica en un punto determinado de una función, al final viene a ser dada esa recta tangente calcular lo que asciende entre dos puntos dividido entre lo que recorre horizontalmente de izquierda a derecha. Formalmente, podemos afirmar que la derivada es la cantidad que crece o decrece una función entre dos puntos, dividido entre la distancia entre esos dos puntos cuando está tiende a cero. Su ecuación asociada viene definida por:

Si quieres ver derivadas con más detalle

Enlaces de interés: <https://youtu.be/V7r4amUPI9k>

## Reglas de las derivadas.

Como vimos anteriormente para funciones relativamente complejas calcular la anterior función puede volverse tedioso, es por todo esto que para las funciones más comunes se tienen lo que se denominan derivadas inmediatas, algunas veces, habrá que jugar con las matemáticas para transformar derivadas no inmediatas en derivadas inmediatas.

* **Regla de la suma:**

Esta regla viene a decir que, si detectamos que nos piden calcular la derivada de la suma de “n” funciones, nos limitemos a calcular las derivadas de cada una de las funciones individualmente y luego las sumemos, simplificando los cálculos.

* **Regla de la multiplicación**:

Esta regla podría extrapolarse para una multiplicación de tres funciones, tal que:

Si aplicamos para cuatro funciones tendríamos que:

Si se estudia con un poco de detallelas 4 funciones podemos afirmar que:

Enlace recomendado: <https://www.youtube.com/watch?v=GEujiBObdRM>

* **Regla de la división**: Existe una reglar específica para la derivada de una función, la cual esta constituida por división de dos funciones:

No obstante, esta regla no es estrictamente necesaria de aprender, dado que cualquier división de funciones puede ser reescrita para ser representada como una multiplicación y así aplicar la regla de la multiplicación, un ejemplo ilustrativo sería:

Teniendo esto en cuenta podemos concluir que está regla no es estrictamente necesaria memorizarla.

* **Regla de la cadena:** La regla de la cadena es aplicable queremos calcular la derivada de una función compuesta, es decir, una función que tiene como entrada otra función, por ejemplo:

En este caso tenemos dos opciones, o extendemos de tal forma que la convertimos en una función extensa y calculamos su derivada aplicando la regla de la suma o aplicamos la regla de la cadena. Que dice así:

## Derivadas inmediatas:

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

## ¿Qué es un gradiente?

Se puede entender como una derivada, pero con varias variables, más formalmente, el gradiente es una generalización del concepto de derivada para funciones multivariables. Cuando trabajas con una función de una sola variable, la derivada y el gradiente son esencialmente lo mismo, ya que el gradiente sería simplemente un vector de una dimensión que contiene la derivada de la función con respecto a su única variable.

## Derivadas parciales.

Una derivada parcial es una medida de cómo cambia una función en relación con una variable específica, mientras se mantienen las demás variables constantes. En otras palabras, calcula la tasa de cambio de una función en relación con una sola variable independiente, sin considerar las otras variables. Esta se denota como donde f es la función y x es la variable con respecto a la cual se está derivando parcialmente.

Las derivadas parciales son especialmente útiles en el análisis de funciones de múltiples variables, ya que permiten estudiar cómo la función cambia en relación con cada una de las variables de forma individual.

## Derivada total

Puede considerarse como una generalización de la derivada parcial. Mientras que la derivada parcial mide la tasa de cambio de una función en relación con una variable específica, la derivada total considera cómo cambia la función en relación con todas las variables de forma simultánea, se suele definir como o simplemente df y se define como:

En esta expresión, (∂f/∂xi) representa la derivada parcial de f con respecto a la variable xi, y dxi representa el diferencial de la variable xi.

## Regla de la cadena en funciones de varias variables.

Antes de entrar en profundidad, para simplificar la representación formal de la regla de la cadena, podemos afirmar que notación matemática, se puede expresar:

Es decir, que la serie de variables puede ser representa como una “x” en negrita. Prosiguiendo si tenemos una función cuyas variables constituyen funciones en si misma, podemos aplicar la regla de la cadena, tal que si cada es a su vez una función que está respecto de una variable “t”, por poner un ejemplo, entonces podemos concluir que:

Podemos resumir que es muy similar al concepto de regla de la cadena aplicado para derivadas de una sola variable, ahora bien, otra manera de calcular esto es con la siguiente formula:

## Vector Jacobiano.

Si tenemos una función de muchas variables:

Podemos definir el Jacobiano como un vector en el cada uno de sus elementos es la derivada parcial de la función con respecto de cada una de las variables, tal que:

Podemos afirmar que aquellos valores de la tupla ( en los que el vector jacobiano (es decir, el gradiente) de la función es cero, son puntos críticos de la función, los cuales podrían representar máximos, mínimos o puntos de silla. Estos puntos son donde la función puede cambiar de dirección, lo que puede indicar un máximo o mínimo local (o incluso global), o un punto de silla.

Sin embargo, el vector jacobiano por sí solo no nos proporciona suficiente información para determinar si un punto crítico es un máximo, un mínimo o un punto de silla. Para clasificar los puntos críticos, se suele recurrir a la matriz hessiana, como se verá más adelante.

## Matriz jacobiana.

La matriz jacobiana se puede construir cuando se tienen varias funciones, de tal manera de que cada fila de la matriz jacobiana se corresponde con el vector jacobiano de una función concreta, por ejemplo, si tenemos que:

## La matriz hessiana.

En este caso volvemos a enfocarnos en una única función, el hessiano de una función consiste en una matriz cuadrada de dimensión “n x n” tal que “n” es el número de variables de la función, la matriz hessiana consiste en la derivada segunda con respecto de cada variable que conforme la función para cada uno de los elementos del vector jacobiano de una función.

Es importante mencionar que para que el hessiano exista, la función debe ser al menos dos veces diferenciable. Además, el hessiano de una función es simétrico si la función es dos veces continuamente diferenciable.

Un ejemplo práctico sería el siguiente, tenemos la función:

Lo primero que hacemos es calcular el jacobiano de esta función tal que:

Luego calculamos la derivada de cada uno de los elementos del vector jacobiana con respecto de cada una de las variables de la función original, es decir “x”, “y” y “z” quedándonos que:

En la primera fila hemos derivado cada uno de los elementos del vector jacobiano con respecto de x, en la segunda fila con respecto de y, y finalmente, en la tercera fila con respecto de z, obteniendo así la matriz hessiana.

Ahora bien, la gran utilidad de la matriz hessiana es que nos permite detectar cuando un punto crítico es un máximo, un mínimo o un punto de silla, hay varias maneras, una es con el análisis de los eigenvalores de la matriz hessiana y otro, que es el que se aborda en este epígrafe, es mediante el análisis del determinante de la matriz hesiana y el elemento en la posición i = 0 y j = 0 de la matriz hessiana, tal que si denotamos el determinante de la matriz hessiana como D, es decir D = det(H), podemos concluir que, para aquellos puntos críticos que hacen el jacobiano igual a cero sí:

* Si D > 0 y > 0, entonces la función tiene un mínimo local en el punto.
* Si D > 0 y < 0, entonces la función tiene un máximo local en el punto.
* Si D < 0, entonces el punto es un punto de silla.
* Si D = 0, entonces la prueba es inconclusa, y se necesita más información para clasificar el punto.

Es importante subrayar que para funciones de tres o más variables, el criterio de la segunda derivada se vuelve más complicado y no se puede determinar simplemente con el signo de D y el primer elemento de la matriz hessiana. En general, para funciones de tres o más variables, se debe examinar los eigenvalores de la matriz hessiana para determinar el tipo de punto crítico.

# Fundamentos matemáticos de redes neuronales.

## Modelos.

Un modelo, en el contexto de la ciencia de datos y la inteligencia artificial, se refiere a una representación matemática o computacional de un sistema o proceso del mundo real, simplificándolo y pudiendo entender mejor dicha realidad. Los modelos se utilizan para entender, analizar y predecir el comportamiento de dicho sistema o proceso.

Para ilustrarlo, pensemos en un mapa como un ejemplo de modelo. Un mapa es una representación simplificada de un mundo tridimensional en un espacio bidimensional. Este modelo simplificado nos permite entender mejor nuestra ubicación y facilita la navegación en el entorno, prescindiendo de detalles innecesarios como la textura del terreno o el color de los edificios. A pesar de que el mapa no captura la totalidad del mundo real, es increíblemente útil porque destaca la información que más nos importa, por ejemplo, las distancias relativas y las ubicaciones de los lugares.

En esencia, la creación de un modelo se trata de encontrar el equilibrio correcto entre simplicidad y utilidad. Un modelo eficaz debe ser lo suficientemente simple para ser comprensible y manejable, pero también debe ser lo suficientemente complejo para capturar los aspectos más importantes del sistema que se está modelando. Los mejores modelos son los que nos permiten hacer predicciones precisas y útiles a pesar de su simplicidad.

Además, a los modelos se les puede dotar de probabilidad para atender ciertos grados de incertidumbre, dando como resultado los modelos probabilísticos.

Si tenemos datos que representan nuestra realidad y poseemos información acerca de los datos “reales” y los datos que capta “nuestro modelo” podemos afirmar que la diferencia entre los datos que capta nuestro modelo y los datos reales se denomina como “error”, la idea es que ajustando los parámetros que definen nuestro modelo acercarnos mucho más a la realidad.

Los tres elementos fundamentales son:

* Los datos: toma de contacto con la realidad, mediciones.
* Los parámetros: son aquellos que nos dan la flexibilidad para ajustar nuestro modelo y hacer que se parezca lo máximo posible a la realidad. Aunque no siempre es bueno dotar al modelo de tanta flexibilidad.
* El error: Definir una función de error que nos diga como nuestro modelo se ajusta o no a los datos de la realidad. Normalmente cuando se usan algoritmos de aprendizaje supervisado esta función de error se calcula a partir de los datos de salida suministrados y en el caso de aprendizaje no supervisado otras medidas se computan a partir de los datos de entrada. Esta función es importante porque nos sirve para hacer “optimización” y para hacer un entrenamiento o ajuste del modelo.

Enlaces:

* https://www.youtube.com/watch?v=Sb8XVheowVQ

## Regresiones lineales.

### Modelo de regresión lineal simple.

### Modelo de regresión lineal con múltiples variables.

## Definición de una neurona a nivel matemático.

### La neurona como función

Una neurona es la unidad básica de procesamiento más simple que nos vamos a encontrar en una red neuronal, en términos simples una neurona recibirá una entrada que denotaremos como “x” y generará una salida denotada como “y” dependiendo de los valores de entrada, comportándose exactamente igual que una función. Ahora bien, la función que representa la neurona viene definida por:

En esta función matemática podemos apreciar una ecuación lineal en el interior de otra función , objeto de estudio más adelante, en la que la variable de entrada tiene como producto una variable “w” que llamaremos peso (**w**eight) a partir de ahora, esta variable nos dice, cuanto de importante es para la neurona el valor de esa entrada, es decir ¿Cuál es su peso o su importancia? Ahora puede que no le veamos mucho el sentido, pero cuando la neurona reciba varias entradas, a cada una de estas entradas se le dará más o menos importancia en función de su peso asociado.

### El papel del Bias

Para entender el papel del 'Bias' (b), consideremos un caso práctico. Imagina que nuestra neurona controla un sistema de calefacción, produciendo una salida de 1 para activar la calefacción y 0 para desactivarla. La entrada 'x' es la temperatura actual de la casa, y queremos que la calefacción se active cuando la temperatura caiga por debajo de los cinco grados. Ahora, si la temperatura llega a 0 grados, nuestra función sin el término de 'Bias', se simplifica a , lo que sugiere que la calefacción debería apagarse, lo cual claramente no es lo que queremos. Para tratar estos casos, introducimos el término de Bias 'b' en la ecuación lineal, resultando en , lo que nos permite ajustar la salida de nuestra neurona incluso cuando la entrada es 0, asegurando así que nuestra calefacción se active cuando la temperatura es críticamente baja.

### La función de activación.

## Definición matemática formal de una red neuronal.

**Calificación de una película:** Imagina que tienes una red neuronal que estás entrenando para calificar películas según tus preferencias. La función de activación podría representar la probabilidad de que te guste una película, y la entrada podría ser características como el género, el director, los actores, etc. Ahora bien, si eres un gran fanático del cine de ciencia ficción, querrás que tu red neuronal esté más predispuesta a calificar positivamente las películas de este género. Aquí es donde el sesgo entra en juego: al ajustar el sesgo, puedes "desplazar" la función de activación hacia la derecha, haciendo que sea más probable que la red neuronal califique positivamente una película incluso antes de considerar las características específicas.

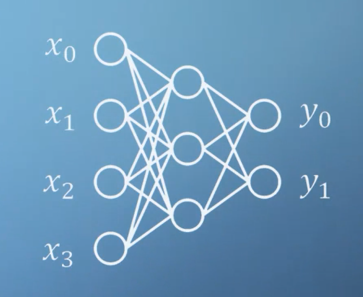
puede definirse con la siguiente expresión matemática:

De tal forma que nos queda que:

* Este es el estado de entrada a la neurona. En otras palabras, es el valor o valores (en forma de vector) que la neurona está recibiendo como entrada.
* **:** Este es el estado de salida de la neurona. Es el resultado de la función de activación aplicada a la suma ponderada de las entradas a la neurona más el sesgo.
* Esta es la función de activación de la neurona. Actúa como una especie de umbral, determinando si la neurona se "activa" o no en base a la entrada ponderada que recibe. Las funciones de activación son fundamentales en las redes neuronales porque permiten introducir no linealidad en el modelo, permitiendo que este aprenda y modele relaciones complejas.
* **W(weight):** Esta es la variable que representa el peso de la conexión entre dos neuronas. En términos simples, el peso determina cuánto contribuye una neurona a la siguiente. Es un parámetro ajustable que la red neuronal aprende durante su entrenamiento.
* **B (Bias):** Este término se refiere al sesgo o "bias". Funciona como una constante que se suma a la entrada ponderada antes de pasar por la función de activación. El sesgo permite ajustar la salida de la neurona junto con el peso, proporcionando un grado adicional de libertad al modelo.

## Definición de una red neuronal a nivel matemático.

Una red neuronal puede entender como una función, que dada unas entradas genera una salida, tanto la entrada como la salida pueden ser vectores de entrada-salida.

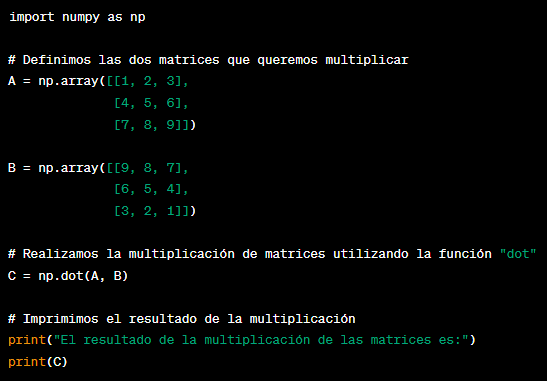


Es decir, la anterior imagen podría representarse como:

**Nota:** recordar que la “x” e “y” en formato negrita representan una tupla de variables, además, el número de variables de entrada “m” puede ser diferente del número de variables de salida “n”

# Python y las matemáticas.

## Multiplicación de producto punto de matrices en numpy.



## Vectores y valores propios en Python.

