Matrices y Machine Learning

Apuntes y Ejercicios

Benja Vera

$$f(x) = \sigma(Wx + b)$$

Versión actualizada al
29 de agosto de 2024

Contenidos

Capítulo 1

Introducción

Este pequeño curso tiene el objetivo de explorar los conceptos principales de la matemática de primeros años de universidad (esto es, cálculo en una variable, álgebra lineal y elementos del cálculo en varias variables) guiado por el ejemplo/objetivo de construir con ellos código que permita entrenar una red neuronal. Así, a nivel transversal, con este curso buscamos lograr varias cosas distintas.

lalalal

1.1. ¿Qué es una red neuronal?

1.2. Aprendizaje y minimización

1.3. Hoja de ruta

Capítulo 2

Cálculo en una Variable

En este primer capítulo, se resumen los elementos principales del cálculo en una variable que son necesarios para desarrollar la teoría de la optimización. Los dos algoritmos principales de minimización que se exploran son el **método de descenso** y el **método de Newton**. Siendo el primero sin duda el más utilizado en el mundo del machine learning hoy en día, aunque algunos autores han sugerido utilizar variaciones del otro (ver por ejemplo [Le+11]). En la sección 2.1 se desarrolla el concepto intuitivo de la derivada de una función, junto con un resumen de las propiedades que esperaríamos que esta cumpliera. Luego, en la sección 2.2 hablamos más bien de cómo realmente calcular la derivada de una función, estas ideas se enlazan con la sección anterior a través de un ejemplo de cómo las primeras dos derivadas de una función se pueden utilizar para esbozar su gráfico completo, el capítulo termina con una discusión sobre la **regla de la cadena**, concepto fundamental para lo que sigue. Después de esto, la sección 2.3 aplica lo discutido sobre la regla de la cadena para construir los llamados **polinomios de Taylor** asociados a una función, los cuales vamos interpretar como buenas aproximaciones de la función cerca de un cierto punto x_0. El capítulo termina en la sección 2.4, la cual expone los dos algoritmos principales de optimización mencionados al inicio.

Las ideas mencionadas en este capítulo se encuentran más detalladas, por ejemplo, en [GOM23]. Pero para una introducción completamente rigurosa al cálculo de una variable, se recomienda ver lo expuesto en [SM88]

2.1. La derivada y sus propiedades

2.2. Las reglas de derivación

2.3. Polinomios de Taylor

2.4. Algoritmos de optimización

2.5. Ejercicios

2.5.1. Preliminares

1. Analice, mediante una tabla, los signos de las siguientes expresiones ya factorizadas
   a) $\frac{x+1}{x-1}$
   b) $(x+2)(x-3)$
   c) $(x+4)(1-x)$
   d) $\frac{(x+1)(x-1)}{x-3}$
2. Factorice las siguientes expresiones y luego analice sus signos como en el item anterior
   a) $\frac{x^2-4}{x+1}$
   b) x^2-4x+3
   c) x^2-3x

2.5.2. Cálculo de derivadas por definición

Entre las dos definiciones de derivada que hemos visto:

$$f'(x) = \lim_{h\to 0} \frac{f(x+h)-f(x)}{h}, \qquad f'(x) = \lim_{w\to x} \frac{f(w)-f(x)}{w-x}$$

Utilice la que más le convenga para calcular las derivadas de las siguientes funciones

1. $f(x) = \frac{1}{x}$
2. $f(x) = \sqrt{x}$
3. $f(x) = x^3$

2.5.3. Reglas de derivación sin regla de la cadena

Utilizando las reglas de derivación, calcule las derivadas de las siguientes funciones

1. $f(x) = x^4$
2. $f(x) = 3x^5 - x^3$
3. $f(x) = 4x^2 - 3\pi^2$
4. $f(x) = (x+1)(x-1)$
5. $f(x) = \frac{x+1}{x-1}$
6. $f(x) = \frac{1}{\sqrt{x}}$

2.5.4. La regla de la cadena

Utilizando las reglas de derivación conocidas, además de la regla de la cadena, obtenga las derivadas de las siguientes funciones.

1. $f(x) = (1+\sqrt{x})^2$
2. $f(x) = \sqrt[3]{2x}$
3. $f(x) = (4+2x)^{2023}$
4. $f(x) = \sqrt{1-x^2}$
5. $f(x) = (\frac{1}{x} + x)^3$
6. $f(x) = (1-x)^5$

2.5.5. Problemas varios sobre rectas tangentes

Para esta sección, recuerde que la recta tangente al gráfico de una función f en el punto $(x_0, f(x_0))$ viene dada por la ecuación

$$y = f(x_0) + f'(x_0)(x - x_0)$$

Además, recordemos también el hecho de que si L_1 y L_2 son dos rectas dadas respectivamente por

$$y = m_1 x + n_1, \qquad y = m_2 x + n_2$$

Entonces estas rectas son perpendiculares cuando $m_1 \cdot m_2 = -1$

1. Considere el gráfico de la función $f(x) = \frac{1}{x}$ y sea $x_0 > 0$ fijo.
   a) Encuentre la ecuación de la recta tangente al gráfico de f en el punto $(x_0, 1/x_0)$. Debería obtener una recta cuyos parámetros vienen escritos en términos de x_0.
   b) Obtenga los puntos de intersección entre la recta obtenida anteriormente y los ejes de coordenadas. Recuerde que esto se puede hacer imponiendo $x = 0$ y $y = 0$ según corresponda. Debería obtener dos puntos cuyas coordenadas vienen dadas en términos de x_0.
   c) Al dibujar lo que está sucediendo, notará que se forma un triángulo (dado por la recta tangente y los ejes de coordenadas). Calcule su área ¿Qué puede decir con respecto a este área?
2. En este problema, vamos a probar el hecho de que las rectas tangentes a los círculos son siempre perpendiculares al radio. Hecho que ya se conoce de la geometría euclideana. Recordemos para esto que la ecuación que define a una circunferencia de radio 1 centrada en el punto $(0, 0)$ viene dada por

   $$x^2 + y^2 = 1$$

   De modo que depejando y, la función que define al semicírculo superior de esta circunferencia viene dada por

   $$f(x) = \sqrt{1 - x^2}$$

   Dado $x_0 \in]-1, 1[$ fijo, proceda como sigue:
   a) Considere el radio que une el origen con el punto $(x_0, f(x_0))$. Conociendo dos puntos, calcule la pendiente de este radio.
   Nota: Es posible calcular también la ecuación de la recta que define a este radio, pero no la necesitamos realmente.
   b) Calcule además la derivada de la función f en el punto x_0.
   c) Interpretando sus resultados anteriores, concluya lo pedido.

2.5.6. Máximos y mínimos

1. Pruebe, utilizando lo expuesto en este capítulo, el hecho ya conocido de que si $f(x) = ax^2 + bx + c$ es una función cuadrática cualquiera con $a \neq 0$, entonces el punto $x_0 = \frac{-b}{2a}$ corresponde a un mínimo si $a > 0$ y un máximo si $a < 0$.

2. Se desea encontrar el punto $P = \left(x, \frac{1}{x}\right)$ con $x > 0$ (construido así de modo que P pertenece al gráfico de la función $f(x) = 1/x$) cuya distancia euclideana al origen de coordenadas sea mínima. Para esto, y teniendo en cuenta la siguiente figura, proceda como sigue:

   [FIGURA]

   a) Pruebe mediante el teorema de pitágoras que para $x > 0$ fijo, la distancia viene dada por

   $$d(x) = \sqrt{x^2 + \frac{1}{x^2}}$$

   b) Obtenga la derivada de esta función $d'(x)$ y resuelva la ecuación $d'(x) = 0$. Llamemos x_0 al valor obtenido.

   c) Con este valor en mano, calcule $d''(x_0)$ y concluya.

2.5.7. Gráficos de funciones

2.5.8. Polinomios de Taylor

Capítulo 3

Álgebra Lineal

3.1. Sesión 1: Introducción

Consideremos el siguiente problema ilustrativo:

Problema introductorio

Sobre una línea recta se desplazan dos móviles A y B siendo observados por un observador O ubicado entre ellos. El móvil A inicia su viaje a 1 [m] de distancia hacia la derecha de O, y desplazándose a 1 [m/s] hacia la derecha, mientras que el movil B inicia su viaje a 1 [m] hacia la izquierda, desplazándose a 2 [m/s] hacia la derecha. Encuentre el instante y ubicación en los cuales los móviles se encuentran.

Para plantear matemáticamente este problema, podemos considerar las distancias hacia la derecha como *positivas* y hacia la izquierda como *negativas*. Estableciendo así lo que en física se conoce como un **sistema de referencia**. De este modo, las ecuaciones de movimiento de los dos móviles se pueden escribir como sigue:

$$x_A(t) = 1 + t$$
$$x_B(t) = -1 + 2t$$

Buscamos entonces t^* tal que $x_A(t^*) = x_B(t^*) := x^*$. Estas dos incógnitas (x^*, t^*) son entonces soluciones al siguiente **sistema de ecuaciones**

$$x - t = 1$$
$$x - 2t = -1$$

Entonces, podemos definir lo siguiente:

Definición 3.1 ($\mathbb{R}^2$, igualdad, suma, ponderación, producto matriz-vector)**.** Los vectores de 2 dimensiones los definimos como sigue:

- Definimos $\mathbb{R}^2$ como el conjunto de los pares ordenados de números, que en adelante denotaremos verticalmente como:

$$\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2$$

- Se define la igualdad entre vectores como sigue:

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} z \\ w \end{pmatrix} \iff (x = z) \wedge (y = w)$$

- Se define la suma entre vectores como:

$$\begin{pmatrix} x \\ y \end{pmatrix} \pm \begin{pmatrix} z \\ w \end{pmatrix} = \begin{pmatrix} x \pm z \\ y \pm w \end{pmatrix}$$

- También se define el producto escalar o *ponderación* de vectores como:

$$\lambda \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \lambda x \\ \lambda y \end{pmatrix}$$

- Y finalmente, se define el *producto matriz-vector* como:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = x \begin{pmatrix} a \\ c \end{pmatrix} + y \begin{pmatrix} b \\ d \end{pmatrix}$$

Con estas definiciones, el sistema de ecuaciones antes mencionado se puede escribir como el problema de encontrar (x, t) tales que:

$$\begin{pmatrix} 1 & -1 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} x \\ t \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

Y esto se puede entender como un problema simple de *text* una función. Es decir, para $f : A \to B$ e $y \in B$, encontrar $x \in A$ tal que $f(x) = y$. El vector al lado derecho juega el rol de y, y la matriz juega el rol de la función f. Esto nose lleva entonces a preguntarnos por las propiedades que puede tener la función $f(x) = Ax$ siendo A una matriz y x un vector. Esto será tema de la siguiente sesión.

3.1.1. Problemas teóricos

Antes de pasar al estudio de las propiedades del producto matriz-vector, haremos las definiciones de una manera más general. Para ello, primero comprenda las siguientes definiciones.

Definición 3.2 ($\mathbb{R}^n$, igualdad, suma, ponderación, producto matriz-vector)**.** Los vectores en n dimensiones los definimos como sigue:

- Definimos $\mathbb{R}^n$ como el conjunto de las n-tuplas ordenadas de números, las cuales denotaremos como:

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$$

- Definimos la igualdad entre vectores en $\mathbb{R}^n$ como:

$$x = y \iff \forall i \in \{1, \ldots, n\} : x_i = y_i$$

- Definimos la suma entre vectores de $\mathbb{R}^n$ como:

$$(x \pm y)_i = x_i \pm y_i$$

- Definimos el producto escalar o *ponderación* para vectores de $\mathbb{R}^n$ como:

$$(\lambda x)_i = \lambda x_i$$

- Las matrices las entendemos como *bloques* de números reales, que en general no serán cuadrados sino rectangulares. Estos bloques los denotaremos como:

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \in \mathbb{R}^{m \times n}$$

Diremos que la matriz A tiene n columnas y m filas. Las cuales frecuentemente querremos aislar y considerar como vectores en su propio mérito (ver la definición dada anteriormente para el producto matriz-vector). Por lo tanto, definiremos los vectores columna $A_{\bullet j} \in \mathbb{R}^m$ para $j \in \{1, \dots, n\}$ por $(A_{\bullet j})_i = a_{ij}$. Además, definimos los vectores fila $A_{i\bullet} \in \mathbb{R}^n$ para $i \in \{1, \dots, m\}$ por $(A_{i\bullet})_j = a_{ij}$. En otras palabras:

$$A = \begin{pmatrix} A_{\bullet 1} & \dots & A_{\bullet n} \end{pmatrix} = \begin{pmatrix} A_{1\bullet} \\ \vdots \\ A_{m\bullet} \end{pmatrix}$$

- Con estas nociones, se define el producto matriz-vector para un vector $x \in \mathbb{R}^n$ y una matriz $A \in \mathbb{R}^{m \times n}$ por

$$Ax = x_1 A_{\bullet 1} + \dots + x_n A_{\bullet n} = \sum_{j=1}^{n} x_j A_{\bullet j} \in \mathbb{R}^m$$

Es decir, $A \in \mathbb{R}^{m \times n}$ convierte vectores de $\mathbb{R}^n$ en vectores de $\mathbb{R}^m$.

Para asegurar que comprendemos estas definiciones, realice los siguientes ejercicios:

1. Considere la siguiente matriz como ejemplo

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

Encuentre:

- a_{13}
- a_{31}
- a_{21}
- a_{12}
- $A_{1\bullet}$
- $A_{\bullet 1}$
- $A_{\bullet 3}$
- $A_{2\bullet}$

2. Realice el siguiente cálculo de multiplicación matriz-vector:

$$\begin{pmatrix} 0 & 1 & 0 \\ 2 & 1 & 1 \\ 3 & 2 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 4 \end{pmatrix}$$

3. **(Matrices elementales)** Realice los siguientes productos:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & \lambda \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}$$

Describa en sus palabras cómo *actúa* la matriz sobre el vector en cada caso. Matrices de este tipo van a ser importantes a la hora de estudiar los sistemas de ecuaciones lineales.

3.1.2. Problemas computacionales

Ver notebook número 4.

3.2. Sesión 2: Linealidad

En base al experimento computacional realizado en la sesión anterior, en que pudimos visualizar con algunos ejemplos cómo las matrices *actúan* sobre el espacio transformando los vectores sobre los cuales se aplican, pudimos observar que esta acción de las matrices sobre el espacio es en algún sentido *uniforme*. Las observaciones geométricas que son relevantes de realizar son las siguientes:

- El orígen está fijo bajo la transformación.
- La transformación convierte rectas paralelas en rectas paralelas.
- Las distancias entre puntos se mantienen proporcionales bajo la transformación.

Estas propiedades geométricas tienen reflejos algebráicos que se pueden resumir en la siguiente proposición.

Proposición 3.3. *Sea $A \in \mathbb{R}^{m\times n}$ y $x, y \in \mathbb{R}^n, \lambda \in \mathbb{R}$. Entonces, se cumple que:*

1. $A0 = 0(\in \mathbb{R}^m)$
2. $A(\lambda x) = \lambda(Ax)$
3. $A(x + y) = Ax + Ay$

Demostración. Recordemos la definición de producto matriz-vector.

$$Ax = \sum_{i=1}^{m} x_i A_{\bullet i}$$

De este modo:

1. Si $x_i = 0 \forall i \in \{1, \ldots, n\}$, entonces $Ax = \sum_{i=1}^{n} x_i A_{\bullet i} = 0$.

2. $A(\lambda x) = \sum_{i=1}^{n} \lambda x_i A_{\bullet i} \overset{(1)}{=} \lambda \sum_{i=1}^{n} x_i A_{\bullet i}$

3. $A(x+y) = \sum_{i=1}^{n}(x+y)_i A_{\bullet i} = \sum_{i=1}^{n}(x_i+y_i)A_{\bullet i} = \sum_{i=1}^{n} (x_i A_{\bullet i} + y_i A_{\bullet i}) \overset{(2)}{=} \sum_{i=1}^{n} x_i A_{\bullet i} + \sum_{i=1}^{n} y_i A_{\bullet i} = Ax + Ay.$

En que se han utiliado las siguientes propiedades conocidas de las sumatorias:

(1) $\sum_{i=1}^{n} cb_i = c \sum_{i=1}^{n} b_i$

(2) $\sum_{i=1}^{n}(a_i + b_i) = \sum_{i=1}^{n} a_i + \sum_{i=1}^{n} b_i$

■

Nota. La primera propiedad $A0 = 0$ puede ser vista como una consecuencia de la segunda tomando $\lambda = 0$, por lo que las propiedades fundamentales son estas últimas dos.

Definición 3.4. (Linealidad) Decimos que una función $f : \mathbb{R}^n \to \mathbb{R}^m$ es *lineal* si cumple

1. $\forall x \in \mathbb{R}^n, \lambda \in R : f(\lambda x) = \lambda f(x)$.
2. $\forall x, y \in \mathbb{R}^n : f(x + y) = f(x) + f(y)$.

En particular, en vista de lo probado anteriormente, si $f : \mathbb{R}^n \to \mathbb{R}^m$ es del tipo $f(x) = Ax$ para alguna $A \in \mathbb{R}^{m\times n}$ dada, entonces f es lineal. Ejemplos de funciones lineales son las siguientes:

- $f : \mathbb{R} \to \mathbb{R}$ por $f(x) = ax \ (a \in \mathbb{R})$
- $f : \mathbb{R} \to \mathbb{R}^2$ por $f(x) = \begin{pmatrix} x \\ -2x \end{pmatrix}$
- $f : \mathbb{R}^2 \to \mathbb{R}$ por $f(x, y) = 2x + y$
- $f : \mathbb{R}^2 \to \mathbb{R}^2$ por $f(x, y) = \begin{pmatrix} x + y \\ x - y \end{pmatrix}$

Una función no lineal es una función que no es lineal. Es decir, tal que alguna de las dos condiciones que definen la linealidad no se cumple, ya sea para algún $x \in \mathbb{R}^n, \lambda \in \mathbb{R}$ o bien para algún par $x, y \in \mathbb{R}^n$. Ejemplos de funciones no lineales son las siguientes:

- $f : \mathbb{R} \to \mathbb{R}$ por $f(x) = x^2$
- $f : \mathbb{R} \to \mathbb{R}^2$ por $f(x) = \begin{pmatrix} 1 \\ x \end{pmatrix}$
- $f : \mathbb{R}^2 \to \mathbb{R}$ por $f(x, y) = (x + y)^2$
- $f : \mathbb{R}^2 \to \mathbb{R}^2$ por $f(x, y) = \begin{pmatrix} x^2 - y^2 \\ 1 - x \end{pmatrix}$

Obsérvese que todos los ejemplos de funciones lineales dados, si bien no los escribimos utilizando matrices, pueden ser escritos utilizándolas. Por ejemplo:

$$f(x,y) = \begin{pmatrix} x+y \\ x-y \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

El resultado principal de esta sesión consiste en afirmar que no existe otra opción. Es decir, que la multiplicación matricial agota todas las posibles funciones lineales. Eso nos lleva al siguiente importante teorema.

Teorema 3.5. *(Representación Matricial) Si $f : \mathbb{R}^n \to \mathbb{R}^m$ es lineal, entonces existe una única $A \in \mathbb{R}^{m\times n}$ tal que para todo $x \in \mathbb{R}^n$:*

$$f(x) = Ax$$

Diremos frecuentemente que esta matriz A *representa* a la función f. Algo que es aún más interesante sobre este teorema es que la matriz A se puede *construir*. Es decir, no es un objeto cuya existencia está garantizada pero al que no podemos acceder explícitamente, la fórmula para obtener esta matriz la veremos en la demostración a continuación.

Demostración. Para facilidad de notación, consideremos el símbolo *delta de kronecker* definido por

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Con esto en mente, definimos el conjunto de n vectores de $\mathbb{R}^n$ $\{e_i\}_{i=1}^n$ que vienen definidos por $(e_i)_j = \delta_{ij}$. En otras palabras, el vector e_i es aquel que tiene ceros salvo en la i-ésima posición en que vale 1. Notemos que cualquier $x \in \mathbb{R}^n$ se puede escribir como

$$x = \sum_{i=1}^n x_i e_i$$

De este modo, si f es lineal, entonces podemos desarrollar lo siguiente:

$$f(x) = f\left(\sum_{i=1}^n x_i e_i\right) = \sum_{i=1}^n f(x_i e_i) = \sum_{i=1}^n x_i f(e_i)$$

Y esto puede ser visto como un producto matriz vector Ax cuando (y solo cuando) definamos A de modo que $A_{\bullet i} = f(e_i)$. En otras palabras, la matriz representante de f tiene como columna i-ésima a $f(e_i)$. De este modo, se obtiene $f(x) = Ax$ y se cumple lo pedido. ∎

Nota. La matriz representante puede ser ilustrada del siguiente modo:

$$A = \begin{pmatrix} f(e_1) & \dots & f(e_n) \end{pmatrix}$$

Notar en particular que esto significa que una función lineal viene únicamente determinada (a través de su matriz representante) por sus valores en los n vectores $\{f(e_i)\}_{i=1}^n$. Esto captura la *uniformidad* geométrica que se observó en la sesión anterior, ya que el valor de la función en un x arbitrario puede ser obtenido en base a *conectar los puntos* a partir de estos n vectores.

3.2.1. Problemas teóricos

1. Considere la función $f : \mathbb{R}^2 \to \mathbb{R}^2$ dada por

$$f(x, y) = \begin{pmatrix} y \\ x \end{pmatrix}$$

   a) Pruebe que f es lineal.
   b) Obtenga la matriz $A \in \mathbb{R}^{2\times 2}$ que representa a f.

2. Haga lo mismo que realizó en el item anterior, pero esta vez con la función $f : \mathbb{R}^n \to \mathbb{R}$ dada por

$$f(x_1, \ldots, x_n) = x_1 + x_2 + \cdots + x_n$$

3. Consideremos $R_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ tal que $R_\theta(x)$ es la rotación del vector x en un ángulo θ con respecto del orígen. Se puede verificar (no lo haga, pero convénzase de que es cierto) que $\mathbb{R}_\theta$ es una función lineal. En consecuencia, tiene una matriz representante. Encuéntrela.

Agregar imagen

3.3. Sesión 3: Operaciones matriciales

La sesión anterior nos ha convencido de que existe un vínculo estrecho entre las matrices y las funciones lineales (es más, son lo mismo). Concretemos un poco ese vínculo en términos de matemática abstracta. Consideremos el conjunto

$$\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) = \{f : \mathbb{R}^n \to \mathbb{R}^m : f \text{ lineal}\}$$

Y consideremos la función

$$\begin{aligned} T :&\mathbb{R}^{m\times n} \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) \\ &A \mapsto T(A) = f_A \end{aligned}$$

En que $f_A(x) = Ax$ es la función lineal asociada a la matriz A.

En términos que ya conocemos, el teorema de representación matricial nos dice dos cosas: Ya que la matriz representante existe, la función T es *epiyectiva*, y ya que es única, es *inyectiva*. Entonces, la función T es *biyectiva* y es por lo tanto una correspondencia uno-a-uno entre las matrices y las funciones lineales. Existe entonces una *función inversa* $T^{-1} : \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m) \to \mathbb{R}^{m\times n}$ tal que $T^{-1}(f) = A$ tal que $T(A) = f$. Esta función inversa es precisamente la fórmula que dimos para la matriz representante. En rigor, aquí no hemos dicho nada nuevo, ya que todo el contenido está en el teorema de representación, pero es importante de todos modos darle un sentido a los términos comunes de funciones inversas.

Recordemos que sobre las funciones lineales (y en realidad sobre todas las funciones) tenemos las siguientes operaciones definidas.

Definición 3.6. (Suma, resta, producto escalar, composición) Sean $f, g : \mathbb{R}^n \to \mathbb{R}^m$ lineales y $\lambda \in \mathbb{R}$. Entonces:

1. Definimos la suma y resta $f \pm g : \mathbb{R}^n \to \mathbb{R}^m$ por

$$(f \pm g)(x) = f(x) \pm g(x)$$

2. Definimos la ponderación escalar $\lambda f : \mathbb{R}^n \to \mathbb{R}^m$ por

$$(\lambda f)(x) = \lambda f(x)$$

3. Si $f : \mathbb{R}^n \to \mathbb{R}^m$ y $g : \mathbb{R}^m \to \mathbb{R}^p$ lineales, entonces definimos $g \circ f : \mathbb{R}^n \to \mathbb{R}^m$ por

$$(g \circ f)(x) = g(f(x))$$

Algo interesante de la linealidad es que se preserva a través de estas definiciones. Es decir, tenemos la siguiente proposición:

Proposición 3.7. *La suma, resta, ponderación y composición de funciones lineales es una función lineal.*

Demostración. Queda propuesta como ejercicio. ■

Esto significa que el conjunto $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ tiene estructura, tiene operaciones definidas sobre sus elementos. Por lo tanto, podemos *trasladar* estas operaciones hacia las matrices. Por ejemplo, nos gustaría definir el objeto $A + B$ siendo A, B matrices. Y nos gustaría que esta suma tuviera alguna relación con la suma de las funciones $f_A + f_B$ asociadas. Esta función suma es lineal por la última proposición. En consecuencia, tiene una matriz representante. Nos gustaría definir esta como la *suma entre las matrices* A y B. De este modo, se cumpliría que *la matriz representante de la suma es la suma de las matrices representantes.* Ya que esta suma de funciones $f_A + f_B$ solo se puede hacer si ambas tiene la forma $\mathbb{R}^n \to \mathbb{R}^m$, entonces la suma solo se puede definir para matrices de iguales tamaños.

Repitamos lo mismo esta vez con la composición: Sean A, B matrices (de tamaños que vamos a descubrir prontamente). Estas dos tienen asociadas funciones f_A y f_B lineales, las cuales se pueden componer formando $f_A \circ f_B$. Esta composición solo tiene sentido si $f_B : \mathbb{R}^n \to \mathbb{R}^m$ y $f_A : \mathbb{R}^m \to \mathbb{R}^p$, de modo que necesitamos $B \in \mathbb{R}^{m\times n}$ y $A \in \mathbb{R}^{p\times m}$. Y definimos el *producto matricial* entre las matrices A y B (denotado como AB) como aquella matriz que representa a la función lineal $f_A \circ f_B$.

Esto nos entrega la siguiente definición:

Definición 3.8. (suma, resta, ponderación, producto de matrices) Para $A, B \in \mathbb{R}^{m\times n}$ y $\lambda \in \mathbb{R}$, definimos

- $A \pm B \in \mathbb{R}^{m\times n}$ como la matriz representante de la función lineal $f_A \pm f_B$.
- $\lambda A \in \mathbb{R}^{m\times n}$ como la matriz representante de la función λf_A.
- Para $A \in \mathbb{R}^{m\times p}, B \in \mathbb{R}^{p\times n}$, definimos $AB \in \mathbb{R}^{m\times n}$ como la matriz representante de la función $f_A \circ f_B$.

Esta definición es natural desde un punto de vista teórico, pero altamente poco amigable. Ya que no nos deja calcular explícitamente ninguna de estas operaciones sin pasar por pensar en funciones y matrices representantes. Por suerte, las podemos explicitar fácilmente utilizando las fórmulas que hemos obtenido hasta el momento. Observar, por ejemplo para la suma y la ponderación, que en términos de la función T introducida al inicio de la sesión, estamos definiendo[1]

$$\begin{aligned} A + B &= T^{-1}(T(A) + T(B)) \\ \lambda A &= T^{-1}(\lambda T(A)) \end{aligned}$$

[1]Esto esconde grandes cosas. Aplicando T a ambos lados, lo que se observa es que estamos definiendo las operaciones para que se cumpla

$$\begin{aligned} T(A + B) &= T(A) + T(B) \\ T(\lambda A) &= \lambda T(A) \end{aligned}$$

Es decir, las operaciones se introducen de modo que la biyección en sí misma T sea *lineal*. A las funciones lineales biyectivas les llamamos *isomorfismos*, y esto permite frasear lo que hemos estado haciendo de la siguiente forma: Estamos dando a $\mathbb{R}^{m\times n}$ la misma estructura algebraica presente en el espacio de funciones $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ ya que el teorema de representación matricial nos decía que estaban en correspondencia uno a uno.

De modo que aplicando la fórmula para la matriz representante, tenemos entonces que

$$A+B=\begin{pmatrix}(f_A+f_B)(e_1) & \dots & (f_A+f_B)(e_n)\end{pmatrix}=\begin{pmatrix}f_A(e_1)+f_B(e_1) & \dots & f_A(e_n)+f_B(e_n)\end{pmatrix}$$

$$=\begin{pmatrix}Ae_1+B_{e_1} & \dots & Ae_n+Be_n\end{pmatrix}=\begin{pmatrix}A_{\bullet 1}+B_{\bullet 1} & \dots & A_{\bullet n}+B_{\bullet n}\end{pmatrix}$$

En términos de cada una de las entradas, esto se ve mucho más simplemente como

$$(A+B)_{ij}=A_{ij}+B_{ij}$$

Similarmente, se puede ver que

$$(\lambda A)_{ij}=\lambda A_{ij}$$

Podríamos comenzar a pensar que todo este asunto de pasar por matemática abstracta y definir a través de matrices representantes fue una pérdida de tiempo solo para definir cosas que de todos modos eran intuitivas, pero ahora consideremos el producto matricial.

De acuerdo a la definición, tenemos entonces

$$AB=\begin{pmatrix}(f_A\circ f_B)(e_1) & \dots & (f_A\circ f_B)(e_n)\end{pmatrix}=\begin{pmatrix}f_A(f_B(e_1)) & \dots & f_A(f_B(e_n))\end{pmatrix}$$

$$=\begin{pmatrix}f_A(B_{\bullet 1}) & \dots & f_A(B_{\bullet n})\end{pmatrix}=\begin{pmatrix}AB_{\bullet 1} & \dots & AB_{\bullet n}\end{pmatrix}$$

En otras palabras, la columna j-ésima del producto AB es la matriz A aplicada sobre el vector columna j-ésimo de B. Explícitamente:

$$(AB)_{\bullet j}=\sum_{k=1}^{p}B_{kj}A_{\bullet k}$$

Y en términos de las entradas individuales,

$$(AB)_{ij}=\sum_{k=1}^{p}A_{ik}B_{kj}$$

Esta fórmula es la más conocida para el producto matricial y definitivamente no es la primera que se nos hubiese ocurrido al definirlo. Lo común en los cursos de álgebra lineal es comenzar dando estas definiciones y luego probar que *mágicamente* se cumplen las cosas que ya hemos explorado (y otras propiedades que veremos más adelante), pero el camino más transparente es el que hemos hecho, en que las definiciones están al servicio de que se cumplan las propiedades.

En la siguiente sesión exploraremos el concepto de una *matriz inversa* y con ello comenzaremos nuestro camino hacia el problema original de despejar x en la ecuación $Ax=b$, camino que llamaremos el de *invertir* una matriz A.

3.3.1. Problemas teóricos

1. Demuestre, la proposición 3.7.

2. Realice los siguientes productos matriciales

1

2

3

Deduzca de los últimos dos que el producto matricial no es conmutativo.

3. Encuentre dos matrices $A, B \in \mathbb{R}^{2\times 2}$ no nulas tales que $AB = 0$.

A lo largo de estos siguientes problemas, vamos a utilizar nuestros resultados abstractos para demostrar de manera breve propiedades aritméticas de matrices que no son obvias y cuyas demostraciones a partir de las definiciones tradicionales resultan engorrosas, técnicas y no muy iluminadoras.

4. Sean $f, g : \mathbb{R}^p \to \mathbb{R}^m$ y $h : \mathbb{R}^n \to \mathbb{R}^p$ funciones lineales. Pruebe que se cumple

$$(f + g) \circ h = (f \circ h) + (g \circ h)$$

5. Sean $f, g : \mathbb{R}^n \to \mathbb{R}^p$ y $h : \mathbb{R}^p \to \mathbb{R}^m$ funciones lineales. Pruebe que se cumple

$$h \circ (f + g) = (h \circ f) + (h \circ g)$$

Nota. Luego de hacer las demostraciones, notará que la linealidad solo se utiliza en el segundo problema y sobre h. Es decir, estas igualdades se cumplen en un contexto de funciones más general que aquel en el que todo es lineal. Esto es interesante, pero en este contexto solo lo utilizaremos para funciones lineales, por lo que el enunciado tiene supuestos más restrictivos.

6. Deduzca de los problemas anteriores que las siguientes identidades matriciales se cumplen

$$A(B + C) = AB + AC$$

$$(A + B)C = AC + BC$$

Siendo A, B, C matrices de los tamaños que correspondan.

7. Pruebe que el producto de matrices es *asociativo*. Es decir,

$$(AB)C = A(BC)$$

Para ello, recuerde que la composición de funciones es evidentemente asociativa (¿Por qué es evidente?).

Nota. Solo para comparar, si hubiésemos hecho esta última demostración utilizando la fórmula clásica del producto matricial, hubieramos tenido que probar la siguiente identidad de sumatorias dobles:

$$\sum_{k=1}^{m} C_{kj} \left(\sum_{l=1}^{p} A_{il} B_{lk} \right) = \sum_{l=1}^{p} A_{il} \left(\sum_{k=1}^{m} B_{lk} C_{kj} \right)$$

Bibliografía

[GOM23] Alejandro González, Harold Ojeda e Iván Morales. *Apunte de matemáticas 1*. Mar. de 2023.

[KF62] A.N. Kolmogorov y S.V Fomin. *Elementos de la Teoría de Funciones y del Análisis Funcional*. First. Editorial MIR, 1962.

[Le+11] Quoc V. Le et al. «On Optimization Methods for Deep Learning». En: *Proceedings of the 28th International Conference on International Conference on Machine Learning*. ICML'11. Bellevue, Washington, USA: Omnipress, 2011, págs. 265-272. ISBN: 9781450306195.

[SM88] M. Spivak y B.F. Marqués. *Cálculo Infinitesimal*. Reverté, 1988. ISBN: 9788429151367. URL: `https://books.google.cl/books?id=mjdXY8rshREC`.

Apéndice A

Consideraciones Topológicas

Este anexo está dedicado a detallar más formalmente algunos de los aspectos de cálculo en una y varias variables que para estas notas optamos por exponer de manera más superficial. Es decir, conceptos tales como lo que significa ser punto interior de un conjnuto, la definición formal de límites, la definición de continuidad y derivadas.

Aquí se exponen algunas de las preguntas que podrían surgir al leer estas notas y que se aboran en este apéndice:

- ¿Qué es lo que entendemos por $\lim_{x\to x_0} f(x)$? ¿Qué significa que esa expresión *exista* o *no exista*?
- ¿Podemos calcular un límite de cualquier función en cualquier punto?
- ¿Cuánto se pueden generalizar los conceptos de derivada y límite?
- ¿Por qué este capítulo se llama *consideraciones topológicas*? ¿Qué es la topología?

Los contenidos a ser tratados en este anexo son típicamente cubiertos en un curso de Análisis Real o bien en un buen curso de Cálculo, y para más información, puede consultarse el segundo capítulo de [KF62]

A.1. En los reales

Partamos por definir y caracterizar los límites de funciones $f : A \subseteq \mathbb{R} \to \mathbb{R}$, y la buena forma de iniciar esa discusión está en primero hablar de algo que pareciera ser completamente diferente, que son las sucesiones. Estas no son estríctamente necesarias, pero a lo largo de todo el anexo nos van a aportar un segundo punto de vista sobre las cosas, uno que siempre es útil tener.

Definición A.1 (sucesiones y convergencia)**.** Una sucesión de términos reales es una función $a : \mathbb{N} \to \mathbb{R}$ cualquiera. Sus términos los denotamos $a_1, a_2, \dots$ en lugar de $a(1), a(2), \dots$ por simplicidad. Decimos que $\lim_{n\to\infty} a_n = L$ si para todo $\varepsilon > 0$, existe un $n_0 \in \mathbb{N}$ tal que para todo $n \geq n_0$, se cumple $|a_n - L| < \varepsilon$. En este contexto, decimos que a_n *converge* o *tiende* a L cuando $n \to \infty$, situación que también se denota $a_n \to L$.

Ejemplo A.2.

- La sucesión $a_n = 1/n$ tiende a 0 cuando $n \to \infty$, hecho que se puede demostrar mediante la llamada *propiedad arquimediana* de los números reales, la cual es consecuencia del axioma del supremo[1].
- La sucesión $b_n = (-1)^n$ es una para la cual ningún número real L satisface la definición, de modo que decimos que el límite de b_n *no existe*.

[1] Para esta demostración, ver este video.

- La convergencia a un límite no necesita ser monótona (es decir, desde un solo lado), como ejemplo, considérese la sucesión $c_n = (-1)^n \cdot \frac{1}{n}$, que tiende a cero de manera alternante.

Teniendo estos objetos en mente, vamos a iniciar nuestra discusión de límites preguntándonos sobre en qué puntos tiene siquiera sentido preguntarnos por el límite de una función. La respuesta a esta pregunta no es inmediata, pensemos por ejemplo en una función como $f(x) = \frac{\sqrt{x}}{x-2}$. Su dominio es $[0, \infty) - \{2\}$, y si bien estaríamos de acuerdo en que no tiene sentido evaluar el límite de esta función cuando $x \to -3$, sí tiene sentido hacernos la pregunta de qué pasa si $x \to 2$, aún si 2 no es parte del dominio. Quisiéramos decir que sí admitimos ese punto ya que está *cerca del dominio*, pero esto es algo que es necesario precisar. Hagámoslo.

Definición A.3 (Adherencia)**.** Sea $A \subseteq \mathbb{R}$ un conjunto de números reales. Un punto $x_0 \in \mathbb{R}$ se dice punto adherente de A si para todo $\varepsilon > 0$, el intervalo[2] $(x_0 - \varepsilon, x_0 + \varepsilon)$ tiene intersección no vacía con A. A la colección de todos los puntos adherentes a un conjunto A se le llama *adherencia* de A y se denota por $\operatorname{adh}(A)$ o bien $\overline{A}$.

Nota. Crucialmente, el punto x_0 no necesita estar en A para poder ser considerado un punto adherente. Pero si está en A, entonces automáticamente cumple con la definición y por lo tanto está en $\overline{A}$. Esto prueba que $A \subseteq \overline{A}$.

Ejemplo A.4. Algunos ejemplos de adherencia de un conjunto son los siguientes:

- $\operatorname{adh}((0,1)) = [0,1]$
- $\operatorname{adh}((0,2] \cup \{4\}) = [0,2] \cup \{4\}$
- $\operatorname{adh}(\mathbb{Q}) = \mathbb{R}$.

Veamos esta noción en términos de sucesiones con el siguiente resultado, cuya demostración queda como ejercicio.

Proposición A.5.

$$\operatorname{adh}(A) = \{x \in \mathbb{R} : \exists (a_n) \subseteq A : a_n \to x\}$$

En otras palabras, la adherencia de un conjunto A consiste precisamente en el conjunto de los números reales aproximables por una secuencia con términos en A.

Nota. Esto nos entrega una definición alternativa para la adherencia. Los resultados de este tipo, que nos permiten escribir nuestras nociones en términos de otros objetos (en este caso sucesiones) son de las cosas más deseables en matemáticas, ya que nos entregan nuevos puntos de vista sobre nuestros conceptos, es común que reciban el nombre de *caracterizaciones*.

Podríamos pensar que la adherencia es precisamente lo que buscamos. Es decir, todo punto adherente al dominio es un punto en el cual nos podemos preguntar por el límite de una función. Pero volvamos a la función que anteriormente motivó nuestro ejemplo y modifiquémosla ligeramente para dejarla de la siguiente forma

$$f(x) = \begin{cases} 0 & x = -3 \\ \frac{\sqrt{x}}{x-2} & x \in [0, \infty) - \{2\} \end{cases}$$

De este modo, agregamos un punto al dominio de nuestra función, la adherencia de este nuevo dominio es $\overline{A} = \{-3\} \cup [0, \infty)$ pero... ¿Tiene sentido realmente preguntarnos por $\lim_{x \to -3} f(x)$?

La respuesta es no. Ya que al cálculo de un límite no le debería interesar lo que está sucediendo en el punto mismo, quisiéramos excluir el punto $x_0 = -3$ por estar *muy lejos del dominio*, pero la adherencia no es capaz de flitrarlo, ya que es miembro del dominio. Es decir, necesitamos un concepto más restrictivo, por lo que introducimos esta definición más exigente

[2] A los intervalos de este tipo es común llamarles *vecindades*.

Definición A.6 (punto de acumulación)**.** Sea $A \subseteq \mathbb{R}$ un conjunto de números reales. Un punto $x_0 \in \mathbb{R}$ se dice punto de acumulación de A si para todo $\varepsilon > 0$, la vecindad **perforada** $(x_0 - \varepsilon, x_0 + \varepsilon) - \{x_0\}$ tiene intersección no vacía con A. A la colección de todos los puntos de acumulación de un conjunto A se le denota por A'.

Ejemplo A.7. Los puntos de acumulación de $(0, 2] \cup \{4\}$ forman el conjunto $[0, 2]$. Es decir, no existe una inclusión entre un conjunto y sus puntos de acumulación.

Así, la noción de punto de acumulación es precisamente la que necesitamos, ya que captura a los puntos *cercanos al dominio de la función pero sin ser aislados*. Queda como ejercicio probar la siguiente caracterización.

Proposición A.8 (Punto de acumulación en términos de sucesiones)**.**

$$A' = \{x \in \mathbb{R} : \exists (a_n) \subseteq A - \{x\} : a_n \to x\}$$

En otras palabras, la adherencia de un conjunto A consiste precisamente en el conjunto de los números reales aproximables por una secuencia ***no constante*** *con términos en A.*

Tenemos ahora los ingredientes que necesitamos para definir el límite de una función en un punto

Definición A.9 (Límite de una función)**.** Sea x_0 punto de acumulación del dominio de $f : A \subseteq \mathbb{R} \to \mathbb{R}$. Decimos que $\lim_{x \to x_0} f(x) = L$ cuando L cumple la condición siguiente

$$\forall \varepsilon > 0, \exists \delta > 0, \forall x \in A : 0 < |x - x_0| < \delta \implies |f(x) - L| < \varepsilon$$

Nota. La definición recién dada (conocida como la definición ε-δ de límites) es famosa por lo difícil que es de interpretar, y por si sola es la razón por la que esto está siendo expuesto en un apéndice y no en el contenido central del libro, por lo que se recomienda mirarla con atención y tratar de descifrar lo que está diciendo. A grandes largos, la idea está siendo que *los valores que toma f se parecen tanto como queramos a L cuando x se encuentra lo suficientemente cerca de x_0 sin ser igual a él*. Nótese para esto la presencia del 0 en la expresión $0 < |x - x_0| < \delta$.

Caractericemos esta noción mediante sucesiones con la siguiente proposición, cuya demostración queda como ejercicio.

Proposición A.10 (Límite en términos de sucesiones)**.** *Sea x_0 punto de acumulación del dominio de $f : A \subseteq \mathbb{R} \to \mathbb{R}$. Así, $\lim_{x \to x_0} f(x) = L$ si y solamente si para cualquier sucesión $(a_n) \subseteq A - \{x_0\}$ tal que $a_n \to x_0$, se tiene que $f(a_n) \to L$.*

Ejemplo A.11. Considérese la función $f : \mathbb{R} \to \mathbb{R}$ dada por

$$f(x) = \begin{cases} 0 & x \neq 0 \\ 1 & x = 0 \end{cases}$$

Demuéstrese independientemente según ambas definiciones posibles de límites que $\lim_{x \to 0} f(x) = 0$.