Introducción a las Redes Neuronales con TensorFlow

Fundamentos matemáticos

Centro de Investigación en Matemáticas, A.C.

CIMAT

manuel.suarez@cimat.mx

8 de agosto de 2022

Outline

Introducción

Aprendizaje automático

El aprendizaje máquina o aprendizaje automatizado es un marco de trabajo donde, a partir de un conjunto de datos, un algoritmo "aprende" patrones de información que le permiten resolver un problema.

- El aprendizaje automático se divide en
  - Aprendizaje Supervisado
  - Aprendizaje no Supervisado
- En ambos casos empleamos técnicas y algoritmos con bases matemáticas para implementar los diferentes tipos de aprendizaje.
- Dado que el aprendizaje automatizado tiene sus bases en las matemáticas en esta sesión haremos un repaso de los principales conceptos necesarios para entender el fundamento de las redes neuronales.

Objetivos

Objetivo de la Sesión

Conocer los conceptos matemáticos que permiten la formulación de un modelo de redes neuronales en el aprendizaje automatizado

Criterio de Calificación

- Resolución de ejercicios en Google Colab

Definición de Función

Definición

Una **función** f es una regla de correspondencia que asocia a cada objeto x en un conjunto —denominado **dominio**— un solo valor $f(x)$ de un segundo conjunto. El conjunto de todos los valores así obtenidos se denomina **rango** de la función.

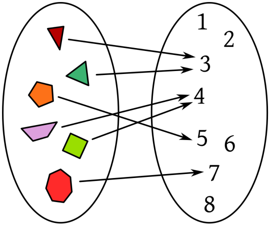

Figura: Representación de una función

Notación funcional y expresiones algebráicas

Notación

Una sola letra como f, g se utiliza para nombrar una función. Entonces $f(x)$, que se lee 'f de x' o 'f en x', denota el valor que la función f asigna a x. Denominamos x el argumento de la función

- Si $f(x) = x^3 - 4$ entonces:
  - $f(2) = (2)^3 - 4 = 8 - 4 = 4$
  - $f(3) = (3)^3 - 4 = 27 - 4 = 23$
  - $f(z) = z^3 - 4$
- La notación algebráica nos permite expresar la regla de correspondencia de una función en términos de operaciones algebráicas.

Límites I

Definición de límite

Decir que $\lim_{x \to c} f(x) = L$ significa que para cada $\varepsilon > 0$ dada (no importa qué tan pequeña) existe una correspondiente $\delta > 0$, tal que $|f(x) - L| < \varepsilon$, siempre que $0|x - c| < \delta$; esto es,

$$0 < |x - c| < \delta| \to |f(x) - L| < \epsilon$$

Límites II

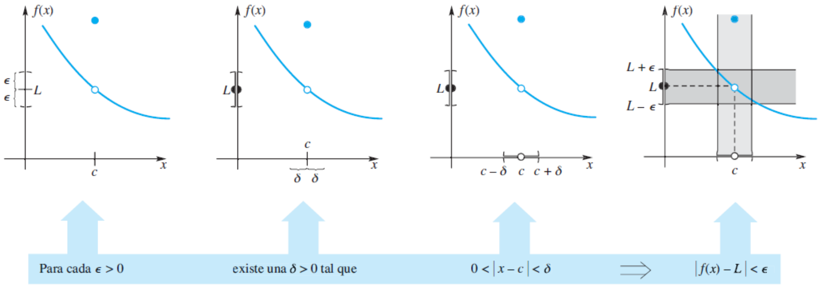

Figura: Representación gráfica de límite

Ejemplo

Demostrar que $\lim_{x\to 4}(3x-7)=5$

Solución: Sea ε cualquier número positivo. Debemos encontrar una $\delta > 0$ tal que

$$0 < |x-4| < \delta \to |(3x-7)-5| < \varepsilon$$

—(3x-7)-5—¡ε

Derivada

Definición de derivada

La **derivada** de una función f es otra función f' cuyo valor en cualquier número x es

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

Si este límite existe, decimos que f es **derivable** en x.

Ejemplos I

Sea $f(x) = 13x - 6$. Encuentre $f'(4)$

Contexto I

- Una red neuronal está compuesta de múltiples elementos denominados **perceptrones** donde cada uno de ellos procesa una serie de datos de entrada para obtener un resultado

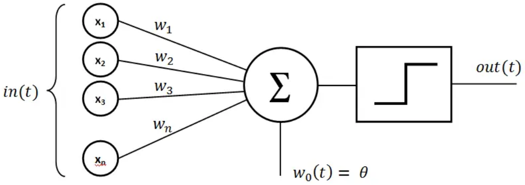

Figura: Perceptrón

Contexto II

Dados los datos de entrada $x_1, x_2, \ldots, x_n$ y los pesos $w_1, w_2, \ldots w_n$ la expresión matemática correspondiente del perceptrón es

$$x_1\dot{w}_1 + x_2\dot{w}_2 + \cdots + x_n\dot{w}_n = \sum_{i=1}^{n} x_i \cdot w_i \tag{1}$$

El resultado de la suma se evalúa en una función de activación la cuál puede tener diferentes expresiones, una de las mas comunes es la función sigmoide

$$f(x) = \frac{1}{1 + e^{-t}}$$

Contexto III

Para realizar el ajuste de los datos se realiza un proceso de minimización donde se requiere la derivada de la función de activación, en el caso de la función sigmoide su expresión es

$$f'(x) = \frac{e^{-x}}{(1 + e^{-x})^2}$$

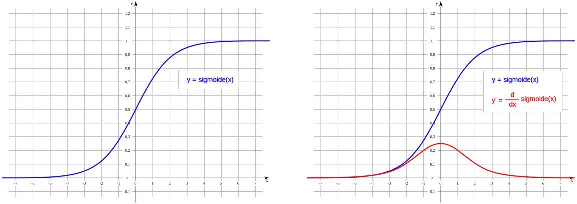

Figura: Función sigmoide y su derivada

Ejercicios

Vectores

Definición de vector

Un vector de tamaño n se define como un conjunto ordenado de n elementos

Un vector puede representarse en modo *columna* o *renglón*

$$(x_1, x_2, \ldots, x_n)$$

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

Matrices I

Definición de Matriz

Una matriz A de $m \times n$ es un arreglo rectangular de mn elementos dispuestos en m renglones y n columnas

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{ij} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ij} & & a_{in} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mj} & \cdots & a_{mn} \end{pmatrix}$$

Matrices II

El símbolo $m \times n$ se lee 'm por n'. El vector renglón $(a_{i1}, a_{i2}, \ldots, a_{in})$ se denomina **renglón** i y el vector columna $\begin{pmatrix} ij \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}$ se llama **columna** j.

El **componente o elemento** ij de A, denotado por a_{ij}, es el elemento que aparece en el renglón i y la columna j de A

Operaciones con vectores y matrices I

Suma de matrices

Sean $A = (a_{ij})$ y $B = (b_{ij})$ dos matrices de tamaño $m \times n$. Se define la operación de la suma de matrices $A \pm B$ como la suma de sus respectivos elementos ij

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

$A \pm B$ es la matriz $m \times n$ que se obtiene al sumar los elementos correspondientes de A y B

Operaciones con vectores y matrices II

Multiplicación de una matriz por un escalar

Si $A = (a_{ij})$ es una matriz de $m \times n$ y si α es un escalar, entonces la matriz $m \times n$, αA, está dada por

$$\alpha A = (\alpha a_{ij}) = \begin{pmatrix} \alpha a_{11} & \alpha a_{12} & \cdots & \alpha a_{1n} \\ \alpha a_{21} & \alpha a_{22} & \cdots & \alpha a_{2n} \\ \vdots & \vdots & & \vdots \\ \alpha a_{m1} & \alpha a_{m2} & \cdots & \alpha a_{mn} \end{pmatrix} \tag{3}$$

$\alpha A = (\alpha a_{ij})$ es la matriz obtenida al multiplicar cada componente de A por α

Operaciones con vectores y matrices III

Producto escalar

Sean $\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$ y $\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$ dos vectores. Se define el **producto escalar** de $\mathbf{a}$ y $\mathbf{b}$ denotado por $\mathbf{a} \cdot \mathbf{b}$ como

$$\mathbf{a} \cdot \mathbf{b} = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n \tag{4}$$

Nótese que el producto escalar (también denominado **producto punto**) entre dos vectores da como resultado un número (escalar).

Operaciones con vectores y matrices IV

Producto de matrices

Sea $A = (a_{ij})$ una matriz $m \times n$, y sea $B = (b_{ij})$ una matrix $n \times p$. Entonces el **producto** de A y B es una matrix mp, $C = (c_{ij})$ donde

$$c_{ij} = a_i \dot{b}_j \tag{5}$$

donde a_i es el vector renglón i de A y b_j es el vector columna j de B

El elemento ij de AB es el producto punto del renglón i de A y la columna j de B, es decir

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj} \tag{6}$$

Para que esta operación pueda realizarse es necesario que el número de columnas de A sea igual al número de renglones de B

Operaciones con vectores y matrices V

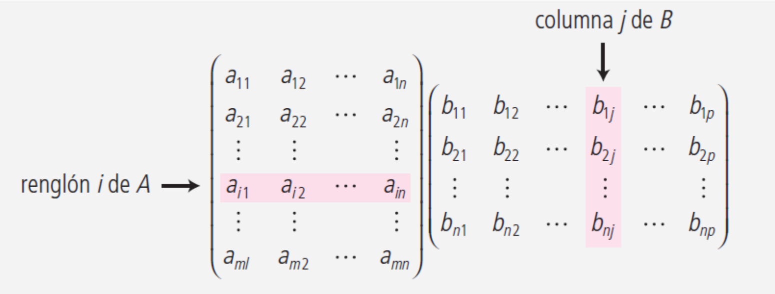

Figura: Multiplicación de matrices

Sistemas de Ecuaciones Lineales

Sistema de Ecuaciones Lineales

Un **sistema de ecuaciones** es un conjunto de **ecuaciones** con las mismas incógnicas. Un sistema de ecuaciones lineales es un sistema de ecuaciones en el que cada ecuación es lineal. Una solución de un sistema es una asignación de valores para las incógnitas que hace verdadera cada una de las ecuaciones

Ejemplo

$$\begin{cases} x_1 + 2x_2 + 3x_3 = 9 \\ 4x_1 + 5x_2 + 6x_3 = 24 \\ 3x_1 + x_2 - 2x_3 = 4 \end{cases}$$

Matricialmente escribimos el sistema anterior como $Ax = b$ donde A es la matriz de coeficientes del sistema, x es el vector de incógnitas y b el vector de términos independientes

Contexto I

De acuerdo con la ecuación 7, la expresión para el cálculo del resultado que un perceptrón tiene sobre una entrada $\mathbf{x}$ en relación con sus pesos $\mathbf{w}$ podemos considerar que cada uno de éstos es un vector por lo que podemos escribir la fórmula del perceptrón como el producto punto entre los vectores de entrada y pesos

$$x_1 \cdot w_1 + x_2 \cdot w_2 + \cdots + x_n \cdot w_n = \sum_{i=1}^{n} x_i \cdot w_i = \mathbf{x} \cdot \mathbf{w} \tag{7}$$

Ya que una red neuronal consiste en la evaluación de múltiples perceptrones cada uno con sus respectivos pesos sobre las entradas de la red la evaluación del resultado de la red sobre los datos de entrada puede representarse mediante un producto matricial Wx donde W es una matriz conformada por los pesos de cada una de las neuronas presentes en una capa de la red.

Contexto II

El mismo razonamiento se aplica sobre las entradas a cualquier capa de una red neuronal.

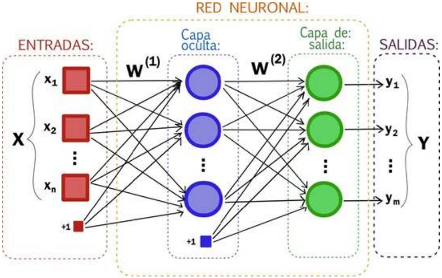

Figura: Red Neuronal

Ejercicios

Conceptos

Variable aleatoria

Una variable x es **variable aleatoria** si el valor que toma, corresponde al resultado de un experimento o evento aleatorio.

Variables discretas y continuas

Una variable discreta puede tomar sólo un número finito o contable de valores. Una variable continua puede tomar infinitamente muchos valores correspondientes a los puntos en un intervalo de recta

Distribución de Probabilidad

La **distribución de probabilidad** para una variable aleatorio es una fórmula, tabla o gráfica que da los posibles valores de x y su probabilidad asociada $p(x)$ con cada valor de x.

Contexto

Las redes neuronales artificiales aproximan funciones de separación dando como resultado una medida de probabilidad dependiendo del contexto en el cuál se apliquen, por ejemplo en el problema de clasificación de Kaggle "Dog vs Cats" https://www.kaggle.com/c/dogs-vs-cats

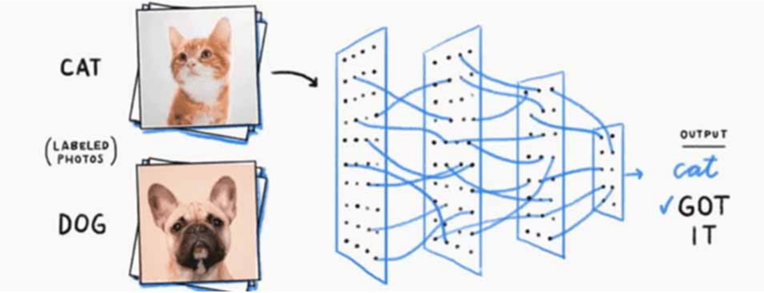

Figura: Dog vs Cats

Ejercicios

Optimización

Optimización

La optimización es la rama de las matemáticas que se encarga de encontrar los puntos óptimos (máximos o mínimos) de una función de manera global o en un intervalo determinado.

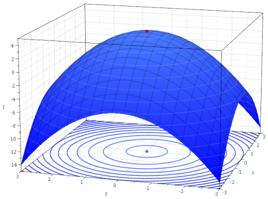

Figura: Máximo de una función

Algoritmos de aproximación I

- Teóricamente podemos encontrar los máximos y mínimos de una función aplicando los criterios de la 1a. y 2a. derivada (derivadas parciales en el caso de funciones vectoriales) sin embargo muchas veces esto no es factible de calcular por lo que recurrimos a los algoritmos de aproximación.
- Los algoritmos mas utilizados para optimización se basan en el cálculo del gradiente de una función como dirección de máximo descenso considerando un tamaño de paso (fijo o adaptable) para iterar sucesivamente en búsqueda del punto óptimo de la función (no se garantiza un óptimo global a menos que la función sea convexa)

Algoritmos de aproximación II

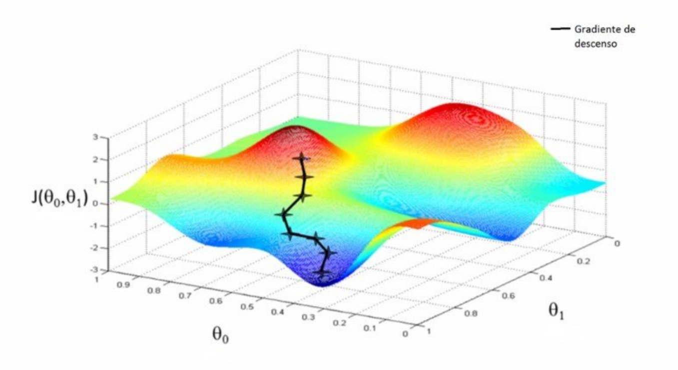

Figura: Descenso de gradiente

Contexto I

En el área del aprendizaje automático se busca realizar el ajuste de una función a un conjunto de datos, este ajuste se realiza por medio de la minimización de una función de costo que establezca una noción de "distancia.entre el resultado del modelo predictivo y los datos a ajustar. Por lo tanto el procedimiento principal para realizar el ajuste de los datos es minimizar esta función de costo que normalmente está dada por el cuadrado de la distancia entre los datos observados y la predicción del modelo

$$min_{\alpha,\mu} F(\alpha,\mu) = \frac{1}{2}||\Phi\alpha - y||_2^2 \tag{8}$$

donde y son los datos observados y $\Phi\alpha$ es el resultado de la predicción del modelo. El ajuste de los parámetros del modelo entonces, es un proceso de optimización de la función de costo.

Ejercicios

Conclusiones

- El aprendizaje automático utilizando redes neuronales es un ajuste de composición de funciones a un conjunto de datos observados sobre el cuál queremos resolver un problema, como clasificación o regresión. El proceso de entrenamiento de la red es un proceso de optimización por medio del ajuste del modelo a los datos observados utilizando una función de costo.
- El planteamiento del modelo y su resolución tiene bases matemáticas en el cálculo, álgebra lineal y probabilidad.

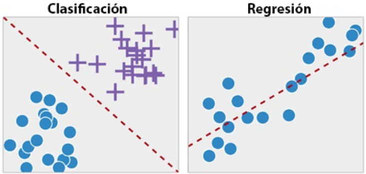

Figura: Clasificación/Regresión