Universidad Autónoma De Nuevo León

Facultad De Ciencias Físico Matemáticas

Introducción al Aprendizaje Automático

Lic. Ángel Adrián Domínguez Lozano

Tarea 1

Resumen

Equipo 14

Nombre del alumno:	Matrícula:
Sofía Alejandra Gaytán Díaz	1845533
Juan Daniel Morón Guajardo	1842834
Humberto Gerardo Peña Páez	1862464

San Nicolás de los Garza, Nuevo León 3 de septiembre de 2021

Tabla de contenido

Algebra Lineal

El álgebra lineal es una herramienta básica para la mayoría de las ramas de la matemática y disciplinas de ciencias exactas, esta rama estudia conceptos como vectores, matrices, espacios, sistemas de ecuaciones, pero su enfoque más formal son los espacios vectoriales y sus transformaciones lineales. En este caso únicamente tomaremos nota de las operaciones que se es posible realizar con matrices.

Operaciones de matrices

Las operaciones con matrices que se pueden realizar son: suma, resta, multiplicación o producto y división.

Suma y Resta

Una de las operaciones básicas que se pueden realizar con matrices. Es necesario que se cuente con dos de estas. Las cuales se definen a continuación.

$$sea\ A_{nxn}\ y\ B_{nxn}\ matrices\ de\ nxn\ de\ la\ forma:$$

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{12} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{pmatrix} \qquad B = \begin{pmatrix} b_{11} & b_{12} & b_{13} & \cdots & b_{1n} \\ b_{21} & b_{22} & b_{23} & \cdots & b_{12} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & b_{n3} & \cdots & b_{nn} \end{pmatrix}$$

Es necesario que las matrices en las que se efectuará la operación sean de la misma dimensión. No estrictamente tienen que ser matrices cuadradas.

Tomadas A y B arriba definidas y efectuamos una suma sobre ellas obtenemos lo siguiente:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{pmatrix} - \begin{pmatrix} b_{11} & b_{12} & b_{13} & \cdots & b_{1n} \\ b_{21} & b_{22} & b_{23} & \cdots & b_{1n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & b_{n3} & \cdots & b_{nn} \end{pmatrix} = \begin{pmatrix} a_{11}-b_{11} & a_{12}-b_{12} & a_{13}-b_{13} & \cdots & a_{1n}-b_{1n} \\ a_{21}-b_{21} & a_{22}-b_{22} & a_{23}-b_{23} & \cdots & a_{2n}-b_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1}-b_{n1} & a_{n2}-b_{n2} & a_{n3}-b_{n3} & \cdots & a_{nn}-b_{nn} \end{pmatrix}$$

De manera textual, sumamos los elementos $A_{ij} + B_{ij}$ lo que nos da como resultado el elemento ij de la nueva matriz.

De manera similar sucede con la resta de matrices, véase a continuación:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{12} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{pmatrix} - \begin{pmatrix} b_{11} & b_{12} & b_{13} & \cdots & b_{1n} \\ b_{21} & b_{22} & b_{23} & \cdots & b_{12} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & b_{n3} & \cdots & b_{nn} \end{pmatrix} = \begin{pmatrix} a_{11}-b_{11} & a_{12}-b_{12} & a_{13}-b_{13} & \cdots & a_{1n}-b_{1n} \\ a_{21}-b_{21} & a_{22}-b_{22} & a_{23}-b_{23} & \cdots & a_{2n}-b_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1}-b_{n1} & a_{n2}-b_{n2} & a_{n3}-b_{n3} & \cdots & a_{nn}-b_{nn} \end{pmatrix}$$

Efectuamos $A_{ij} - B_{ij}$ lo que nos como resultado el elemento ij de la nueva matriz.

Multiplicación

1. Multiplicación por escalar

Para multiplicar una matriz por un escalar, no es necesario seguir muchas restricciones, solo hay que contar con un escalar y una matriz los cuales multiplicar. A continuación, se definen:

$$\alpha \text{ es un escalar cualquiera}, A^{nxm} \text{ es una matriz de } nxm.$$

Entonces, se realiza $\alpha(a_{ij})$ y este es el elemento ij de la matriz resultante.

$$\alpha \cdot \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{pmatrix} = \begin{pmatrix} \alpha(a_{11}) & \alpha(a_{12}) & \alpha(a_{13}) & \cdots & \alpha(a_{1n}) \\ \alpha(a_{21}) & \alpha(a_{22}) & \alpha(a_{23}) & \cdots & \alpha(a_{2n}) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha(a_{n1}) & \alpha(a_{n2}) & \alpha(a_{n3}) & \cdots & \alpha(a_{nn}) \end{pmatrix}$$

2. Multiplicación de matrices

Para multiplicar dos matrices es necesario que cumplan con la condición tal que la primera matriz debe tener el mismo número de columnas que de filas de la segunda matriz, dando como matriz resultante una matriz con el mismo número de filas de la primera y el número de columnas de la segunda.

$$Sean\ A_{n\times m}\ y\ \ B_{m\times s}\ dos\ matrices\ de\ n\ \times m\ y\ m\ \times s, respectivamente$$

La multiplicación se define de la siguiente manera:

$$C_{ij} = \sum_{k=1}^{m} A_{ik} B_{kj} \quad ,1 \le i \le n, 1 \le j \le r.$$

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1m} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{1m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nm} \end{pmatrix} \cdot \begin{pmatrix} b_{11} & b_{12} & b_{13} & \cdots & b_{1s} \\ b_{21} & b_{22} & b_{23} & \cdots & b_{2s} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & b_{m3} & \cdots & b_{ms} \end{pmatrix}$$

$$= \begin{pmatrix} a_{11} \cdot b_{11} + a_{12} \cdot b_{21} + \cdots + a_{1m} \cdot b_{m1} & a_{11} \cdot b_{12} + a_{12} \cdot b_{22} + \cdots + a_{1m} \cdot b_{m2} & a_{11} \cdot b_{13} + a_{12} \cdot b_{23} + \cdots + a_{1m} \cdot b_{m3} & \cdots & a_{11} \cdot b_{1s} + a_{12} \cdot b_{2s} + \cdots + a_{1m} \cdot b_{ms} \\ a_{21} \cdot b_{11} + a_{22} \cdot b_{21} + \cdots + a_{2m} \cdot b_{m1} & a_{21} \cdot b_{12} + a_{22} \cdot b_{22} + \cdots + a_{2m} \cdot b_{m2} & a_{21} \cdot b_{13} + a_{22} \cdot b_{23} + \cdots + a_{2m} \cdot b_{m3} & \cdots & a_{21} \cdot b_{1s} + a_{22} \cdot b_{2s} + \cdots + a_{2m} \cdot b_{ms} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} \cdot b_{11} + a_{n2} \cdot b_{21} + \cdots + a_{nm} \cdot b_{m1} & a_{n1} \cdot b_{12} + a_{n2} \cdot b_{22} + \cdots + a_{nm} \cdot b_{m2} & a_{n1} \cdot b_{13} + a_{n2} \cdot b_{23} + \cdots + a_{nm} \cdot b_{m3} & \cdots + & a_{n1} \cdot b_{1s} + a_{n2} \cdot b_{2s} + \cdots + a_{nm} \cdot b_{ms} \end{pmatrix}$$

División

1. División por escalar

Al igual que la multiplicación por escalar, no es necesario ninguna restricción, solo contar con un escalar y una matriz a los cuales efectuar dicha operación.

$$\alpha \; es \; un \; escalar \; cualquiera, A^{nxm} \; es \; una \; matriz \; de \; nxm.$$

Se realiza $\frac{1}{\alpha} \cdot a_{ij}$ y el resultado obtenido es el elemento ij de la matriz resultante.

$$\alpha \div \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{pmatrix} = \begin{pmatrix} \frac{1}{\alpha}(a_{11}) & \frac{1}{\alpha}(a_{12}) & \frac{1}{\alpha}(a_{13}) & \cdots & \frac{1}{\alpha}(a_{1n}) \\ \frac{1}{\alpha}(a_{21}) & \frac{1}{\alpha}(a_{22}) & \frac{1}{\alpha}(a_{23}) & \cdots & \frac{1}{\alpha}(a_{2n}) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\alpha}(a_{n1}) & \frac{1}{\alpha}(a_{n2}) & \frac{1}{\alpha}(a_{n3}) & \cdots & \frac{1}{\alpha}(a_{nn}) \end{pmatrix}$$

2. División de matrices

El término de división de matrices no existe como tal, pero coloquialmente se le llama así a la multiplicación de una matriz por la inversa de una segunda matriz. Para este apartado es necesario que al menos la matriz divisora sea cuadrada o su determinante sea 0, ya que, si cumple con esto, la matriz es invertible.

$$Sean \; A^{n \times n} \; y \; B^{n \times n} \; dos \; matrices \; de \; n \times n \; y \; B^{n \times n} \; es \; invertible$$

Se efectúa la siguiente multiplicación de matrices para obtener la matriz resultante.

$$A \div B = A \cdot \frac{1}{B} = A \cdot B^{-1}$$

Probabilidad

La probabilidad se define como aquel porcentaje de seguridad de que ocurra una situación teniendo en cuenta diversos factores, donde el valor puede variar entre 0 y 1, donde:

- Si es cercano a 0, hay menor seguridad.
- Si es cercano a 1, hay más seguridad.

Mientras la probabilidad es la posibilidad de que ocurra un hecho, la estadística se encarga de extraer las conclusiones de un fenómeno observado, para que, a partir de dichas conclusiones, se realicen predicciones. Sin embargo, la probabilidad se clasifica como un concepto estadístico, donde, por ello, puede usarse en distintos ámbitos, y dependiendo de la situación, se puede utilizarse un diferente tipo de probabilidad, como es el caso de:

- Cuando se trata de la posibilidad de que ocurra un solo suceso, se le conoce como probabilidad simple.
- Cuando se trata de la posibilidad de que ocurran dos sucesos al mismo tiempo, se le conoce como probabilidad compuesta.
- Cuando se trata de una probabilidad en base a un experimento acreditado, se le conoce como probabilidad objetiva, mientras que cuando se trata de una probabilidad en base a experiencia y/o creencias, se le conoce como probabilidad subjetiva.
- Dependiendo del espacio muestral, se le puede conocer como probabilidad de espacio muestral, de la unión y de la intersección.

<u>Distribuciones de probabilidad</u>

Una distribución de probabilidad es la que determina todos los resultados probables que se pueden obtener de un experimento para conocer la factibilidad de cada uno de ellos. Además, se le puede conocer como una herramienta que ayuda a predecir acontecimientos considerando los resultados de un experimento.

Una distribución de probabilidad se genera a partir de una variable aleatoria, conocida así porque el valor que se toma es totalmente aleatorio, la cual es una función que asigna un valor numérico a cada uno de los elementos del espacio muestral. Existen dos tipos de variables aleatorias:

- Variables aleatorias discretas. Se cuenta con un número finito de valores, donde las probabilidades de sus valores deben de sumar 1.
- Variables aleatorias continuas. Se cuenta con un número infinito de valores, es decir, un número no contable de valores, donde se puede tomar cualquier valor dentro de un intervalo sin repetirse. Estas variables cuentan con una función de probabilidad conocida como función de densidad, donde la función es mayor o igual que cero en cualquier punto y el área bajo la curva de la función debe de tener un valor de 1.

Donde, dependiendo del tipo y cantidad de las variables aleatorias, se genera un diferente tipo de distribución de probabilidad, clasificándose de la siguiente manera:

Distribuciones discretas

Una distribución discreta es la que se basa en probabilidades de los valores de una variable aleatoria discreta, donde cada uno de los valores puede asociase a una probabilidad diferente de cero.

Distribuciones continuas

Una distribución continua es la que se basa en las probabilidades de los valores de una variable aleatoria continua, las cuales son el área por debajo de la curva y donde los rangos de valores pueden tener una probabilidad diferente de cero y la probabilidad de que una variable aleatoria continua equivalga a un valor siempre es 0.

Distribuciones multivariadas

Una distribución multivariada, o también conocida como multivariante, es la que se basa en las probabilidades de los valores de más de una variable aleatoria, donde pueden o no estar correlacionadas.

Binary Cross Entropy

La entropía de una variable aleatoria es el nivel de incertidumbre de los posibles resultados que pueden existir donde se trata de predecir qué tan lejos están los valores de los valores reales. La entropía cruzada es la función de perdida entre dos distribuciones de probabilidad para conocer la diferencia entre un valor determinado con el valor verdadero.

Binary Cross Entropy, en español conocido como Entropía Cruzada Binaria, es una función de pérdida dentro de la entropía cruzada, siendo la forma más simple y fácil de interpretar, ya que se utiliza en tareas de clasificación binaria, el cual se utiliza para entrenar un modelo para resolver problemas de clasificación donde la clasificación puede ser entre dos valores.

Para esta función de perdida, la función sigmoide es la única función compatible, donde se necesita calcular logaritmos que solo existen si los valores varían entre 0 y 1, teniendo una escala univariante y es simétrica.

Optimización

La optimización es la rama de la matemática que tiene como finalidad resolver problemas con premisas como: encontrar áreas mínimas, menor coste, la forma optima de realizar algo, la menor resistencia, el máximo beneficio, el mayor alcance, todo esto es lo que engloba en general a la optimización y estos problemas pueden ser resueltos aplicando el calculo diferencial y técnicas de optimizaciónl.

Mínimos máximos locales/globales

Los máximos y mínimos de una función son conocidos como los extremos de una función, respectivamente son los valores más grandes (máximos) o más pequeños (mínimos) que esta misma toma en un punto situado dentro de una región especifica de la función (local) o en el total dominio de esta misma (global).

Al menos los mínimos máximos relativos o locales se pueden obtener aplicando el criterio de la primera derivada que pudimos estudiar en cálculo diferencial.

Gradient descent

También conocido como gradiente descendiente, es la base de aprendizaje en muchas técnicas de machine learning. Por ejemplo, es fundamental en el Deep Learning para entrenar redes neuronales. También necesario para la regresión logística.

Para entender un poco sobre el algoritmo, definiremos qué es una gradiente: "El gradiente es el conjunto de todas las derivadas parciales de una función". Si la utilizamos para machine learning, en la mayoría de los casos se interesa en el gradiente de la función de coste y sea para obtener el mínimo o el máximo.

El gradiente descendiente es un método de optimización númerica que sirve para estimar los mejores coeficientes para minimizar la función de coste. Lo que vamos buscando con este método es encontrar el conjunto de parámetros que minimizan la función de coste (MSE).

Stochastic Gradient Descent

El Descenso de Gradiente Estocástico (SDG) es un enfoque simple pero muy eficiente para la regresión logística y más. Este método es utilizado en su mayoría para problemas de aprendizaje a gran escala. En esencia, la diferencia con el algoritmo de descenso de gradiente radica en la aplicación del algoritmo original a problemas de datos en la vida real.

Algunas ventajas de este método son: La eficiencia, la fácil implementación ya que cuenta con muchas oportunidades para el ajuste del código. Mientras que de desventajas tenemos: Requiere una serie de hiperparámetros como el parámetro de regularización y el número de iteraciones, es sensible a la escala de las características.

Aprendizaje máquina

El aprendizaje máquina, también conocido como aprendizaje automático, o en inglés, Machine Learning, es un campo de la Inteligencia Artificial que hace posible que una máquina, ordenador o computadora trabaje de forma autónoma sin la necesidad de ser programados, ya que, trabajan en base a patrones y predicciones que se elaboran de conclusiones de un evento pasado. El aprendizaje máquina se le denomina a la ciencia que estudia el comportamiento que tiene una máquina con el fin de analizar los resultados de esta y ver si serán beneficiosos o no, ya que, el comportamiento de las máquinas debe de ser analizado para tener los resultados esperados, analizar errores que cometió la máquina y si es posible mejorarlos.

Tipos de aprendizaje máquina

Los tipos de aprendizaje maquina se dividen en tres categorías:

Aprendizaje supervisado

El aprendizaje supervisado es el basado en etiquetas previas, donde se crean patrones para poder decidir y proveer predicciones para eventos futuros que cuenten con la misma etiquetación. Existen 2 tipos de modelos de aprendizaje supervisado:

1. *Modelos de clasificación.* Se usan para clasificar 2 clases y producen como salida un dato discreto de un conjunto finito.
2. *Modelos de regresión.* Producen como salida un valor real.

Aprendizaje no supervisado

El aprendizaje no supervisado no está basado en conocimientos previos, sino va aprendiendo en base de a los datos y su organización para encontrar patrones. El aprendizaje supervisado se suele usar en 3 casos:

1. *Problemas de Clustering.* Se agrupan los datos en grupos ya sea por su similitud y/o características.
2. *Agrupamientos de Co-ocurrencias.* Se usa una matriz para analizar la relación entre los datos.
3. *Perfilado (Profiling).* Técnica que consiste en analizar y entender un conjunto de datos para determinar qué información se obtuvo de los datos y hacer predicciones del aprendizaje.

Aprendizaje semi-supervisado

El aprendizaje semi-supervisado es una mezcla entre el aprendizaje supervisado y no supervisado. Se utiliza principalmente cuando se quiere analizar un conjunto de datos no etiquetados para luego etiquetarlos y luego usar un algoritmo de aprendizaje no supervisado.

Aprendizaje reforzado

El aprendizaje reforzado es aquel que se basa en los conocimientos que la maquina a aprendido en base a su experiencia.

Overfitting

El Overfitting, o Sobreajuste que un problema del aprendizaje máquina, donde ocurre cuando la máquina ha analizado muchos datos y clasifica mal los datos debido a que encontró una característica que no siempre está presente entre los datos.

Para evitar este problema, existen diversas acciones, entre ellas se encuentran las siguientes:

- Cantidad moderada de muestras, donde es recomendable que sea una cantidad mínima.
- Uso diverso de clases.
- Uso de variantes entre los datos.
- Ajuste de parámetros.

- Creación de subconjuntos de datos.

Además, se puede requerir el uso de la regularización matemática, técnica que consiste en introducir más información para solucionar un problema mal planteado y evitar Overfitting, donde se necesita reducir el error de generalización, el cual es una medida para determinar la efectividad del algoritmo en base a predicciones. Existen varias fórmulas para usar la regularización, las más importantes son:

- Regularización Tikhonov.
- Parada anticipada, o Early stopping.
- Regularizadores para escasez.
- Regularizadores para aprendizaje semi-supervisado.
- Regularizadores para el aprendizaje multitarea.

<u>Underfirtting</u>

El Underfitting, o Subajuste es un problema del aprendizaje máquina similar al Sobreajuste o Overfitting, que ocurre cuando una máquina no puede clasificar los datos debido a que ha analizado muy pocos datos para llegar a clasificarlos.

<u>Validación Cruzada</u>

La validación cruzada se refiere a una técnica que consiste en usar varios modelos para un conjunto de datos entrada y analizarlos, donde se repite el proceso y se obtiene un promedio de las medidas de evaluación de los modelos. Sus usos principales consisten en detectar sobreajuste, y sobre todo para predecir cómo será el entorno y analizar su precisión en un modelo.

Bibliografía

Análisis de grupos. (12 de abril de 2021). Obtenido de Wikipedia: https://es.wikipedia.org/wiki/An%C3%A1lisis_de_grupos

Calvo, D. (10 de diciembre de 2018). *Función de coste – Redes neuronales*. Obtenido de Diego Calvo Web Site: https://www.diegocalvo.es/funcion-de-coste-redes-neuronales/

Chapter 4. Co-occurrence and Recommendation. (s.f.). Obtenido de OREILLY: https://www.oreilly.com/library/view/practical-machine-learning/9781491915707/ch04.html

Conoce las principales distribuciones de probabilidad. (10 de octubre de 2016). Obtenido de Conexión ESAN: https://www.esan.edu.pe/apuntes-empresariales/2016/10/conoce-las-principales-distribuciones-de-probabilidad/

Data Profiling. (s.f.). Obtenido de DataRobot: https://www.datarobot.com/wiki/data-profiling/

Descubre los principales beneficios del 'Machine Learning'. (s.f.). Obtenido de IBERDROLA: https://www.iberdrola.com/innovacion/machine-learning-aprendizaje-automatico

Función de pérdida de entropía cruzada. (s.f.). Obtenido de ICHI.PRO: https://ichi.pro/es/funcion-de-perdida-de-entropia-cruzada-267783942726718

Gabriela Jeronimo, J. S. (2008). *Fasículo 2 Álgebra Lineal.*

Heras, J. M. (09 de 2020). *Gradiente Descendiente para aprendizaje automatico.* Obtenido de https://www.iartificial.net/gradiente-descendiente-para-aprendizaje-automatico/

ICHI.PRO. (s.f.). *Comprensión del descenso de gradiente estocástico.* Obtenido de https://ichi.pro/es/comprension-del-descenso-de-gradiente-estocastico-189809026705555

Khan Academy. (s.f.). *Repaso sobre máximos y mínimos.* Obtenido de https://es.khanacademy.org/math/ap-calculus-ab/ab-diff-analytical-applications-new/ab-5-4/a/relative-minima-and-maxima-review

Kunin, D., Guo, J., Devlin, T. D., Xiang, D., & Orozco, C. (s.f.). *Distribuciones de Probabilidad*. Obtenido de Viendo la Teoría: https://seeing-theory.brown.edu/probability-distributions/es.html

MathWorks. (s.f.). *Distribuciones multivariadas*. Obtenido de MathWorks: https://es.mathworks.com/help/stats/multivariate-distributions.html

McCrea, N. (s.f.). *Introducción a la Teoría de Aprendizaje de Máquina y sus Aplicaciones: Un Tutorial Visual con Ejemplos*. Obtenido de Developers: https://www.toptal.com/machine-learning/introduccion-a-la-teoria-de-aprendizaje-de-maquina-y-sus-aplicaciones-un-tutorial-visual-con-ejemplos

Minitab 18. (s.f.). *Distribuciones de probabilidad continuas y discretas*. Obtenido de Soporte de Minitab 18: https://support.minitab.com/es-mx/minitab/18/help-and-how-to/probability-distributions-and-random-data/supporting-topics/basics/continuous-and-discrete-probability-distributions/#what-is-a-discrete-distribution

Na8. (12 de Diciembre de 2017). *Qué es overfitting y underfitting y cómo solucionarlo*. Obtenido de Aprende Machine Learning en Español: https://www.aprendemachinelearning.com/que-es-overfitting-y-underfitting-y-como-solucionarlo/

Peltarion. (s.f.). *Binary crossentropy loss function*. Obtenido de Peltarion Platafform: https://peltarion.com/knowledge-center/documentation/modeling-view/build-an-ai-model/loss-functions/binary-crossentropy

Quintela del Rio, A. (4 de septiembre de 2019). *Estadística Básica Edulcorada*. Obtenido de Bookdown: https://bookdown.org/aquintela/EBE/

Regularization (mathematics). (20 de mayo de 2021). Obtenido de Wikipedia: https://en.wikipedia.org/wiki/Regularization_(mathematics)

Sanchez, J. A. (31 de agosto de 2020). *¿Cómo aprenden las máquinas? Machine Learning y sus diferentes tipos*. Obtenido de datos.gob.es: ¿Cómo aprenden las máquinas? Machine Learning y sus diferentes tipos

teatista, stackoverflowuser2010, Malra dura, & Sandipan Dey. (s.f.). *¿Qué es la entropía cruzada?* Obtenido de QA Stack: https://qastack.mx/programming/41990250/what-is-cross-entropy

Unipython. (s.f.). *Descenso de Gradientes Estocástico "SGD".* Obtenido de https://unipython.com/descenso-gradientes-estocastico-sgd/

Universidad Veracruzana. (s.f.). *Algebra de Matrices.* Obtenido de https://www.uv.mx/personal/aherrera/files/2014/08/11a.-ALGEBRA-DE-MATRICES-1.pdf

Unsupervised learning: aprendizaje automático sin restricciones. (14 de septiembre de 2020). Obtenido de Digital Guide IONOS: https://www.ionos.mx/digitalguide/online-marketing/marketing-para-motores-de-busqueda/unsupervised-learning/

Validación cruzada. (s.f.). Obtenido de Amazon Web Services: https://docs.aws.amazon.com/es_es/machine-learning/latest/dg/cross-validation.html

Validación Cruzada. (5 de marzo de 2020). Obtenido de Wikipedia: https://es.wikipedia.org/wiki/Validaci%C3%B3n_cruzada

Westreicher, G. (25 de agosto de 2020). *Probabilidad*. Obtenido de Economipedia.com: https://economipedia.com/definiciones/probabilidad.html