

REVISTA

TIA

Tecnología, Investigación y Academia -Red Avanzada – RITA
Publicación Facultad de Ingeniería y Red de Investigaciones de Tecnología

Técnicas de aprendizaje automático aplicadas en los sistemas de predicción

Citar este documento:

Ingry-Nathaly Salamanca R, Edgar Junior C., (2021), Técnicas de aprendizaje automático aplicadas en los sistemas de predicción. Technol. Investig. Academia TIA, ISSN: 2344-8288, 8 (1), pp. 39-53. Bogotá-Colombia

Técnicas de aprendizaje automático aplicadas en los sistemas de predicción

Machine learning techniques applied to forecasting systems

Ingrý Nathaly Salamanca Rativa ¹, Edgar Junior Castro Escorcía ²

Resumen: En este artículo se realiza una revisión conceptual del aprendizaje automatizado su definición modelos y algoritmos que son principalmente usados en sistemas de predicción, recopilando de diferentes autores sus definiciones y perspectivas de la aplicación de algoritmos según los objetivos de investigación. Se seleccionaron los algoritmos de redes neuronales y regresión usados en la predicción de patrones lo más acertados posibles, se da a conocer el hecho de como realmente puede llegar una máquina a aprender y que el centro de su operación es un algoritmo que revisa los datos y es capaz de predecir comportamientos futuros. Automáticamente, también en este contexto, implica que estos sistemas se mejoran de forma autónoma con el tiempo, sin intervención humana.

Palabras clave: Machine Learning (ML), Aprendizaje supervisado, Aprendizaje no supervisado, Dataset, Predicción, Estadística, función lineal, Función polinómica, Sesgo variante.

Keywords: Machine Learning (ML), Supervised learning, Unsupervised learning, Data-set, Prediction, Statistics, linear function, Polynomial function, Variant bias.

Abstract: This article delves into the field of machine learning as a science for machine learning, they glimpse that differentiating factor in the analysis of the information they pay, in the context of what should be formulated as a tool for prediction and the use of statistics In different contexts, being able to predict facts based on data obtained categorically and systematically for periods of time, allow; In conjunction with the mathematical and statistical equations to conclude that an uncertain event will occur with a fairly acceptable level of veracity, these machine learning techniques are already categorized and their use has been recognized in the scientific and technological field, achieving with this determine the effectiveness of these techniques in commercial, scientific and social sectors.

¹ Facultad de Ingeniería. Universidad Distrital Francisco José de Caldas, e-mail: insalamancar@correo.udistrital.edu.co

² Facultad de Ingeniería. Universidad Distrital Francisco José de Caldas, e-mail: jcastroe@correo.udistrital.edu.co

I. Introducción

La inteligencia artificial ya hace parte de nosotros sin siquiera saberlo, determinar cuándo va a llover basado en datos debidamente medidos por herramientas tecnológicas ya es un hecho cierto, las predicciones basados en las probabilidades de ocurrencia de un hecho, determinan el comportamiento del ser humano en un futuro, podríamos empezar a pensar que el ser humano a través de las ciencias informáticas, matemáticas y estadísticas empezaron a conocer como si se tratará de una bola de cristal acontecimientos del futuro, ¿pero cómo es esto posible?

Basta con conocer la ciencia del Machine Learning y sus fundamentos en las matemáticas que caracterizan de forma determinística y apoyados en el método científico que una máquina puede emular aprendizaje, serialización, cuantificación y predicción de valores casi como lo haría una persona, sumado al hecho de una gran capacidad de procesamiento las máquinas en ciertos aspectos pueden superar al ser humano en tareas muy específicas, las máquinas al igual que los seres humanos pueden aprender con la experiencia para poder caracterizar las propiedades físicas, químicas, biológicas de la materia junto con factores de aprendizaje supervisado llegar a concluir cual es la mejor opción para una tarea específica, también se pueden emular comportamientos de decisión basado en redes neuronales como lo haría un cerebro; permitiendo tomar decisiones sin titubear basado en hechos cuantificables que no dan cabida a la subjetividad, poder obtener datos de entrada, procesarlos para tomar una decisión y hacerlo parte del acervo de la experiencia de la máquina para futuras decisiones de iguales características basados en los valores de verdad de las mismas lo que permite crear patrones de comportamiento que garantizan que la opción decidida es la mejor.

II. Aprendizaje Automatizado

Tom M. Mitchell define machine learning como el campo del aprendizaje automático se ocupa de la cuestión de cómo construir programas informáticos que mejoran automáticamente con la experiencia [1]. El campo del aprendizaje automático se ocupa de la cuestión de cómo construir programas informáticos que mejoran automáticamente con la experiencia [2].

El término aprendizaje automático se refiere a la detección automatizada de patrones en los datos. En las últimas dos décadas se ha convertido en una herramienta común en casi cualquier tarea que requiera la extracción de información de grandes conjuntos de datos. Estamos rodeado por una tecnología basada en el aprendizaje automático: los motores de búsqueda aprenden cómo para brindarnos los mejores resultados (mientras coloca anuncios rentables), software antispam aprende a filtrar nuestros mensajes de correo electrónico, y las transacciones con tarjeta de crédito están protegidas por un software que aprende a detectar fraudes [3].

El aprendizaje automático se define como un proceso automatizado que extrae patrones de los datos. Construimos los modelos utilizados en aplicaciones de análisis de datos predictivos, utilizamos supervisados aprendizajes automáticos. Aprendizaje automático supervisado 1 técnicas aprenden automáticamente un modelo de la relación entre un conjunto de características descriptivas y una característica objetivo basado en un conjunto de ejemplos históricos o instancias. Luego se puede usar este modelo para hacer predicciones para nuevas instancias [4]. El área de aprendizaje automatizado por

máquinas (ML) está interesada en responder a la pregunta de cómo una computadora puede “aprender” y si es posible medir su progreso y veracidad, su aplicación está determinada en diferentes ámbitos dependiendo de las necesidades del problema, permitiendo a los científicos y especialistas tomar decisiones referentes a temas tan puntuales como si un tipo de material puede ser separado por características de resistencia a la fragmentación o capacidad de absorción de energía pudiendo determinar con esto si es recomendable su uso en una u otra tarea específica.

Pero antes de responder a tales preguntas que se consideran como el ser humano “aprende”. El problema de las relaciones entre el lenguaje y el pensamiento puede ser planteado, entonces, a propósito de estas operaciones concretas en los siguientes términos: ¿es el lenguaje la única fuente de las clasificaciones, de las seriaciones [5], etc., que caracterizan a la forma de pensamiento relacionada con estas operaciones, o bien, al contrario, ¿son estas últimas relativamente independientes del lenguaje? .

Por ejemplo muy simplificado: todos los Pájaros (= clase A) son Animales (clase B). Pero todos los Animales no son Pájaros puesto que existen Animales-no-Pájaros (clase A'). El problema consiste entonces en saber si las operaciones $A + A' = B$ y $A \sim B - A'$ provienen únicamente del lenguaje, que permite agrupar los objetos en clases A, A', y B, o si estas operaciones tienen raíces más profundas que el lenguaje. Puede plantearse un problema análogo respecto a las series $A < B < C < \dots$ etc.

Pero el estudio del desarrollo de las operaciones en el niño permite hacer una constatación muy instructiva: esta constatación es que las operaciones que permiten reunir (+) o disociar (-) las clases o las relaciones son acciones propiamente dichas antes de ser operaciones del pensamiento. Antes de ser capaz de poder reunir o disociar las clases relativamente generales y abstractas, como por ejemplo las clases de Pájaros o de Animales [5].

Si analizamos como un niño determina que un objeto es tal y no otro. ¿Cómo encuentra esos rasgos que caracterizan al objeto? (forma, color, densidad, tamaño) que lo diferencian de algún otro y que son únicos a este.

Analizando este simple ejemplo, notamos que los seres humanos dependen de características (también llamadas atributos o incluso variables) para construir algún tipo de función de clasificación (o simplemente clasificadora) f para mapear un objeto x en una etiqueta y , es decir, $f(x) = y$. [6].

Como seres pensantes, estamos constantemente observando diferentes escenarios e intentando extraer las características asociadas que permitan tomar decisiones a situaciones específicas. Por ejemplo, Basándonos en las variables del clima y la estación del año, podríamos concluir sobre la viabilidad de tomar vacaciones en la playa o en la montaña. Nos referimos a esas variables como entrada, variables porque se tomarán como insumo para aprender algún concepto (por ejemplo, la viabilidad de tomar vacaciones en un lugar u otro). En este ejemplo, las variables de entrada sólo están asociadas al clima condiciones (temperatura y humedad, estación del año), pero también podríamos incluir las condiciones de factores de cercanía o económicos (por ejemplo, a cuantos kilómetros de distancia se encuentra cada uno, dinero disponible para gastar). Como en esta situación, simplemente se quiere predecir cuál es la mejor opción para tomar vacaciones, la variable de salida (también conocida como clase o etiqueta) es una (playa) o dos (montaña). Por lo tanto, basado en nuestras experiencias pasadas y dado las variables

de entrada actuales, se debe decidir si la salida es uno o dos. Hay situaciones más complejas en las que la variable de salida puede asumir más de dos etiquetas.

Por ejemplo, un mapa de bits “La imagen en mapa de bits se encuentra totalmente determinada por dos parámetros derivados de la estructuración en píxeles: la resolución y la profundidad de color.” [7] que representa una letra del alfabeto por ejemplo la letra “A”, se considera ahora la variable de salida como un valor entero, que asume 0 para “A” y 1 para “B”, y así sucesivamente. En este tipo de situación se puede tener más de dos clases. ¿Pero cómo se podría aprender los mapas de bits representados para todo el alfabeto? para hacer que la computadora aprenda la representación del alfabeto, primero se debe construir un conjunto de entrenamiento que contiene n ejemplos.

derivados de la estructuración en píxeles: la resolución y la profundidad de color.” [7] que representa una letra del alfabeto por ejemplo la letra “A”. Se considera la variable de salida como un valor entero, que asume 0 para “A” y 1 para “B”, y así sucesivamente. En este tipo de situación se tienen más de dos clases. ¿Pero cómo podríamos aprender esos mapas de bits representados para todo el alfabeto? para hacer que la computadora aprenda la representación del alfabeto, primero se debe construir un conjunto de entrenamiento que contiene n ejemplos.

III. Modelos

Modelos geométricos

El espacio de instancias es el conjunto de todas las instancias posibles o descriptibles, estén presentes en nuestro conjunto de datos o no. Por lo general, este conjunto tiene alguna estructura geométrica. Por ejemplo, si todas las características son numéricas, entonces se puede usar cada característica como una coordenada en un sistema de coordenadas cartesianas. Los modelos geométricos se construyen directamente en el espacio de instancias, utilizando conceptos geométricos como líneas, planos y distancias.

Modelos probabilísticos

segundo tipo de modelos son de naturaleza probabilística, como el clasificador bayesiano que consideramos antes. Muchos de estos modelos se basan en la siguiente idea. Sea X las variables que se conoce por ejemplo, los valores de las características de nuestra instancia; y deje que, Y denote las variables de destino que nos interesan, por ejemplo, la clase de la instancia. La pregunta clave en el aprendizaje automático es cómo modelar la relación entre X e Y.

Modelos lógicos

El tercer tipo de modelo que distinguimos es de naturaleza más algorítmica, y se inspira en la informática y la ingeniería. A este tipo lo llamo “lógico” porque los modelos de este tipo pueden traducirse fácilmente en reglas comprensibles para los humanos, como si

·if Viagra=1 then Class=Y=spam· [8].

IV. Tipos De Aprendizaje

Principales tipos de aprendizaje

El área de Aprendizaje Automático está típicamente organizada en tres ramas principales: aprendizaje supervisado, aprendizaje no supervisado y aprendizaje por refuerzo. [4]. El aprendizaje supervisado: en este tipo de aprendizaje el algoritmo de ML recibe ejemplos de entrada preetiquetados y pretende converger con el mejor clasificador posible $f: X \rightarrow Y$ [6], así que uno puede predecir las etiquetas de los ejemplos no vistos con gran precisión y cualquier divergencia entre f y su DPC implica una pérdida.

Por lo tanto, se puede formular el riesgo $R(f)$ como [25]. El aprendizaje no supervisado está asociado al proceso de construcción de modelos después de analizar las similitudes entre los datos de entrada (data-set). El aprendizaje por refuerzo es aquel que funciona con intervención humana mediante el proceso, motivo por el cual no se analizara este tipo.

Aprendizaje supervisado

Como ya se ha dicho, el aprendizaje supervisado consiste en encontrar el mejor clasificador posible $f: X \rightarrow Y$ para un problema determinado. El algoritmo responsable de encontrar este mapeo se denomina algoritmo de clasificación, que infiere un modelo de cada ejemplo de entrada $x \in X$ y su respectiva clase $y \in Y$. Este modelo es una aproximación para la distribución de probabilidad conjunta (DPC) [10] de las variables X e Y .

Como un clasificador puede ser visto como una función $f: X \rightarrow Y$, la forma más fácil de entender cómo obtener la mejor cartografía posible es probablemente considerando un problema de regresión. En una regresión, (Y) no es un conjunto de etiquetas discretas, pero a menudo un rango de valores reales, y, por lo tanto, será más fácil describir un ejemplo para f y comprender lo bien o mal que un clasificador se comporte referente a un problema específico.

Por ejemplo, se considera que se quiere encontrar una función para mapear la variable de entrada presión atmosférica a la clase de salida probabilidad de lluvia (que se ve como una variable continua en el rango $[0, 1]$). Por lo tanto, nuestro problema es responder si lloverá.

Para proceder, tomamos un conjunto de entrenamiento de observaciones previas basado en data-sets que nos cuantifiquen y cualifiquen las variables de entrada (presión atmosférica), para producir la probabilidad de lluvia como salida.

Para este tipo de problemas. Observamos que hay algún tipo de relación lineal entre esas variables, por lo que podríamos considerar un conjunto de funciones lineales para estimar el clasificador f . Pero, al usar una función f , algunos puntos de este conjunto de entrenamiento no coinciden con la función, es decir, no se evitan errores en el entrenamiento de la muestra.

Entonces, ¿qué tal si usamos una función polinómica g . La función g encajaría perfectamente en esta muestra. Esta pregunta motiva el dilema sesgo variante de la estadística, que está aquí básicamente traducido a: ¿Cuál función es la más apropiada para el entrenamiento de esta muestra? Para responder a esa pregunta, necesitamos cuantificar cual se adecúa mejor.

Pérdida de función.

Esta medida de adecuación requiere lo que se denomina pérdida [6], que se encarga de cuantificar el error o el riesgo [1, 2].

Dos de las funciones de pérdida más comúnmente utilizadas son:

la función de pérdida 0-1

$$l_{0-1}(X, Y, f(X)) = \begin{cases} 1, & \text{si } f(X) \neq Y \\ 0, & \text{si no} \end{cases} \quad l_{0-1}(X, Y, f(X)) = \begin{cases} 1, & \text{si } f(X) \neq Y \\ 0, & \text{si no} \end{cases}$$

(1) la función de pérdida por error al cuadrado.

$$l^2(X, Y, f(X)) = (Y - f(X))^2$$

La primera consiste básicamente en contar cuántas veces la clase de salida está mal clasificada por f y la pérdida por error al cuadrado cuantifica los errores de manera continua, por lo que es adecuado cuando la salida del clasificador está en \mathbb{R} , como en los problemas de regresión.

Después de seleccionar la función de pérdida más adecuada, ahora se tiene una forma de cuantificar apropiadamente y entonces se puede volver a la pregunta principal: ¿es la función lineal o la polinómica más apropiada para abordar el problema descrito anteriormente? Analizando sólo la pérdida de los datos de entrenamiento, uno podría decir que la función polinómica es mejor. Sin embargo, para responder adecuadamente a esto se debe evaluar qué tan bien están ambos clasificadores al predecir lo que no se ve.

¿Por qué lo que no se ve? Porque eso nos permite confirmar si los clasificadores están sesgados para representar simplemente los datos de entrenamiento. Por ejemplo, si se tienen dos estudiantes tratando de aprender programación orientada a objetos por sí mismos. Asumiendo que el primer estudiante memoriza tanto como sea posible el libro A, mientras que el segundo tiene una estrategia: evitará la memorización y estudiará utilizando una variedad de libros, así como codificar algunas aplicaciones. Ahora se consideran dos pruebas aplicadas para evaluar los conocimientos adquiridos por ambos estudiantes:

- La primera prueba se basa sólo en los ejemplos detallados en el libro A, así que el primer estudiante probablemente obtendría una puntuación perfecta, mientras que el segundo obtendría una calificación justa;
- La segunda prueba se basa en varios libros no vistos (excluyendo todos los libros que los estudiantes habían visto antes), programas y problemas generales, el segundo estudiante probablemente volvería a obtener una puntuación justa, mientras que el primero fallaría. Entonces ¿qué pasó?, probablemente observamos este escenario varias veces, pero ¿cuál es el mejor estudiante?

En términos de aprendizaje supervisado, la capacidad de generalización de conocimientos es lo mejor. Cuando generalizamos, el estudiante aprende de hecho un tema y cómo aplicarlo en el mundo real a situaciones en lugar de sólo memorizarlo. ¿Cómo se podría medir tal logro?

Riesgo de la misma manera, se espera que un clasificador f proporcione resultados justos para ejemplos no vistos. Se puede pensar en una prueba usando preguntas de los libros, de los estudiantes que fallaron la prueba. Para calcular lo bueno que es un clasificador [24], se debe considerar el valor esperado de la función de pérdida como se define (expresado por: $=$) en la siguiente ecuación:

$$R(f) := E(\lim(X, Y, f(X))).$$

Esta medida da como resultado el riesgo esperado $R(\cdot)$ de algún clasificador f , dado cualquier posible ejemplo $x \in X$ (incluyendo los no vistos) que debería ser correctamente etiquetado como $y \in Y$.

El valor esperado puede ser interpretado aproximadamente como el valor promedio de la pérdida, dados todos los ejemplos posibles de entrada. Obsérvese que esto no es factible para problemas del mundo real que casi nunca proporcionan acceso a todos los ejemplos que cubren todo un rango de posibilidades. Para calcular el valor esperado, necesitamos la distribución de probabilidad conjunta (DPC) de las variables X e Y para que haya un acceso completo a cada combinación de $x \in X$ frente a sus correspondientes clases $y \in Y$.

La distribución de probabilidad conjunta (DPC) $P(X \times Y)$ se encarga de asociar los ejemplos de entrada $x \in X$ a sus correspondientes clases de salida $y \in Y$. La DPC real requeriría infinitas observaciones de datos una vez que cada variable está en R .

De hecho, los algoritmos supervisados por ML intentan encontrar el mejor clasificador que se ajuste a $P(X \times Y)$ con el menor error posible. Por lo tanto, se puede inferir lo que probablemente ocurre cuando se observa una presión atmosférica determinada.

Por ejemplo, la presión atmosférica actual es de 65 kPa (Kilo Pascales), por lo que analizamos el rango $[64,5, 65,5]$ (centrado a 65 kPa) para comprobar el valor más probable de Y . En la práctica, los algoritmos de clasificación intentan estimar el DPC basándose en los ejemplos de entrada. El mejor clasificador f es el que mejor se ajusta a $P(X \times Y)$, y cualquier divergencia entre f y su DPC implica una pérdida. Por lo tanto, se puede formular el riesgo $R(f)$ como [25]

(3) la integral de la divergencia:

$$\int_{X \times Y} l(X, Y, f(X)) d(X \times Y).$$

Por supuesto, hay problemas para los cuales el DPC $P(X \times Y)$ tiene un finito y un número accesible de posibilidades, por consiguiente $R(f)$ es computable.

Regresión lineal

La regresión lineal es una técnica de aprendizaje supervisado que normalmente se utiliza para predecir, pronosticar y encontrar relaciones entre datos cuantitativos. Es una de las primeras técnicas de

aprendizaje, que todavía se usa ampliamente. Por ejemplo, esta técnica se puede aplicar para examinar si existe una relación entre el presupuesto publicitario de una empresa y sus ventas. También puede usarlo para determinar si existe una relación lineal entre una radioterapia en particular y el tamaño del tumor.

Técnicas de clasificación

Las técnicas de clasificación que se discutirán en esta sección son aquellas que se enfocan en predecir una respuesta cualitativa mediante el análisis de datos y el reconocimiento de patrones. Por ejemplo, este tipo de técnica se utiliza para clasificar si una transacción con tarjeta de crédito es fraudulenta o no. Existen muchas técnicas de clasificación o clasificadores diferentes, pero algunas de las más utilizadas incluyen:

- Regresión logística,
- Análisis discriminante lineal,
- K vecinos más cercanos,
- Árboles,
- Redes neuronales y
- Máquinas vectoriales de soporte.[11]

Aprendizaje no supervisado

El aprendizaje no supervisado es aprender sin un maestro. Una cosa básica que puede querer hacer con los datos es visualizarlos. Lamentablemente, es difícil visualizar las cosas en más de dos (o tres) dimensiones, y la mayoría de los datos se encuentran en cientos de dimensiones (o más). La reducción de la dimensionalidad es el problema de tomar datos de alta dimensión y anidarlos en un espacio de menor dimensión. Otra cosa que podría querer hacer es derivar automáticamente una partición de los datos en grupos [12],

Muchos problemas caen naturalmente en los paradigmas de aprendizaje supervisado o no supervisado. Sin embargo, a veces la cuestión de si un análisis debe considerarse supervisado o no supervisado es menos clara. Por ejemplo, suponga que se tiene un conjunto de observaciones. Forma de las observaciones, donde $m < n$, se tienen medidas predictoras como una medida de respuesta. Para el resto de $n-m$ observations, se cuenta con mediciones de predicción, pero no medición de respuesta. Tal escenario puede surgir si los predictores se pueden medir de forma relativamente barata pero las respuestas correspondientes son mucho más caras de recopilar. Nos referimos a esta configuración como problema de aprendizaje semi-supervisado. En este contexto, deseamos utilizar un método de aprendizaje estadístico semi-supervisado que pueda incorporar las observaciones para las que se dispone de medidas de respuesta, así como las observaciones de entonces para las que no lo están [13].

Aprendizaje semisupervisado

A veces llamado aprendizaje semisupervisado, en el que el objetivo es la clasificación, pero la entrada contiene datos etiquetados y sin etiquetar. No se puede hacer una clasificación sin datos etiquetados, por supuesto, porque solo las etiquetas indican cuáles son las clases. Pero a veces es atractivo aumentar una pequeña cantidad de datos etiquetados con una gran cantidad de datos no etiquetados. Resulta que los datos sin etiquetar pueden ayudarlo a aprender las clases. ¿Cómo puede ser esto? Primero, ¿por qué lo querrías? Muchas situaciones presentan grandes volúmenes de datos sin procesar, pero la asignación de clases es costosa porque requiere conocimientos humanos. La minería de texto proporciona algunos ejemplos clásicos.[14].

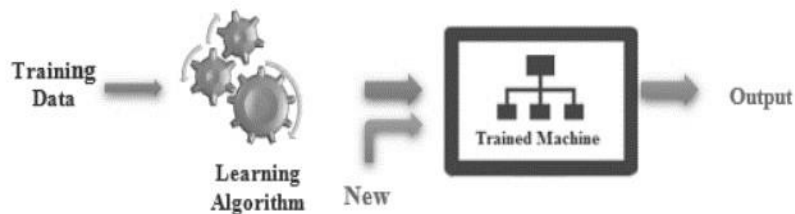


FIGURE 1. Different stages of machine learning.

Figura1. Diferentes estados del aprendizaje supervisado [15].

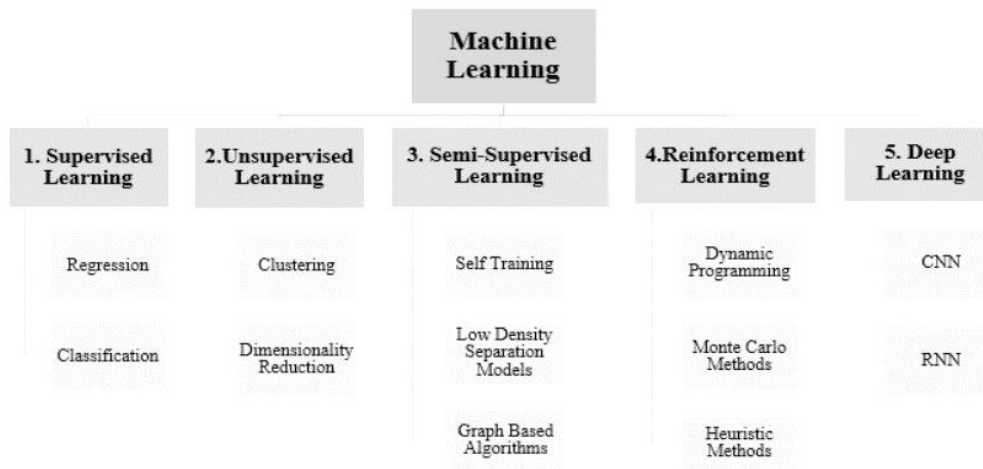


Figura 2. Diferentes tipos de aprendizaje automático y algoritmos [16].

V. Métodos De Predicción

- Redes Neuronales

Una red neuronal es un sistema que imita el cerebro humano en la transformación de entradas en una o más salidas con un conjunto de neuronas. Para encontrar el modelo más eficiente en este estudio se utilizaron las redes multinivel siempre hacia adelante (MLFN) y las redes neuronales probabilísticas (PNN). De modo específico, una función tangente hiperbólica se usa como función de activación en las neuronas de la capa oculta para la red MLFN. La neurona de salida emplea la identidad como función de activación, la cual simplemente devuelve la suma ponderada de sus entradas. Las redes PNN fueron propuestas primero por Specht (1991) y aproximan cualquier función arbitraria entre los vectores de entrada y salida con el trazo de la función estimada de manera directa desde los datos de entrenamiento [17] Las características fundamentales de las RNA son: Aprenden de la experiencia: Las RNA pueden modificar su comportamiento como respuesta a su entorno. Dado un conjunto de entradas (quizá con las salidas deseadas), las RNA se ajustan para producir respuestas consistentes. Una amplia variedad de algoritmos de entrenamiento se ha desarrollado, cada uno con sus propias ventajas e inconvenientes.

Generalizan de ejemplos anteriores a los ejemplos nuevos: Una vez que la RNA esté entrenada, la respuesta de la red puede ser, hasta un cierto punto, insensible a pequeñas variaciones en las entradas, lo que las hace idóneas para el reconocimiento de patrones. Abstracción de la esencia de las entradas: Algunas RNA son capaces de abstraer información de un conjunto de entradas. Por ejemplo, en el caso de reconocimiento de patrones, una red puede ser entrenada en una secuencia de patrones distorsionados de una letra. Una vez que la red sea correctamente entrenada será capaz de producir un resultado correcto ante una entrada distorsionada, lo que significa que ha sido capaz de aprender algo que nunca había visto [18] Una RNA, en el contexto del método propuesto, es en esencia un modelo matemático no lineal que contiene parámetros conocidos como “pesos”. Cuántos más pesos tenga una RNA, mayor no linealidad se podrá representar con ella. Para encontrar estos pesos se utilizan algoritmos de optimización que minimizan una función de errores cuadrados de aproximación. En el área de redes neuronales, varios de estos algoritmos de optimización se clasifican bajo el nombre de “algoritmos de retro propagación”. Por otro lado, encontrar un conjunto de pesos que permitan una aproximación adecuada a datos conocidos se le denomina “entrenamiento” [19].

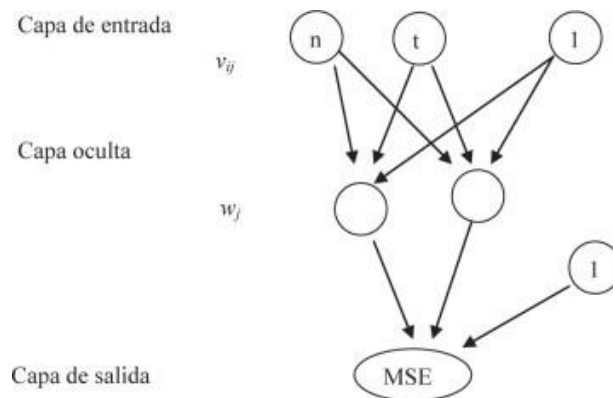


Figura 3. RNA con tres capas, entrenada por retro propagación del error

- **Regresión**

Desde el punto de vista matemático, la Regresión tiene dos significados: uno surge de la distribución conjunta de probabilidades de dos variables aleatorias mientras que el otro significado es empírico y nace de la necesidad de ajustar alguna función a un conjunto de datos. [19]

En la regresión simple se tiene una única variable predictora. Algunas veces se tiene interés en dos o más variables regresoras o predictoras. En esos casos, se debe recurrir al uso de regresión múltiple. A partir de la regresión lineal es posible hacer predicciones sobre la respuesta con base en valores de la variable predictora. La ecuación para una línea recta es donde y es la variable respuesta; x es la variable predictora es la intersección, determina el valor de y cuando x es cero, y b es la pendiente, determina la cantidad en la que cambia y cuando x se incrementa en una unidad. Las distancias entre los puntos y la línea de regresión se llaman residuos. Ellos representan la porción de la respuesta que no es explicada por la ecuación de regresión; es decir que la diferencia entre el valor observado y el valor aproximado es el residuo. En cualquier análisis de regresión se observará que algunos puntos están más cerca de la línea y otros mucho más lejos de ella. Entre más cerca se encuentren los puntos a la línea, mejor será el ajuste entre la línea de regresión y el dato. Los residuos permiten verificar la ecuación con el fin de comprobar cuan bien se ajusta la línea a los datos. [20]

Las Regresiones también son una subcategoría de los algoritmos de aprendizaje supervisado, pero se utilizan en la predicción de valores continuos. En el análisis de regresiones, se entrega un número o números (predictor) que explican el valor de salida, es decir, los valores de entrada guardan relación con la variable a explicar. Dentro de las regresiones, la más común es la regresión lineal, que mediante una ecuación de primer orden busca explicar un valor de salida [21].

- **Clasificación**

Como se ha comentado previamente, el aprendizaje automático puede ser usado en varios campos. Uno de ellos muy usado es el de la clasificación, donde se pueden usar diferentes técnicas o algoritmos que nos ayudarán a extraer información relevante y utilizar dependiendo del objetivo buscado. Un ejemplo comúnmente conocido es el de la clasificación binaria para la detección de correo spam. Se pretende detectar y clasificar un correo entrante para saber si lo consideramos como spam o no. Deberíamos tener en cuenta que este tipo de información es más sofisticada y humanamente no seríamos capaces de identificar si es o no malintencionado. Así, gracias al aprendizaje automático es posible procesar las distintas características que conforman un correo de este tipo y por lo tanto usarlo como filtro para que sea adecuadamente clasificado antes que llegue a nuestro buzón. El uso del aprendizaje automático nos ayudaría a una clasificación más acertada de la que seríamos capaz de realizar por nosotros mismos [22].

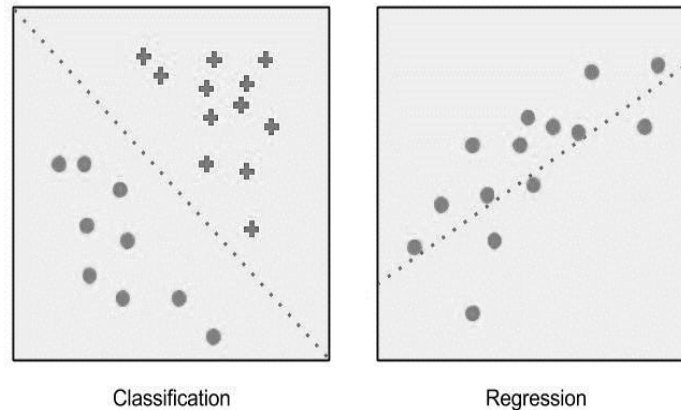


Figura 4. Comparación grafica entre regresión y clasificación [23].

VI. Conclusiones

En conclusión, las técnicas de aprendizaje automático aplicadas en los sistemas de predicción permiten a un sistema aprender de los datos en lugar de aprender mediante la programación explícita. Machine learning no es un proceso sencillo. Conforme el algoritmo recibe datos de entrenamiento, es posible producir modelos más precisos basados en datos. Un modelo de machine learning es la salida de información que se genera cuando entrena su algoritmo de machine learning con datos. Después del entrenamiento, al proporcionar un modelo con una entrada, se le dará una salida. Por ejemplo, un algoritmo predictivo creará un modelo predictivo.

VII. Recomendaciones Para Trabajos Futuros

En los diferentes documentos analizados se pudo evidenciar que se tiene resultados positivos y determinante en los estudios realizados, en la mayoría de los casos se cumple con los objetivos iniciales que se establecieron, se encuentra información patrones relevantes que aportan a la toma de decisiones de casa caso en cuestión, la selección de un algoritmo se puede realizar de acuerdo al tipo de objetivo que se requiere alcanzar.

Debido a este planteamiento en muchos de los documentos se plantea el mismo escenario con la aplicación de dos técnicas diferentes para establecer cual tuvo mejor resultado, otro factor determinante es la selección de variables independientes.

IX. Referencias Bibliograficas

- [1]. Mitchell, T., 1997. Machine Learning. 1st ed. New York: MacGraw-Hill, p.15.
- [2]. Murphy, K., 2012. Machine Learning A Probabilistic Perspective. 3rd ed. New York: The MIT Press, p.1.
- [3]. Murphy, K., 2012. Machine Learning. 3rd ed. New York: The MIT Press, p.1. Shalev-Shwartz, S. and Ben-David, S., 2014. Understanding Machine Learning: From Theory to Algorithms. 1st ed. Cambridge: Cambridge University Press, p.19.
- [4]. Kelleher, J., Mac Namee, B. and D'Arcy, A., 2015. Fundamentals Of Machine Learning For Predictive Data Analytics. 1st ed. Cambridge (Massachusetts): MIT Press, p.43. Mitchell, T., 1997. Machine Learning. 1st ed. New York: MacGraw-Hill, p.15.
- [5]. J.Piaget, Seis Estudios de Psicología (Editorial Labor, S.A, p117, España, 1991)
- [6]. R Jiménez, Funciones matemáticas 4 (Pearson Prentice Hall, México,p5, 2006, 168)
- [7]. Gráficos de mapa de bits [Online]. Available: http://cv.uoc.edu/annotation/02dada81f121a9c9bf2f770a98d79c82/573715/PID_00236882/modul_4.html
- [8]. P. Flach, Machine Learning The Art and Science of Algorithms that Make Sense of Data, 1st ed. New York: Cambridge University Press, 2012, pp. 20-32
- [9]. JJ Beunza, E Puertas, E Condés, Manual práctico de inteligencia artificial en Entornos Sanitarios (Elsevier, p35, Madrid España, 2019)
- [10]. Walpole, Ronald E, Probabilidad y estadística para ingenieros, 6ª. Ed. (prentice-hallhispanoamericana, s.a., p69 méxico, 1999)
- [11]. M. Talabis, R. McPherson, I. Miyamoto and J. Martin, Information Security Analytics, 1st ed. Syngress, 2014, p. 35.
- [12]. H. Daumé III, A Course in Machine Learning, 2nd ed. self -published, 2013, p. 178.
- [13]. G. James, D. Witten, T. Hastie and R. Tibshirani, An introduction to statistical learning, 1st ed. New York: Springer, p. 28.
- [14]. I. Witten, E. Frank and M. Hall, Data Mining Practical Machine Learning Tools and Techniques, 3rd ed. San Diego: Morgan Kaufmann, 2011, p. 294.
- [15]. A. B. Nassif, I. Shahin, I. Attili, M. Azzeh and K. Shaalan, "Speech Recognition Using Deep Neural Networks: A Systematic Review," in IEEE Access, vol. 7, pp. 19143-19165, 2019, doi: 10.1109/ ACCESS.2019.2896880.
- [16]. A. B. Nassif, I. Shahin, I. Attili, M. Azzeh and K. Shaalan, "Speech Recognition Using Deep Neural Networks: A Systematic Review," in IEEE Access, vol. 7, pp. 19143-19165, 2019, doi: 10.1109/ ACCESS.2019.2896880.
- [17]. aplicación de redes neuronales en la clasificación de arcillas, 17th ed. Medellin: Revista EIA, 2012, p. 185 [Online]. Available: <http://www.scielo.org.co/pdf/eia/n17/n17a14.pdf>. [Accessed: 08-Nov- 2020]
- [18]. Www2.ulpgc.es, 2020. [Online]. Available: https://www2.ulpgc.es/hege/almacen/download/38/38584/practica_ia_2.pdf. [Accessed: 14- Nov- 2020]
- [19]. M.C. Acosta-Cervantes, M.G. Villarreal-Marroquín, M. Cabrera-Ríos, "Estudio de validación de un método para seleccionar técnicas de pronóstico de series de tiempo mediante redes neuronales artificiales", Ingeniería, Investigación y Tecnología, Volume 14, Issue 1, 2013, Pages 53-63, ISSN 1405-7743, [https://doi.org/10.1016/S1405-7743\(13\)72225-X](https://doi.org/10.1016/S1405-7743(13)72225-X).
- [20]. Carrasquilla-Batista, A; Chacón-Rodríguez, A; NúñezMontero, K, "Regresión lineal simple y múltiple: aplicación en la predicción de variables naturales relacionadas con el crecimiento microalgal". Tecnología en Marcha. Encuentro de Investigación y Extensión 2016. Pág 33-45
- [21]. "Introducción a Machine Learning: Algunos conceptos básicos – Llpe.com", Llpe.com, 2020. [Online]. Available: <https://www.llpe.com/2017/03/13/introduccion-a-machine-learning/>. [Accessed: 16- Nov- 2020]

- [22]. J. Ruiz, “Comparativa y análisis de algoritmos de aprendizaje automático para la predicción del tipo predominante de cubierta arbórea”, máster en ingeniería en informática, Universidad Complutense de Madrid, 2018.
- [23]. Searchengineland.com, 2020. [Online]. Available: <https://searchengineland.com/figz/wp-content/uploads/2016/07/class-regress.png>. [Accessed: 19- Nov- 2020].
- [24]. V. Vapnik, *The Nature of Statistical Learning Theory*. Information Science and Statistics (Springer, New York, 1999)
- [25]. U. von Luxburg, B. Schölkopf, *Statistical Learning Theory: Models, Concepts, and Results*, vol. 10 (Elsevier North Holland, Amsterdam, 2011), pp. 651–706 [26]. D. Forsyth, *Probability and Statistics for Computer Science* (Springer, New York, 2018)
- [27]. Z. Hu, Y. Jin, Q. Hu, S. Sen, T. Zhou y MT Osman, “Predicción del consumo de combustible para barcos en ruta basada en aprendizaje automático”, en *IEEE Access* , vol. 7, págs.119497-119505, 2019, doi: 10.1109 / ACCESS.2019.2933630.
- [28]. L. Yang y Z. Zhou, “Sistema experto basado en reglas de creencias para problemas de clasificación con la activación de nuevas reglas y procedimientos de cálculo de peso”, *Inf. Sci.* , vol. 336, págs. 75-91, abril de 2016.
- [29]. S. Ameer et al., “Comparative Analysis of Machine Learning Techniques for Predicting Air Quality in Smart Cities,” in *IEEE Access*, vol. 7, pp. 128325-128338, 2019, doi: 10.1109/access.2019.2925082.
- [30]. D. Zhang y SS Woo, “Monitoreo y predicción de la calidad del aire localizada en tiempo real a través de una red de detección de IoT fija y móvil”, en *IEEE Access* , vol. 8, págs. 89584-89594, 2020, doi: 10.1109 / ACCESS.2020.2993547.
- [31]. L. Zhang, D. Li y Q. Guo, “Aprendizaje profundo de datos espacio-temporales utilizando CNN residual de regularización ortogonal para la predicción aérea”, en *IEEE Access* , vol. 8, págs.6603766047, 2020, doi: 10.1109 / ACCESS.2020.2985657.
- [32]. L. Zhang, D. Li y Q. Guo, “Aprendizaje profundo de datos espacio-temporales utilizando CNN residual de regularización ortogonal para la predicción aérea”, en *IEEE Access* , vol. 8, págs.6603766047, 2020, doi: 10.1109 / ACCESS.2020.2985657.
- [33]. D. Zhang and S. S. Woo, “Real Time Localized Air Quality Monitoring and Prediction Through Mobile and Fixed IoT Sensing Network,” in *IEEE Access*, vol. 8, pp. 8958489594, 2020, doi: 10.1109/ACCESS.2020.2993547.
- [34] C. Zhang, H. Zhang y D. Liu, “Un estudio contrastante del aprendizaje automático sobre la predicción del valor de la empresa energética”, en *IEEE Access*, vol. 8, págs.11635-11643, 2020, doi: 10.1109 / ACCESS.2019.2953807.
- [35]. J. Zhang, Z. Li, Z. Pu and C. Xu, “Comparing Prediction Performance for Crash Injury Severity Among Various Machine Learning and Statistical Methods,” in *IEEE Access*, vol. 6, pp. 60079-60087, 2018, doi: 10.1109/ACCESS.2018.2874979.

Publicación Facultad de Ingeniería y Red de Investigaciones de Tecnología Avanzada – RITA

REVISTA

TIA