

Algoritmos De Machine Learning

En 1940 fue creado el primer sistema informático operado manualmente llamado ENIAC (Electronic Numerical Integrator and Computer). Con el paso del tiempo se desarrollaron más sistemas informáticos complejos, por lo tanto, el desarrollo de la inteligencia artificial iba en aumento. “Gracias a la estadística, el Machine Learning se hizo muy famoso en 1990. La intersección de la informática y la estadística dio lugar a enfoques probabilísticos en la inteligencia artificial. Esto desplazó el campo aún más hacia enfoques basados en datos [...] los científicos comenzaron a construir sistemas inteligentes que podían analizar y aprender de grandes cantidades de datos”. [1]

Según Arthur Samuel, los algoritmos de Machine Learning permiten a las computadoras aprender de los datos e incluso mejorarse a sí mismos, sin ser programados explícitamente.

El Machine Learning es una categoría de algoritmo que permite que las aplicaciones de software sean más precisas para predecir resultados sin ser programadas explícitamente.

En el libro *Foundations of Machine Learning* mencionan que el Machine Learning puede definirse, en términos generales, como métodos computacionales que utilizan la experiencia para mejorar el rendimiento o hacer predicciones precisas”. [2] Al decir experiencias se refieren a los datos disponibles para que la maquina los analice.

En el aprendizaje supervisado, los algoritmos trabajan con datos “etiquetados” (labeled data), intentado encontrar una función que, dadas las variables de entrada (input data), les asigne la etiqueta de salida adecuada. El algoritmo se entrena con un “histórico” de datos y así “aprende” a asignar la etiqueta de salida adecuada a un nuevo valor, es decir, predice el valor de salida. [3]

Ejemplos de Técnicas de Machine Learning:

A continuación, hablaré sobre las técnicas que yo he usado para aplicaciones de comparación supervisadas con datos CSV, son las que más conozco y las explicaré de forma histórica y el funcionamiento de cada uno.

KNeighbors

“La clasificación K-nearest-neighbor (KNN) se desarrolló a partir de la necesidad de realizar análisis discriminantes cuando las estimaciones paramétricas fiables de densidades de probabilidad son desconocidas o difíciles de determinar. En un informe [...] de la Escuela de Medicina de Aviación de la Fuerza Aérea de los Estados Unidos en 1951, Fix y Hodges introdujeron un método no paramétrico para la clasificación de patrones que desde entonces se ha dado a conocer la regla de K-nearest neighbor”. [4]

El clasificador k-nearest-neighbor se basa comúnmente en la distancia euclidiana entre una muestra de prueba y las muestras de entrenamiento especificadas. KNeighbors se puede usar para clasificación y regresión.

KNeighbors es un método que simplemente busca en las observaciones más cercanas a la que se está tratando de predecir y clasificar el punto de interés basado en la mayoría de datos que le rodean [5]. Este algoritmo es de tipo supervisado, es decir, tenemos etiquetado nuestro conjunto de datos de entrenamiento, con la clase o resultado esperado dado un conjunto de datos.

También es un algoritmo basado en instancia, esto quiere decir que son modelos de aprendizaje para problemas de decisión con instancias o muestras de datos de entrenamiento que son importantes o requeridos por el modelo [6].

El funcionamiento de KNeighbors se basa en calcular la distancia entre el elemento a clasificar del resto de conjunto de datos y con esto selecciona los “K” elementos más cercanos dependiendo de la función que se use. La forma más usada para medir estas distancias o la cercanía entre elementos es mediante la distancia euclidiana que básicamente es la distancia entre dos puntos de un espacio euclídeo.

Decision Tree Classifier

Ross Quinlan fue quien desarrolló el método de Árbol de Decisión (Decision Tree) en la universidad de Sydney y lo presentó en su libro Machine Learning en 1975, a este algoritmo de Árbol de Decisión lo llamó Iterative Dichotomiser 3 (ID3).

“El análisis de árbol de decisiones es una herramienta de modelado predictivo general que tiene aplicaciones que abarcan varias áreas diferentes [7]

En el libro Decision Trees for Analytics Using SAS Enterprise Miner mencionan que: “Los árboles de decisión son una forma de análisis de múltiples variables (o efectos múltiples). Todas las formas de múltiples análisis de variables nos permiten predecir, explicar, describir o clasificar un resultado u objetivo. [...] Esta capacidad de análisis de múltiples variables de árboles de decisión le permite ir más allá de las simples relaciones de un solo efecto y descubrir y describir cosas en el contexto de múltiples influencias. El análisis de variables múltiples es particularmente importante en la resolución de problemas actual porque casi todos los resultados críticos que determinan el éxito se basan en múltiples factores” [8]

En general, los árboles de decisión se construyen a través de un enfoque algorítmico que identifica formas de dividir un conjunto de datos basado en diferentes condiciones.

El algoritmo de árbol de decisión modela la toma de decisión basado en los valores de los atributos que tienen nuestros datos.

Cabe destacar que este algoritmo usa el modelo de Decision Tree (Árbol de Decisión) que básicamente va modelando y bifurcando a partir de los datos dados caminos con posibles escenarios para llegar al resultado.

REFERENCIAS

- [1] Pant, A. (07 de 01 de 2019). Medium. Obtenido de <https://towardsdatascience.com/introduction-to-machine-learning-for-beginners-eed6024fdb08>
- [2] Mehryar, M., Afshin, R., & Ameet, T. (2012). Foundations of Machine Learning. E.U.A: MIT.
- [3] Simeone, O. (05 de 11 de 2018). Cornell University. Obtenido de <https://arxiv.org/pdf/1808.02342.pdf>
- [4] Peterson, L. E. (2009). Scholarpedia. Recuperado el 28 de 06 de 2020, de http://www.scholarpedia.org/article/K-nearest_neighbor
- [5] Na8. (10 de 07 de 2018). Aprende Machine Learning. Recuperado el 28 de 06 de 2020, de <https://www.aprendemachinelearning.com/clasificar-con-k-nearest-neighbor-ejemplo-en-python/>
- [6] Na8. (04 de 11 de 2017). Aprende Machine Learning. Recuperado el 28 de 06 de 2020, de <https://www.aprendemachinelearning.com/principales-algoritmos-usados-en-machine-learning/>
- [7] Advanced Management Science - Decision Tree Analysis. (s.f.). Recuperado el 28 de 06 de 2020, de <http://amsdecisiontreeanalysis.weebly.com/>
- [8] De Ville, B., & Neville, P. (2013). Decision Trees for Analytics Using SAS Enterprise Miner. Carolina del Norte, E.U.A: SAS.
- [9] Desai, R. (s.f.). Medium. Recuperado el 06 de 07 de 2020, de <https://towardsdatascience.com/top-10-algorithms-for-machine-learning-beginners-149374935f3c>