

Lección 5: Aprendizaje supervisado: Algoritmos de clasificación y regresión

A la hora de elegir un algoritmo también es importante tener en cuenta que tipo de aprendizaje se está utilizando dentro del machine learning, es decir, si se está hablando de aprendizaje supervisado o no supervisado. De acuerdo con el tipo de aprendizaje se utilizan diferentes algoritmos.

Una de las modalidades de aprendizaje del machine learning, es la de **aprendizaje supervisado**. A través de esta, se entrena un algoritmo otorgándole las preguntas, que se denominan características, y las respuestas conocidas como etiquetas. Lo anterior se realiza con el fin de que el algoritmo las combine y se puedan hacer predicciones. La técnica es comúnmente utilizada para deducir una función a partir de los datos de entrenamiento, que consisten en partes de objetos dentro de los cuales uno de los componentes son los datos de entrada y el componente restante son los resultados deseados.

El objetivo fundamental del aprendizaje supervisado es la creación de funciones con la posibilidad de predecir valores correspondientes a objetos de entrada luego de haberse familiarizado con una serie de ejemplos que son los datos de entrenamiento.

X – Conjunto de objetos

Y – Conjunto de etiquetas o respuestas

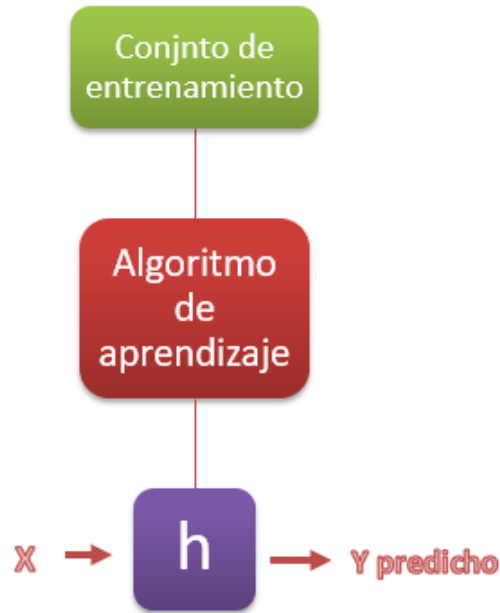
$\gamma: X \rightarrow Y$ – Valores conocidos de la función

$$\gamma^i = \gamma(x^{(i)}), i = 1, \dots, \iota$$

El problema es encontrar la hipótesis h que permita acercar el conjunto de objetos al de respuestas. Por ejemplo, es posible contar con información de casas (conjunto de ejemplos), y los precios que le incumbe a cada una (respuesta). Con base en ello, se quiere aprender alguna regla de correspondencia entre casas y sus precios que permita predecir el costo cuando se tenga información de casas nuevas.

El aprendizaje supervisado consta de dos pasos esenciales:

1. El entrenamiento, donde se utiliza el conjunto de entrenamiento con un algoritmo que permita crear una hipótesis.
2. La prueba, en donde se utiliza la hipótesis obtenida en el paso anterior para generar nuevas predicciones con objetos nuevos.



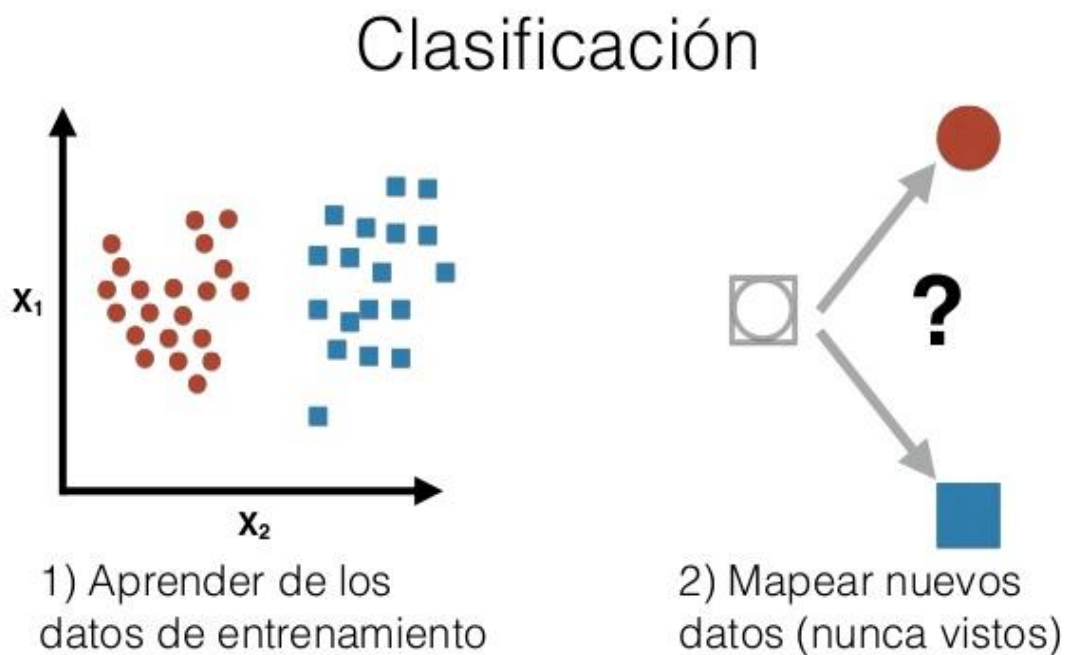
El proceso es conocido como aprendizaje supervisado, pues al conocer las respuestas de cada ejemplo del conjunto de entrenamiento, es posible corregir la hipótesis generada por el algoritmo. Se supervisa el entrenamiento del algoritmo mediante la corrección de parámetros del mismo, según sean los resultados que se obtienen de forma iterativa.

Es posible que la salida de la función sea de valor numérico, en cuyo caso serán de regresión, o una etiqueta de clase, de clasificación.

Los **algoritmos de clasificación** se utilizan cuando el resultado deseado es una etiqueta discreta. Es decir, son útiles cuando la respuesta a la pregunta sobre la empresa se aloja dentro de un conjunto finito de resultados posibles. Como, por ejemplo, la determinación de si un correo electrónico es deseado o no deseado, en dicho caso solo se tienen dos opciones y se conoce como **clasificación binaria**.

Por otra parte, **la clasificación de múltiples categorías** logra capturar todo lo demás. Es muy útil para segmentar al cliente, categorizar imágenes y audios y analizar texto para optimizar el sentimiento del cliente.

Los algoritmos enfocados a la clasificación trabajan generalmente sobre la información entregada por un conjunto de muestras, patrones, ejemplos o prototipos de entrenamiento que son tomados como representantes de las clases, y los mismos conservan una etiqueta de clase correcta. A este conjunto de prototipos correctamente etiquetados se les llama conjunto de entrenamiento, y es el conocimiento disponible para la clasificación de nuevas muestras. El objetivo de la clasificación supervisada es determinar, según lo que se tenga conocimiento, cual es la clase a la que debería concernir una nueva muestra, teniendo en cuenta la información que se pueda extraer.



Fuente: Phytion Diario

Algunos de los ejemplos de tareas de clasificación son:

- Predecir tumores malignos o benignos
- Clasificar transacciones de tarjeta de crédito como legítimas o fraudulentas
- Clasificar estructuras secundarias de proteínas
- Categorizar artículos de noticias como financieros, clima, entretenimiento, deportes, entre otros.

La clasificación es muy similar al proceso de aprendizaje de las personas, ya que poseemos la capacidad de clasificar alimentos, libros, animales, planetas, es decir, todo lo que nos rodea.

Por otro lado, los **algoritmos de regresión** son útiles para predecir productos que son continuos, lo que significa que las respuestas a los interrogantes se representan mediante una cuantía que puede ser definitiva de forma flexible en función de las entradas del modelo, en lugar de restringirse a un conjunto de posibles etiquetas. Los problemas de regresión con entradas ordenadas por tiempo se designan como problemas de pronóstico de series temporales, como el pronóstico ARIMA, que permite exponer patrones estacionales en aspectos como ventas, evaluar el impacto de campañas de marketing, entre otros.

Aunque existen muchos modelos de regresión, el más común es la regresión lineal. Esta consiste en un modelo matemático utilizado para aproximar la relación de dependencias entre una variable dependiente Y , una variable independiente X , y un término aleatorio ε .

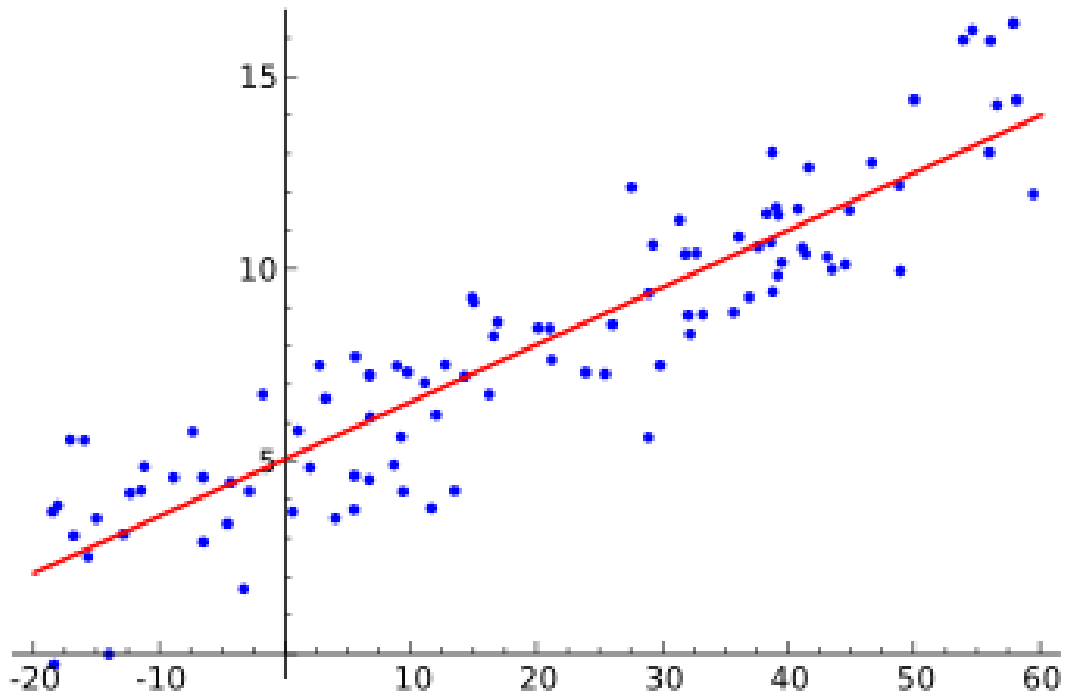
$$Y_t = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon$$

Donde:

Y_t : variable dependiente o regresado

X_1, X_2, \dots, X_p : variables independientes o regresoras

$\beta_1, \beta_2, \dots, \beta_p$: parámetros que miden la influencia de las variables independientes sobre la dependiente



Fuente: Python Diario

La línea roja representa el modelo de acción que trata de ajustarse lo mejor posible.

La regresión lineal como modelo estadístico posee muchos usos y aplicaciones, como determinar si un conjunto de datos ha presentado aumento o decremento en un periodo establecido, como el precio del petróleo, el valor de acciones, PIB de un país, ventas de una empresa, entre otros.

Referencias:

- Cassis, A. (2015). Inteligencia artificial 101. Recuperado el 5 de octubre de 2018, de <https://inteligenciaartificial101.wordpress.com/2015/10/20/aprendizaje-supervisado/>
- EcuRed. (s.f.). EcuRed. Recuperado el 5 de octubre de 2018, de https://www.ecured.cu/Algoritmos_de_clasificaci%C3%B3n_supervisada
- Héroes del dato. (2018). Héroes del dato. Recuperado el 5 de octubre de 2018, de <http://heroesdeldato.com/tipos-de-machine-learning-clasificacion-vs-regresion>
- Salcedo, L. (2018). Mi diario Python. Recuperado el 5 de octubre de 2018, de <http://www.pythondiario.com/2018/01/introduccion-al-machine-learning-8.html>