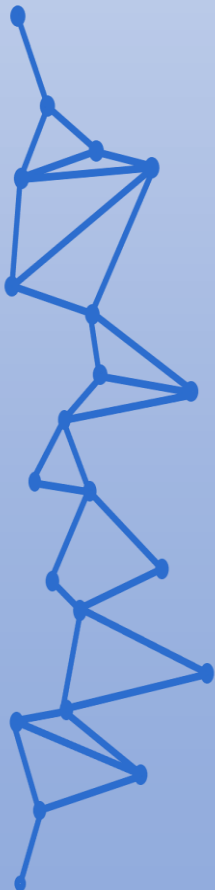




## Curso de Especialización de Inteligencia Artificial y Big Data (IABD)



### Sistemas de aprendizaje Automático

UD02. Tipos de algoritmos en relación con el  
Aprendizaje Automático.  
Resumen.

JUAN ANTONIO GARCIA MUELAS

---

Existen, principalmente, **tres tipos de Aprendizaje Automático: supervisado, no supervisado, y por refuerzo.**

### Técnicas o algoritmos de Aprendizaje Automático (Supervisado y No Supervisado):

#### REGRESIÓN LINEAL.

La regresión, **es más un proceso de ajuste que un algoritmo.** Genera un modelo que relaciona las variables de forma iterativa, a través de la media del error con la función que maneja la predicción. Ejemplo: Para predecir una enfermedad, se entrena con miles de radiografías etiquetadas(enfermo o no). Construye una función a base de probar sus predicciones con la realidad de los casos etiquetados.

La **Regresión Lineal** se usa para **estimar valores reales de variables con distribución continua.** Por **ejemplo** el **precio de viviendas**, el **número de llamadas** que hay que hacer para conseguir un **cliente nuevo**, **las ventas totales** que habrá el **trimestre** que viene...

**Hay unas variables de entrada** (número de habitaciones de una vivienda, ascensores, metros totales) **y un campo objetivo o valor de salida** (precio).

La relación entre las variables de entrada y el valor de salida (o la predicción) se produce a través de la **generación de una línea que se ajuste lo mejor posible a la distribución de resultados.**

$$y = a \cdot x + b$$

Para entendernos, **lo que hace el algoritmo de regresión lineal es ir inclinando más o menos la línea para que al final pase lo más cerca posible de cada punto.**

#### REGRESIÓN LOGÍSTICA.

El algoritmo de regresión logística **es más bien un algoritmo de clasificación que de regresión**, pues **va a estimar valores discretos** (SI/NO, 1 ó 0, verdadero o falso) en función de las variables de entrada (y no una distribución continua de valores como en el caso de la Regresión Lineal).

Es decir, que este tipo de algoritmo **nos va a permitir predecir la probabilidad de que se produzca un evento.**

La regresión logística resulta **útil para los casos en los que se desea predecir la presencia o ausencia de una característica o resultado según** los valores de un conjunto de predictores

**Matemáticamente** la función logística **se representa por una curva en forma de S** (conocida como **función sigmoide**), que va, en el eje Y, desde el valor cero ( 0% probabilidad de que se produzca el evento) al valor 1 (100% probabilidad de que se produzca el evento).

La regresión logística es una técnica **muy empleada por los científicos de datos** debido a su eficacia y simplicidad.

#### ÁRBOL DE DECISIÓN.

Los algoritmos de árbol de decisión construyen un modelo de decisiones basadas en los atributos que presentan los datos que entran en el modelo. **Trabajan muy bien tanto con datos del tipo de regresión como de clasificación.**

**¿Cómo funcionan los algoritmos de Árbol de Decisión? Crea bifurcaciones, separando los datos en función de los diferentes atributos, hasta que se llega a una decisión concluyente en base a la predicción buscada.**

**Construyen un modelo de decisiones basadas en los atributos que presentan los datos que entran en el modelo.**

**Son los grandes favoritos en aprendizaje automático, porque son rápidos y precisos.** Así que **cuando no se tiene claro qué tipo de algoritmo usar** para algún caso concreto, lo mejor es empezar aplicando este tipo de algoritmo.

Los más utilizados son:

- ✓ Classification and Regression Tree (CART).
- ✓ Iterative Dichotomiser 3 (ID3).
- ✓ C4.5 and C5.0 (different versions of a powerful approach).
- ✓ Chi-squared Automatic Interaction Detection (CHAID).
- ✓ Decision Stump.
- ✓ M5.
- ✓ Conditional Decision Trees.

**El más popular es el CART, que se conoce directamente como algoritmo de árbol de decisión.**

#### **MÁQUINAS DE VECTOR SOPORTE.**

Las Máquinas de Vector Soporte (en inglés Support Vector Machines - SVM), es otro método de **Aprendizaje Automático Supervisado que se aplica a problemas de clasificación**. Podemos decir que este método "dibuja cada instancia" o caso como un punto en un espacio n-dimensional (donde n es el número de atributos que tomaremos como coordenadas para cada punto).

Dentro de las técnicas de Aprendizaje Automático **el uso de las Máquinas de Vector Soporte como clasificador se ha visto incrementado en los últimos años** debido a que **sirven para resolver problemas de clasificación y regresión y que su rendimiento en los diferentes campos en los que se utilizan suele ser bastante alto, resultando ser una de las técnicas más precisas.**

La técnica consiste en calcular la recta que separa ambos grupos de forma que esté equidistante de los dos puntos más cercano.

Algunos **campos de aplicación exitosos** de este tipo de algoritmos han sido:

- ✓ **Reconocimiento óptico de caracteres.**
- ✓ **Detección de caras** para que las cámaras digitales enfoquen correctamente.
- ✓ **Filtros de spam** para correo electrónico.
- ✓ Reconocimiento de **imágenes a bordo de satélites** (saber qué partes de una imagen tienen nubes, tierra, agua, hielo, etc.)

#### **CLUSTERING.**

Los algoritmos de tipo clustering, **también llamados "de agrupación", son típicos de Aprendizaje Automático No Supervisado.** Es decir, cuando en el entrenamiento de la Inteligencia Artificial no determinamos ningún campo objetivo. **Se utilizan para agrupar datos existentes de los que, a priori, no intuimos sus características en común.**

Funcionan creando unos puntos centrales o "centroides" y jerarquías para diferenciar los grupos y descubrir características comunes por cercanía.

**Se utilizan muy frecuentemente en herramientas de negocios que te recomiendan productos o servicios que "intuyen" te pueden interesar,** pues tienes características y comportamientos

similares a los de clientes que ya han comprado o consumido dicho producto o servicio. Ejemplo: recomendación de películas en una plataforma.

Como todos algoritmos de aprendizaje no supervisado, **el algoritmo k-Means se utiliza** cuando tenemos una buena **cantidad de datos sin etiquetar con el objetivo de encontrar “k” grupos** entre los datos con características similares. El valor de "k" lo decidimos nosotros.

Algunos de sus **usos más comunes** son:

- ✓ **Detección de datos anormales.**
- ✓ **Agrupación en clústeres** de documentos de texto.
- ✓ **Análisis de conjuntos de datos** antes de utilizar otros métodos de clasificación o Regresión.

El **algoritmo G-Means** recibe su nombre de Gaussian-Means (que vendría a traducirse por medias gaussianas). Su objetivo es **descubrir por sí mismo la cantidad de agrupaciones o clústeres óptimos** en un problema de **Aprendizaje Automático No Supervisado**. Realiza sucesivas iteraciones, partiendo de un único grupo o clúster, dividiendo en cada una de ellas los datos en dos nuevos grupos mientras que no detecte que sus datos adquieren una distribución gaussiana.

**En los algoritmos de Clustering tipo G-Means el número de centroides lo calcula de forma automática el propio algoritmo.**