**TECNOLÓGICO NACIONAL DE MÉXICO**

**INSTITUTO TECNOLÓGICO DE TIJUANA**

**DEPARTAMENTO DE SISTEMAS Y COMPUTACIÓN**

**ENERO - JUNIO 2020**

**INGENIERÍA EN TECNOLOGÍAS DE LA INFORMACIÓN Y COMUNICACIONES**

**DATOS MASIVOS**

**BDD-1704IF9A**

**“Principales tipos de algoritmos de Machine Learning”**

**INTEGRANTES:**

García Bautista Ana Laura No. Control: 15210793

**PROFESOR:**

Romero Hernández José Christian

Principales tipos de algoritmos de Machine Learning

[**Algoritmos de Regresión** 1](#_Toc39500124)

[**Algoritmos basados en Instancia** 1](#_Toc39500125)

[**Algoritmos de Árbol de Decisión** 2](#_Toc39500126)

[**Algoritmos Bayesianos** 2](#_Toc39500127)

[**Algoritmos de Clustering (agrupación)** 3](#_Toc39500128)

[**Algoritmos de Redes Neuronales** 3](#_Toc39500129)

[**Algoritmos de Aprendizaje Profundo** 4](#_Toc39500130)

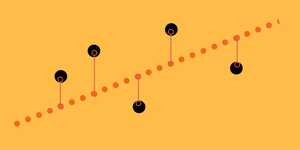
[**Algoritmos de Reducción de Dimensión** 4](#_Toc39500131)

[**Procesamiento del Lenguaje Natural (NLP)** 5](#_Toc39500132)

# **Algoritmos de Regresión**

Los algoritmos de Regresión modelan la relación entre distintas variables (features) utilizando una medida de error que se intentará minimizar en un proceso iterativo para poder realizar predicciones «lo más acertadas posible». Se utilizan mucho en el análisis estadístico. El ejemplo clásico es la predicción de precios de Inmuebles a partir de sus características: cantidad de ambientes del piso, barrio, distancia al centro, metros cuadrados del piso, etc.

Los Algoritmos más usados de Regresión son:

* Regresión Lineal.
* Regresión Logística.

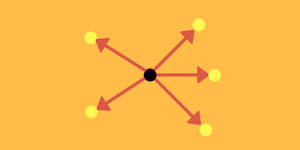
# **Algoritmos basados en Instancia**

Son Modelos de Aprendizaje para problemas de decisión con instancias o ejemplos (muestras) de datos de entrenamiento que son importantes o requeridos por el modelo.

También son llamados Algoritmos «Ganador se lleva todo» y aprendizaje basado-en-memoria en el que se crea un modelo a partir de una base de datos y se agregan nuevos datos comparando su similitud con las muestras ya existentes para encontrar «la mejor pareja» y hacer la predicción.

Los Algoritmos basados en instancia más usados son:

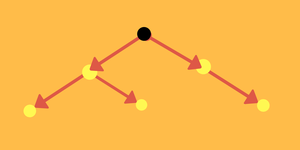
* k-Nearest Neighbor (kNN).
* Self-Organizing Map.



# **Algoritmos de Árbol de Decisión**

Modelan la toma de Decisión basado en los valores actuales (reales) de los atributos que tienen nuestros datos. Se utilizan sobre todo para clasificación de información, bifurcando y modelando los posibles caminos tomados y su probabilidad de ocurrencia para mejorar su precisión. Una vez armados, los arboles de decisión ejecutan muy rápido para obtener resultados.

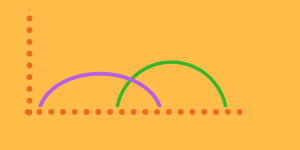
Los Algoritmos de árbol de decisión más usados son:

* Arboles de Clasificación y Regresión (CART)
* Decisión de Arbol condicional
* Random Forest

# **Algoritmos Bayesianos**

Son algoritmos que utilizan explícitamente el Teorema de Bayes de probabilidad para problemas de Clasificación y Regresión.

Los más utilizados son:

* Naive Bayes
* Gaussian Naive Bayes
* Multinomial Naive Bayes
* Bayesian Network

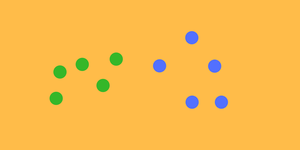
# **Algoritmos de Clustering (agrupación)**

Se utilizan para agrupar datos existentes de los que desconocemos sus características en común o queremos descubrirlas.

Estos métodos intentan crear «puntos centrales» y jerarquías para diferenciar grupos y descubrir características comunes por cercanía.

Los más utilizados son:

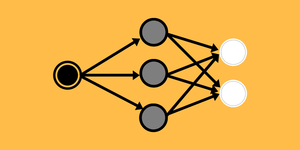
* K-Means
* K-Medians
* Hierarchical Clustering



# **Algoritmos de Redes Neuronales**

Son algoritmos y estructuras inspirados en las funciones biológicas de las redes neuronales. Se suelen utilizar para problemas de Clasificación y Regresión, pero realmente tienen un gran potencial para resolver multitud de problemáticas. Son muy buenas para detectar patrones. Las Redes Neuronales Artificiales requieren mucha capacidad de procesamiento y memoria y estuvieron muy limitadas por la tecnología del pasado hasta estos últimos años en los que resurgieron con mucha fuerza dando lugar al Aprendizaje Profundo (se detalla más adelante).

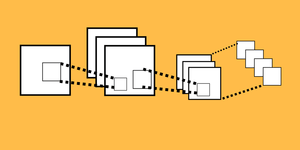
Las redes neuronales básicas y clásicas son:

* Compuerta XOR.
* Perceptron.
* Back-Propagation.
* Hopfield Network
* MLP: Multi Layered Perceptron.

# **Algoritmos de Aprendizaje Profundo**

Son la evolución de las Redes Neuronales Artificiales que aprovechan el abaratamiento de la tecnología y la mayor capacidad de ejecución, memoria y disco para explotar gran cantidad de datos en enormes redes neuronales interconectarlas en diversas capas que pueden ejecutar en paralelo para realizar cálculos. Comprende mejor Deep Learning con esta guía rápida que escribí.

Los algoritmos más populares de Deep Learning son:

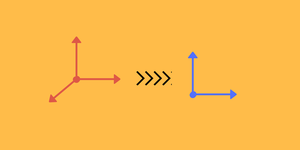
* Convolutional Neural Networks.
* Long Short Term Memory Neural Networks.

# **Algoritmos de Reducción de Dimensión**

Buscan explotar la estructura existente de manera no supervisada para simplificar los datos y reducirlos o comprimirlos.

Son útiles para visualizar datos o para simplificar el conjunto de variables que luego pueda usar un algoritmo supervisado.

Los más utilizados son:

* Principal Component Analysis (PCA).
* t-SNE (próximamente artículo).

# **Procesamiento del Lenguaje Natural (NLP)**

El Natural Language Processing es una mezcla entre DataScience, Machine Learning y Lingüística. Tiene como objetivo comprender el lenguaje humano. Tanto en textos como en discurso/voz. Desde analizar sintáctica ó gramáticamente miles contenidos, clasificar automáticamente en temas, los chatbots y hasta generar poesía imitando a Shakespeare. También es común utilizarlo para el Análisis de Sentimientos en redes sociales, (por ejemplo, con respecto a un político) y la traducción automática entre idiomas. Asistentes como Siri, Cortana y la posibilidad de preguntar y obtener respuestas, o hasta sacar entradas de cine.