

Historia de las matemáticas Machine Learning

Daniel López García
Lothar Soto Palma
Elena María Toro Pérez
Universidad de Granada

January 12, 2017

- 1 Introducción
- 2 Tipos de aprendizaje
- 3 Una idea inicial
- 4 Problema de clasificación
 - Redes neuronales
 - Vecino más cercano
 - Árboles de decisión
 - SVM
- 5 Otras técnicas para el aprendizaje
 - Clustering
 - Algoritmos genéticos
 - Colonias de hormigas
- 6 Máquinas y juegos

Introducción

Machine learning es una disciplina científica del ámbito de la Inteligencia Artificial que crea sistemas que aprenden automáticamente.

Hoy en día, podemos observar la aplicación concreta de Machine Learning en varios campos que nos rodean. Algunos de ellos son:

- **Automóviles:** sistemas de piloto automático.
- **Finanzas:** asesoría de préstamos (aceptación/denegación de un crédito).
- **Internet:** compras on-line, anuncios on-line.
- **Medicina:** diseño de prediagnósticos médicos basados en síntomas del paciente.
- **Recursos Humanos:** predicción de qué empleados serán más rentables el año que viene.
- **Transporte:** sistemas de rutas y seguimiento de flotas.

Antecedentes Estadísticos

Para que los sistemas puedan aprender por ellos mismos, se utilizan una serie de técnicas y algoritmos capaces de crear modelos predictivos, patrones de comportamiento, etc.

Todo este proceso está basado en la teoría estadística.

Los 3 métodos más importantes son:

- **Teorema de Bayes.**
- **Método de los mínimos cuadrados.**
- **Cadenas de Markov.**

Un ejemplo ilustrativo

Una empresa de telefonía quiere saber qué clientes están en “peligro” de darse de baja de sus servicios para evitar que se vayan a la competencia.

- La empresa tiene **muchos datos** de los clientes, pero seguramente los usa solo para facturar y para hacer estadísticas.
- Esta cantidad inmensa de datos son **imposibles de analizar por una persona** para sacar conclusiones y menos aún para hacer predicciones.
- Los **algoritmos** en cambio sí pueden detectar patrones de comportamiento contando con las variables que le proporcionamos y descubrir cuáles son las que le han llevado, en este caso, a darse de baja como cliente.

Un ejemplo ilustrativo

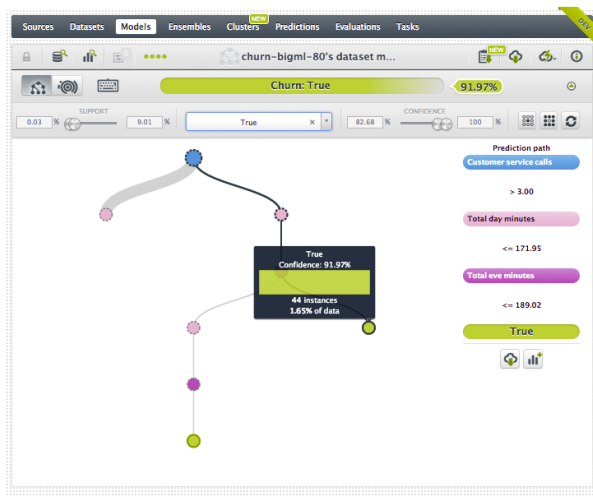


Figure: Ejemplo de una predicción de una compañía de telefonía ficticia

Tipos de aprendizaje

Las diferentes formas de aprender se pueden clasificar en función de la salida que produzcan y de su forma de tratar con los ejemplos. Los principales tipos son:

- **Aprendizaje supervisado:** se conoce a priori una clasificación de los datos del problema.
- **Aprendizaje no supervisado:** no se tiene conocimiento a priori sobre el problema, no hay instancias etiquetadas.
- **Aprendizaje semisupervisado:** estos algoritmos combinan los dos algoritmos anteriores (se tienen en cuenta ejemplos clasificados y no clasificados).
- **Aprendizaje por refuerzo:** el sistema aprende a base de prueba-error, reforzando las acciones que reciben una respuesta positiva.
- **Aprendizaje multi-tarea:** métodos que usan conocimiento previamente aprendido para resolver problemas parecidos.

Una idea inicial

En **1950**, Alan Turing publicó uno de sus artículos más importantes para la revista *Mind* donde propuso **el Test de Turing**.

Se basa en una hipótesis: “si una máquina se comporta en todos los aspectos como inteligente, entonces, dicha máquina debe ser inteligente”.

El artículo de Turing comenzaba con una frase que era toda una declaración de intenciones de lo que evaluaría el test:

”Propongo considerar la siguiente cuestión: ¿Pueden pensar las máquinas?”

Por lo que el resultado que se obtiene de este test es **intentar medir si una máquina puede ser inteligente** con un método que, aún hoy, sigue estando vigente.

¿En qué consistía el Test de Turing?

- Se basaba en el Juego de la Imitación, en el cual se ubicaban en una habitación un **hombre** y a una **mujer** con terminales con algún sistema de comunicación. El objetivo del interrogador era descubrir quién era la mujer y quién era el hombre, mientras que el de los otros dos era convencer al interrogador de que son la mujer.
- Turing proponía realizar un cambio: coger a uno de los dos sujetos y sustituirlo por una **máquina**. Así, cambiaba el objetivo de reconocer el sexo por el de reconocer la máquina.

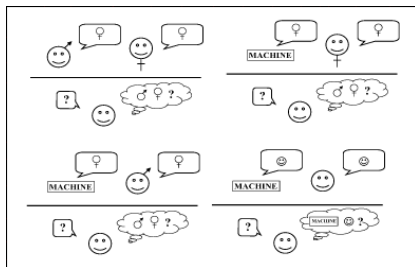


Figure: Test de Turing

Problema de Clasificación

En el aprendizaje automático y la estadística en problema de las clasificación consiste en identificar un conjunto de categorías a la que pertenece una observación o conjunto de instancias de una población. La clasificación es un ejemplo de reconocimiento de patrones en una población.

Conceptos previos

- **Dato o instancia:** Los datos son representación simbólica de atributos o un conjunto de atributos, estos son la principal materia prima para el aprendizaje.
- **Conjunto de entrenamiento:** Partición de un conjunto de datos que es usado para descubrir relaciones en los datos que posteriormente nos puedan servir para predecir
- **Conjunto de prueba:** Partición de un conjunto de datos que se usa para evaluar el modelo generado usando el conjunto anterior.

Problemas de los métodos de clasificación

Uno de los grandes problemas de los algoritmos de clasificación es el sobreajuste o sobreentrenamiento

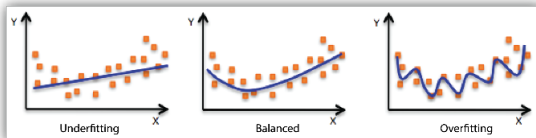


Figure: Ejemplo de sobreajuste

- **¿Cuándo se produce?:** Se produce cuando un modelo se ajusta muy bien al conjunto de datos que usamos para entrenar y que de alguna manera elimina la generalidad que debería de tener el modelo para poder tratar con él.
- **¿Cómo tratar de evitarlo?:** Una de las formas es la validación, es un conjunto de técnicas que se aplican normalmente antes de evaluar el modelo obtenido con el aprendizaje sobre un conjunto de entrenamiento y sirve para obtener resultados que ilustren una hipotética relación entre los atributos.

Validación Cruzada

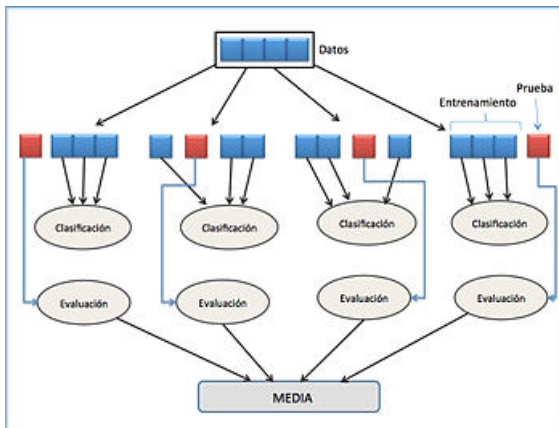


Figure: Ejemplo de Validación cruzada

Red Neuronal

Las redes neuronales son un conjunto de neuronas artificiales interconectadas, son un paradigma de aprendizaje y procesamiento automático inspirado en el sistema nervioso, que es objeto de estudio en el campo de la inteligencia artificial.

- **¿Cuándo surgió esta idea?**
- **¿Qué ideas han dado lugar a las redes neuronales tal y como son en la actualidad?**

Marco histórico (Redes neuronales)

Los hechos más destacados:

- En el año **1943**, Warren McCulloch y Walter Pitts construyen un modelo simple de red neuronal usando circuitos electrónicos.
- En el año **1949**, Donald Hebb introduce la teoría de la asamblea celular.
- En los años 50, Nathaniel Rochester hizo un primer intento fallido de simular una red neuronal.
- En el año **1951**, Marvin Lee Minsky y Dean Edmonds construyen la primera máquina que ejecutaría una red neuronal.
- En el año **1958**, Frank Rosenblatt desarrolla el perceptrón.

Perceptrón

Es un tipo de red neuronal artificial. El modelo más simple de un perceptrón es una neurona.

Se pudo demostrar que el perceptrón era capaz de clasificar patrones correctamente. La mayor desventaja de este tipo de redes es su incapacidad para solucionar problemas que no sean linealmente separables.

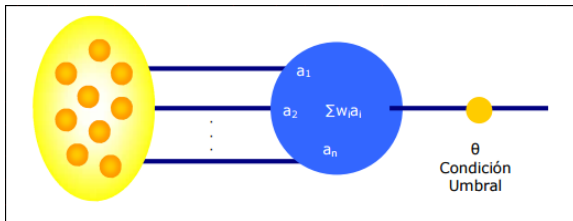


Figure: Estructura del perceptrón.

Modelo matematico del perceptron

Un **perceptrón simple** es un dispositivo de computación con umbral θ y N entradas reales x_1, \dots, x_N a través de arcos con pesos w_1, \dots, w_N y que tiene salida 1 cuando $\sum_i w_i x_i \geq \theta$ y -1 en caso contrario.

Es decir, supongamos la función de clasificación f en \mathbb{R}^n tal que:

$$f(x_1, x_2, \dots, x_N) = \begin{cases} 1 & \text{si } w_1 x_1 + w_2 x_2 + \dots + w_N x_N \geq \theta \\ -1 & \text{si } w_1 x_1 + w_2 x_2 + \dots + w_N x_N < \theta \end{cases}$$

Dicha función realiza una partición en el espacio \mathbb{R}^n de patrones de entrada: por una parte estarían los patrones con salida $+1$ y por otra parte los patrones con salida -1 . Por lo tanto, diremos que la función f clasifica a los patrones de entrada en dos clases. En este caso, se expresa la función f mediante la función signo, es decir:

$$f(x_1, x_2, \dots, x_N) = \text{sign}(x_1, x_2, \dots, x_N)$$

donde la función signo es:

$$\text{sign}(x_1, x_2, \dots, x_N) = \begin{cases} 1 & \text{si } x \geq 0 \\ -1 & \text{si } x < 0 \end{cases}$$

Vecino más cercano

- El algoritmo de clasificación K-Nearest Neighbor o K-vecinos más cercanos es uno de los métodos de clasificación más fundamental y sencillo.
- Fue introducido en 1951 por Fix y Hodge para resolver el problema de clasificación.
- No fue hasta el año 1967 que este método no tuvo constancia.
- Este hito está considerado como el nacimiento del campo de reconocimiento de patrones en computadores.
- **¿Cómo es su funcionamiento?**

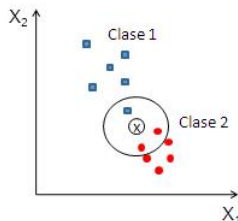


Figure: Ejemplo de KNN

Árboles de decisión

Un **árbol de decisión** es un algoritmo que se basa en la creación de un grafo dirigido donde los nodos son reglas sobre los atributos y las hojas del árbol son el atributo clase o sobre el que se va a clasificar. Algunos de los algoritmos más reconocidos que usan este modelo de aprendizaje son:

- **Algoritmo ID3 y C4.5:** Desarrollado por John Ross Quinlan en los años 60.
- **Algoritmo Random Forest:** Desarrollado en 1995 por Tinkam Ho.

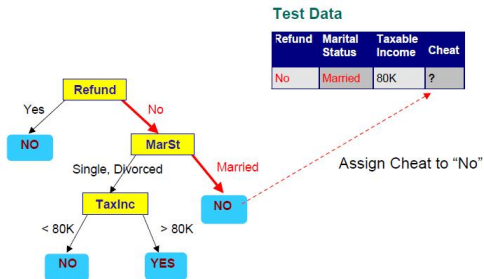


Figure: Ejemplo de árbol de decisión

SVM (Support Vector Machine)

- El **algoritmo SVM** es un conjunto de técnicas de aprendizaje supervisado desarrollados en el año 1995 por Corinna Cortes y Vladimir Vapnik.
- La idea principal del método es la construcción de un hiperplano o conjunto de hiperplanos que separen de forma óptima las clases.

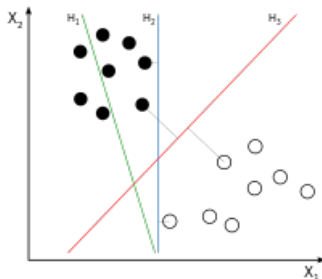


Figure: Ejemplo de Support Vector Machine

Clustering

El **clustering** es una técnica de aprendizaje no supervisado que identifica de forma automática agrupaciones o clústeres de elementos de acuerdo a una medida de similitud entre ellos.

Existen principalmente dos tipos diferentes de técnicas de clustering:

- **Clústering Jerárquico**
 - Clustering jerárquico aglomerativo
 - Clustering jerárquico de división
- **Clústering de Partición: K-means**

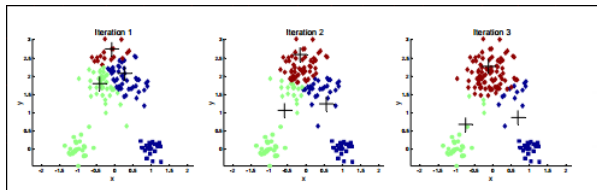


Figure: Ejemplo de K-means

Algoritmos genéticos

Los algoritmos genéticos son una técnica de búsqueda basada en la teoría de la evolución de Darwin. Fue desarrollado en 1975 por John Holland.

La capacidad de aprendizaje del algoritmo radica en la evolución de los individuos de la población, es decir, a medida que ocurren cruces y mutaciones se va recopilando información relevante que se traslada generación a generación.

Hay diferentes tipos de criterio de selección de soluciones:

- Modelo Generacional
- Modelo Estacionario

Una aplicación directa de este algoritmo es sobre el problema denominado selección de características que pretende disminuir el espacio del conjunto de datos, es decir, reducir su número de atributos en base a patrones.

Colonias de hormigas

El algoritmo de la colonia de hormigas es una técnica probabilística para solucionar problemas computacionales que pueden reducirse a buscar los mejores caminos o rutas en grafos. Fue propuesto por Marco Dorigo en **1992**.

Dicho algoritmo está basado en el comportamiento de las hormigas cuando estas están buscando un camino entre la colonia y una fuente de alimentos.

¿Cómo funciona el método?

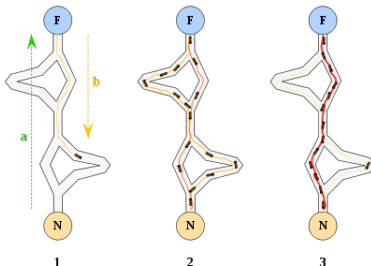


Figure: Ejemplo de colonias de hormigas

Máquinas y juegos