Electiva Aprendizaje automático

Maestría en Gestión de Tecnologíias de Información y Conocimiento
Facultad de Ingeniería
Universidad de Nariño)

MSc Jimmy Mateo Guerrero Restrepo[1]
jimaguere@udenar.edu.co
3168211698

[1]Departamento de Sistemas
Facultad de Ingeniería

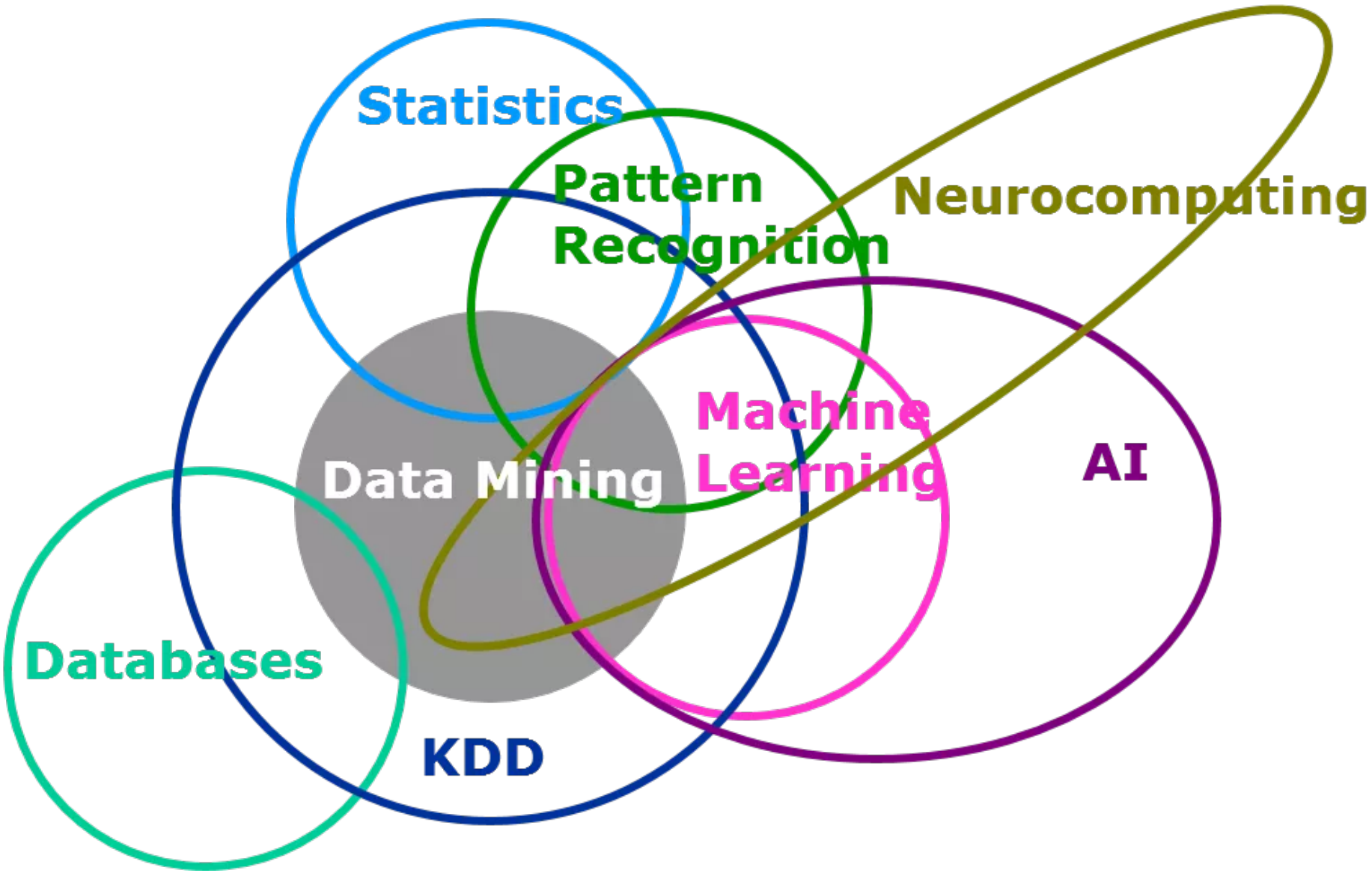
Aprendizaje automático, 2021

Contenido

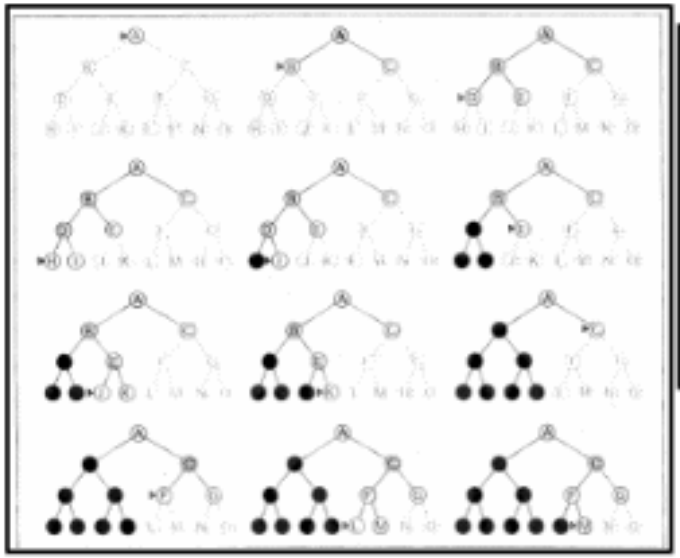
Aprendizaje automático

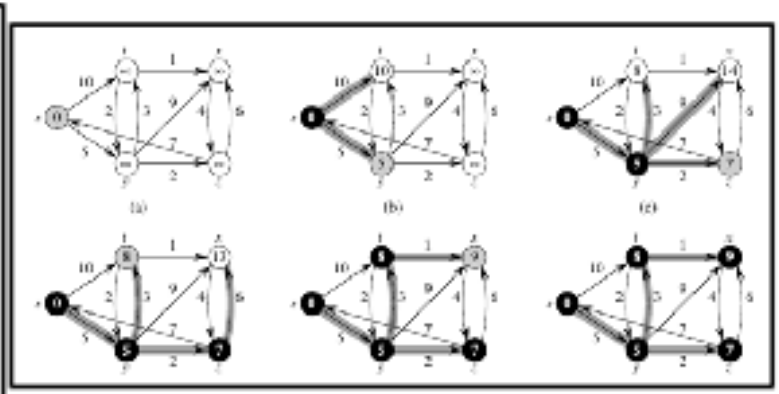
- Hay problemas en Informática que se pueden "definir" concretamente y son simples de convertir en un algoritmo:
  - Ordenar alfabéticamente una lista, calcular el balance de una cuenta.
- Hay otros que son simples de "entender" pero muy difíciles de "definir" y convertir en algoritmo:
  - Detectar una sonrisa en una cara, disntiguir entre un gato y un perro, interpretar un gesto,....
- El Aprendizaje Automatizado introduce métodos que pueden resolver esas tareas "aprendiendo" la solución a partir de ejemplos de como se realiza la misma.

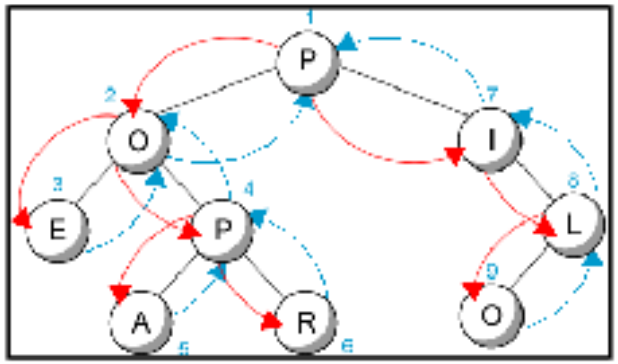
Aprendizaje automático

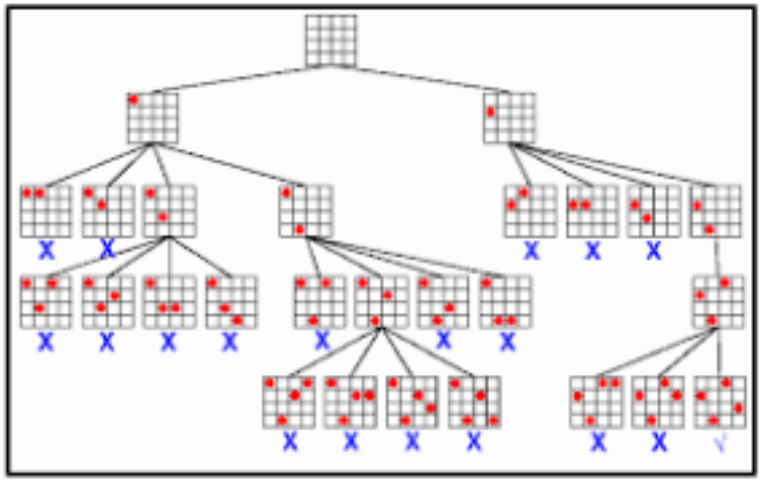


Resolver problemas como búsqueda en un grafo









Resolver problemas como búsqueda en un grafo

- Recorrer el grafo
- Evaluar solución
- Seleccionar mejor solución
- Proponer solución

Resolver problemas como búsqueda en un grafo

- Desafortunadamente, recorrer un grafo y encontrar la solución óptima es un problema intratable.
- Es preciso buscar soluciones no óptimas pero tratables (Subdividir el problema,Usar algoritmos aproximativos,Usar heurísticas)

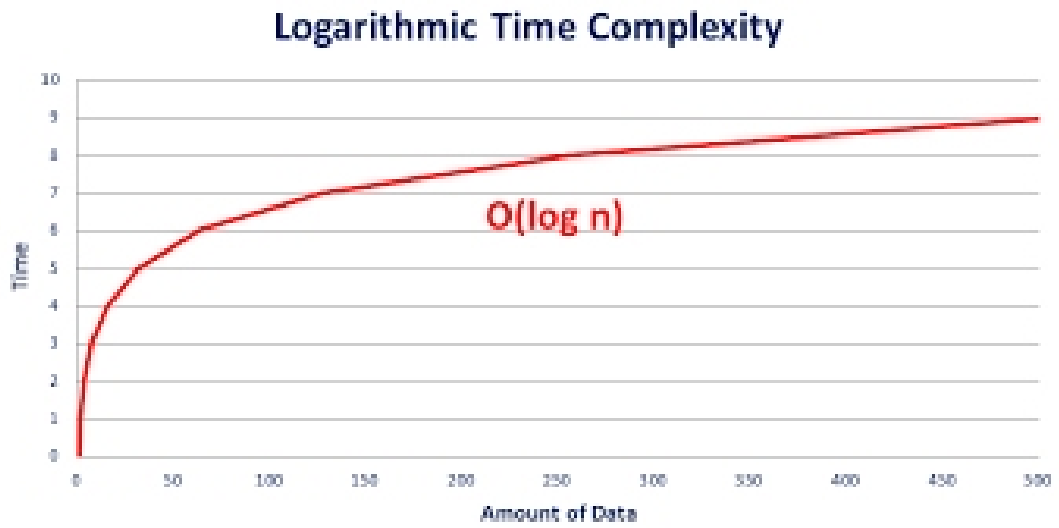
¿Problema intratable?

- Cualquier algoritmo cuya función de complejidad temporal no pueda ser acotada a un polinomio O(p(n)), se denomina algoritmo de tiempo exponencial.
- La mayoría de los algoritmos de tiempo exponencial son simples variaciones de una búsqueda exhaustiva, mientras que los algoritmos de tiempo polinomial, usualmente se obtienen mediante un análisis más profundo de la estructura del problema.
- En la teoría de la complejidad computacional, existe el consenso de que un problema no está "bien resuelto" hasta que se conozca un algoritmo de tiempo polinomial que lo resuelva.
- Por tanto, nos referiremos a un problema como intratable, si es tan difícil que no existe algoritmo de tiempo polinomial capaz de resolverlo.

¿Problema intratable?

[redacted]

Data	Time
1	0
2	1
4	2
8	3
16	4
32	5
64	6
128	7
256	8
512	9



[redacted]

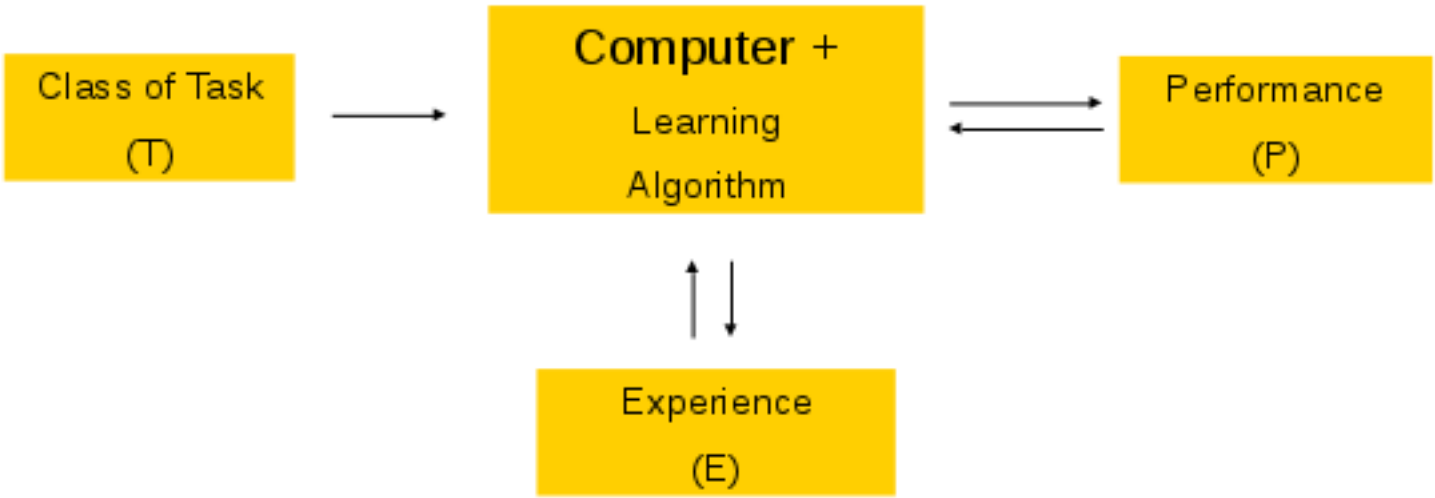
¿Problema intratable?

- ¿Heurística?
  - Como metodología científica, la heurística es aplicable a cualquier ciencia e incluye la utilización de: medios auxiliares, principios, reglas, estrategias o algoritmos que faciliten la búsqueda de vías de solución a problemas.

Inferir

- Inferencia
  - Establecer una relación entre razones y conclusión.
  - Establecer conclusiones en base a información conocida.
  - Deducción.
    - Inferencia lógica (desde lo universal hacia lo particular).
    - Ejemplo: Todas las aves tienen plumas. Las palomas son aves.
    - Conclusión: Las palomas tienen plumas.
  - Inducción
    - Es lo contrario a la deducción y establece conclusiones a partir de observaciones específicas (particular hacia lo universal)
    - Forma las inferencias basándose en muchos datos y va de lo específico a lo general
    - Ejemplo: Pedro Juega fútbol y se cansa, María Juega fútbol y se cansa, Juan Juega fútbol y se cansa, Mario Juega fútbol y se cansa, Ana Juega fútbol y se cansa.
    - Conclusión: Si una persona juega al futbol se va a cansar.

Sistema de aprendizaje: estructura



"Se dice que un programa aprende si mejora su performance en una cierta tarea al incorporar experiencia"

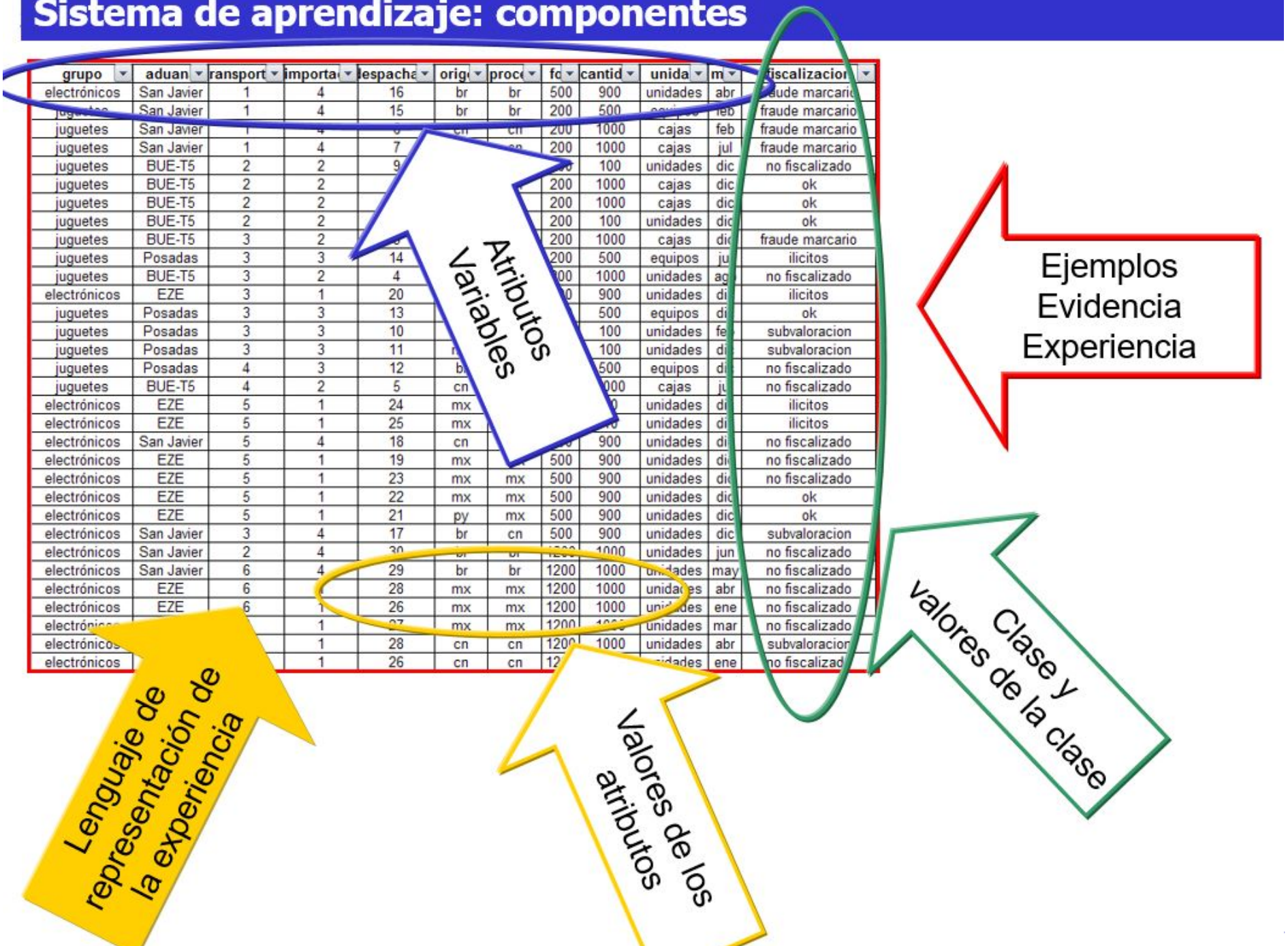
Sistema de aprendizaje: componentes

Para definir un sistema de aprendizaje debemos establecer:

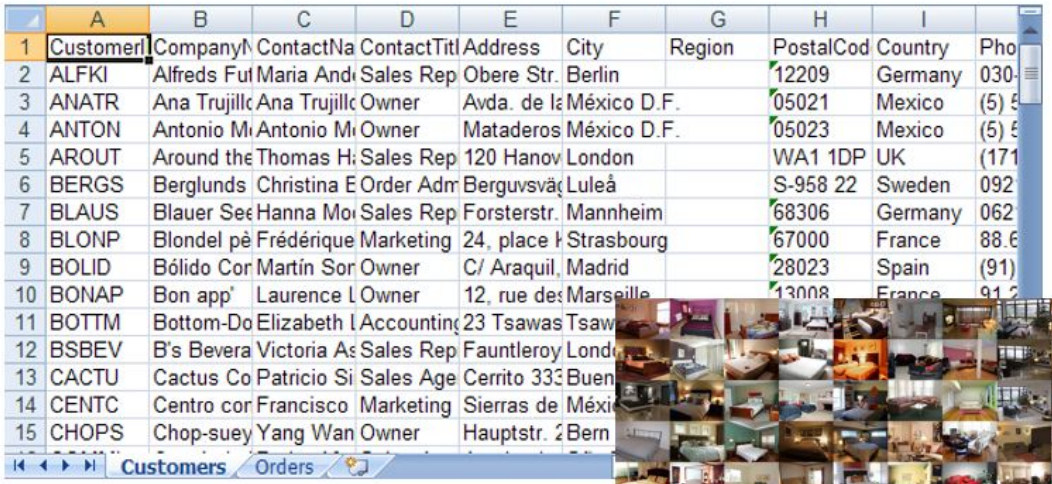
- La experiencia.
- El conocimiento a aprender.
- El modelo de aprendizaje.
- La medida de performance.

- El modelo de monitoreo.

Sistema de aprendizaje: componentes



grupo	aduan	ransport	importa	despacha	orig	proc	fc	cantid	unida	m	fiscalizacion
electrónicos	San Javier	1	4	16	br	br	500	900	unidades	abr	fraude marcari
juguetes	San Javier	1	4	15	br	br	200	500	equipos	feb	fraude marcario
juguetes	San Javier	1	4	8	cn	cn	200	1000	cajas	feb	fraude marcario
juguetes	San Javier	1	4	7	cn	cn	200	1000	cajas	jul	fraude marcario
juguetes	BUE-T5	2	2	9	cn	cn	200	100	unidades	dic	no fiscalizado
juguetes	BUE-T5	2	2	8	cn	cn	200	1000	cajas	dic	ok
juguetes	BUE-T5	2	2	8	cn	cn	200	1000	cajas	dic	ok
juguetes	BUE-T5	2	2	8	cn	cn	200	100	unidades	dic	ok
juguetes	BUE-T5	3	2	8	cn	cn	200	1000	cajas	dic	fraude marcario
juguetes	Posadas	3	3	14	mx	mx	200	500	equipos	ju	ilicitos
juguetes	BUE-T5	3	2	4	cn	cn	200	1000	unidades	ag	no fiscalizado
electrónicos	EZE	3	1	20	mx	mx	500	900	unidades	di	ilicitos
juguetes	Posadas	3	3	13	mx	mx	200	500	equipos	di	ok
juguetes	Posadas	3	3	10	mx	mx	200	100	unidades	fe	subvaloracion
juguetes	Posadas	3	3	11	mx	mx	200	100	unidades	di	subvaloracion
juguetes	Posadas	4	3	12	br	br	200	500	equipos	di	no fiscalizado
juguetes	BUE-T5	4	2	5	cn	cn	200	1000	cajas	ju	no fiscalizado
electrónicos	EZE	5	1	24	mx	mx	500	900	unidades	di	ilicitos
electrónicos	EZE	5	1	25	mx	mx	500	900	unidades	di	ilicitos
electrónicos	San Javier	5	4	18	cn	cn	500	900	unidades	di	no fiscalizado
electrónicos	EZE	5	1	19	mx	mx	500	900	unidades	di	no fiscalizado
electrónicos	EZE	5	1	23	mx	mx	500	900	unidades	di	no fiscalizado
electrónicos	EZE	5	1	22	mx	mx	500	900	unidades	di	ok
electrónicos	EZE	5	1	21	py	mx	500	900	unidades	di	ok
electrónicos	San Javier	3	4	17	br	cn	500	900	unidades	dic	subvaloracion
electrónicos	San Javier	2	4	30	br	br	1200	1000	unidades	jun	no fiscalizado
electrónicos	San Javier	6	4	29	br	br	1200	1000	unidades	may	no fiscalizado
electrónicos	EZE	6	1	28	mx	mx	1200	1000	unidades	abr	no fiscalizado
electrónicos	EZE	6	1	26	mx	mx	1200	1000	unidades	ene	no fiscalizado
electrónicos			1	27	mx	mx	1200	1000	unidades	mar	no fiscalizado
electrónicos			1	28	cn	cn	1200	1000	unidades	abr	subvaloracion
electrónicos			1	26	cn	cn	1200	1000	unidades	ene	no fiscalizado



	A	B	C	D	E	F	G	H	I	
1	CustomerI	CompanyN	ContactNa	ContactTitl	Address	City	Region	PostalCod	Country	Pho
2	ALFKI	Alfreds Fut	Maria And	Sales Rep	Obere Str.	Berlin		12209	Germany	030-
3	ANATR	Ana Trujillo	Ana Trujillo	Owner	Avda. de la	México D.F.		05021	Mexico	(5) 5
4	ANTON	Antonio Mo	Antonio Mo	Owner	Mataderos	México D.F.		05023	Mexico	(5) 5
5	AROUT	Around the	Thomas Ha	Sales Rep	120 Hanov	London		WA1 1DP	UK	(171
6	BERGS	Berglunds	Christina E	Order Adm	Berguvsväg	Luleå		S-958 22	Sweden	092
7	BLAUS	Blauer See	Hanna Mo	Sales Rep	Forsterstr.	Mannheim		68306	Germany	062
8	BLONP	Blondel pè	Frédérique	Marketing	24, place k	Strasbourg		67000	France	88.6
9	BOLID	Bólido Cor	Martín Sor	Owner	C/ Araquil,	Madrid		28023	Spain	(91)
10	BONAP	Bon app'	Laurence L	Owner	12, rue des	Marseille		13008	France	91 2
11	BOTTM	Bottom-Do	Elizabeth L	Accounting	23 Tsawas	Tsaw				
12	BSBEV	B's Bevera	Victoria As	Sales Rep	Fauntleroy	Londo				
13	CACTU	Cactus Co	Patricio Si	Sales Age	Cerrito 333	Buen				
14	CENTC	Centro cor	Francisco	Marketing	Sierras de	Méxi				
15	CHOPS	Chop-suey	Yang Wan	Owner	Hauptstr. 2	Bern				

Customers | Orders

WIKIPEDIA
The Free Encyclopedia

Sistema de aprendizaje: tipos

- Aprendizaje supervisado:
  - A partir de un conjunto de ejemplos clasificados obtener una función que permita predecir casos no vistos.
  - Si X es el espacio de ejemplos posibles
  - Y es el espacio de clases posibles:
  - Aprender F: X —) Y
- Por ejemplo:
  - Aprendiendo a jugar al ajedrez:
    - X: espacio de configuraciones de tableros.
    - Y: espacio de movida válidas.
  - Conduciendo:
    - X: espacio de contextos posibles.
    - Y: espacio de acciones de manejo.

Sistema de aprendizaje: tipos

- Ejemplo: decidir si un paciente que entra a la guardia de un hospital debe ser atendido por urgencia cardíaca.
  - X1 : temperatura corporal
  - X2 : presión sanguínea
  - X3 : tipo de sangre
  - X4 : edad
  - X5 : peso
  - Y:(Con problema cardíaco) (Sin problema cardíaco).

Sistema de aprendizaje: tipos

- Clasificación:
  - Dado un objeto (conjunto de características medidas de alguna forma) asignarle una (o varias) etiqueta de un conjunto finito.
- Regresión:
  - Dado un objeto asignarle un número real.
  - Predecir la relación peso-dolar de mañana.
  - Predecir niveles de stock/ventas a futuro.
- Búsqueda y ranking:
  - Dado un objeto, asignarle y ordenar las respuestas más probables dentro de una base de datos.
  - Buscadores en Internet
  - Sistemas de recomendación
- Detección de novedades:
  - Detectar outliers, objetos que son diferentes a los demás.
  - Alarmas de comportamiento en compras con tarjeta.
  - Detección de fallas en equipos críticos.

Sistema de aprendizaje: tipos

- Aprendizaje no supervisado (Clustering) :
  - El objetivo básico del análisis de clusters es descubrir grupos en los datos, de modo tal que los objetos del mismo grupo sean similares, mientras que los objetos de diferentes grupos sean tan disímiles como sea posible.
  - Segmentación de consumidores/clientes a partir de sus patrones de consumo.

Sistema de aprendizaje: tipos

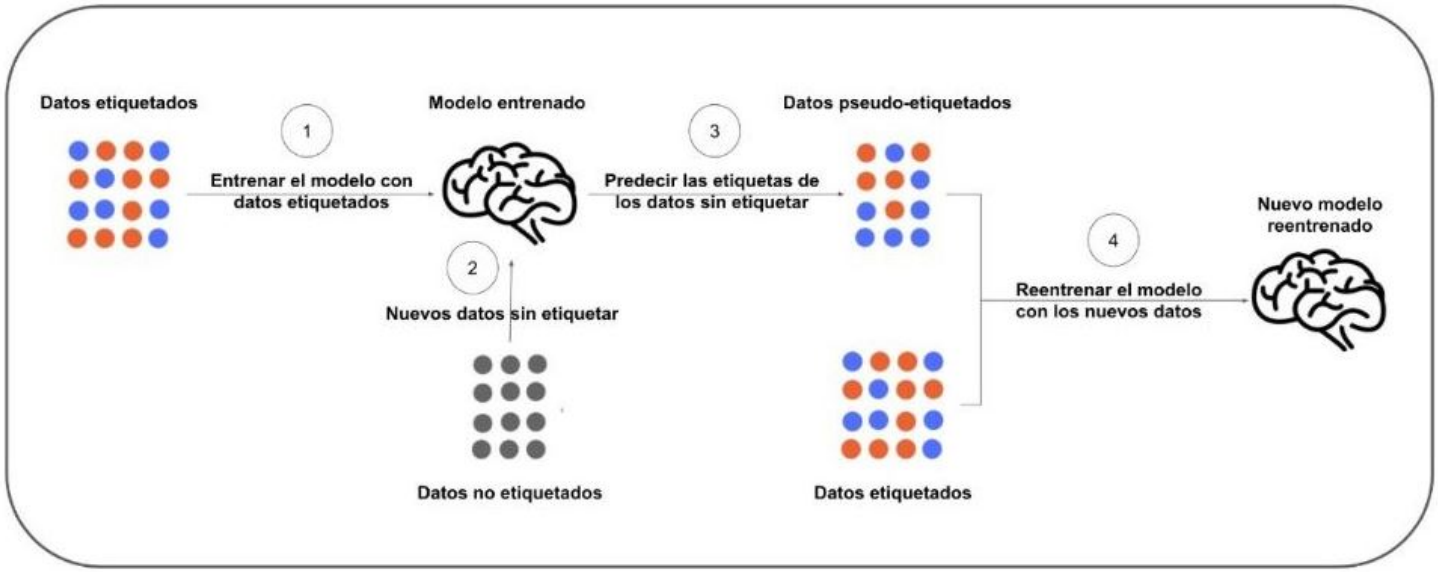
- Aprendizaje no supervisado (Clustering) :
  - Las aplicaciones son muy numerosas, por ejemplo la clasificación de plantas y animales, en ciencias sociales la clasificación de personas considerando sus costumbres y preferencias, en marketing la identificación de grupos de consumidores con necesidades parecidas, etc.
  - Cluster de documentos recuperados (por ejemplo, Tema) para presentar resultados más organizados y comprensibles para el usuario
  - Detección de duplicados P.ej. Thorsten Joachims == Thorsten B Joachims

Sistema de aprendizaje: tipos

- Aprendizaje no supervisado (Clustering) :
  - Generalmente lo que se pretende agrupar son individuos, pero existen algunas circunstancias en las que es interesante agrupar variables para intentar buscar variables de comportamiento similar. Para ello, la metodología es la misma que para el análisis cluster por individuos y simplemente tendremos que transponer la matriz de datos y aplicar el método general.

Sistema de aprendizaje: tipos

- Aprendizaje semi-supervisado.



Sistema de aprendizaje: tipos

- Aprendizaje por refuerzo :
  - Aprender cómo tomar secuencias de decisiones óptimas para lograr una meta.
  - Si el resultado de esa decisión es beneficioso, el agente aprende automáticamente a repetir esa decisión en el futuro, mientras que si el resultado fuera perjudicial evitará volver a tomar la misma decisión.
  - En los modelos de Aprendizaje Supervisado (o no supervisado) como redes neuronales, árboles, knn, etc, se intenta "minimizar la función coste", reducir el error.
  - En cambio en el RL se intenta "maximizar la recompensa". Y esto puede ser, a pesar de a veces cometer errores ó de no ser óptimos.

Sistema de aprendizaje: tipos

El Reinforcement Learning postula los siguientes 2 componentes:

- Agente: será nuestro modelo que queremos entrenar y que aprenda a tomar decisiones.
- Ambiente: será el entorno en donde interactúa y "se mueve" el agente. El ambiente contiene las limitaciones y reglas posibles a cada momento.

Entre ellos hay los siguientes nexos:

- Acción: las posibles acciones que puede tomar en un momento determinado el Agente.
- Estado (del ambiente): son los indicadores del ambiente de cómo están los diversos elementos que lo componen en ese momento.
- Recompensas (ó castigos!): a raíz de cada acción tomada por el Agente, podremos obtener un premio ó una penalización que orientarán al Agente en si lo está haciendo bien ó mal

Sistema de aprendizaje: tipos

Aprendizaje profundo

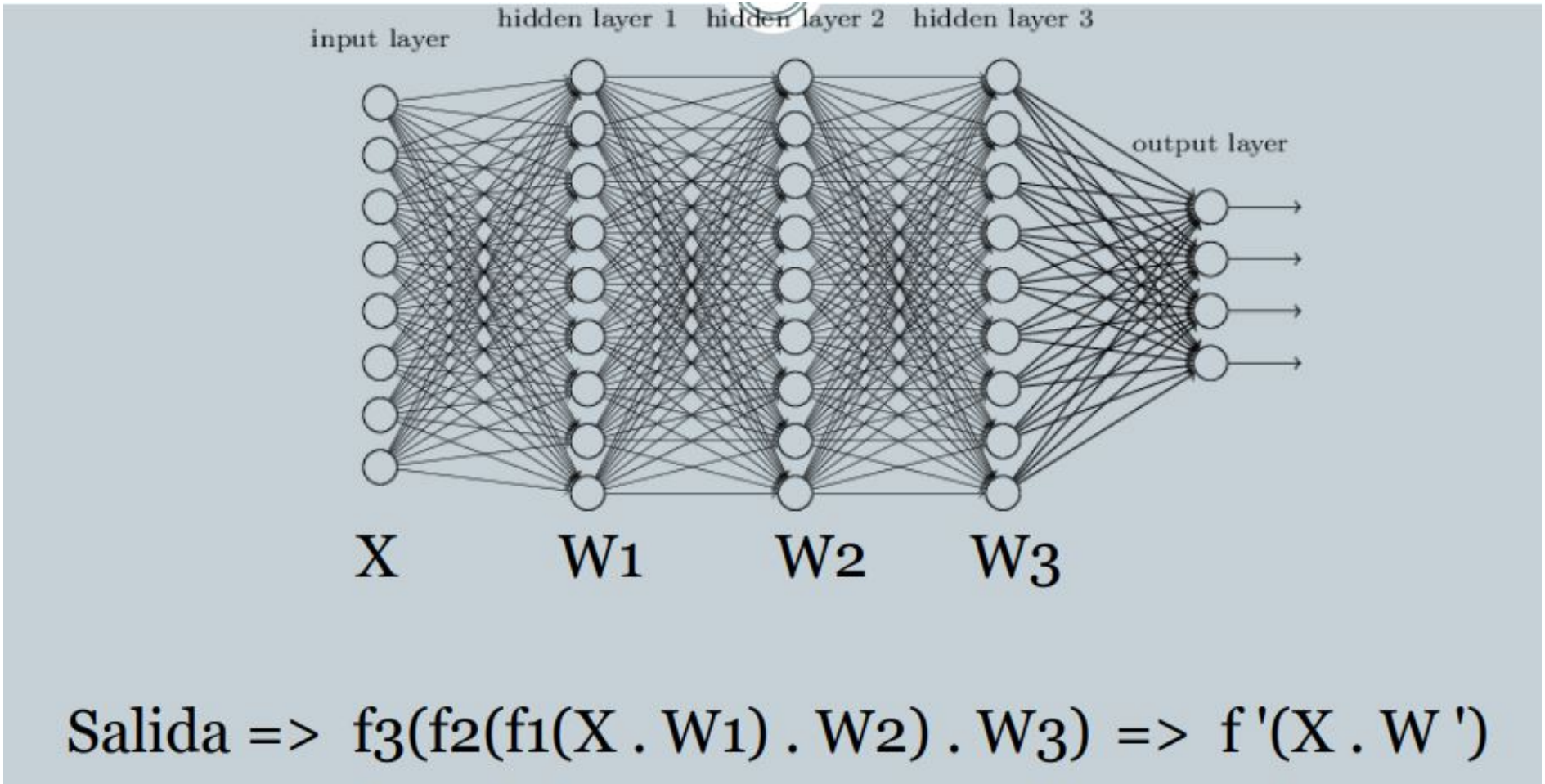
- Desde 2006 varios grupos de investigación comenzaron a trabajar en nuevos modelos de redes neuronales creando un campo que se denominó "Deep learning".
- La profundidad de una red se establece como el número de capas de entrenamiento.
- Aparecen los primeros resultados exitosos con redes de profundidad 3 (dos ocultas más la capa de salida).

Aprendizaje profundo

Las redes de aprendizaje profundo han sido utilizadas en tareas de:

- Clasificación.
- Regresión.
- Reducción de la dimensionalidad.
- Modelado de movimiento.
- Segmentación de imágenes.
- Recuperación de información.
- Robótica.
- Procesamiento del lenguaje natural.

Aprendizaje profundo



Aprendizaje profundo

- Basado en el comportamiento del cerebro humano lo que se busca es que las primeras capas de estas redes aprendan por si solas a detectar patrones y ejemplos similares.
- Las últimas capas se centran en detectar "que son"los patrones aprendidos (aprendizaje supervisado).

Aprendizaje profundo

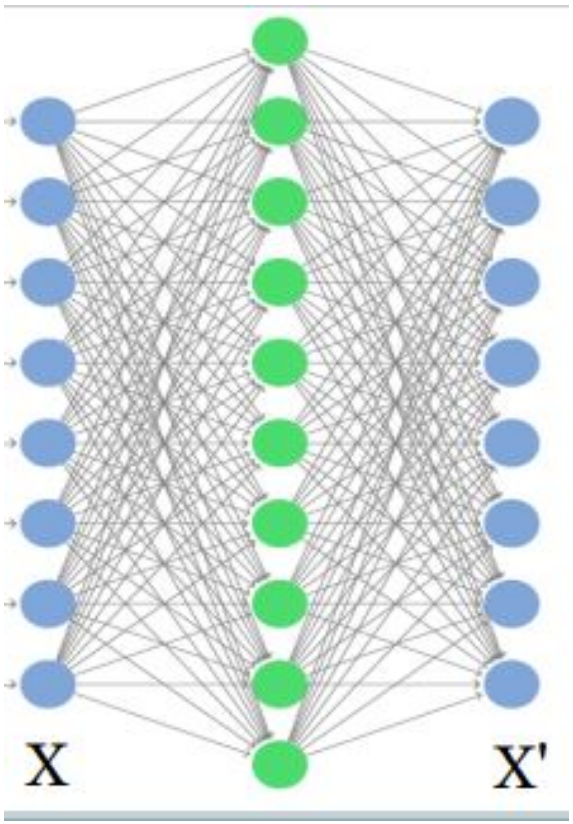
Auto-encoders

- El objetivo es entrenar un modelo que sea capaz de reconstruir la entrada.
- Es útil en problemas de clases desbalanceadas.
- Pueden ser entrenadas solo con los datos de una única clase.

Aprendizaje profundo

Auto-encoders

- La red se entrena presentando un ejemplo X para obtener una salida Y. Esta salida es comparada con el propio X. Se espera que la salida de la red sea la misma que la entrada.



Auto-encoders

- Al momento de presentar un ejemplo X, se compara la salida obtenida (X') con la propia entrada (X). El error cometido se utiliza para ajustar los vectores de pesos de las neuronas.

$$e(X) = \sum_{i=1}^{n} (X_i - X'_i)^2$$

- Cada ejemplo es reconstruido con un cierto error.

Aprendizaje profundo

Auto-encoders

- El entrenamiento consiste en minimizar el error promedio de todos los ejemplos utilizados.
- Para el entrenamiento solo se utilizan un subconjunto X de ejemplos (por ejemplo los pertenecientes a una clase)
- Para el testeo se utiliza la base de datos completa.
- Se espera que los ejemplos pertenecientes a X tengan un error de reconstrucción bajo, mientras que los otros tengan un error alto.
- Al momento de presentar un nuevo ejemplo (desconocido) a la red, se calcula su error de reconstrucción, si es más bajo que un cierto umbral entonces se afirma que pertenece al mismo grupo de X.

Aprendizaje profundo

Red neuronal convolucional

- Las redes convolucionales están inspiradas en la estructura del sistema visual.
- Tienen un amplio uso en el reconocimiento de imágenes.
- Las redes convolucionales por lo general están formadas por cinco o más capas. Estas capas se dividien en:
  - Capas convolucionales
  - Capas de subsampling

Red neuronal convolucional

W x H

W x H

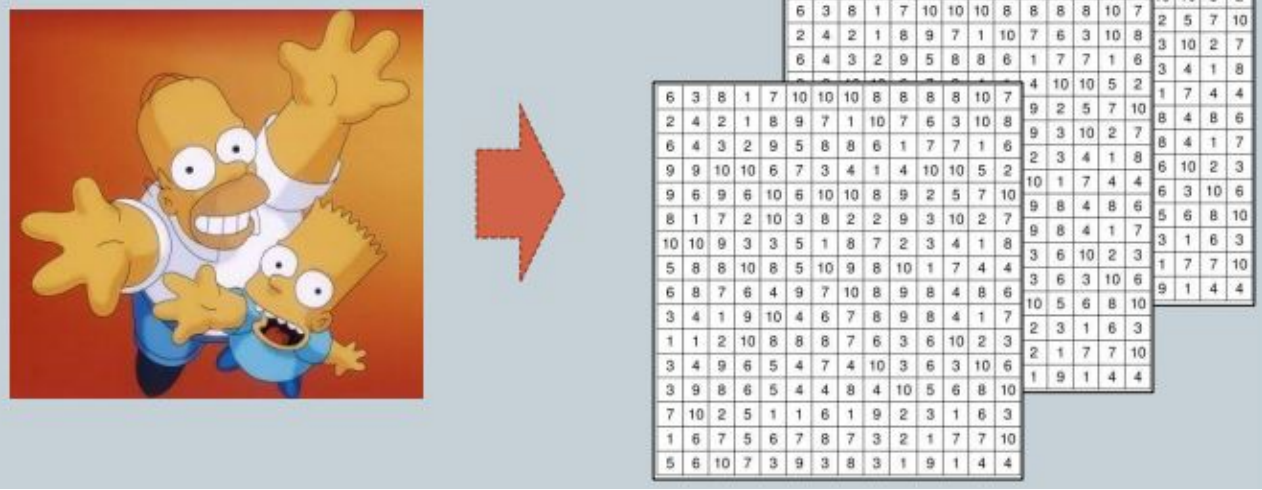
6	3	8	1	7	10	10	10	8	8	8	8	10	7
2	4	2	1	8	9	7	1	10	7	6	3	10	8
6	4	3	2	9	5	8	8	6	1	7	7	1	6
9	9	10	10	6	7	3	4	1	4	10	10	5	2
9	6	9	6	10	6	10	10	8	9	2	5	7	10
8	1	7	2	10	3	8	2	2	9	3	10	2	7
10	10	9	3	3	5	1	8	7	2	3	4	1	8
5	8	8	10	8	5	10	9	8	10	1	7	4	4
6	8	7	6	4	9	7	10	8	9	8	4	8	6
3	4	1	9	10	4	6	7	8	9	8	4	1	7
1	1	2	10	8	8	8	7	6	3	6	10	2	3
3	4	9	6	5	4	7	4	10	3	6	3	10	6
3	9	8	6	5	4	4	8	4	10	5	6	8	10
7	10	2	5	1	1	6	1	9	2	3	1	6	3
1	6	7	5	6	7	8	7	3	2	1	7	7	10
5	6	10	7	3	9	3	8	3	1	9	1	4	4

Los valores de la matriz suele ser un número entre 0 y 255

Aprendizaje profundo

Red neuronal convolucional

- En una imagen en color cada pixel tiene tres componentes.
  - RGB
  - HSL
  - HSV
  - RYB
  - CMYK



Aprendizaje profundo

Red neuronal convolucional

- En el procesamiento digital de imágenes un filtro convolucional es una matriz (por lo general cuadrada) de valores.

$$\begin{pmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{pmatrix} \begin{pmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{pmatrix} \begin{pmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{pmatrix}$$

- Los filtros se utilizan para obtener una imagen resultado con las características más relevantes para tareas como clasificación o búsqueda de patrones.
- El kernel en las redes convolucionales se considera como el filtro que se aplica a una imagen para extraer ciertas características importantes o patrones de esta.

Aprendizaje profundo

Red neuronal convolucional

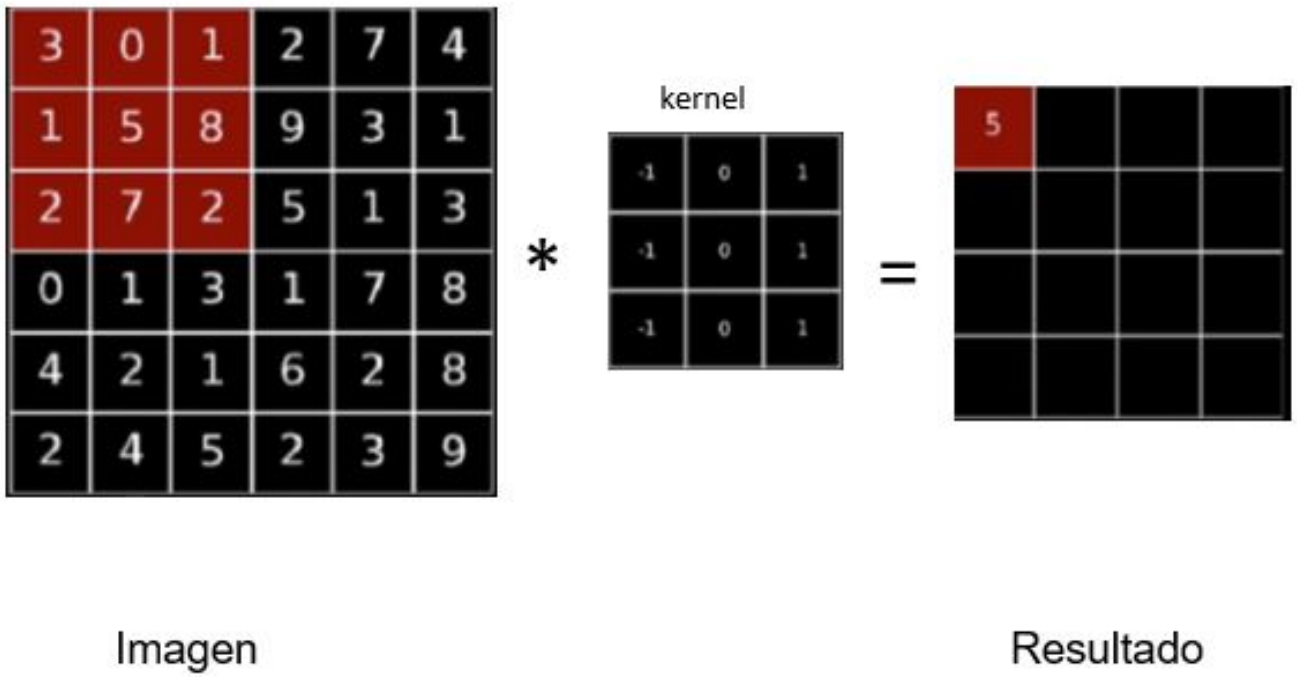
- Filtros para la detección de bordes.

- Filtros para la erosión o dilatación

Aprendizaje profundo

Red neuronal convolucional

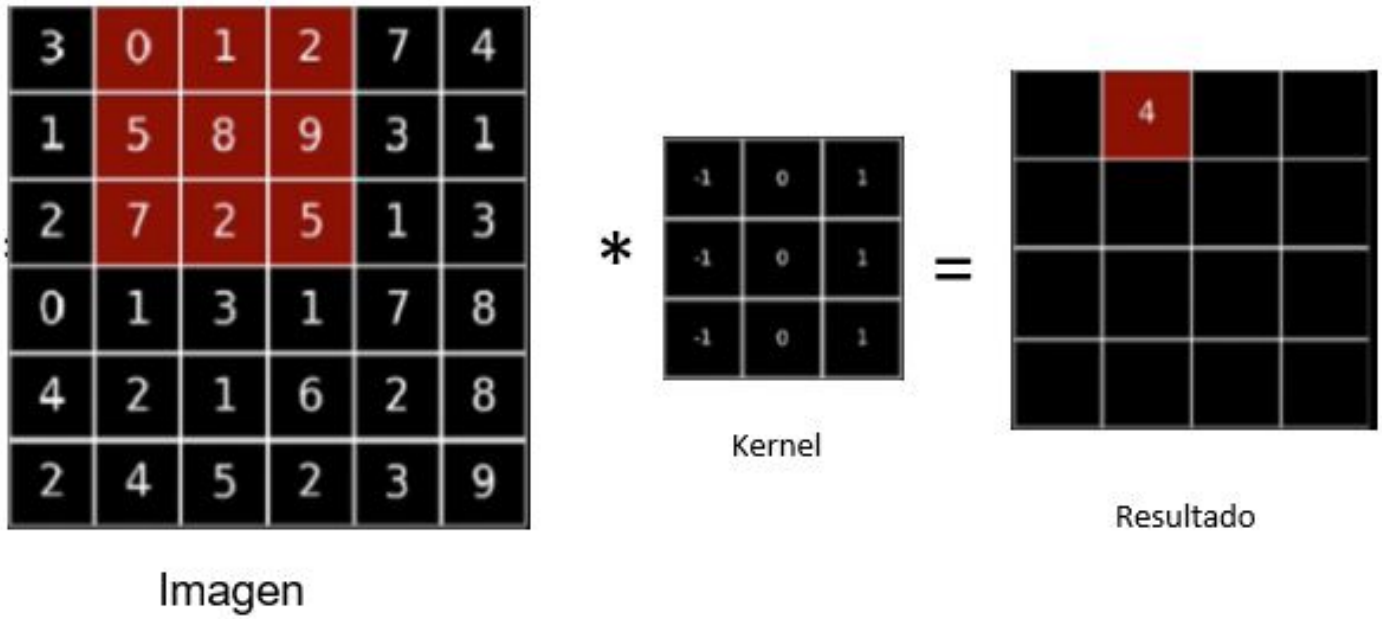
- (3*-1)+(0*0)+(1*1) +
  (1*-1)+(5*0)+(8*1) +
  (2*-1)+(7*0)+(2*1) = 5



Aprendizaje profundo

Red neuronal convolucional

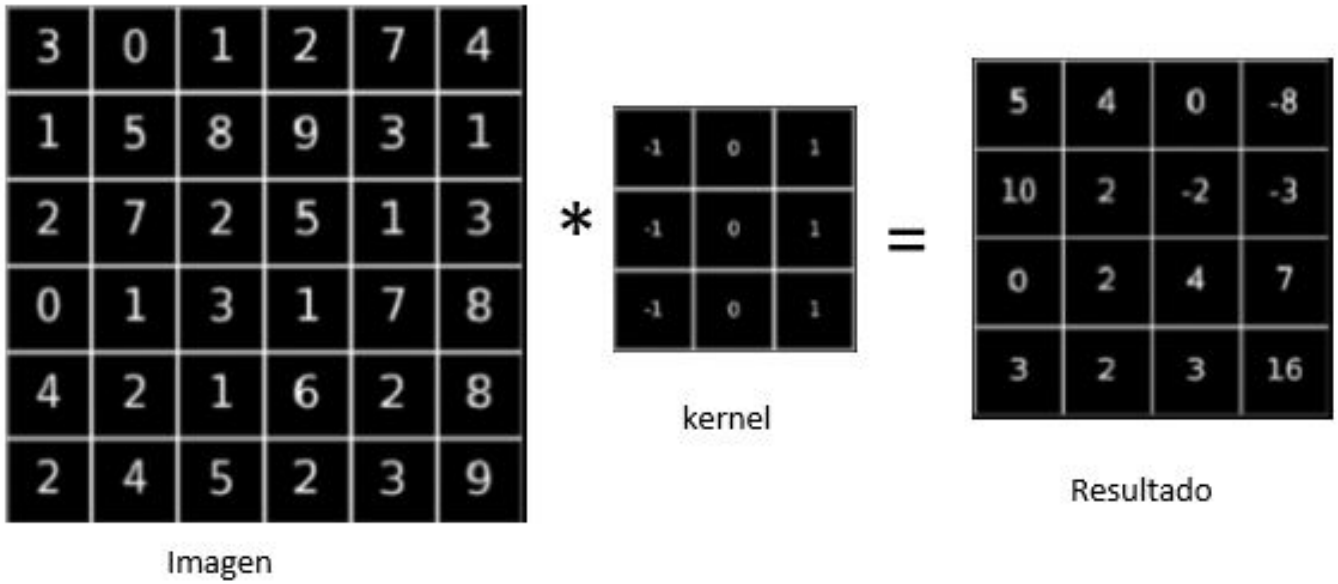
- (0*-1)+(1*0)+(2*1) +
  (5*-1)+(8*0)+(9*1) +
  (7*-1)+(2*0)+(5*1) = 4



Aprendizaje profundo

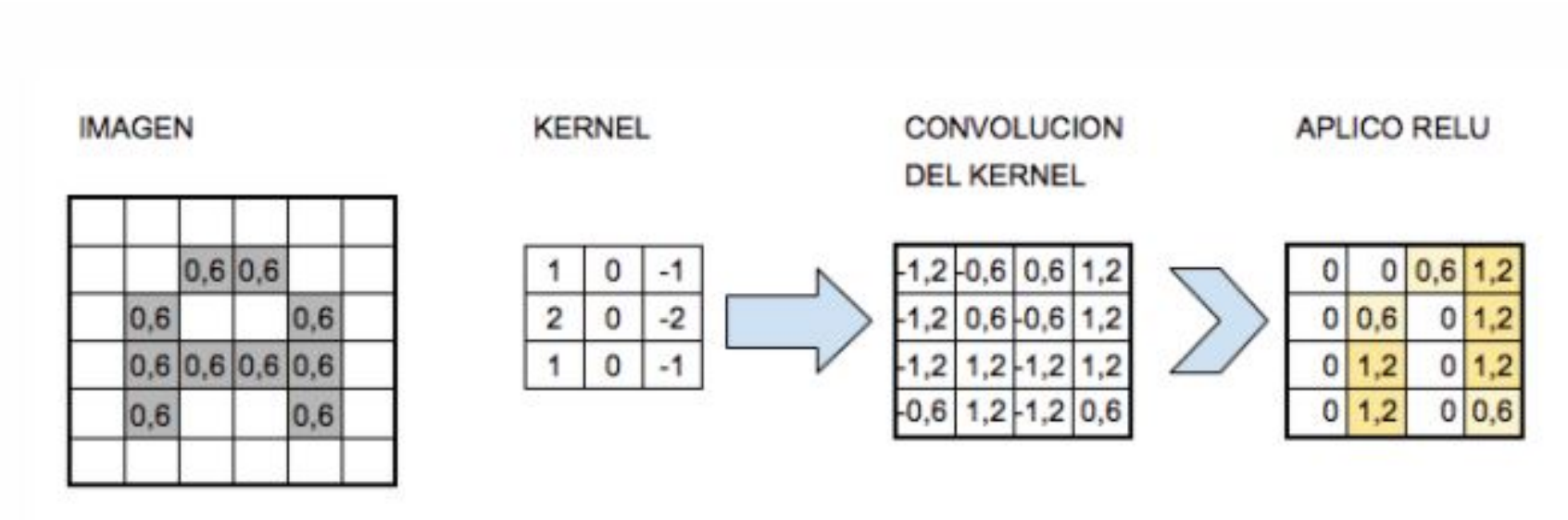
Red neuronal convolucional

- A la matriz resultado se lo conoce como mapa de activación



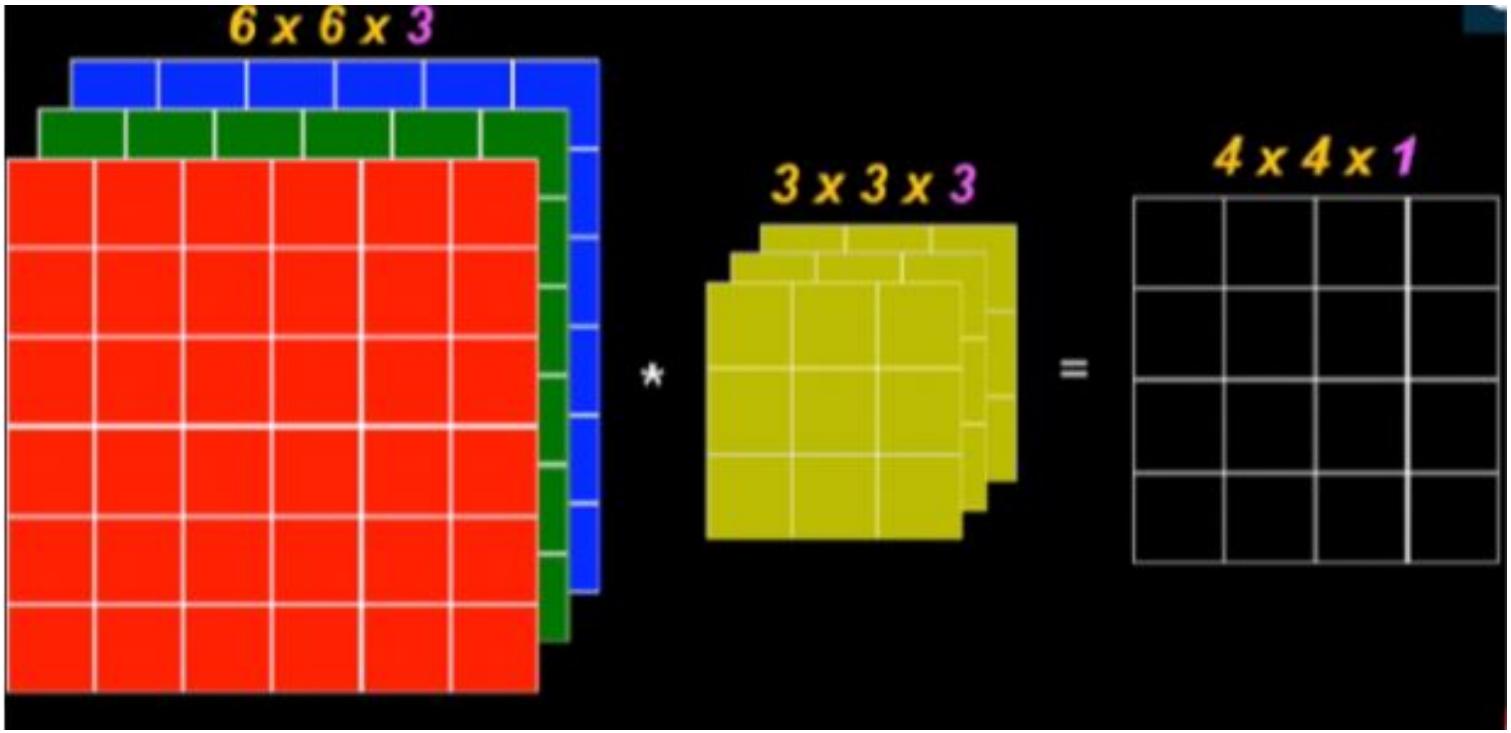
Red neuronal convolucional

- Capa ReLU
- f(x) = max(0, x)



Aprendizaje profundo

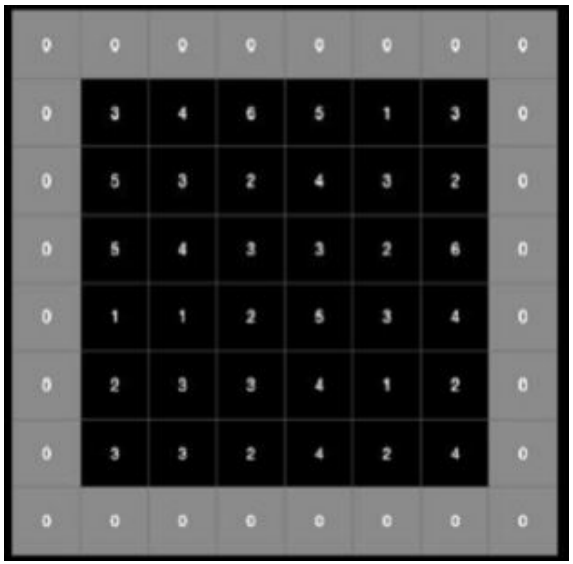
Red neuronal convolucional



Aprendizaje profundo

Red neuronal convolucional

- Padding
  - El padding se aplica agregando píxeles de valor cero alrededor de la imagen original tiene dos usos.
  - El primero es para que al realizar la convolución la imagen resultante sea de igual tamaño que la imagen original.
  - El segundo es cuando se tiene información relevante en las esquinas de la imagen.

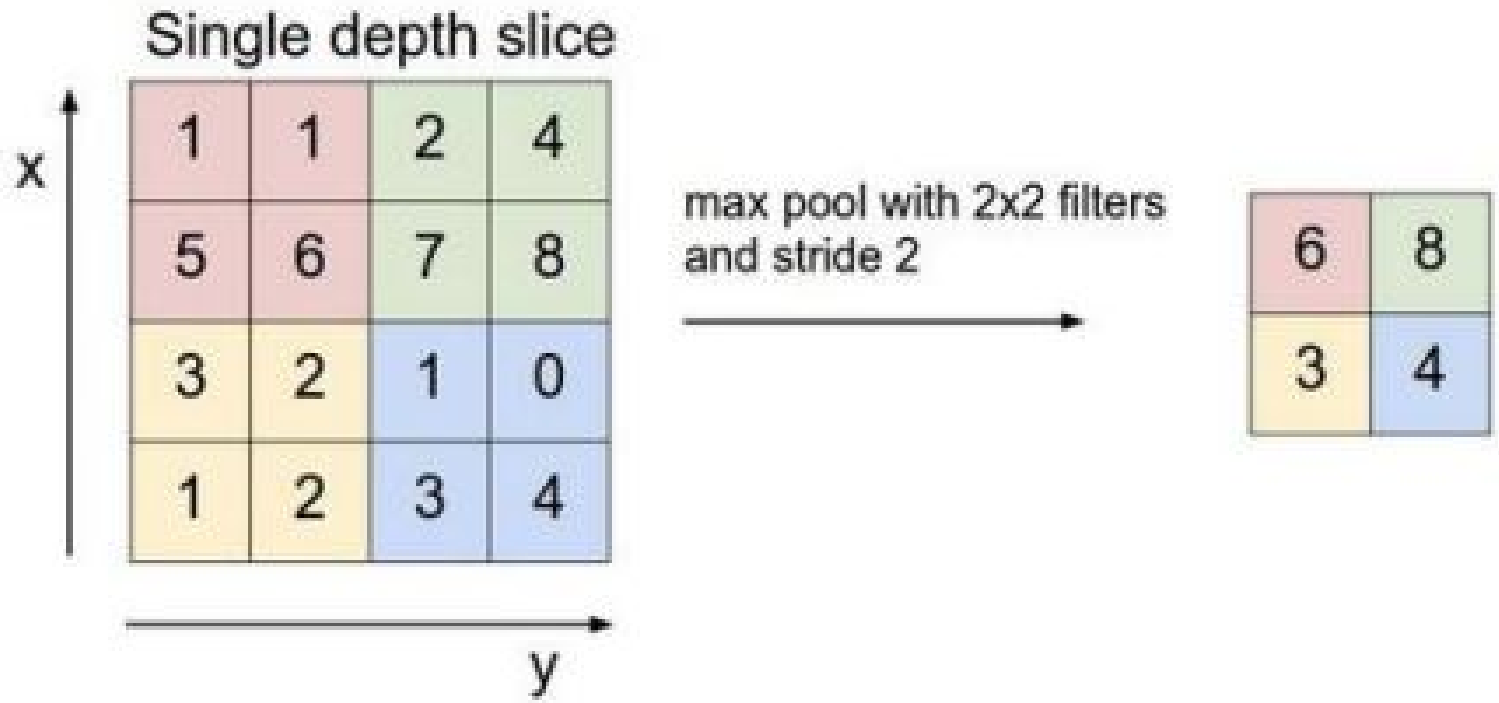


- https://ringa-tech.com/vision01/imagen.html

Aprendizaje profundo

Red neuronal convolucional

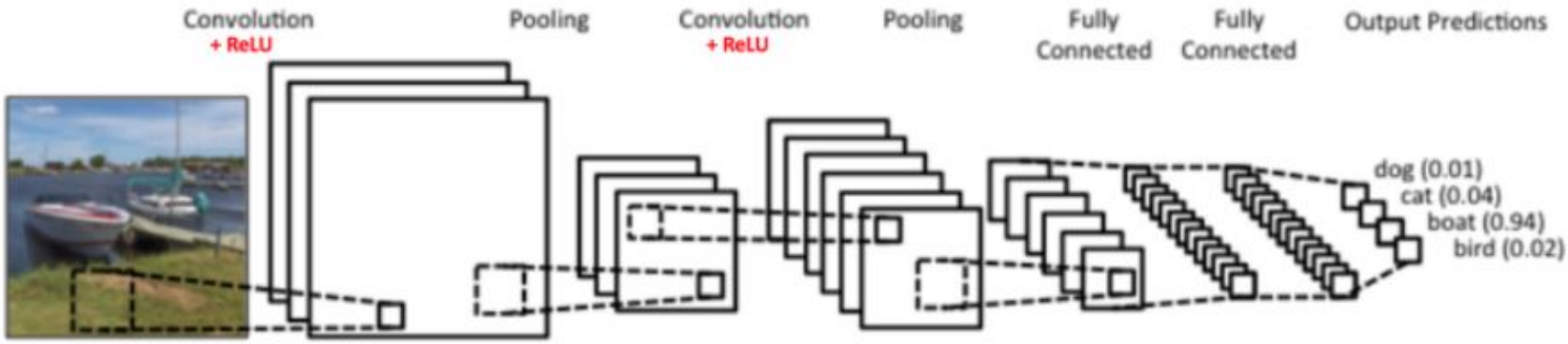
- Capa subsampling (downsampling o max pooling)
  - Se encargan de reducir el tamaño de la imagen.
  - Se define un tamaño de división para la imagen. De cada porción de la imagen se toma la característica más importante.
  - Por lo general la caracterísitca más importante es el valor máximo.



Aprendizaje profundo

Red neuronal convolucional

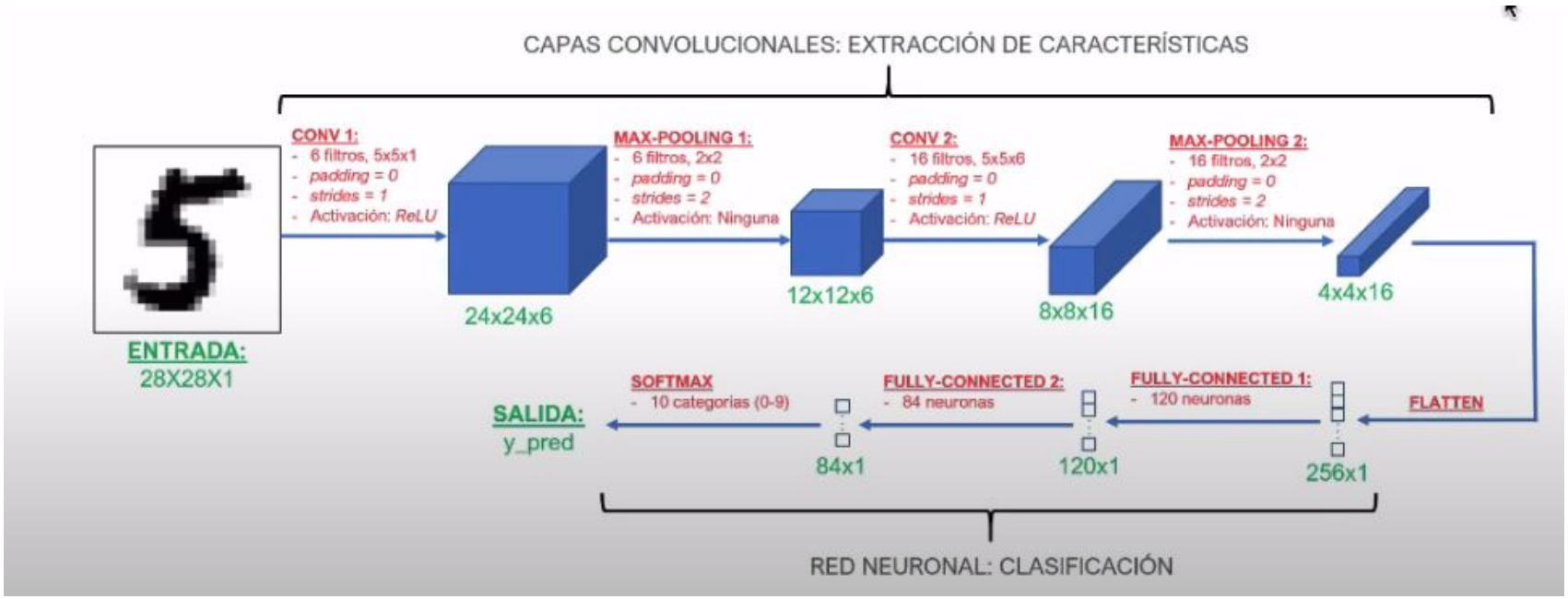
- Arquitectura completa de una red convolucional con dos capas convolucionales



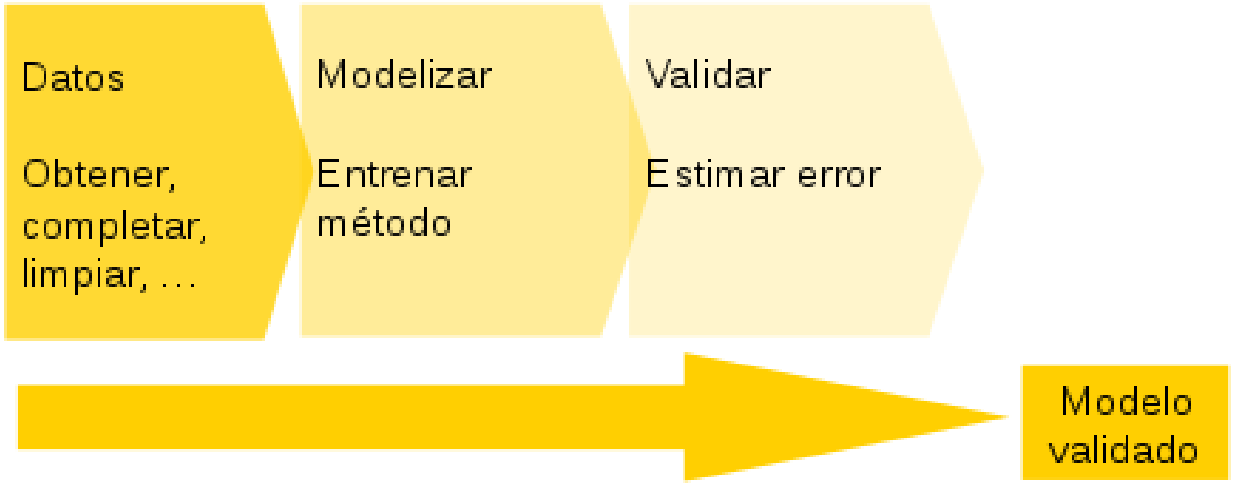
Aprendizaje profundo

Red neuronal convolucional

- Arquitectura completa de una red convolucional con dos capas convolucionales

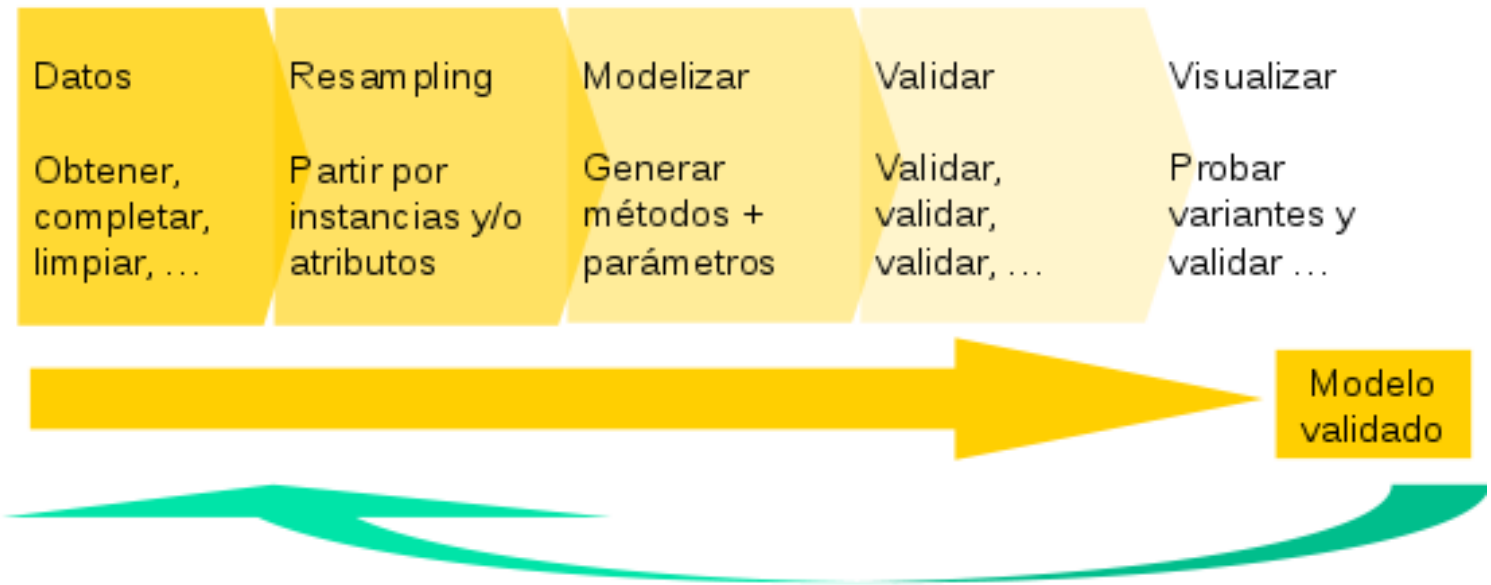


Metodología tradicional para crear modelos



- Identificar el problema y conseguir conocimiento experto.
- Conseguir muchos datos, limpiarlos y completarlos.
- Elegir un método adecuado.
- Entrenar el método con el conjunto de entrenamiento.
- validar el modelo generado con el conjunto de validación.
- Estimar error.
- Proponer hipótesis.

Metodología actual para crear modelos



Metodología actual para crear modelos

- Identificar el problema y conseguir conocimiento experto.
- Conseguir muchos datos, limpiarlos y completarlos.
- Particionar por filas y/o columnas el dataset.
- Elegir varios métodos adecuados.
- Entrenar muchos métodos con diferentes parámetros con varios subconjuntos de entrenamiento para generar varios modelos simples.
- Validar individualmente los modelos simples.
- Ensamblar los mejores modelos con varios ensambles y validarlos.
- Visualizar, validar con experto y proponer hipótesis.
- Volver a empezar.

Selección de variables

- Muchos problemas actuales tienen cientos o miles de variables medidas (sobre pocos ejemplos)
- Modelar esos problemas "directamente" suele ser subóptimo.
- Para mejorar la performance de los métodos de aprendizaje:
  - Algunos métodos trabajan mucho mejor con menos variables (Aunque los métodos modernos de ML suelen ser muy resistentes al problema de la dimensionalidad).
  - En ciertos casos muchas variables no son informativas del problema (ruido o redundancias). (Al eliminarlas reducimos el riesgo de sobreajuste, correlaciones).

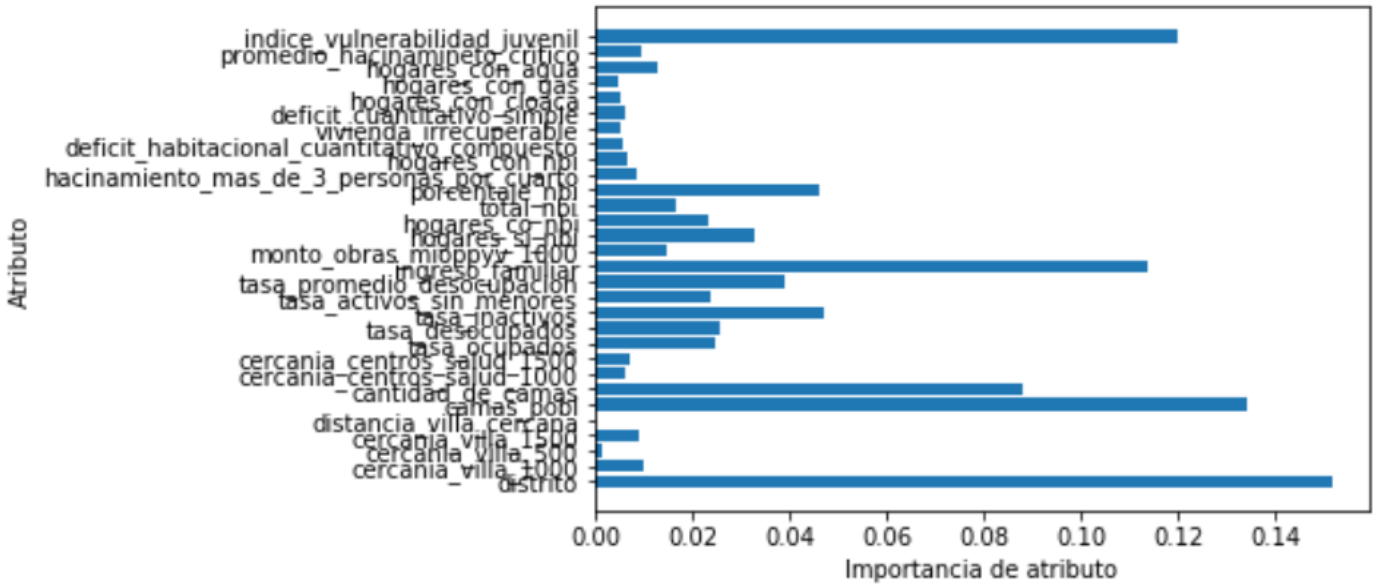
Selección de variables

- Para descubrir:
  - Cuáles son las variables más importantes en un problema.
  - Cuáles variables están correlacionadas, co-reguladas, o son dependientes y cuáles no.
- La selección de variables no es más una técnica de pre-procesado, actualmente es una herramienta para descubrir información de un problema.

Selección de variables

- Univariados: consideran una variable a la vez.
- Multivariados: consideran subconjuntos de variables al mismo tiempo.
- Filtros: Ordenan las variables con criterios de importancia independientes del predictor.
- Wrappers: Usan el predictor final para evaluar la utilidad de las variables.

Selección de variables



Matriz de confusión

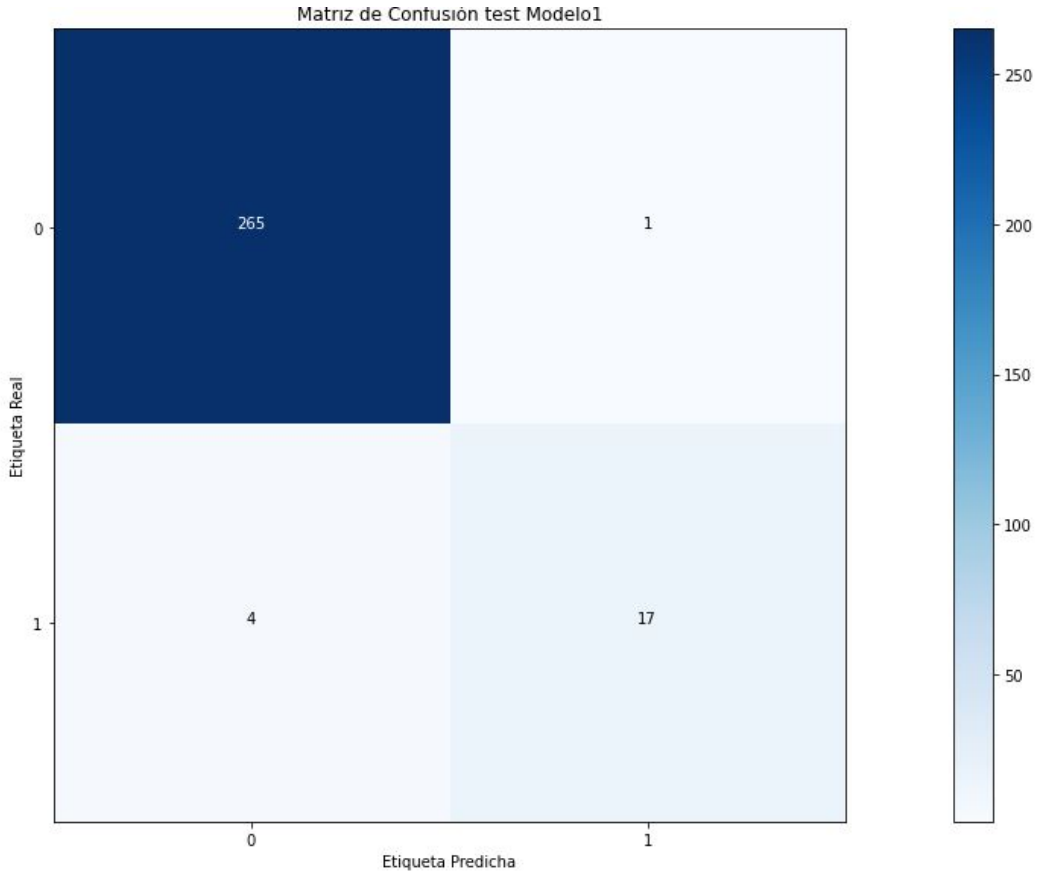
Conocer la distribución según clases de aciertos y errores de un modelo.

	N (modelo)	S (modelo)
n (real)	Negativos Reales	Falsos Positivos
p (real)	Falsos Negativos	Positivos reales

Matriz de confusión

Conocer la distribución según clases de aciertos y errores de un modelo.

- (VN verdaderos negativos: 265) (FP falsos positivos: 1)



- (VP verdaderos positivos: 17) (FN falsos negativos: 4)

Medidas de Performance

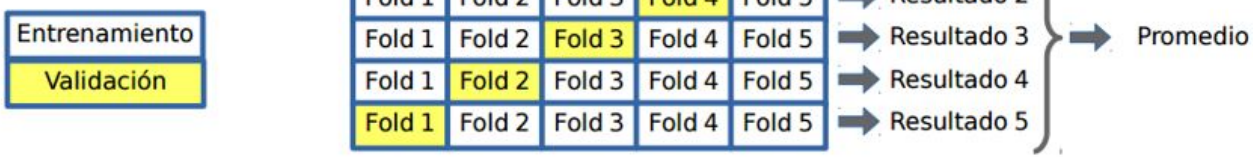
- Accuracy(exactitud) = (VN+VP)/(VN+FP+VP+FN)
- Recall = VP/VP+FN
- Precision (precisión) = VP/VP+FP

Medidas de Performance

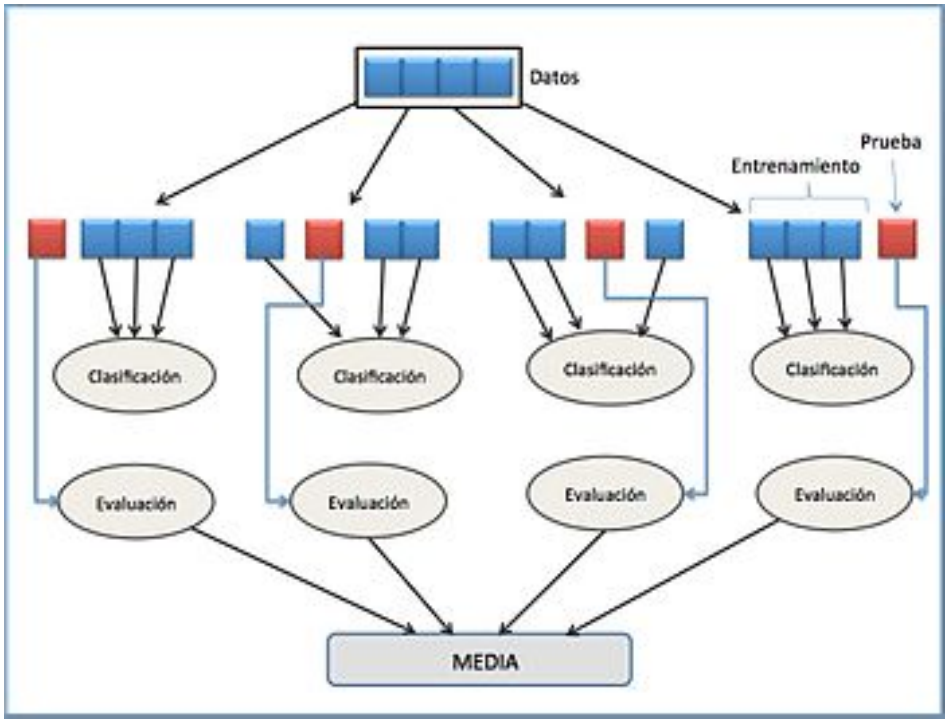
	precision	recall	f1-score	support
clase 0 No Enfermo	0.9851	0.9962	0.9907	266
clase 1 Enfermo	0.9444	0.8095	0.8718	21
accuracy			0.9826	287

Validación Cruzada

- ¿Qué puede pasar si tenemos mala suerte al separar los datos para entrenamiento/validación?
- *k*-Fold Cross Validation:
  1) Desordenar los datos.
  2) Separar en *k* folds del mismo tamaño.
  3) Para $i = 1..k$:
     Entrenar sobre todos los folds menos el *i*.
     Evaluar sobre el fold *i*.
- Ej. para $k=5$:

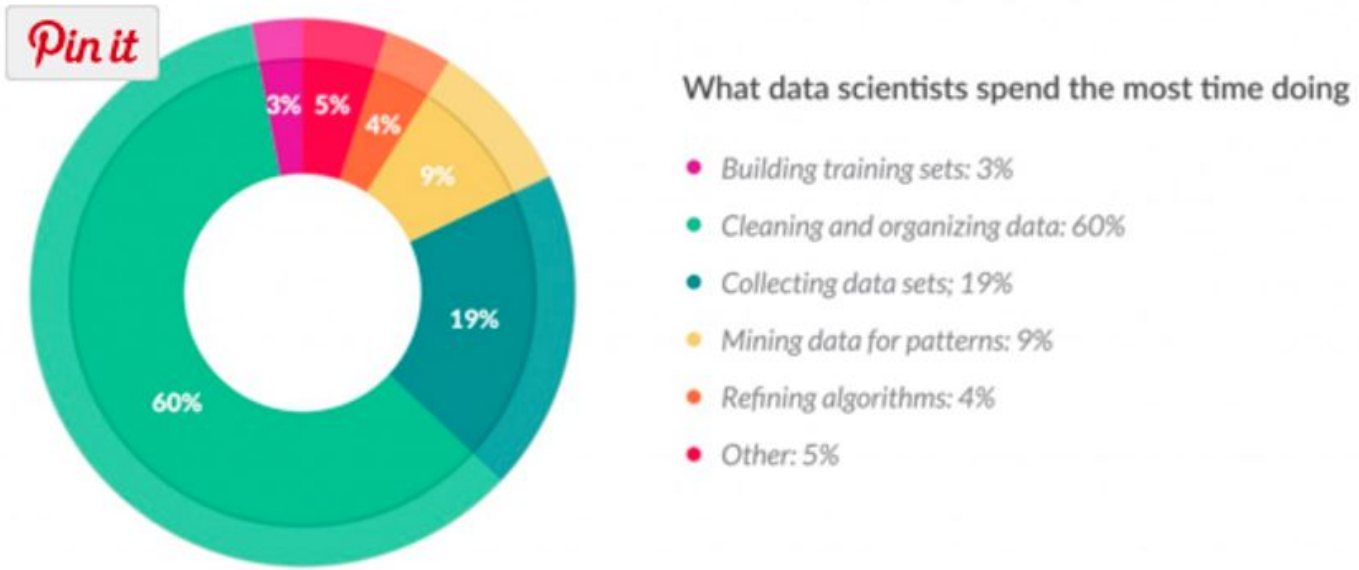


Cross Validation



Esfuerzo para crear modelos

Los científicos de datos dedican el 60 % de su tiempo a limpiar y organizar datos. La recopilación de conjuntos de datos ocupa el segundo lugar con el 19 % de su tiempo.

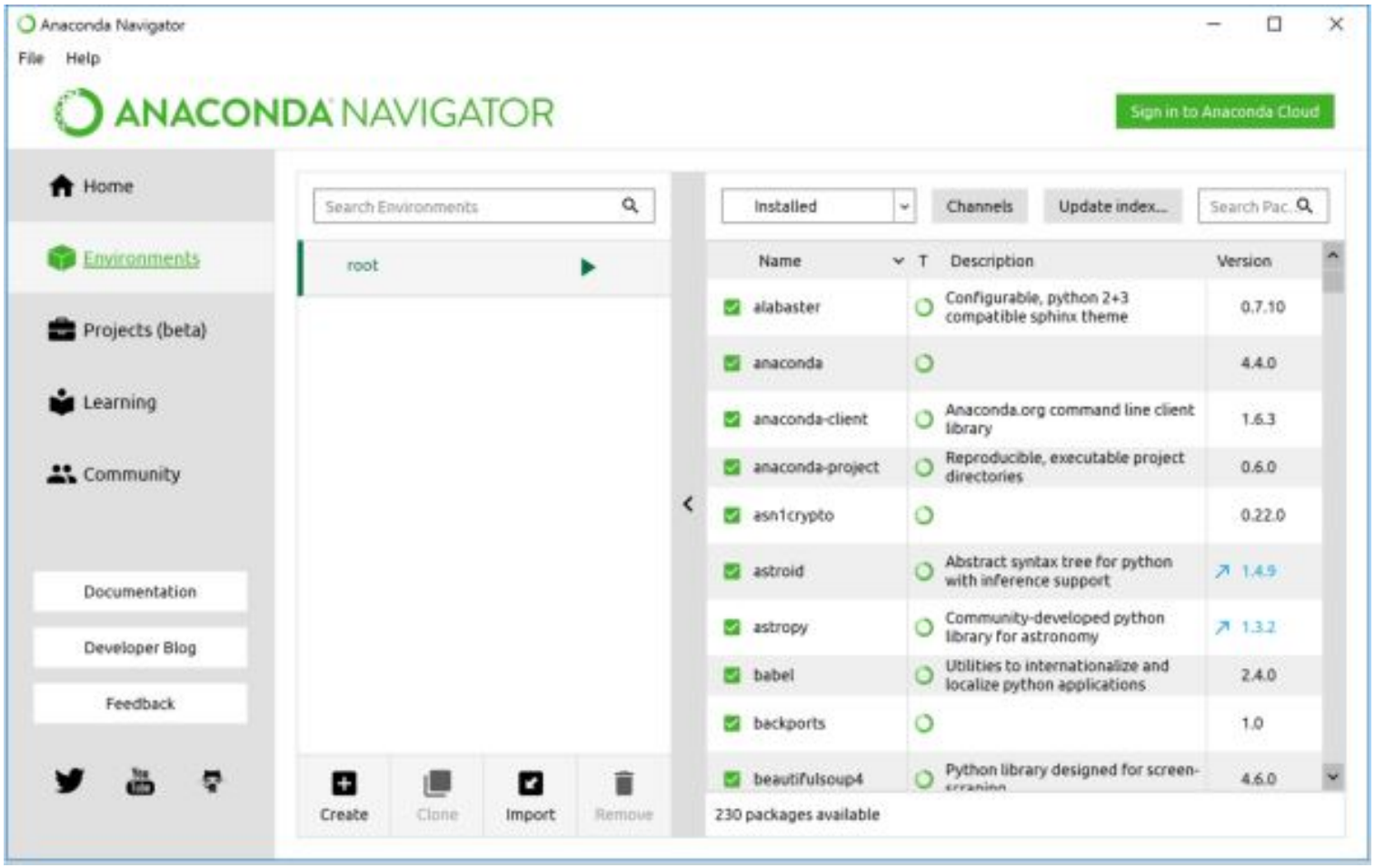


Anaconda

- Utilizaremos la version de Python provista por Anaconda y su administrador de paquetes.

Anaconda

- Utilizaremos la version de Python provista por Anaconda y su administrador de paquetes.



Librerías

- NumPy
- Pandas
- Scikit-learn
- TensorFlow
  - https://www.tensorflow.org/
  - APIs en Python y C++ son las más completas
  - API en Java y Go
  - JavaScript, Lua, R
- Keras
- Torch
  - http://torch.ch/
  - Framework con scripts en Lua
  - JavaScript, Lua, R
  - Interface con C
  - PyTorch

Bibliografía I

Aurélien Géron

Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems.

Publisher: O'Reilly Media, Year: 2017

Tom M. Mitchel, Cap I

Machine Learning

McGraw-Hill Science/Engineering/Math; (March 1, 1997)

North, Matthew

Data Mining for the Masses

Global Text Project, 2012