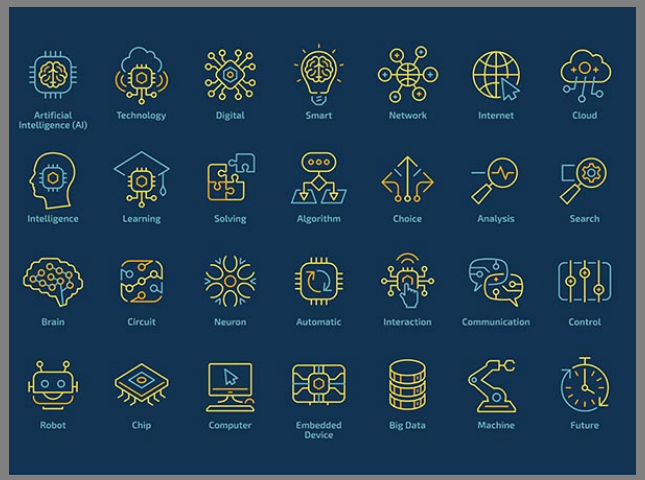
Modulo I
INCEPTIONS MACHINE LEARNING

Abraham Zamudio

VI Programa de Especialización en Machine Learning con Python

2020

CTIC-UNI
BUSINESS SCHOOL

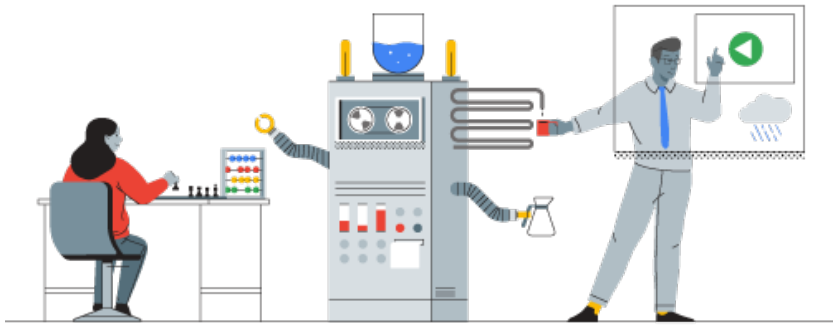


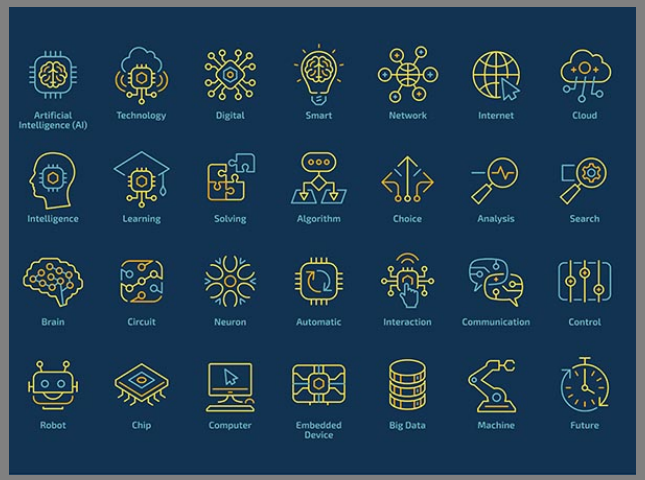
Machine Learning : Aspectos Históricos

Aspectos históricos

Veamos algunos enlaces para ello :

1. ▸ Machine Learning: The complete history in a timeline
2. ▸ A history of machine learning
3. ▸ A History of Machine Learning and Deep Learning
4. ▸ Brief History of Machine Learning
5. ▸ History of Machine Learning



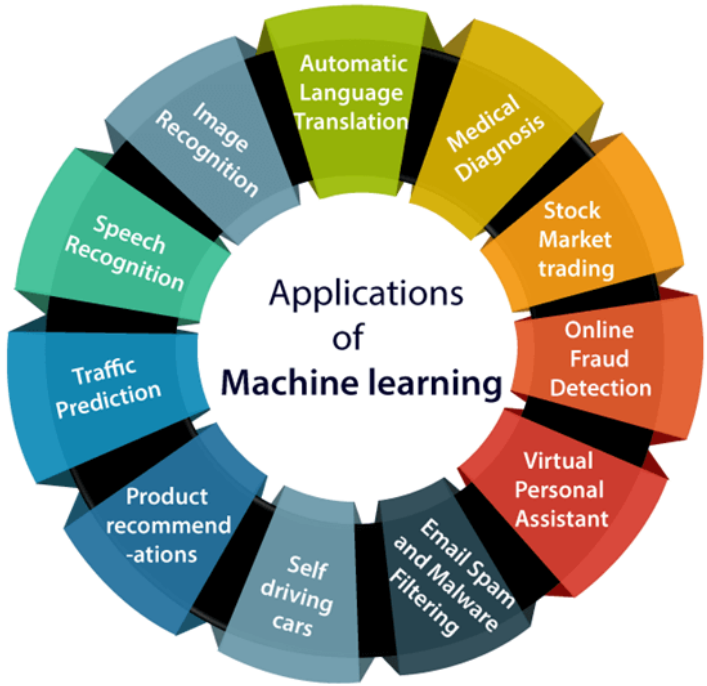


Machine Learning : Un overview

Introducción : Aplicaciones

El machine learning es un tema muy candente por muchas razones, y porque proporciona la capacidad de obtener automáticamente conocimientos profundos, reconocer patrones desconocidos y crear modelos predictivos de alto rendimiento a partir de datos, todo sin requerir instrucciones explícitas de programación.

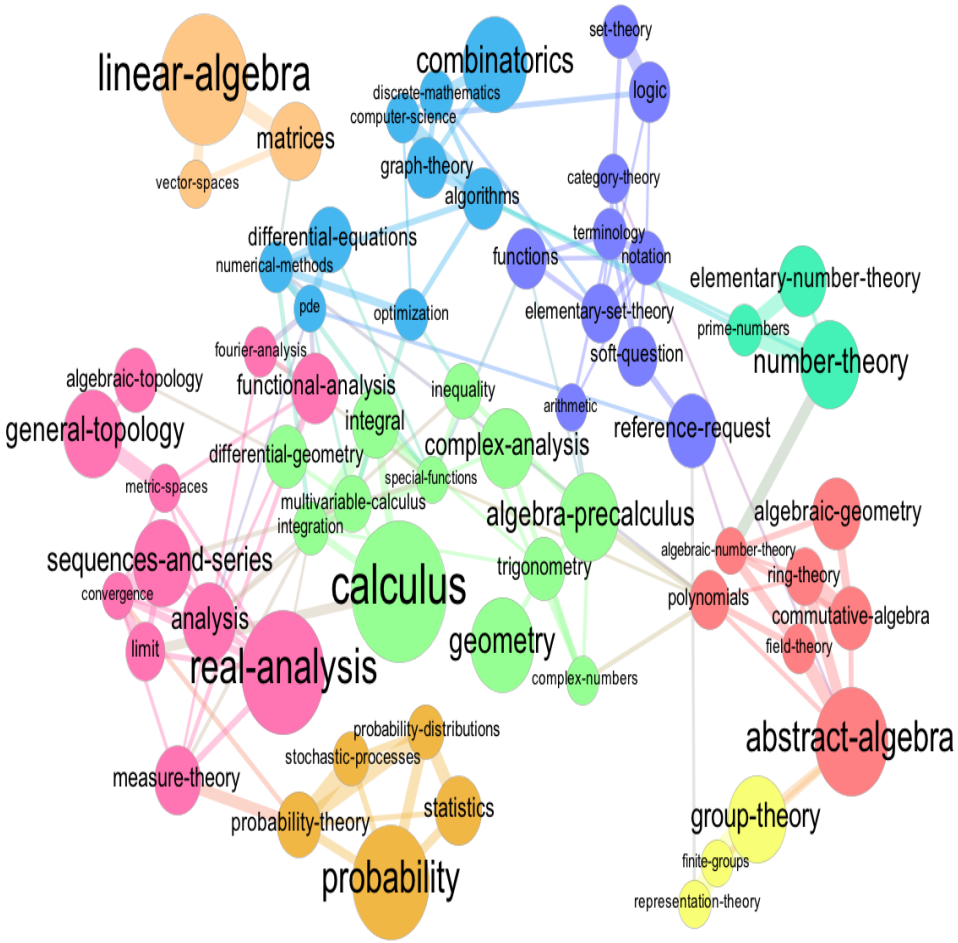
A pesar de la popularidad del tema, el verdadero propósito y los detalles del machine learning no se comprenden bien, excepto por personas muy técnicas y/o científicos de datos con basta experiencia.



Introducción : Aplicaciones

1. ▸ 27 Incredible Examples Of AI And Machine Learning In Practice
2. ▸ StarGAN (Yunjey Choi, et. al.)
    - ▸ StarGAN - Official PyTorch Implementation
3. ▸ Machine Learning Is The Future Of Cancer Prediction
4. ▸ How artificial intelligence is changing drug discovery
5. ▸ Automatic Instrument Segmentation in Robot-Assisted Surgery Using Deep Learning
6. ▸ 14 Ways Machine Learning Can Boost Your Marketing
7. ▸ Tracing outbreaks with machine learning
8. ▸ How Artificial Intelligence is Revolutionizing Food Processing Business?
9. ▸ 4 Applications of Artificial Intelligence in the Food Industry
10. ▸ Machine learning and its radical application to severe weather prediction

Introducción : Fundamento



Conforme pasa el tiempo y las aplicaciones se hacen cada vez mas comunes se mostró que aprender el lado teórico junto con el lado de la programación hace que sea más fácil aprender ambos, Por lo que a lo largo de este curso se comentara tanto las matemáticas **fáciles de entender** como los algoritmos implementados en Python.

Definición de Machine Learning

La definición formal de aprendizaje automático citada y ampliamente aceptada, según lo afirma el pionero de campo Tom M. Mitchell[1] es:

Mitchell, T. (1997)

A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E

[1] https://bit.ly/2EjvDiH

Definición adecuada del problema a resolver (I)

La primera y más crítica tarea que hay que realizar es encontrar cuáles son las entradas y las salidas esperadas. Habría que responder a las siguientes cuestiones:

1. ¿Cuál es el principal objetivo?. ¿Qué vamos a intentar predecir?
2. ¿Cuáles son las características objetivo?
3. ¿Cuáles son los datos de entrada?, ¿están disponibles?
4. ¿A qué clase de problema nos enfrentamos?, ¿clasificación binaria?, ¿agrupamiento?
5. ¿Cuál es la mejora esperada?
6. ¿Cuál es el estado actual de la variable objetivo?
7. ¿Cómo se va a medir la variable objetivo?

Es fundamental tener presente que el Machine Learning solo puede ser usado para memorizar pautas que están presentes en los datos de entrenamiento, por lo tanto solo podemos reconocer lo que hemos visto antes. Cuando usamos Machine Learning estamos asumiendo que el futuro se comportará como el pasado, lo que no siempre es cierto.

Definición adecuada del problema a resolver (II)

No todos los problemas se pueden resolver, solo podemos hacer ciertas hipótesis hasta que tengamos un modelo funcionando:

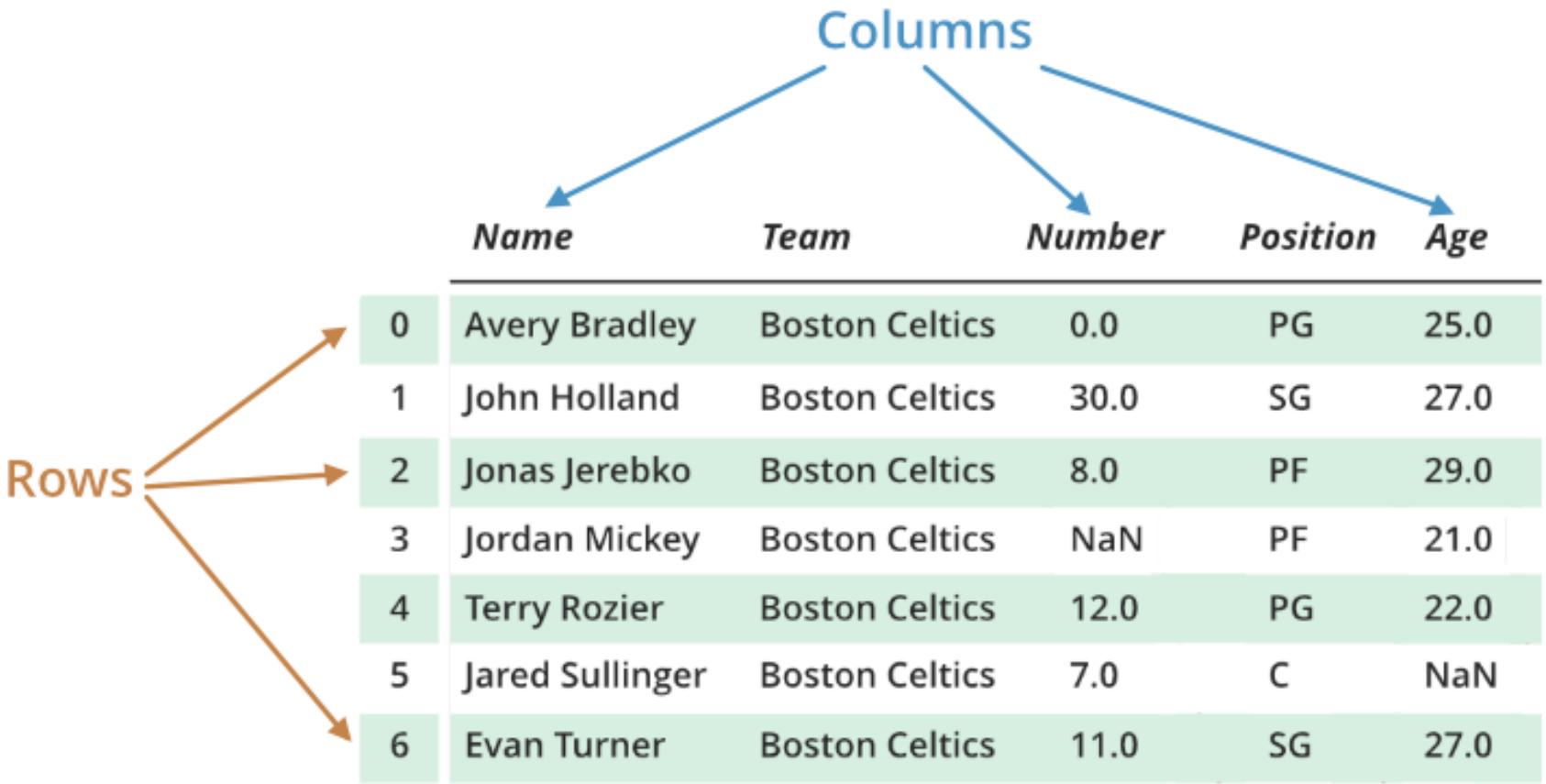
1. Nuestras salidas se pueden predecir proporcionando las entradas.
2. Los datos disponibles contienen la información suficiente para aprender la relación entre las entradas y las salidas.

Nota.

Es fundamental tener presente que el Machine Learning solo puede ser usado para memorizar pautas que están presentes en los datos de entrenamiento, por lo tanto solo podemos reconocer lo que hemos visto antes. Cuando usamos Machine Learning estamos asumiendo que el futuro se comportará como el pasado, lo que no siempre es cierto.

Descripción general del proceso de Machine Learning (I)

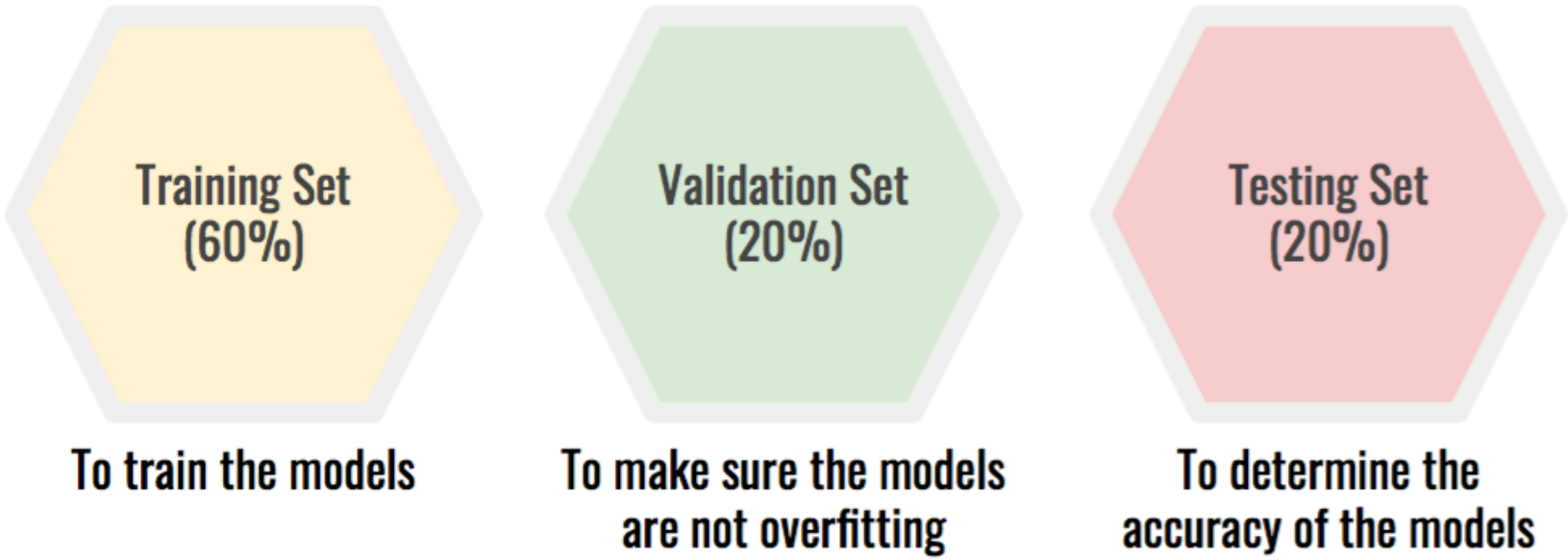
Imagine un conjunto de datos como una tabla, donde las filas son cada observación, y las columnas para cada observación representan las características de esa observación.



	Name	Team	Number	Position	Age
0	Avery Bradley	Boston Celtics	0.0	PG	25.0
1	John Holland	Boston Celtics	30.0	SG	27.0
2	Jonas Jerebko	Boston Celtics	8.0	PF	29.0
3	Jordan Mickey	Boston Celtics	NaN	PF	21.0
4	Terry Rozier	Boston Celtics	12.0	PG	22.0
5	Jared Sullinger	Boston Celtics	7.0	C	NaN
6	Evan Turner	Boston Celtics	11.0	SG	27.0

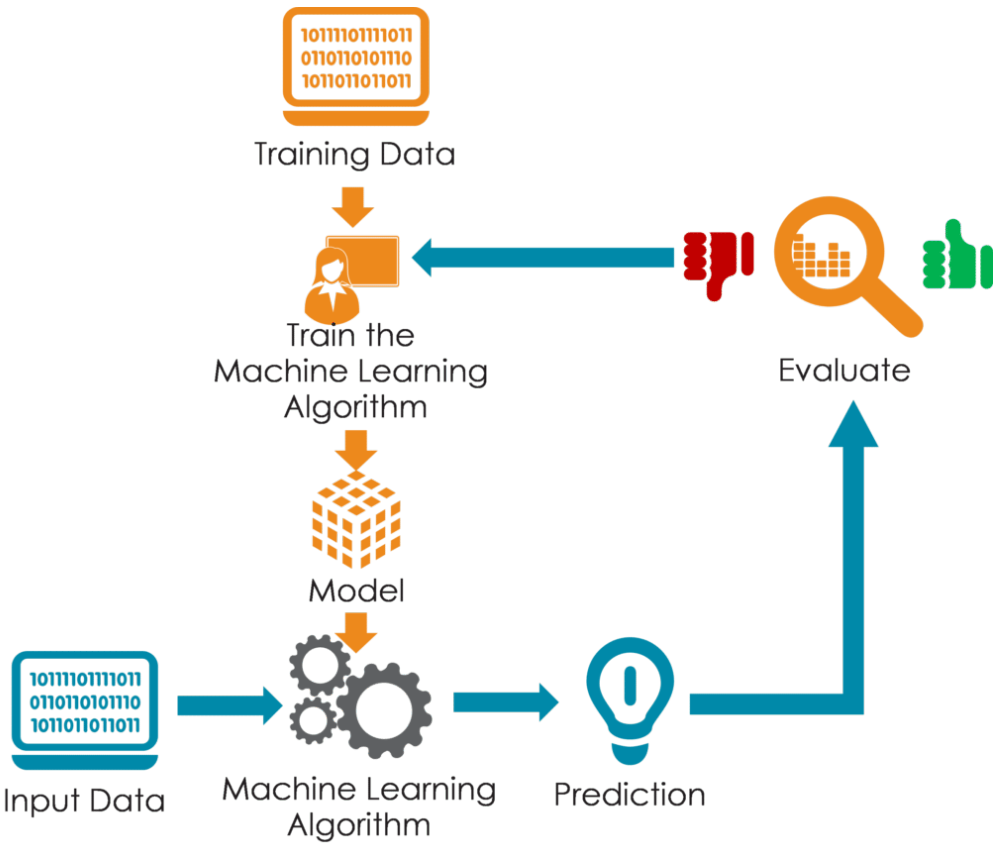
Descripción general del proceso de Machine Learning (II)

Al comienzo de un proyecto de machine learning, un conjunto de datos generalmente se divide en dos o tres subconjuntos. Los subconjuntos mínimos son los conjuntos de datos de entrenamiento y prueba, y a menudo también se crea un tercer conjunto de datos de validación opcional.



Descripción general del proceso de Machine Learning (III)

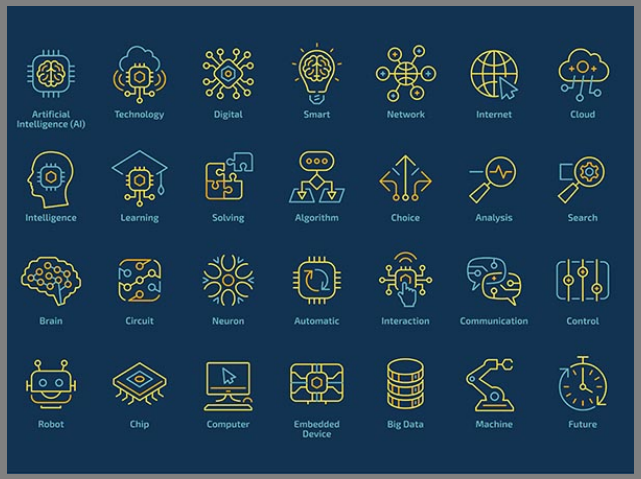
Una vez que estos subconjuntos de datos se crean a partir del conjunto de datos original, se entrena un modelo predictivo o clasificador utilizando los datos de entrenamiento, y luego se determina la precisión predictiva del modelo utilizando los datos de prueba.



Como se mencionó, el machine learning aprovecha los algoritmos para modelar y encontrar patrones automáticamente en los datos, generalmente con el objetivo de predecir algún resultado o respuesta objetivo. Estos algoritmos se basan en gran medida en estadísticas y optimización matemática.

Lecturas

1. ▸ Cómo comenzar con Machine Learning: consejos de expertos de vanguardia
2. ▸ 7 pasos del Machine Learning para construir tu máquina
3. ▸ Machine Learning: Análisis y pasos a seguir para crear tu propio modelo
4. ▸ Applied Machine Learning Process
5. ▸ Six Important Steps to Build a Machine Learning System
6. ▸ Developing a Machine Learning Model from Start to Finish



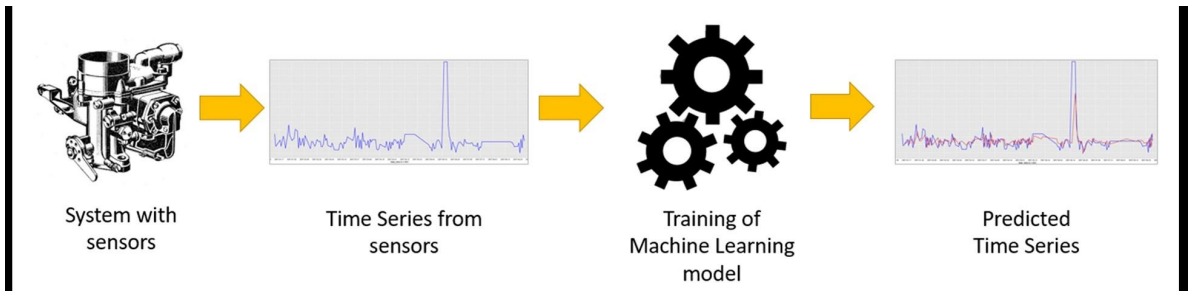
Machine Learning : una guía detallada de selección, preparación y modelado

Introducción

En este sección, presentaremos brevemente los conceptos de rendimiento del modelo y luego nos centraremos en las siguientes partes del proceso de machine learning: selección de datos, preprocesamiento, selección de características, selección de modelos y consideraciones de compensación de modelos.

Introducción al rendimiento del modelo

El rendimiento del modelo se puede definir de muchas maneras, pero en general, se refiere a la eficacia con la que el modelo puede alcanzar los objetivos de solución para un problema determinado (por ejemplo, predicción, clasificación, detección de anomalías, sistemas de recomendación ,etc).



Dado que los objetivos pueden diferir para cada problema, la medida del rendimiento también puede diferir. Algunas medidas de rendimiento comunes incluyen accuracy, precision, recall, receiver operator characteristic (ROC),etc.

Selección de datos y preprocesamiento

Algunos dicen **garbage in equals garbage out** , y este es definitivamente el caso. Básicamente, esto significa que puede haber creado un modelo predictivo, sin importar si los datos utilizados para construir el modelo son no representativos, de baja calidad, con errores, etc. La calidad, cantidad, preparación y selección de datos es crítica para el éxito de una solución de machine learning.

El primer paso para garantizar el éxito es evitar el sesgo (bias) de selección. El sesgo de selección se produce cuando las muestras utilizadas para producir el modelo no son totalmente representativas de los casos en los que el modelo puede utilizarse en el futuro, particularmente con datos nuevos.

Los datos suelen ser desordenados y a menudo consisten en valores faltantes, valores inútiles (por ejemplo, NA), valores atípicos, etc. Antes del modelado y análisis, los datos sin procesar deben analizarse, limpiarse, transformarse y procesarse previamente. Los terminos que se utilizan para estas tareas son : *data munging* o *data wrangling*.

El sesgo de selección es una distorsión en una medida de asociación (como una relación de riesgo) debido a una selección de muestra que no refleja con precisión la población objetivo. El sesgo de selección puede ocurrir cuando los investigadores usan procedimientos inadecuados para seleccionar una muestra de población, pero también puede ocurrir como resultado de factores que influyen en la participación continua de los sujetos en un estudio. En cualquier caso, la población de estudio final no es representativa de la población objetivo: la población general para la que se calcula la medida del efecto y de la que se seleccionan los miembros del estudio.

Sesgo de selección : Ejemplo 1

En un estudio de casos y controles sobre el tabaquismo y la enfermedad pulmonar crónica, la asociación de la exposición con la enfermedad tenderá a ser más débil si los controles se seleccionan de una población hospitalaria (porque fumar causa muchas enfermedades que resultan en hospitalización) que si los controles se seleccionan de la comunidad .

En este ejemplo, los controles hospitalarios no representan la prevalencia de exposición (tabaquismo) en la comunidad de la que surgen los casos de enfermedad pulmonar crónica. La asociación exposición-enfermedad ha sido distorsionada por la selección de controles hospitalarios.

El estudio, publicado en Science el 25 de octubre del 2019 , concluyó que el algoritmo tenía menos probabilidades de referir a personas negras que a personas blancas que estaban igualmente enfermas a programas que apuntan a mejorar la atención de pacientes con necesidades médicas complejas. Los hospitales y las aseguradoras usan el algoritmo y otros similares para ayudar a administrar la atención de aproximadamente 200 millones de personas en los Estados Unidos cada año.

Sesgo de selección : Ejemplo 3

Como hemos visto en los dos primeros ejemplos el sesgo de selección viene en diversas formas.

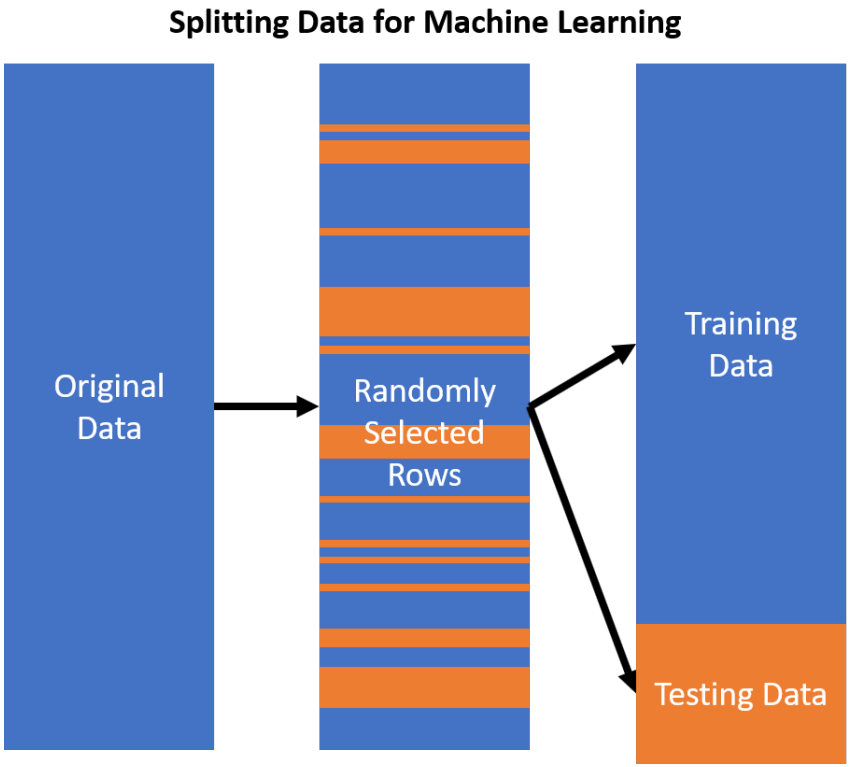
Aquí otro ejemplo. Los encuestadores políticos podrían realizar una encuesta sobre las preferencias de voto de las personas en una elección mediante un muestreo aleatorio de números de teléfono.Este enfoque sub-representa a las personas que renuncian a los teléfonos fijos por teléfonos celulares (nota: ¡los encuestadores modernos también llaman a los teléfonos celulares, exactamente por este problema!). Si estos usuarios que prefieren los teléfonos celulares también tienden a preferir a un candidato sobre otro que difiere de la población general, entonces esta subrepresentación conducirá a un sesgo.

Nos gustaría una forma de corregir rigurosamente este problema. En general, las muestras de encuestas no representarán perfectamente a la población general.

Sesgo de selección : Ejemplo 4

Aquí hay otro ejemplo ligeramente diferente. Es posible que esté evaluando conjuntos de habilidades dentro de una universidad. En esa universidad, los estudiantes serán admitidos si tienen fuertes habilidades matemáticas o fuertes habilidades sociales (o ambas). Si admite personas con esta política, encontrará una relación negativa entre las matemáticas y las habilidades sociales dentro de la población universitaria, incluso si no están asociadas en la población general. Esto se debe estrictamente al proceso de admisión. Intuitivamente, si sé que alguien fue admitido, y sé que no tienen habilidades matemáticas fuertes, deben tener habilidades sociales fuertes. En otras palabras, condicional a la admisión, las matemáticas y las habilidades sociales son estadísticamente dependientes. Si solo muestra a estudiantes (personas que fueron admitidas), entonces su muestra implícitamente condiciona la admisión.

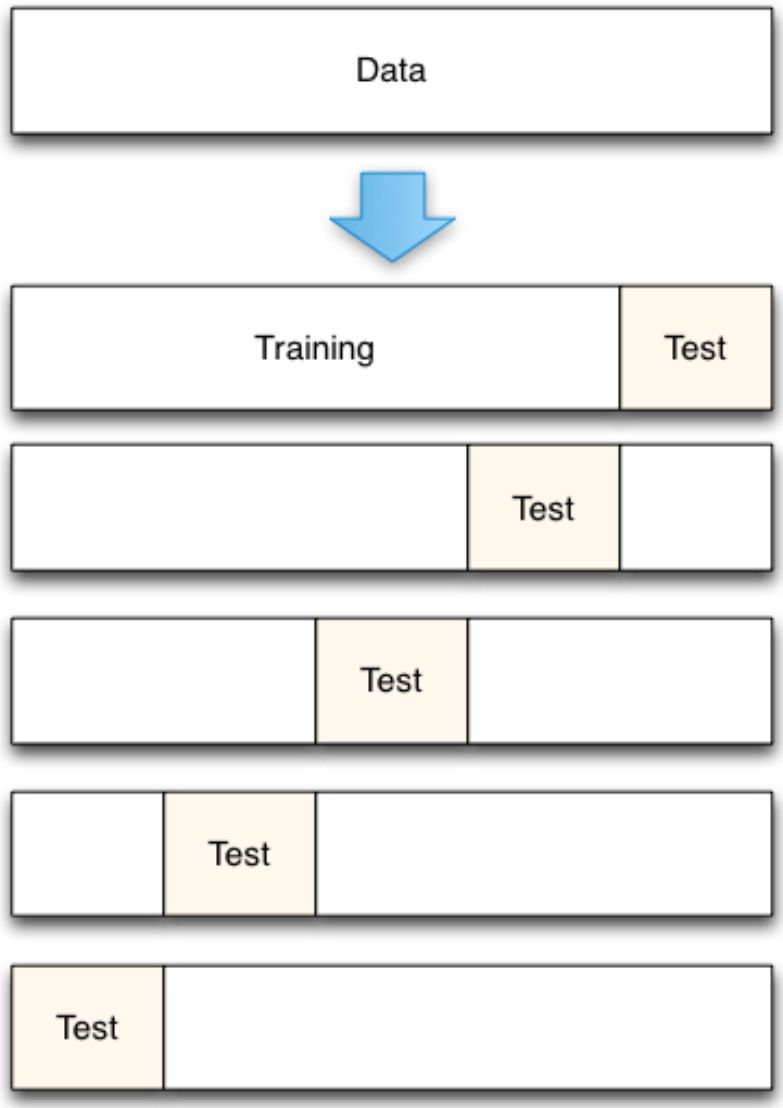
Particionado (splitting) de datos



Recuerde de lo comentado anteriormente que los datos utilizados para el aprendizaje automático deben dividirse en conjuntos de datos de entrenamiento y prueba, así como un tercer conjunto de datos de validación opcional para la validación y ajuste del modelo.Elegir el tamaño de cada conjunto de datos puede ser algo subjetivo y depender del tamaño total de la muestra, y una discusión completa está fuera de nuestro interés en este momento. Sin embargo, como ejemplo, dado solo un conjunto de datos de entrenamiento y prueba, algunas personas pueden dividir los datos en 80 % de entrenamiento y 20 % de prueba.

Particionado (splitting) de datos

En general, más datos de entrenamiento dan como resultado un mejor modelo y un rendimiento potencial, y más datos de prueba dan como resultado una mayor evaluación del rendimiento del modelo y la capacidad de generalización.



Selección de características e ingeniería de características

Feature Selection and Feature Engineering

Una vez que tenga un conjunto de datos representativo, imparcial (insesgado), limpio y totalmente preparado, los siguientes pasos típicos incluyen la selección de características (Feature Selection) y la ingeniería de características (Feature Engineering) de los datos de entrenamiento. Tenga en cuenta que, aunque se discute aquí, ambas técnicas también se pueden utilizar más adelante en el proceso para mejorar el rendimiento del modelo.

Selección de características e ingeniería de características

Feature Selection and Feature Engineering

La selección de características es el proceso de seleccionar un subconjunto de características a partir del cual se pueda construir un modelo de regresión predictiva o clasificador . Esto generalmente se hace para simplificar el modelo y aumentar la capacidad de interpretación, reduciendo los tiempos de entrenamiento y el costo computacional, y para ayudar a reducir el riesgo de sobreajuste, y así mejorar la generalización del modelo.

Las técnicas básicas para la selección de características, particularmente para problemas de regresión, implican estimaciones de los parámetros del modelo (es decir, coeficientes del modelo) y su importancia, y estimaciones de correlación entre las características.

Selección de características e ingeniería de características

Feature Selection and Feature Engineering

1. An Introduction to Variable and Feature Selection
2. Principal component analysis using QR decomposition

La generación de un modelo bien afinado puede ofrecer resultados competitivos pero los sistemas que ofrecen mejores resultados emplean combinaciones de múltiples modelos. Una vuelta de tuerca habitual consiste en apilar múltiples clasificadores con diferentes ajustes de parámetros (los llamados hiperparámetros) y selección de características, y combinar sus predicciones con el fin de reducir el error de generalización y con ello aumentar la precisión.

Veamos un caso real

Caso real.

Se trata de un problema en `MyEmpresa` la cual busca predecir fallos en sus líneas de montaje a partir de miles de medidas sobre las piezas fabricadas. Para empezar, nos enfrentamos a un problema verdaderamente Big Data donde cada pieza lleva asociadas alrededor de 4200 variables, siendo necesario analizar los datos de casi 1,2 millones de piezas para anticipar el destino de otras tantas piezas. Conforme vayamos avanzando en el curso veremos que para el entrenamiento de los algoritmos hubo que hacer frente a una gran cantidad de valores missing, existiendo además un fuerte desbalanceo entre las clases, ya que solo un 0.6 % de las piezas resultaban ser defectuosas.

Selección del modelo

Si bien el algoritmo o modelo que elijamos puede no importar tanto como otras cosas discutidas en esta presentación (por ejemplo, cantidad de datos, selección de características, etc.), aquí hay una lista de cosas a tener en cuenta al elegir un modelo.

1. Interpretabilidad
2. Simplicidad (también conocida como parsimonia)
3. Exactitud (Precisión)
4. Velocidad (entrenamiento, pruebas y procesamiento en tiempo real)
5. Escalabilidad

Selección del modelo

Un buen enfoque es comenzar con modelos simples y luego aumentar la complejidad del modelo según sea necesario, y solo cuando sea necesario. En general, se debe preferir la simplicidad a menos que pueda lograr mayores ganancias de precisión mediante la selección de un modelo mas abstracto.

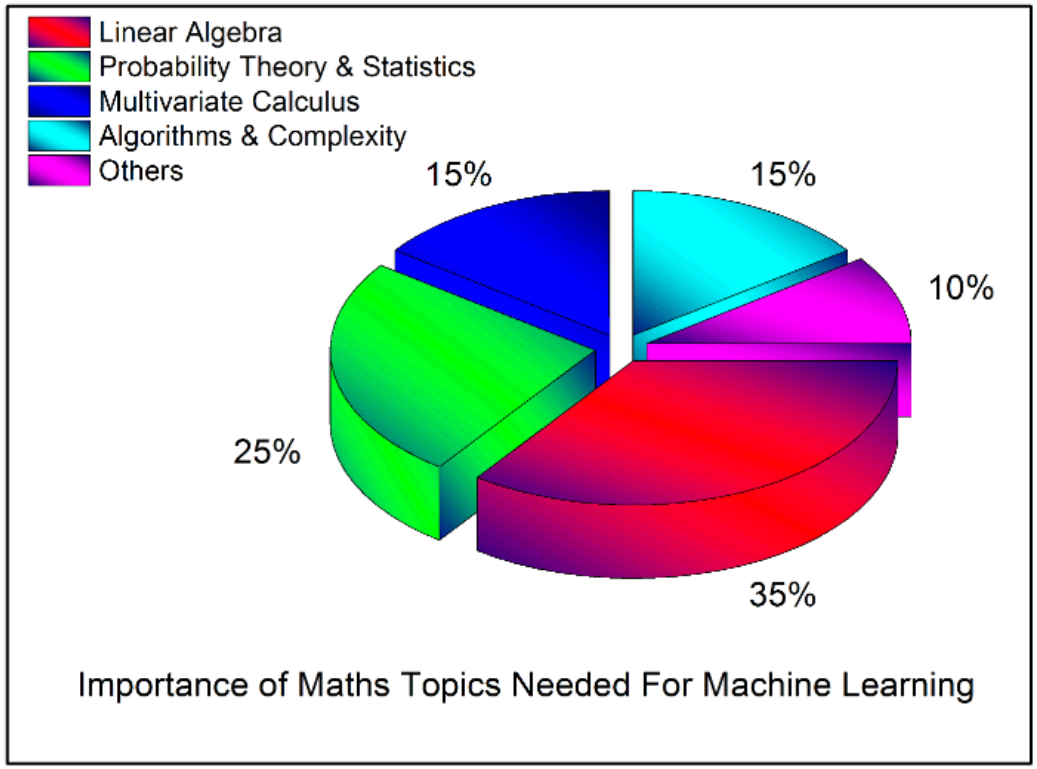
Los modelos relativamente simples incluyen regresión lineal simple y múltiple para problemas de regresión, y regresión logística y multinomial para problemas de clasificación.

Selección del modelo

Más allá de los modelos lineales básicos, las variaciones en la variable de respuesta también pueden deberse a efectos de interacción, lo que significa que la respuesta depende no solo de
ciertas características individuales (efectos principales), sino también de la combinación de
ciertas características (efectos de interacción). Una vez que se incluyen los términos de interacción, la importancia de las interacciones para explicar la respuesta, y si incluirlas, se puede determinar a través de los métodos habituales, como la estimación del p-value.

Fundamentos matemáticos

Para hacer las cosas bien hay que tener una base analítica mínima. En particular, para desarrollar de manera correcta modelos con Machine Learning hay que tener conocimientos básicos de Álgebra Lineal y Teoría de la Probabilidad y Estadística, además de algo de Cálculo y Algoritmia centrada sobre todo en Optimización. Vamos a ver un poquito qué es cada una de estas cosas:



Así dicho, más de uno se habrá llevado las manos a la cabeza. Sin embargo, en muchas de las cosas que se hacen en cualquier departamento de investigación de una empresa, se aplica Álgebra Lineal, por ejemplo al hacer un análisis de Componentes Principales. El álgebra lineal es una rama de las matemáticas que estudia conceptos tales como vectores, matrices, sistemas de ecuaciones lineales, espacios vectoriales y sus transformaciones lineales. Es un área que tiene conexiones con muchas áreas dentro y fuera de las matemáticas, como el análisis funcional, las ecuaciones diferenciales, la investigación de operaciones, las gráficas por computadora, la ingeniería, etc.

1. Linear Algebra - Ian Goodfellow, Yoshua Bengio and Aaron Courville
2. Linear Algebra for Machine Learning By AppliedAICourse
3. Linear Algebra By MIT OpenCourseware

El Machine Learning y la Estadística son campos bastante parecidos. En realidad, el machine Learning es Estadística hecha por las máquinas. Por lo tanto, hay muchas cosas que el analista en este campo tiene que conocer: Combinatoria, Reglas de Probabilidad, Teorema de Bayes, Variables Aleatorias, Varianza, Distribuciones Condicionales y Conjuntas, Distribuciones Estándar (Bernoulli, Binomial, Multinomial, Uniforme y Gaussiana), Estimación de Máxima Verosimilitud, Estimación Máxima a Posteriori, Métodos de Muestreo, etc.

1. Probability and InformationTheory - Ian Goodfellow, Yoshua Bengio and Aaron Courville
2. Statistics for Applications By MIT OpenCourseware
3. Fundamentals of Probability By MIT OpenCourseware

Que es la extensión del cálculo infinitesimal a funciones escalares y vectoriales de varias variables, y que será clave para temas de optimización. En el cálculo multivariante, pasamos de trabajar con números en una línea a puntos en el espacio. Nos brinda las herramientas para liberarse de las limitaciones de una dimensión, usar funciones para describir el espacio y espacio para describir funciones. Cosas que hay que saber hacer o al menos conocer: cálculo diferencial e integral, derivadas parciales, funciones de valores vectoriales, gradiente direccional, matriz Hessiana, Jacobiano, Laplaciano y función Lagragiana.

1. Multivariable Calculus By MIT OpenCourseware
2. Multivariable Calculus By George Cain & James Herod
3. Single and Multivariable Calculus - Early Transcendentals

Importante para comprender la eficiencia computacional y la escalabilidad de nuestro algoritmo de Machine Learning. Se necesitan conocimientos de Estructuras de Datos, Programación Dinámica, Algoritmos Aleatorizados y Sublineales, Gráficos, Gradientes/Descendientes Estocásticos y Métodos Primal-Dual.

1. Data Structures And Algorithms
2. Graph Theory Algorithms By Udemy
3. Convex Optimization: Modeling and Algorithms By Lieven Vandenberghe
4. Dynamic Optimization Methods with Applications By MIT OpenCourseware
5. Numerical Optimization By Jorge Nocedal & Stephen J. Wright