



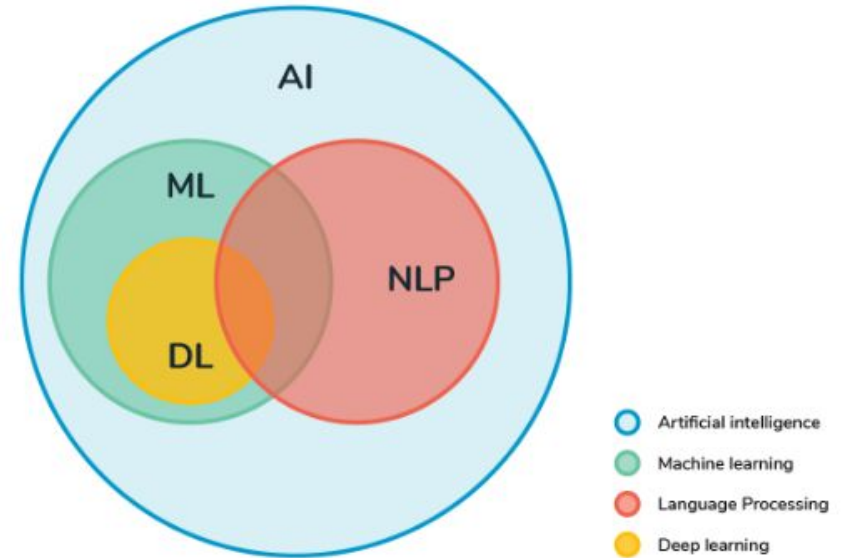
UBA-FCEyN
Departamento de Computación

Laboratorio de Datos
Introducción al
Aprendizaje Automático (Machine Learning)

2do Cuatrimestre 2025

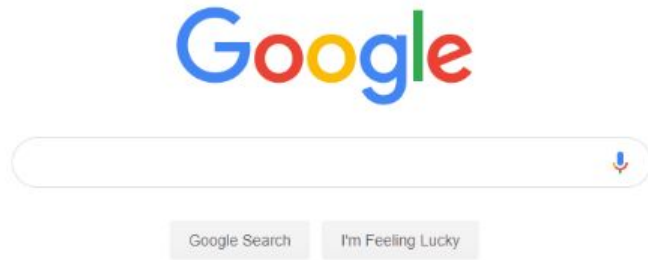


NLP is a subset of AI and uses ML/DL techniques. Source: Sathiyakugan 2018.



NLP is a subset of AI and uses ML/DL techniques. Source: Sathiyakugan 2018.

Algunos ejemplos



Introducing ChatGPT

We've trained a model called ChatGPT which interacts in a conversational way. The dialogue format makes it possible for ChatGPT to answer followup questions, admit its mistakes, challenge incorrect premises, and reject inappropriate requests.

[Try ChatGPT ↗](#)

[Read about ChatGPT Plus](#)

Algunos ejemplos II

- Robots de limpieza



Venii N1 China

- Robots que aprenden a desplazarse

<https://www.youtube.com/watch?v=gn4nRCC9TwQ>

¿Cómo se programan?

¿Qué programas podemos escribir?

Algoritmo: secuencia de instrucciones que transforma una entrada en una salida.

- Ordenamiento de una lista de valores
- Camino más corto entre dos puntos
- ...
- Reconocimiento de caras (humanos vs. computadoras)



The mirror. Diferentes imágenes de Johnny Depp

Aprendizaje automático

Rama de la Inteligencia Artificial, que **estudia algoritmos** para que las **computadoras aprendan a resolver problemas a partir del uso de datos sin ser programadas específicamente.**

Aprendizaje automático - Ejemplos

Extracción de información

(detección de entidades y términos de negación)

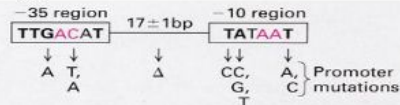
An example of an ultrasonography report (in English)

27518 —14y 11m—20070103—950051 Normal **kidney** echostructure **implant**. **Dilation** **not detected** in the **Urinary tract**. Plenified of normal characteristics. Color Doppler examination: normal characteristics. IR: 0.67. **liver** preserved homogeneous echostructure. **Spleen** homogeneous of 7.8 cm. Both **kidneys** native small echogenic . Evidence in **retroperitoneal** of **solid mass** allready known with **calcifications** and lobular extending left flank **suggesting** It measures approximately 6.3 x 6.8 x 5 cm.

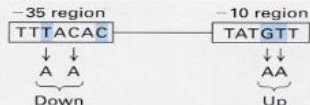
(a)

```
tyr tRNA  TCTCAACGTAACAC TTTACAGCGGCG • CGTCATTTGATATGATGC • GCCCGCTTCCCGATAAGG
rm D1     GATCAAAAAAATAC TTGTGCAAAAAA • TTGGGATCCCTATAATGCGCCTCCGTTGAGACGACACG
rm X1     ATGCAATTTTCCGCTTGCTTCCTGA • GCCGACTCCCTATAATGCGCCTCCATCGACACGGCGGAT
rm (DXE)2 CCTGAAATTCAGGGTTGACTCTGAAA • GAGGAAAGCTAAATATAC • GGCACCTCGCGACAGTGAAG
rm E1     CTGCAATTTTCTATTGCGGCTGCG • GAGAACTCCCTATAATGCGCCTCCATCGACACGGCGGAT
rm A1     TTTTAAATTTCTCTTTGTCAGGCGCG • AATAACTCCCTATAATGCGCCACCATCGACACGGGAACA
rm A2     GCAAAAAATAATGCTTTGACTCTGTAG • CGGGAAGCGGTATATGC • ACACCCCGCGCGCTGASAA
λ PR      TAACACCGTGCGTGTGACTATTTA • CCTCTGGCGGTGATAATGG • TTGCATGTACTAAGGAGGT
λ PL      TATCTCTGGCGGTGTGACATAAATA • CCACTGGCGGTGATACGTA • GCACATCAGCAGGACGCGC
T7 A3     GTGAACAAACAGGGTTGACACATGA • AGTAAACACGCTACGATGT • ACCACATGAACACGACATGA
T7 A1     TATCAAAAAGAGTATTGACTTAAAGT • CTAACCTATAGGATACTTA • CAGCCATCGAGAGGGACACG
T7 A2     ACGAAAAACAGGTATTGACACATGAAGTAACATGCAGTAAAGTAC • AAATCGCTAGGTAAACACTAG
fd VIII   GATACAAATCTCCGTTGACTTTGTT • TCGCGCTTGGTATAATCG • CTGGGGTCAAGATGAGGT
```

(b)



(c)



Detección de promotores

Otras aplicaciones

Minería de datos (registros médicos, bioinformática, clickstream analysis)

Soluciones que no se pueden programar específicamente (autos autónomos, reconocimiento de escritura a mano, visión)

Procesamiento del lenguaje natural: Google Translate, ChatGPT

Reconocimiento del habla: Siri, Cortana, Google Now, Alexa

Detección de fraude: bancos, PayPal, Mercado Libre

Publicidad online: Google Ads

Índice

Introducción

Definición

Principales tipos de Aprendizaje Automático

Aprendizaje automático - Definición

No hay definición única

Samuel (1959)

Campo de estudio que le da a las computadoras la habilidad de aprender sin ser programadas de manera explícita.

Mitchell (1998)

Un programa de computadora se dice que aprende de una **experiencia E** con respecto a una clase de **tareas T** y una **medida de performance P**, si su performance en las tareas T, medidas por P, mejoran con la experiencia E.

Aprendizaje automático - Definición

Un programa de computadora **aprende** una tarea si su performance en la tarea mejora con la experiencia.

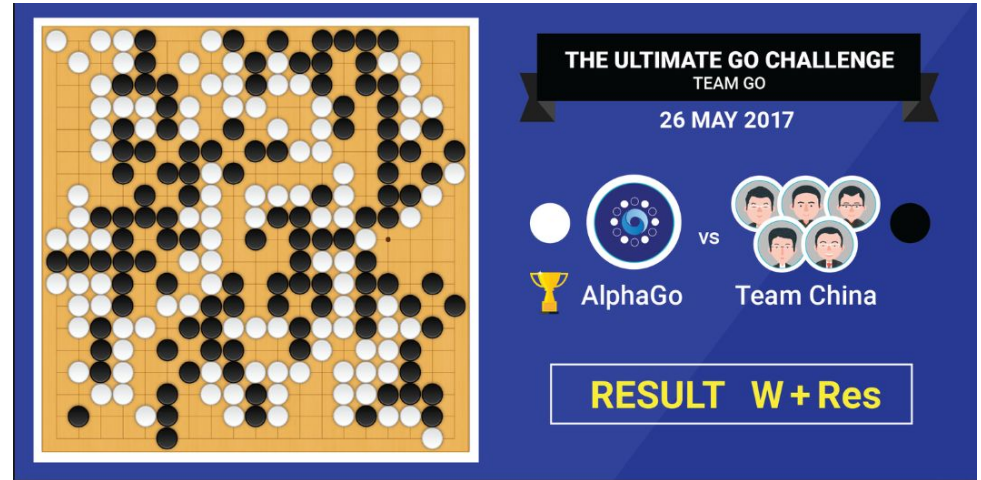
Hay que definir:

- tarea T ,
- medida de performance P ,
- experiencia E

Aprendizaje Automático - Ejemplos I

Juego de Go.

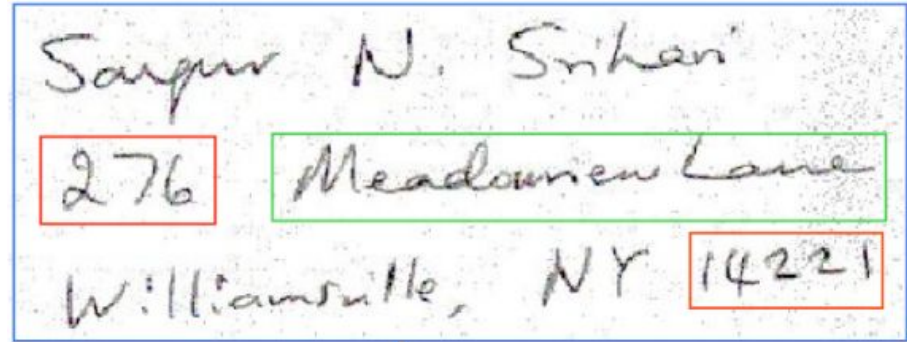
- **Tarea T:** jugar al go
- **Medida de performance P:** probabilidad de que el programa gane el siguiente partido
- **Experiencia E:** la experiencia de jugar muchos juegos de go



Aprendizaje automático - Ejemplos II

Reconocimiento de escritura a mano (handwriting recognition)

- **T:** reconocer la escritura
- **P:** porcentaje de palabras bien reconocido
- **E:** repositorio de imágenes de palabras manuscritas y sus transcripciones



ZIP Code: 14221
Primary number: 276

Imagen tomada de Wikipedia

Aprendizaje Automático - Ejemplos III - Hacer

Clasificación de Spam

- Tarea:
- Medida de performance:
- Experiencia:



Índice

Introducción

Definición

Principales tipos de Aprendizaje Automático

Tipos de aprendizaje automático

Aprendizaje automático:

- supervisado
- no supervisado

Otros:

- aprendizaje por refuerzos

Aprendizaje supervisado

Se le dan datos al algoritmo para que pueda realizar la tarea.



¿Cuánto sale una vivienda de 175 metros cuadrados?

Aprendizaje supervisado

¿Cuánto sale una vivienda de 175 metros cuadrados? Un posible algoritmo de aprendizaje:



Aprendizaje supervisado

¿Cuánto sale una vivienda de 175 metros cuadrados? Otro posible algoritmo de aprendizaje:



Aprendizaje supervisado

Aprendizaje supervisado

Se le dio al algoritmo un conjunto de datos y sus respuestas correctas

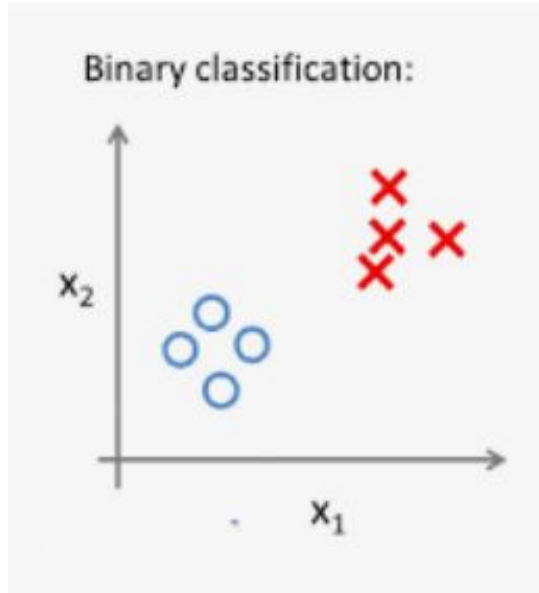
Tarea del algoritmo: producir respuestas ante nuevos datos.

Problema de regresión

Salida: continua (numérica..)



Aprendizaje supervisado



Problema de clasificación de deudores

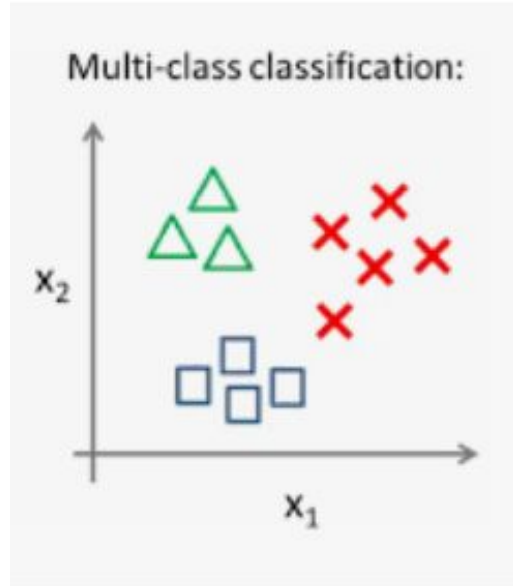
Features o atributos:

x_1 : ingresos, x_2 ahorros.

Clases:

o: riesgo alto, x: riesgo bajo

Aprendizaje supervisado



Features o atributos:

x_1 : tamaño quiste, x_2 : edad paciente

Puede haber muchos otros!

Clases:

 : benigno,

 :maligno tipo 1,

 :maligno tipo 2

Problema de clasificación

Aprendizaje supervisado

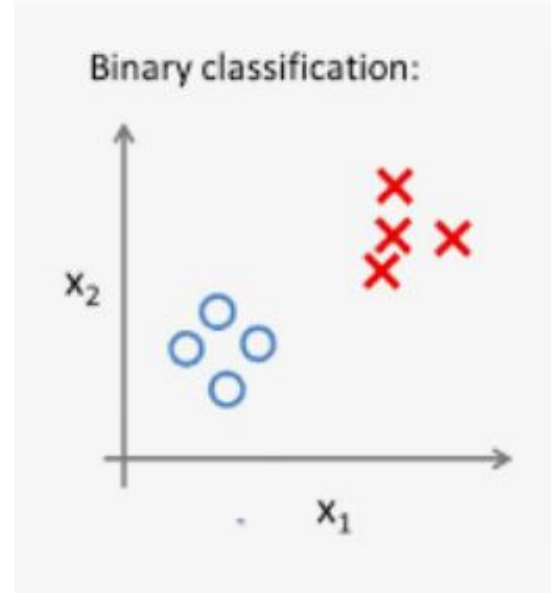
Aprendizaje supervisado

Se le dio al algoritmo un conjunto de datos y sus respuestas correctas

Tarea del algoritmo: producir nuevas respuestas.

Problema de clasificación

Salida: discreta



Aprendizaje supervisado

Datos anotados, requiere:

- **obtener datos**
 - disponibilidad, ¿se pueden publicar?, sensibilidad de la información
 - datos estructurados vs. no estructurados

Aprendizaje supervisado

Datos anotados, requiere:

- **obtener datos**

- disponibilidad, ¿se pueden publicar?, sensibilidad de la información
- datos estructurados vs. no estructurados

- **anotar datos**

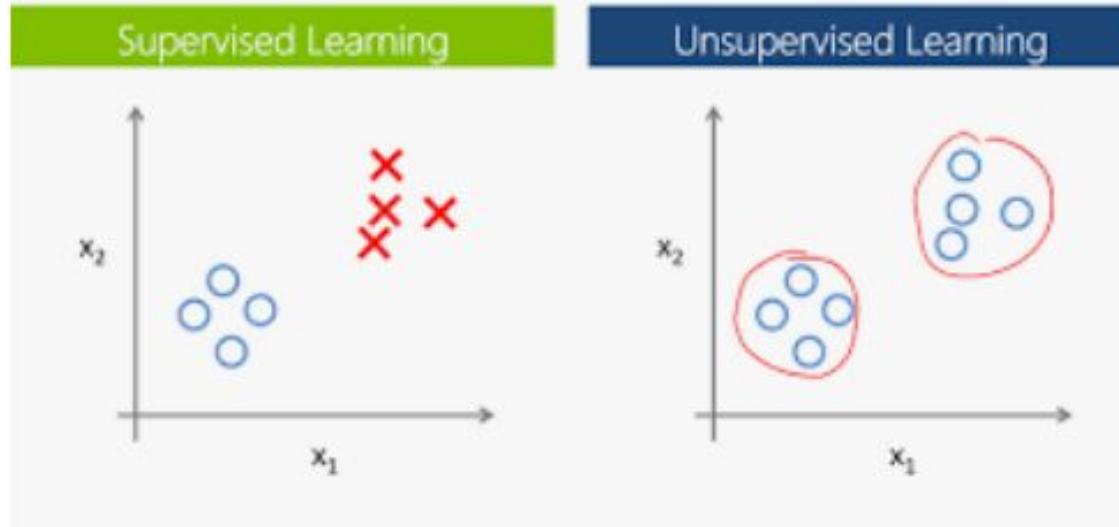
- tarea costosa (tiempo, recursos -> dinero)
- esquema de anotación
- lineamientos de anotación
- anotación de calidad:
 - crowdsourcing (crowdfower, mechanical turk)
 - gratis (facebook)
 - medición de calidad de las anotaciones

Aprendizaje supervisado - Esquema de aprendizaje

- Definición de tarea a aprender
 - Definición (e implementación) de medidas de Performance.
- Recolección de datos
 - Tareas con datos (limpieza, anonimización, armado de conjuntos de entrenamiento, validación y test).
 - Estudio de atributos o features
- Experimentación
 - Selección de algoritmo de aprendizaje automático
 - Selección de atributos o features
 - Elección de parámetros del modelo.
- Evaluación del modelo sobre nuevos datos

Aprendizaje no supervisado

Los datos de entrenamiento no están anotados.



Aprendizaje no supervisado: ejemplos

Análisis de redes sociales



Fuente: Wikimedia commons

Segmentación del mercado.



Segmentación del mercado.

Fuente: internet

Aprendizaje no supervisado

Clustering

detección de patrones ocultos



Reducción de la dimensionalidad

ej. PCA

Aprendizaje por refuerzos (reinforcement learning)

Aprendizaje de un agente autónomo (con sensores) para elegir acciones óptimas que le permitan lograr sus objetivos.

Ej: juegos de mesa, robots autónomos.

Premio y castigos, en función de satisfacción del objetivo. Con efecto atrasado.

Agente tiene que aprender qué secuencias de acciones realizar para producir el mejor premio posible.

Recomendación: **ver película AlphaGo**

Aprendizaje supervisado

Algunos modelos:

- regresión lineal
- regresión logística
- Naïve Bayes
- support vector machines (SVM)
- árboles de decisión
- ensambles de modelos (ej. random forest, gradient boosting, bagging)
- redes neuronales artificiales (RN)
 - aprendizaje profundo (Deep Learning, DL)

Repaso

- Definición de Aprendizaje Automático y ejemplos
- Tipos de aprendizaje automático

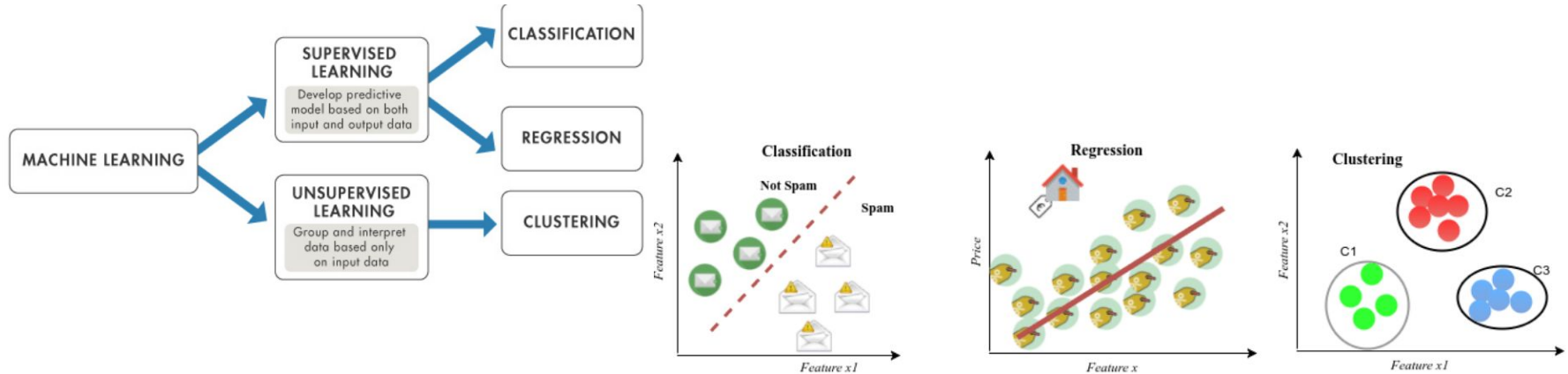
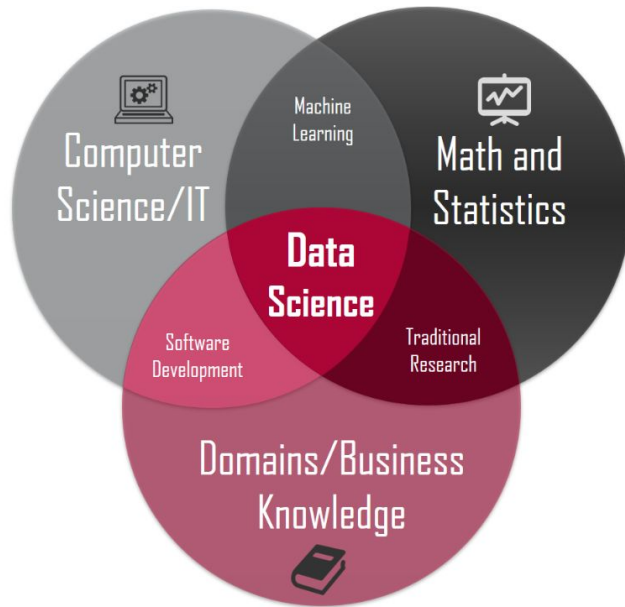


Imagen de Sinan Kaplan

Algunas definiciones - Data science



Para aplicar técnicas de machine learning correctamente, debe haber interacción con conocedores del dominio:

- ciencias sociales y humanidades
- medicina
- género
- leyes
- ...

¿Qué habilidades requiere un data scientist?

- recolectar datos,
- saber interpretar los datos,
- organizar, resumir y analizar datos,
- sacar conclusiones válidas...

Posibles problemas en aplicación de técnicas de ML

- Datos:** mala calidad (no disponibilidad, cantidad inadecuada, mal balanceo, no actuales, formato, sesgos, no correctos, no consistentes), interpretación errónea (ej, correlación no implica causalidad)
- Algoritmos:** parametrización, mala elección
- Capacidad de procesamiento:** inadecuada
- Interpretabilidad**
- Cuestiones éticas**

Bibliografía introducción a ML

Mitchell, Cap. 1

Alpaydin, Cap. 1

Marslan, Cap. 1

Bibliografía básica de ML

- Mitchell, ["Machine Learning"](#), McGraw-Hill, 1997.
- James, Witten, Hastie & Tibshirani, ["An Introduction to Statistical Learning with Applications in R"](#), 6th ed, Springer, 2015.
- Marsland, ["Machine Learning, an Algorithmic Perspective"](#), CRC Press, 2015
- Alpaydin, ["Introduction to Machine Learning"](#), 2010.
- Müller & Guido, ["Introduction to Machine Learning with Python"](#), O'Reilly, 2016. [Código](#).

Bibliografía avanzada de ML

- Hastie, Tibshirani & Friedman, ["The Elements of Statistical Learning"](#), 2nd ed, Springer, 2009.
- Bishop, ["Pattern Recognition and Machine Learning"](#), Springer, 2006.
- Seni, Elder, ["Ensemble Methods in Data Mining: Improving Accuracy Through Combining Predictions"](#), Morgan & Claypool, 2010.
- Leskovec, Rajaramán, Ullman, ["Mining of Massive Datasets"](#), 2010.