

Conceptos básicos en aprendizaje de máquina

A. M. Alvarez-Meza, Ph.D.
amalvarezme@unal.edu.co

Departamento de ingeniería eléctrica, electrónica y computación
Universidad Nacional de Colombia-sede Manizales



Contenido

- 1 Todos quieren aprendizaje de máquina
- 2 Fundamentos del aprendizaje de máquina
- 3 Tipos de aprendizaje de máquina
- 4 Tipos de entrenamiento
- 5 Desafíos en aprendizaje de máquina

Contenido

- 1 Todos quieren aprendizaje de máquina
- 2 Fundamentos del aprendizaje de máquina
- 3 Tipos de aprendizaje de máquina
- 4 Tipos de entrenamiento
- 5 Desafíos en aprendizaje de máquina

El renacer de la inteligencia artificial (Premio Turing 2019)

'Godfathers of AI' honored with Turing Award, the Nobel Prize of computing

Yoshua Bengio, Geoffrey Hinton, and Yann LeCun laid the foundations for modern AI

By James Vincent | Mar 27, 2019, 6:02am EDT



En 2006, Geoffrey Hinton et al. publicaron un artículo ¹ que mostraba como un algoritmo de aprendizaje profundo podía reconocer dígitos a mano con una precisión > 98%, llamándolo “Deep Learning.”

¹ver <http://www.cs.toronto.edu/~hinton/>

El renacer de la inteligencia artificial

- Entrenar un modelo de deep learning era considerado imposible en los 90s.
- Hinton y los demás investigadores en redes neuronales empezaron a destronar a los algoritmos clásicos de aprendizaje de máquina.
- La clave: mucho poder de cómputo y muchos datos.
- En la actualidad: aprendizaje de máquina como corazón de muchos productos de tecnología de punta (búsqueda web, teléfonos inteligentes, reconocimiento de habla, autos que se conducen solos, etc...)

Contenido

- 1 Todos quieren aprendizaje de máquina
- 2 Fundamentos del aprendizaje de máquina
- 3 Tipos de aprendizaje de máquina
- 4 Tipos de entrenamiento
- 5 Desafíos en aprendizaje de máquina

Qué es aprendizaje de máquina?

Básicamente...programar computadores para **aprender desde datos!**

Después de entender la importancia de la ciencia de los datos y su conexión con el aprendizaje de máquina, se busca entonces:

- Entender los modelos básicos de aprendizaje de máquina.
- Avanzar hasta los modelos más avanzados (Deep learning).
- Fortalecer las competencias en estadística y programación.
- Utilizar herramientas libres y reconocidas en Python (Pandas, SciKitlearn, TensorFlow, Keras, PyTorch).

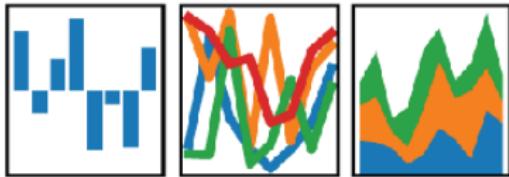
Temas principales del curso

- Conceptos básicos en aprendizaje de máquina.
- Pasos principales en un proyecto de aprendizaje de máquina.
- Funciones de costo y optimización.
- Preparación y preprocesso de los datos.
- Sintonización de hyperparámetros usando validación cruzada.
- Sobreajuste (overfitting) y subajuste (underfitting)
- Reducción de dimensión.
- Algoritmos tradicionales en aprendizaje de máquina: i) regresión polinomial, ii) regresión logística, iii) k-nn, iv) SVM, v) árboles de decisión, vi) métodos de ensamble.
- Fundamentos en redes neuronales y deep learning (para más información ver curso tópicos avanzados en aprendizaje de máquina).

Nuestras librerías amigas en Python - Pandas

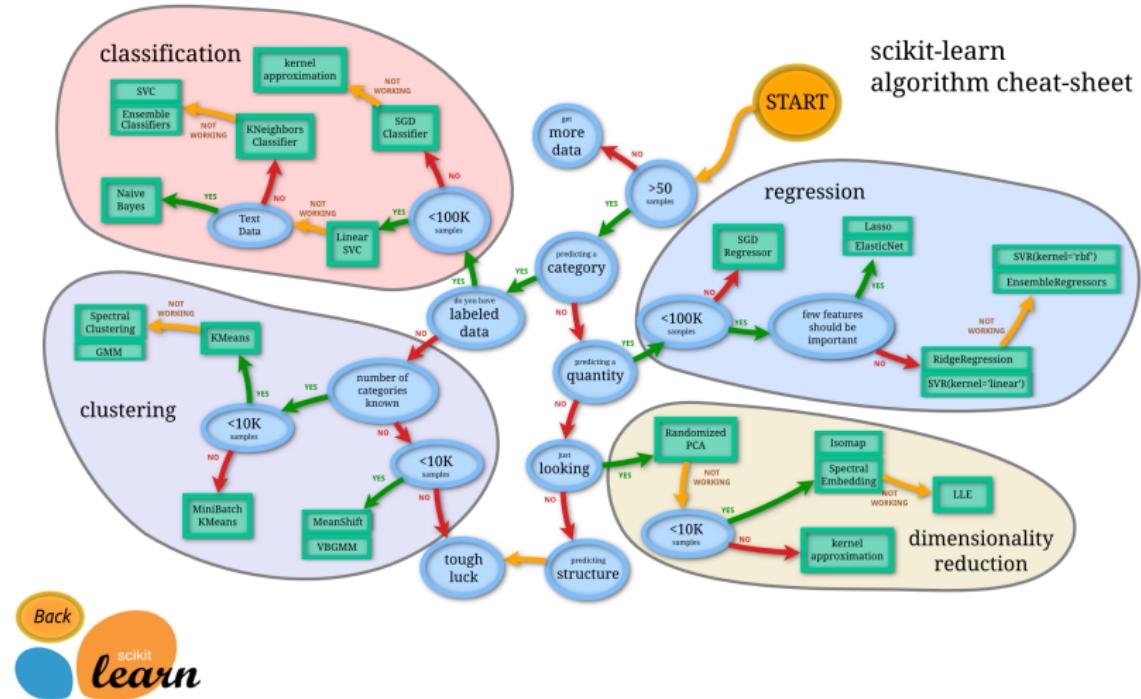
pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



	BandName	WavelengthMax	WavelengthMin
0	CoastalAerosol	450	430
1	Blue	510	450
2	Green	590	530
3	Red	670	640
4	NearInfrared	880	850
5	ShortWaveInfrared_1	1650	1570
6	ShortWaveInfrared_2	2290	2110
7	Cirrus	1380	1360

Nuestras librerías amigas en Python - Scikit-learn



Nuestras librerías amigas en Python - TensorFlow, Keras, PyTorch



Cómputo de alto desempeño Gratis!

No quemes más tu PC!

kaggle™



Free! Cloud Server
TPU & GPU



TensorFlow

Google



colab

Aprendiendo por reglas impuestas (rule by hand-handcraft)

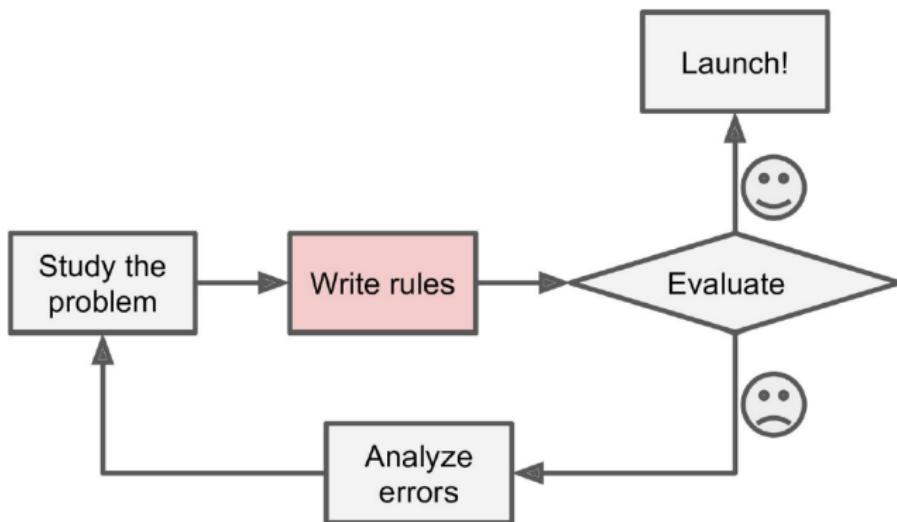


Figure: Aprendizaje por reglas impuestas. fuente: Hands on machine learning book.

Una larga lista de reglas, difíciles de mantener y definir

Aprendizaje estadístico - Aprendizaje de máquina

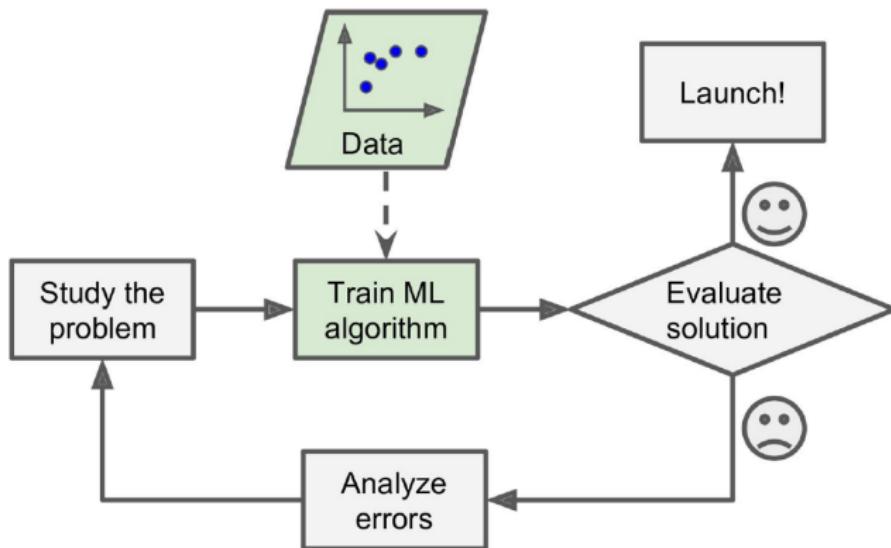


Figure: Aprendizaje de máquina. fuente: Hands on machine learning book.

Aprendiendo desde los datos!

El aprendizaje de máquina y la adaptabilidad

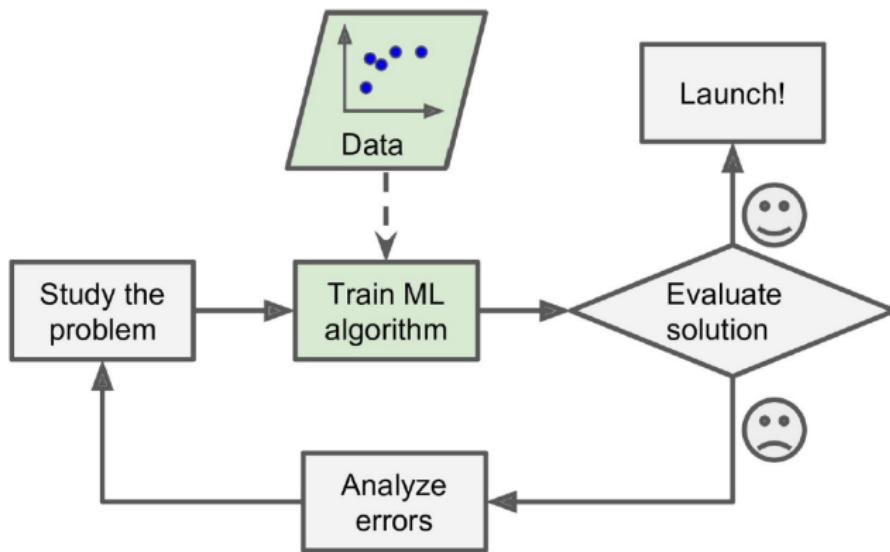


Figure: Adaptabilidad a cambios. fuente: Hands on machine learning book.

Se pueden descubrir patrones (relaciones ocultas) en los datos:
minería de datos.

El aprendizaje de máquina apoyando el entendimiento humano

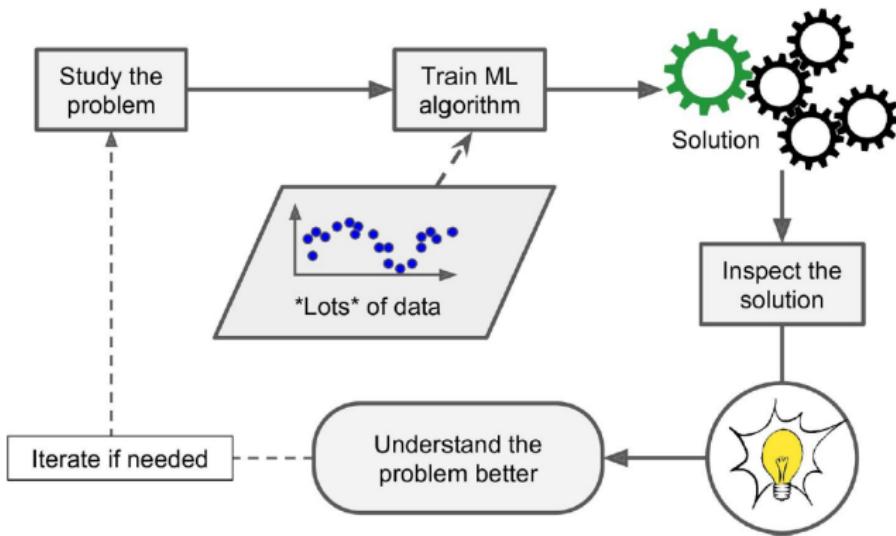


Figure: Humanos aprendiendo del aprendizaje de máquina. fuente: Hands on machine learning book.

Contenido

- 1 Todos quieren aprendizaje de máquina
- 2 Fundamentos del aprendizaje de máquina
- 3 Tipos de aprendizaje de máquina
- 4 Tipos de entrenamiento
- 5 Desafíos en aprendizaje de máquina

Con supervisión humana: clasificación

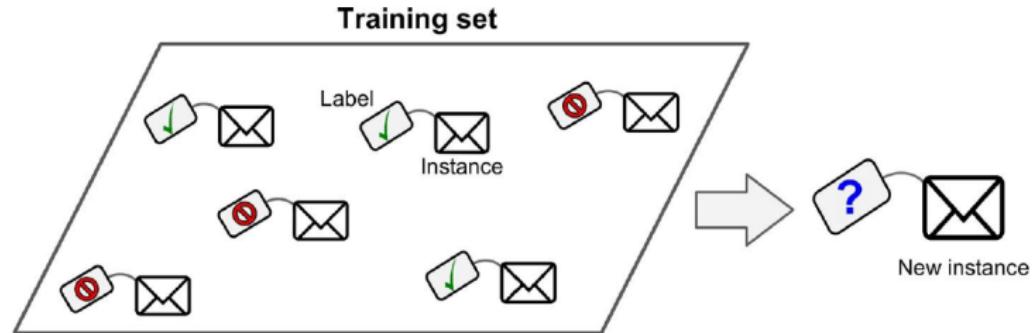
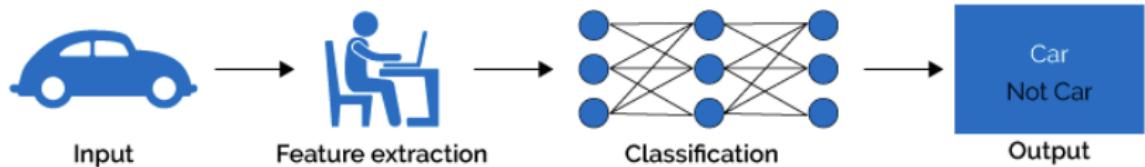


Figure: Aprendizaje supervisado en clasificación. fuente: Hands on machine learning.

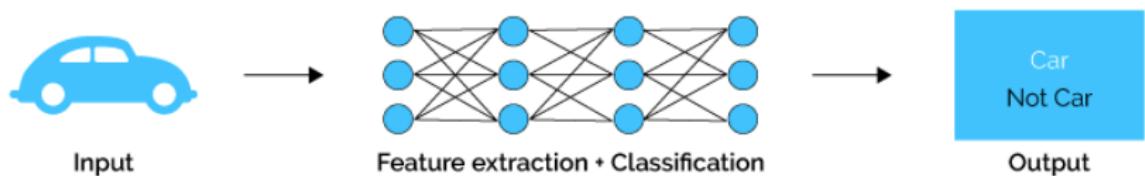
- Instancia u observación: muestra del fenómeno en estudio.
- Atributo: propiedad que codifica la instancia.
- Característica: atributo con valor (cardinal o nominal).
- Etiqueta (nominal): membresía de grupo para tareas de clasificación.

Aprendizaje clásico y aprendizaje profundo

Machine Learning



Deep Learning



Con supervisión humana: regresión

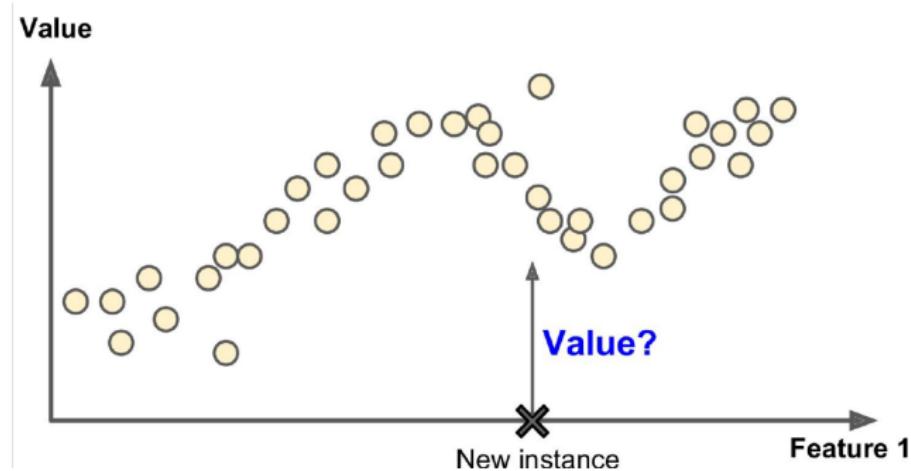


Figure: Aprendizaje supervisado en regresión. fuente: Hands on machine learning.

Se mantiene el mismo concepto que en clasificación, cambiando el tipo de variable etiqueta por variable continua.

Sin supervisión humana: agrupamiento

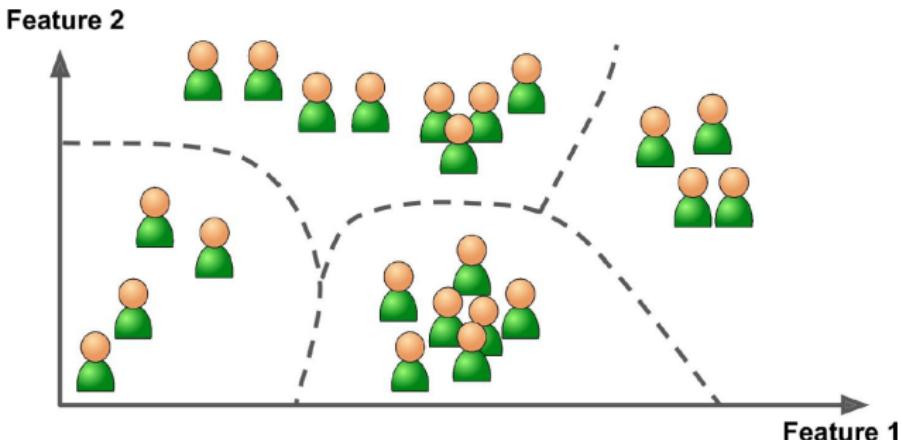


Figure: Aprendizaje no supervisado - agrupamiento (conglomerados). fuente: Hands on machine learning.

Se buscan grupos a partir de las relaciones entre las instancias (regularidades entre datos).

Sin supervisión humana: reducción de dimensión

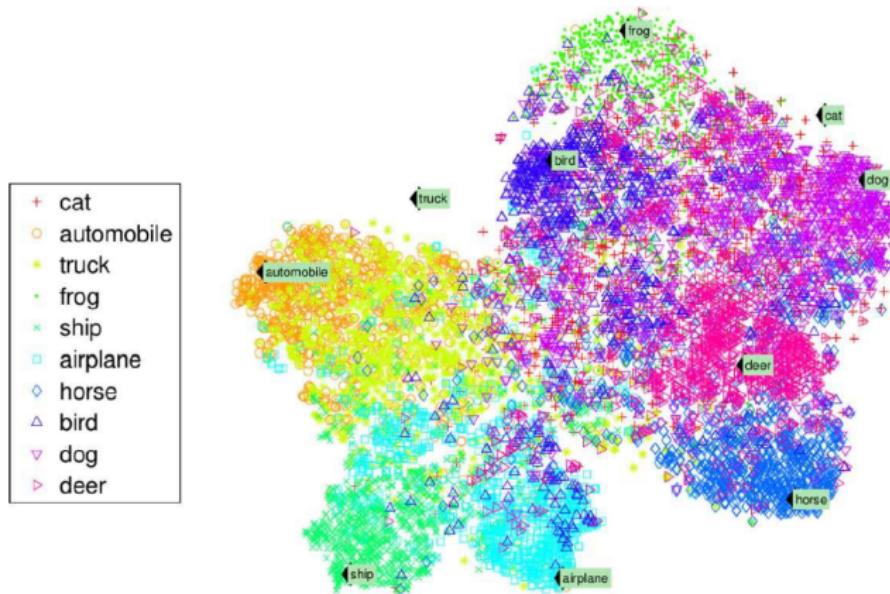


Figure: Aprendizaje no supervisado - reducción de dimensión (visualización de datos). fuente: Hands on machine learning.

Se buscan preservar relaciones de alta dimensión (espacio original de instancias) en un espacio de baja dimensión.

Sin supervisión humana: detección de anómalos



Figure: Aprendizaje no supervisado - detección de anómalos. fuente: Hands on machine learning.

La nueva instancia sigue las regularidades encontradas en el espacio de entrenamiento?

Semi supervisado

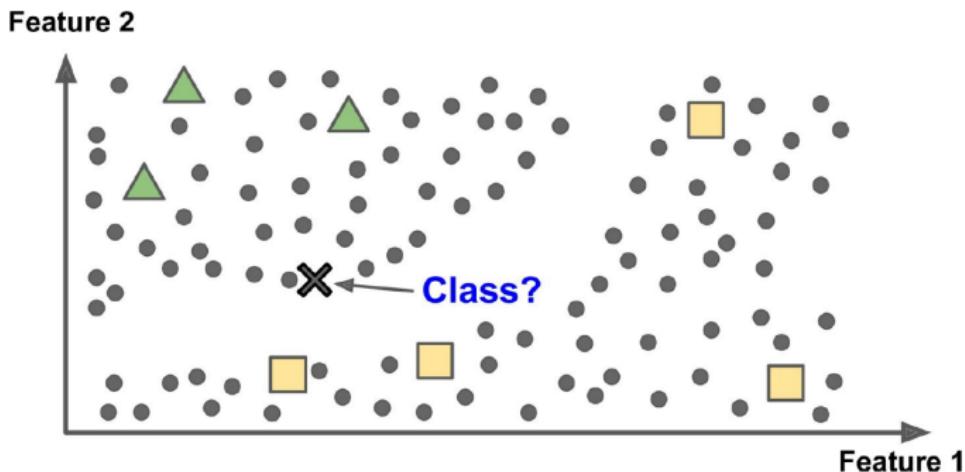
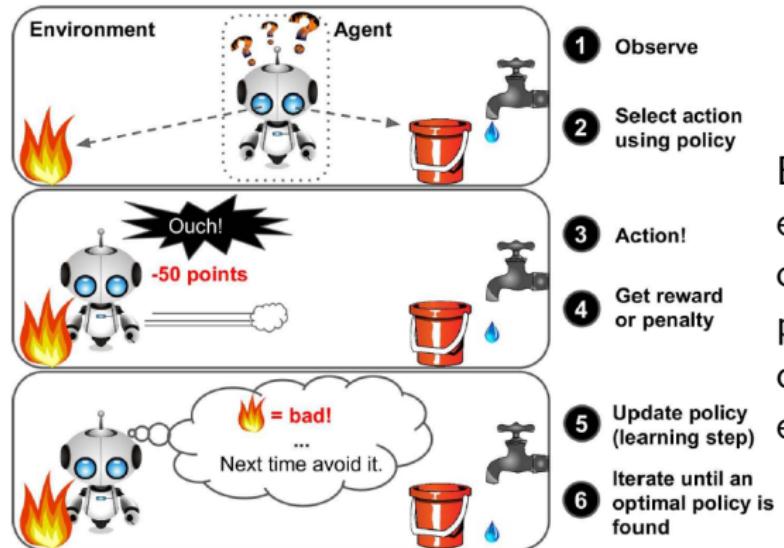


Figure: Aprendizaje semi supervisado. fuente: Hands on machine learning.

Algunas instancias poseen etiqueta (con supervisión humana) pero la mayoría no (sin supervisión humana).

Aprendizaje por refuerzo



El sistema (agente), observa el ambiente y toma decisiones obteniendo recompensas o penalizaciones, a partir de las cuales debe tomar la mejor estrategia (política).

Figure: Aprendizaje por refuerzo. fuente: Hands on machine learning.

Contenido

- 1 Todos quieren aprendizaje de máquina
- 2 Fundamentos del aprendizaje de máquina
- 3 Tipos de aprendizaje de máquina
- 4 Tipos de entrenamiento
- 5 Desafíos en aprendizaje de máquina

Entrenamiento por lote (batch learning)

- También conocido como aprendizaje offline.
- El sistema no se puede entrenar de forma incremental (se utilizan todos los datos disponibles).
- Se entrena una vez y se empieza a correr (predecir) con lo aprendido.
- Si se quiere aprovechar instancias nuevas se debe entrenar todo el sistema completa (datos viejos y datos nuevos).

Aprendizaje en línea

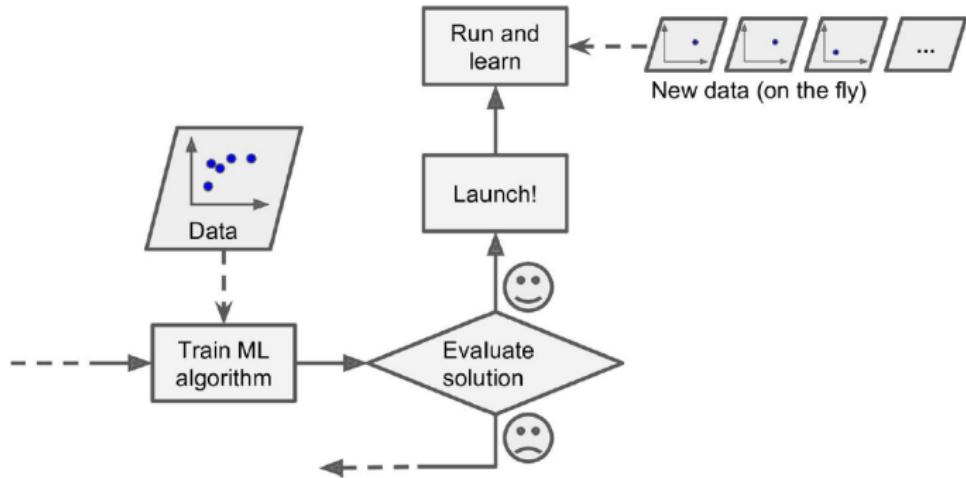


Figure: Aprendizaje en línea. fuente: Hands on machine learning.

El sistema se entrena de forma incremental (por muestra o por mini lotes).

Aprendizaje en línea por lotes

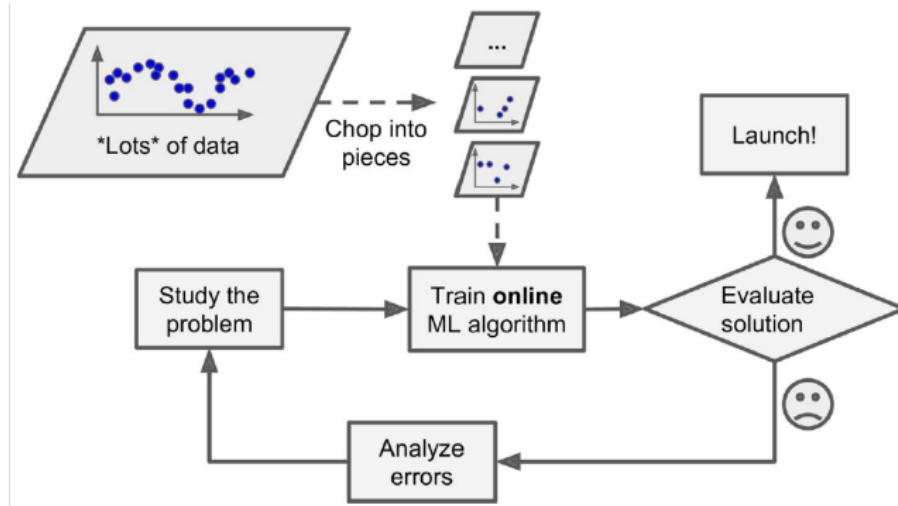


Figure: Aprendizaje en línea por lotes. fuente: Hands on machine learning.

Out-of-core: aprendizaje en línea para entrenar sobre bases de datos muy grandes.

Aprendizaje por instancias

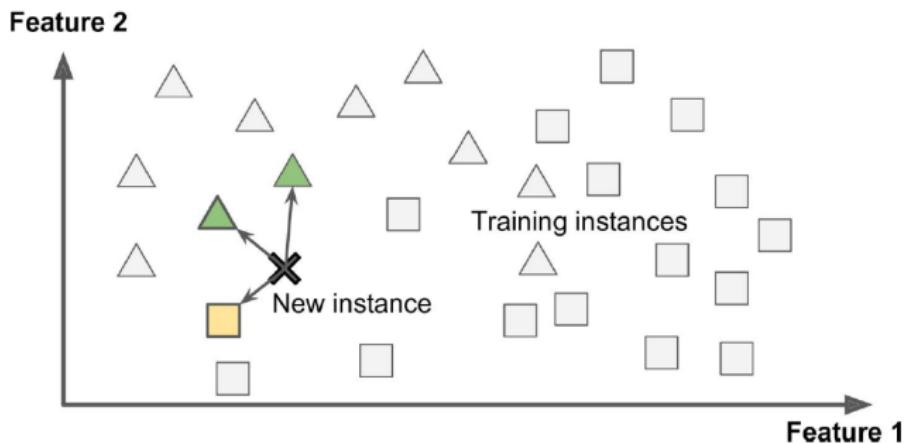


Figure: Aprendizaje por instancias. fuente: Hands on machine learning.

El sistema generaliza a partir de nociones de similitud/disimilitud entre instancias (muestras).

Aprendizaje por modelo

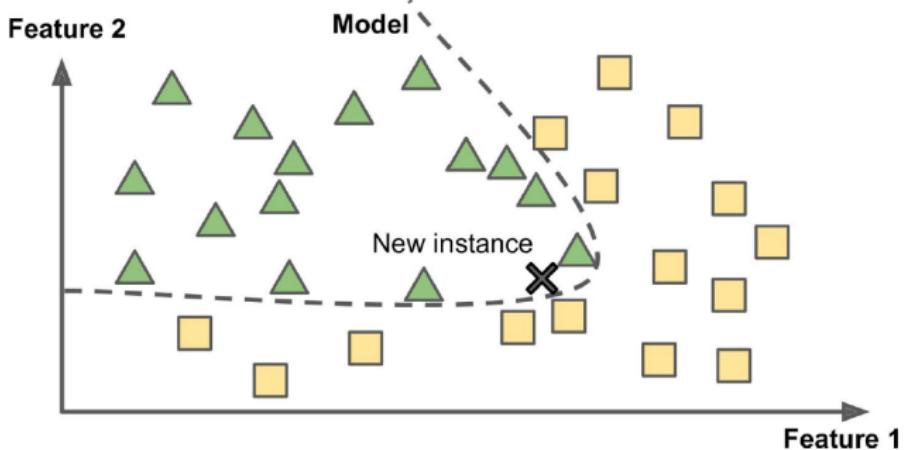


Figure: Aprendizaje por instancias. fuente: Hands on machine learning.

El sistema generaliza a partir de modelo parametrizado.

Contenido

- 1 Todos quieren aprendizaje de máquina
- 2 Fundamentos del aprendizaje de máquina
- 3 Tipos de aprendizaje de máquina
- 4 Tipos de entrenamiento
- 5 Desafíos en aprendizaje de máquina

Desafíos en aprendizaje de máquina

- Pocos datos para entrenar.
- Muestras no representativas
- Datos de poca calidad, ej: atípicos y perdidos.
- Características no relevantes.
- Overfitting y Underfitting
- Testing - Validation: estimar correctamente el error de generalización.

Testing - Validation

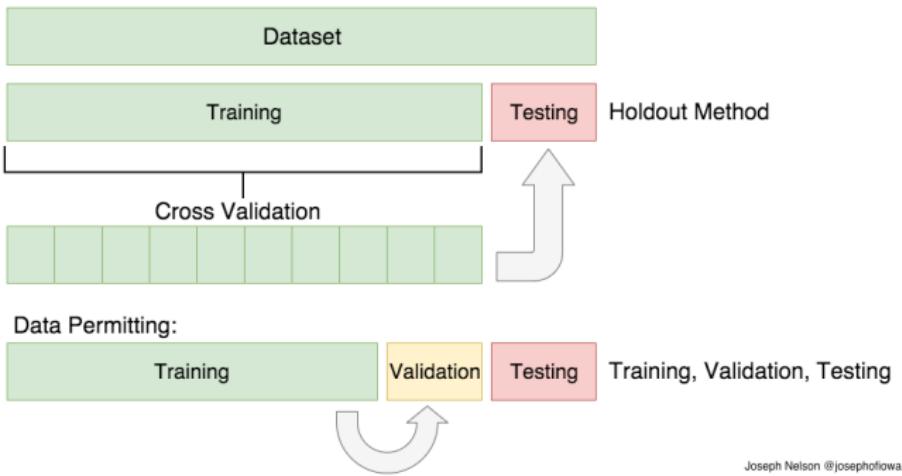


Figure: Evaluación de sistemas de aprendizaje. fuente: ²

La forma de validar es vital a la hora de generalizar el rendimiento del sistema de aprendizaje.

²<https://towardsdatascience.com/train-test-split-and-cross-validation-in-python-80b61beca4b6>

Validación anidada

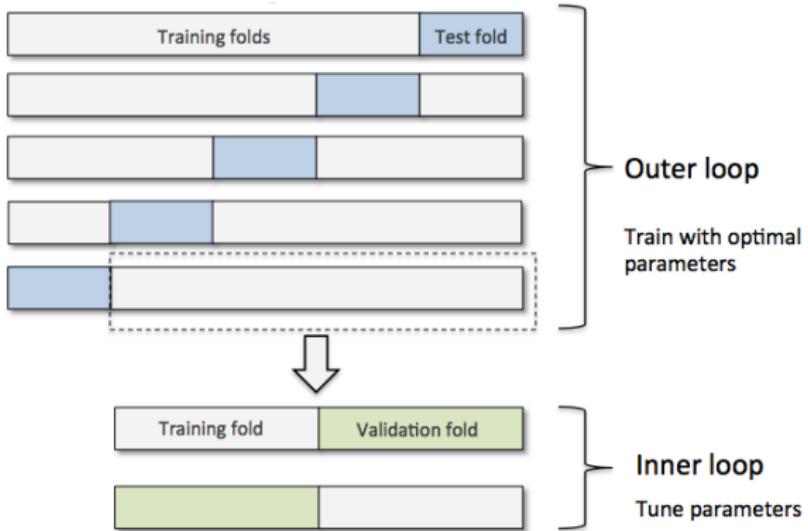


Figure: Validación anidada. fuente: ³

Si se requieren sintonizar hyperparámetros es mejor utilizar validación anidada.

³<https://sebastianraschka.com/faq/docs/evaluate-a-model.html>

Las cuatro claves del éxito reciente

- Cómputo de alto desempeño - incluso Gratis!.
- Diferenciación automática - No más derivadas difíciles de estimar.
- Optimización por mini-lotes - resuelve problemas con muchos datos de forma eficiente.
- Transferencia de conocimiento - Utiliza modelos pre-entrenados sobre grandes bases de datos, solo reajusta en tu problema.

Tip: Escoge/construye un buen modelo, utiliza/define una buena función de pérdida y aprovecha la computación en grafos mediante tensores (Tensorflow/Pytorch/MXNet).