



Programa de formación MACHINE LEARNING AND DATA SCIENCE MLDS

Facultad de
INGENIERÍA





Módulo 2

Introducción al Machine

Learning con Python

Unidad 1

Introducción al Aprendizaje Computacional y
Scikit-learn

Clase sincrónica

Facultad de
INGENIERÍA



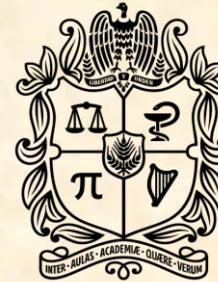


Bienvenida

Fabio Augusto Gonzalez, PhD.

<https://dis.unal.edu.co/~fgonza/>

fagonzalezo@unal.edu.co



UNIVERSIDAD
NACIONAL
DE COLOMBIA

Departamento de Ingeniería de Sistemas e Industrial

Facultad de Ingeniería

Universidad Nacional de Colombia

Sede Bogotá



Tabla de contenidos

1

Introducción al Aprendizaje Computacional

2

Scikit-learn

3

Taller

Objetivos de aprendizaje



Unidad 1 - Introducción al aprendizaje computacional y scikit-learn

Al finalizar la unidad usted deberá ser capaz de:



1

Describir de manera precisa los elementos fundamentales de la solución de problemas mediante el uso del aprendizaje computacional, teniendo en cuenta los diferentes tipos de problemas y los métodos de solución.



2

Conocer la arquitectura general de scikit-learn y su empleo para entrenar y probar modelos de Machine Learning.

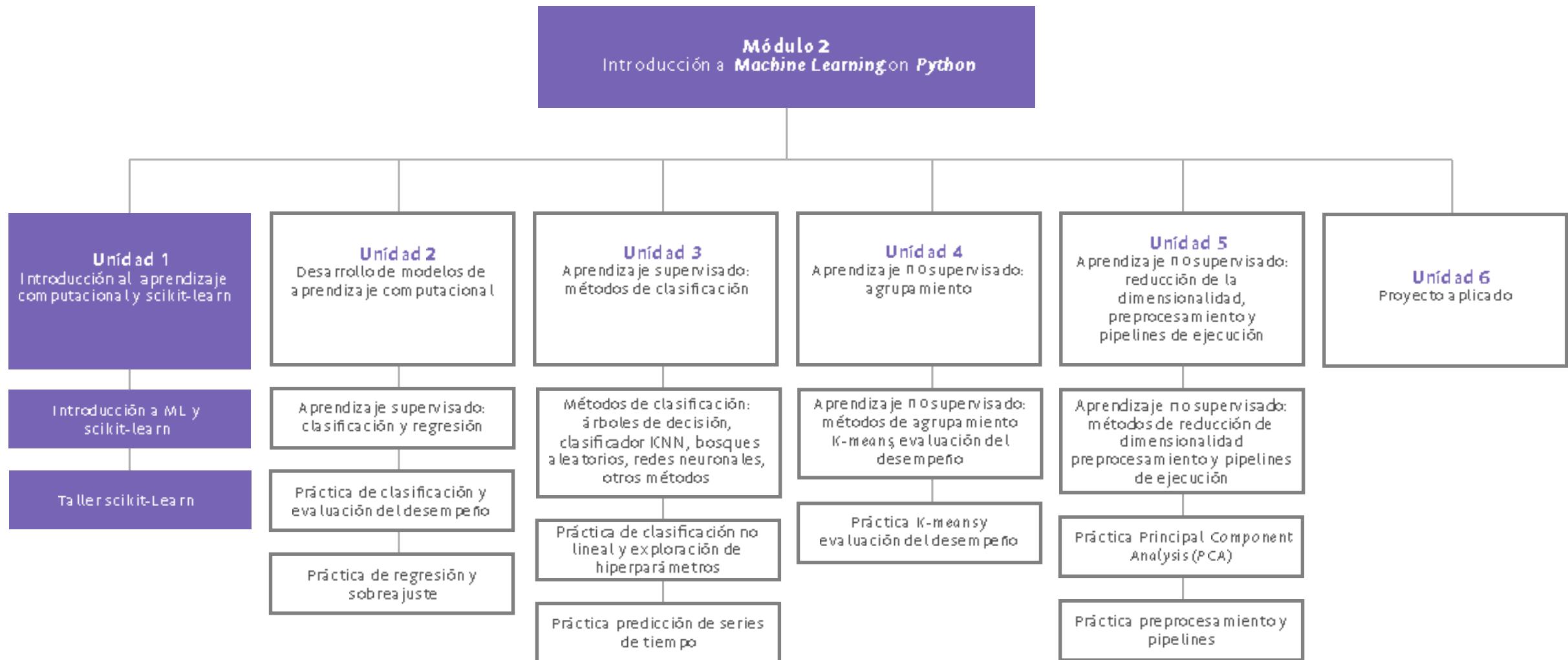


3

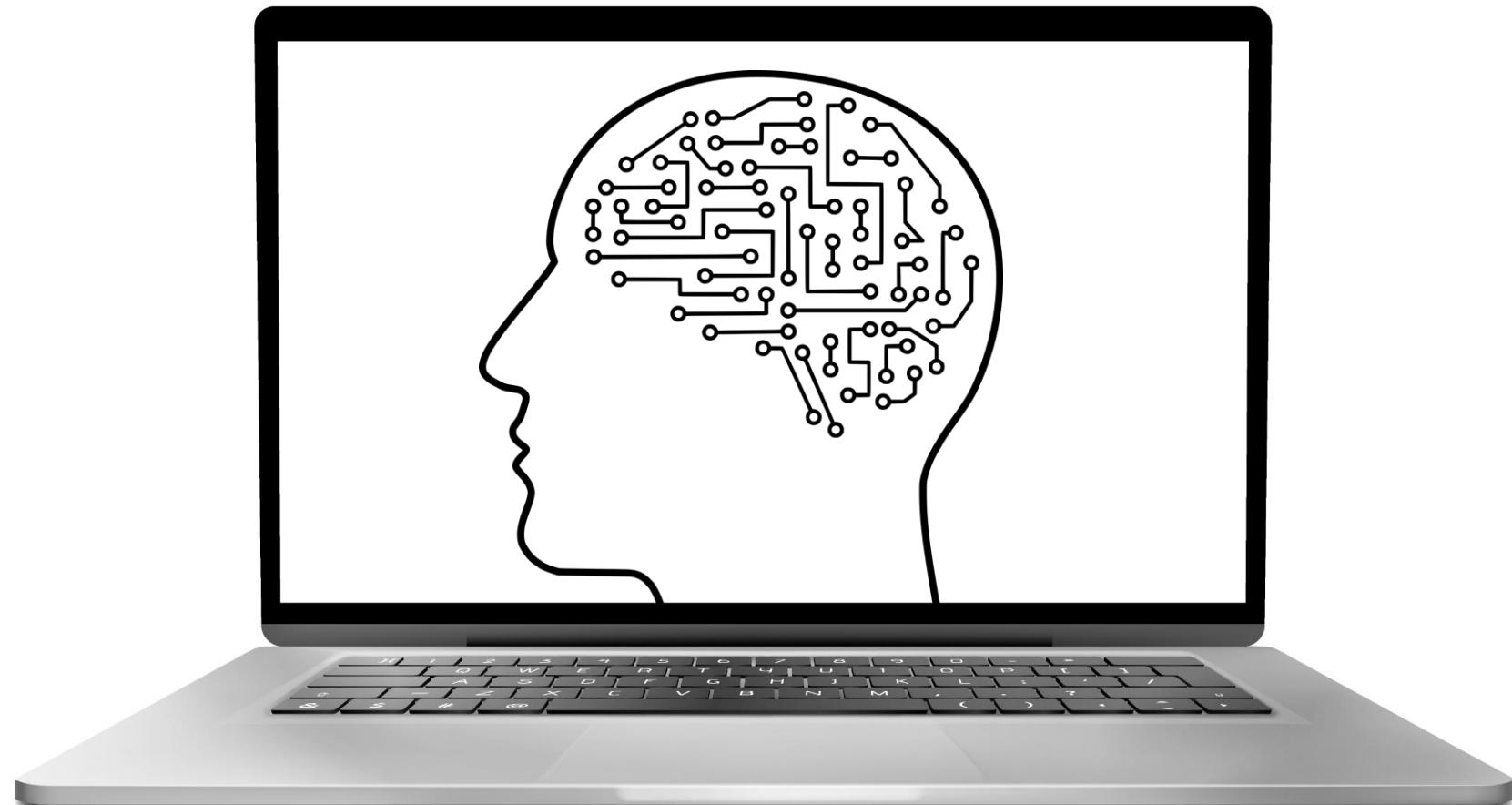
Crear, dentro de programas de Python, instancias de algoritmos de aprendizaje provistos por la librería scikit-learn, entrenar los modelos y aplicarlos sobre nuevos datos.



Mapa de contenidos de la unidad

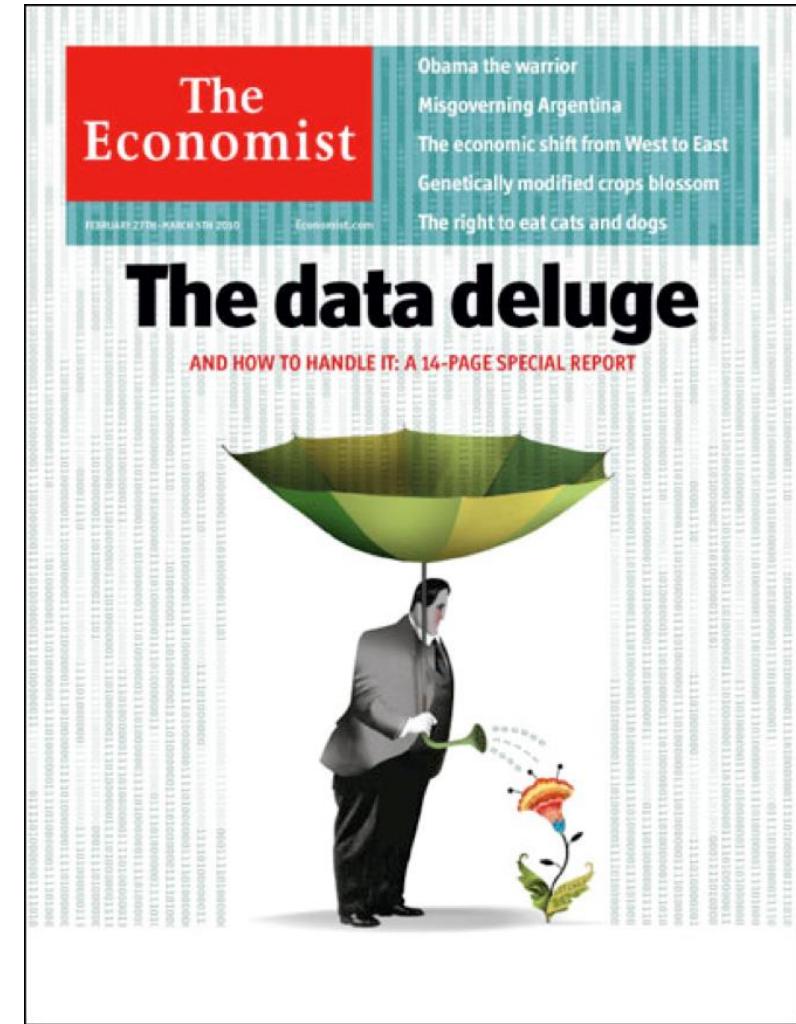


Introducción al Aprendizaje Computacional



Introducción al Aprendizaje Computacional

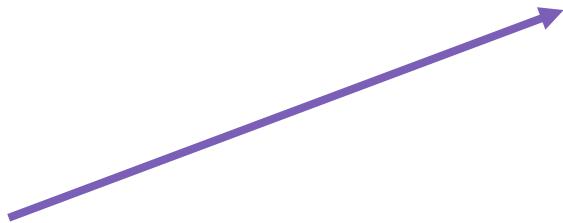
Big Data



Introducción al Aprendizaje Computacional

 Retos

- Almacenar
- Procesar
- Acceder
- Entender
- Seguridad y privacidad

- 
- Analizar
 - Cuantificar
 - Comparar
 - Visualizar
 - Predecir

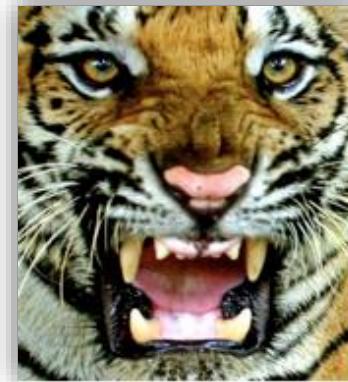
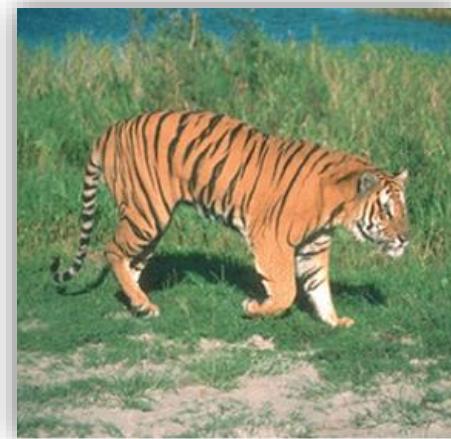
Introducción al Aprendizaje Computacional

/ ¿Qué queremos de los datos?



Introducción al Aprendizaje Computacional

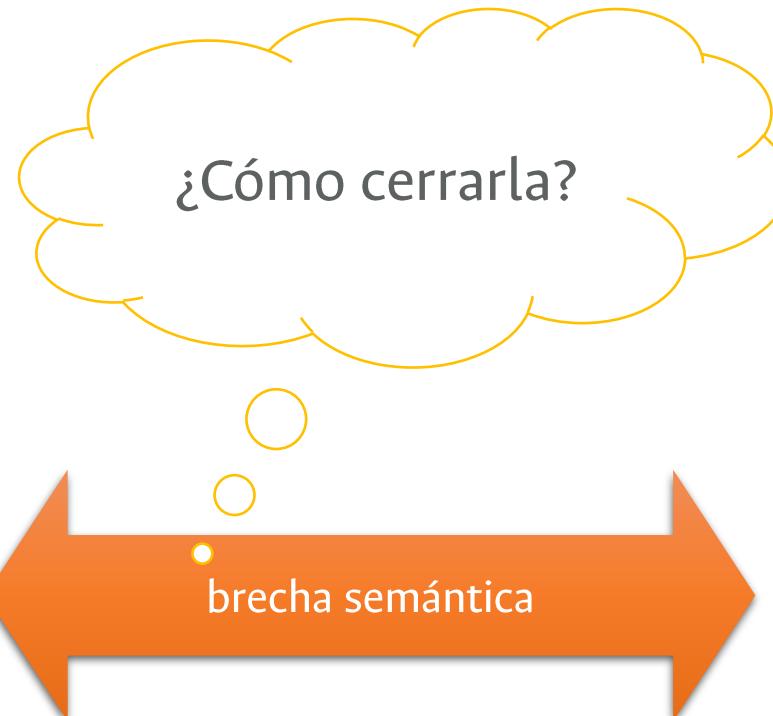
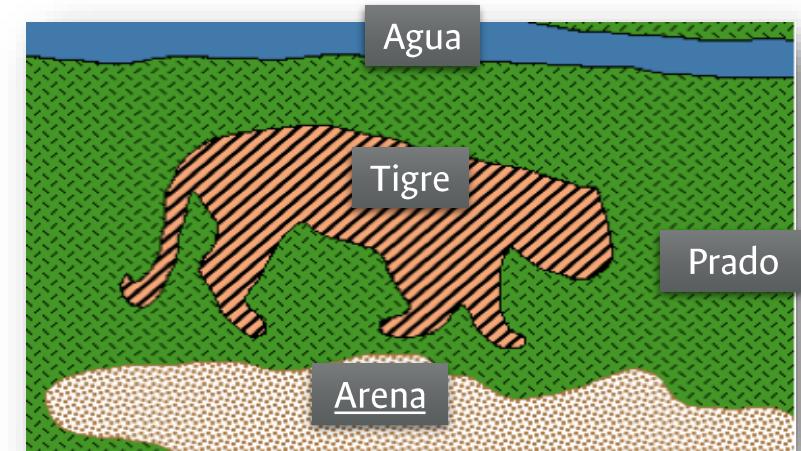
Similitud
semántica



Similitud
visual



Introducción al Aprendizaje Computacional

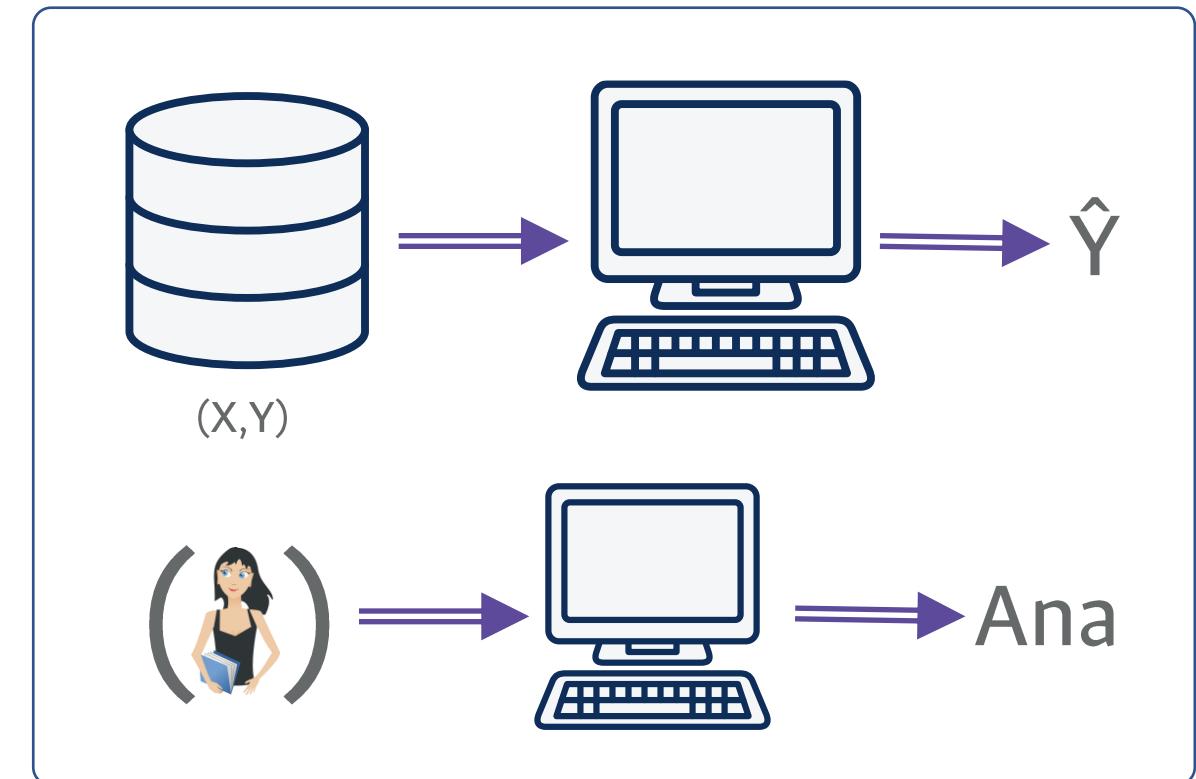
Apariencia
visualContenido
semántico

Introducción al Aprendizaje Computacional

Aprendizaje Computacional

El **aprendizaje computacional** permite abordar tareas complejas que son difíciles de resolver usando otro tipo de técnicas; entre estas tareas, por ejemplo, se presentan las siguientes:

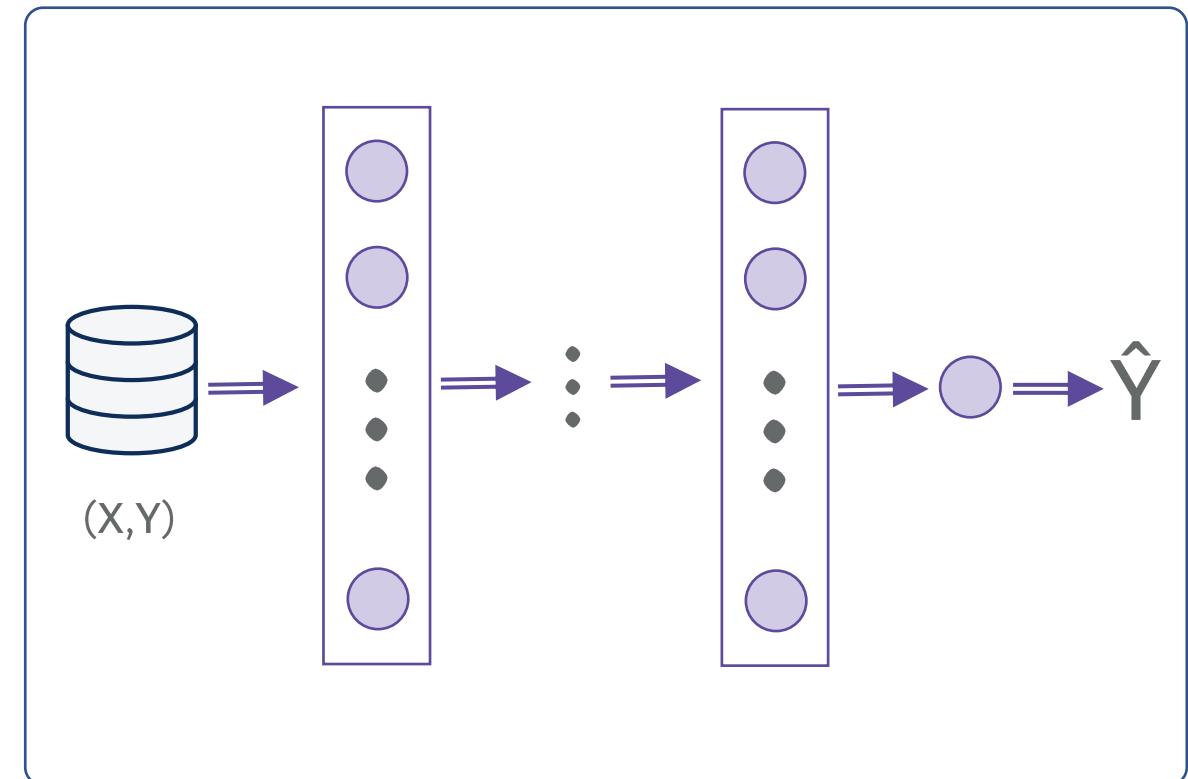
- Reconocimiento de objetos en imágenes
- Entendimiento del habla
- Detección de spam
- Traducción automática



Introducción al Aprendizaje Computacional

Aprendizaje Computacional

Aprendizaje computacional se refiere a la programación de computadoras con el objetivo de optimizar un cierto criterio de desempeño. Para ello, se define un modelo junto con sus parámetros; el aprendizaje consiste en la ejecución de un programa que optimiza estos parámetros del modelo, a través del empleo de datos o experiencia pasada.



Introducción al Aprendizaje Computacional

Tipos de Aprendizaje Computacional

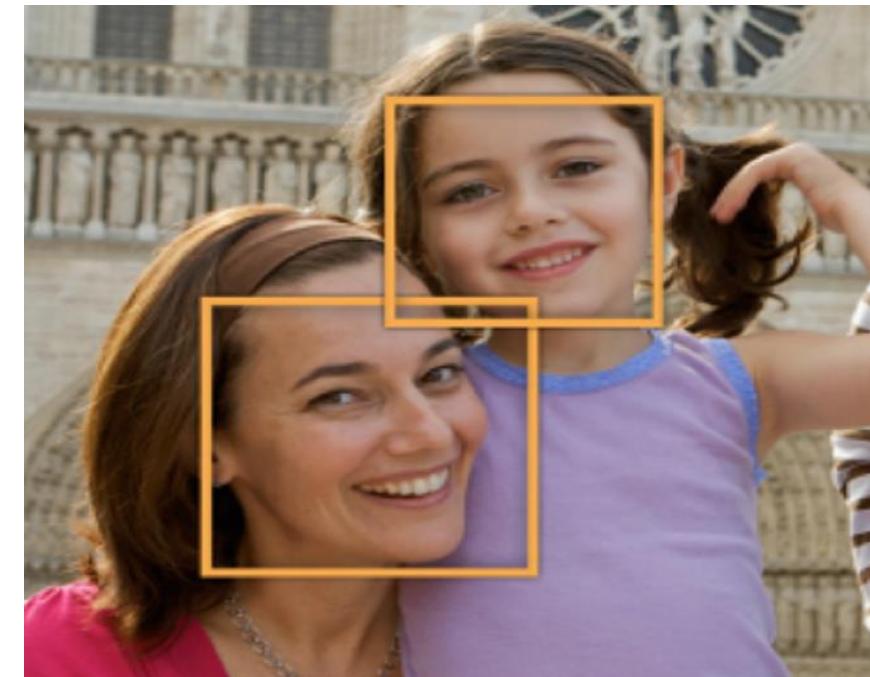
Aprendizaje supervisado

Un instructor o experto proporciona retroalimentación a los modelos, a través de datos especiales conocidos como **etiquetas u objetivos**.

Problema fundamental: encontrar una función que relacione las entradas con las salidas

Tareas típicas:

- Clasificación
- Regresión



Introducción al Aprendizaje Computacional

Tipos de Aprendizaje Computacional

Aprendizaje no supervisado

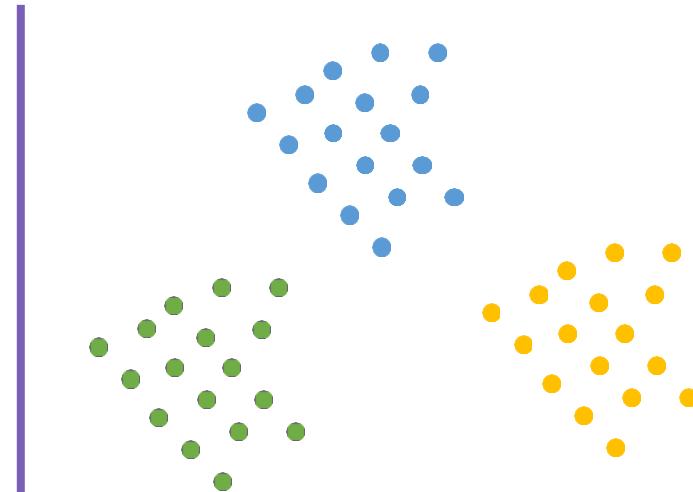
Un algoritmo de aprendizaje no supervisado aprende de un conjunto de datos sin ninguna etiqueta.

Problema fundamental: encontrar la estructura subyacente del conjunto de datos.

Tareas típicas:

- Agrupamiento
- Reducción de la dimensionalidad

K-means

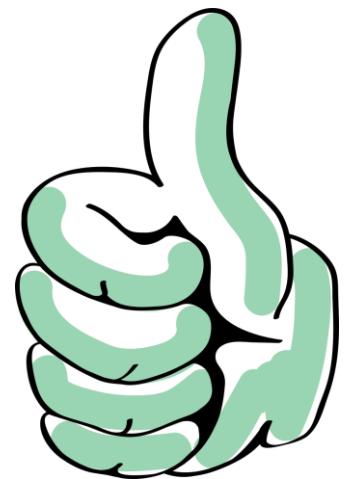


Introducción al Aprendizaje Computacional

Tipos de Aprendizaje Computacional

- Aprendizaje semi-supervisado
- Aprendizaje por refuerzo
- Aprendizaje en línea

Aprendizaje por refuerzo



Introducción al Aprendizaje Computacional

Ejemplos de tareas de aprendizaje

Problema de aprendizaje computacional	Descripción	Ejemplo	Tipo
Clasificación	Predecir la clase a la que pertenece un ejemplo.	Decidir si una imagen contiene un perro o un gato.	Supervisado
Regresión	Predecir un valor numérico en función de un ejemplo.	Predecir el precio de un inmueble.	Supervisado
Agrupamiento	Agrupar ejemplos similares.	Recomendación de películas.	No supervisado
Reducción de dimensionalidad	Reducir la dimensión de los ejemplos de un conjunto de datos.	Convertir muchísimas características sin poder predictivo en pocas características con alto poder predictivo.	No supervisado

Introducción al Aprendizaje Computacional



Introducción al Aprendizaje Computacional

Proceso del aprendizaje computacional

Proceso del aprendizaje computacional



Introducción al Aprendizaje Computacional

Proceso del aprendizaje computacional

Paso**01****Definir el problema****Entender el problema**

Debe enfocarse en su problema, es decir, identificar en qué consiste y establecer los objetivos que pretende alcanzar de manera clara, con lo cual podrá decidir si es apropiado usar el aprendizaje computacional, de tal forma que pueda establecer los componentes de este sistema.

Definir objetivos claros

Si hace un mal planteamiento de los objetivos, afectará el diseño del aprendizaje computacional, lo que le puede generar efectos colaterales involuntarios.

Definir el problema de aprendizaje computacional

¿Cuál es el tipo de problema que debe solucionar? Puede ser de clasificación, regresión o, tal vez, ¿algo más complejo?

Introducción al Aprendizaje Computacional

Proceso del aprendizaje computacional

Paso**01****Definir el problema**

Determinar métricas de éxito y de fracaso

Diseñar la estructura de la salida

Las métricas de éxito y de fracaso deben estar relacionadas con el efecto deseado de su sistema de aprendizaje computacional. Estas se deben definir de manera independiente a la forma en la que se evalúa el desempeño de su modelo de aprendizaje computacional. No se deben expresar en términos del error o de la precisión.

¿Son sus métricas medibles?

¿Qué tipo de salida desearía que produjera su algoritmo?, ¿qué tipo de problema de aprendizaje computacional desea resolver?

¿Puede obtener ejemplos de la salida deseada fácilmente?

Introducción al Aprendizaje Computacional

Proceso del aprendizaje computacional

Paso**Recolectar
datos**Buscar conjuntos
de datos
etiquetados

Es posible que ya existan conjuntos de datos que pueda usar para resolver su problema.

Definir el origen de
los datos

Debe verificar la calidad de los datos y validar que sean características obtenibles en el tiempo de producción.

En caso de que no existan datos previos, usted debe construir su propio conjunto de datos; no tenga miedo de desplegar una solución sin aprendizaje computacional, puede tratar de pensar cómo podría aproximar su problema; por ejemplo, usar heurísticas que le permitan interactuar con sus usuarios y construir un conjunto de datos, para, después, incluirlos en aprendizaje computacional.

Introducción al Aprendizaje Computacional

Proceso del aprendizaje computacional

Paso**03****Preparar los datos****Análisis exploratorio de datos**

En el análisis, debe buscar posibles redundancias de las características o correlaciones entre las características y la etiqueta. También, es posible usar técnicas de aprendizaje no supervisado para explorar los datos (este tema es tratado con profundidad en el módulo anterior).

Selección de características

Para empezar, debe escoger pocas características y luego, incluir algunas adicionales; primero debe centrarse en aquellas con mayor poder predictivo.

Ingeniería de características

La ingeniería de características o *feature engineering* es el conjunto de prácticas para crear nuevas características a partir de las ya seleccionadas. Pueden ser creadas de forma manual o por algoritmos de aprendizaje computacional avanzados, con el fin de automatizar el aprendizaje de características

Introducción al Aprendizaje Computacional

Proceso del aprendizaje computacional

Paso

04

Entrenar
modelos



Introducción al Aprendizaje Computacional

Proceso del aprendizaje computacional

Paso**04****Entrenar
modelos****Definir el protocolo
de evaluación****¿Qué métrica de desempeño usará para evaluar su modelo?**

Es probable que usted se encuentre familiarizado con métricas como la precisión, la sensibilidad o la especificidad. Sin embargo, uno de los protocolos más comunes se conoce como validación cruzada, concepto que se explicará con mayor profundidad en unidades siguientes.

Escoger un modelo

La selección de un modelo debe estar justificada de acuerdo con la salida esperada en el paso uno. Ciertos modelos se desempeñan mejor que otros en una tarea en específico, por lo tanto, usted no debe preocuparse por utilizar un modelo (sencillo o complejo) para comenzar, ya que es posible que alguno de estos modelos sencillos de aprendizaje computacional cumpla con sus necesidades.

Introducción al Aprendizaje Computacional

Proceso del aprendizaje computacional

Paso



Entrenar
modelos

Afinación del modelo

Un modelo se define junto con sus parámetros. Algunos de estos son conocidos como *hiperparámetros*, ya que el algoritmo de aprendizaje no los optimiza. En este caso, para afinarlos, usted debe controlar los hiperparámetros de manera manual.

Introducción al Aprendizaje Computacional

Proceso del aprendizaje computacional

Paso**05****Evaluar
modelos**

Interpretación de los resultados

Comparar el desempeño con el estado del arte

Usted debe cuestionarse sobre el desempeño de su modelo:
¿Por qué este podría estar fallando?
¿Cómo se pueden explicar los resultados obtenidos?
Por lo tanto, es fundamental entender cada uno de los pasos del proceso.

Puede ser que su modelo funcione para algún caso de uso o podría concluir que es mejor utilizar el mejor modelo del estado del arte.

Introducción al Aprendizaje Computacional

Proceso del aprendizaje computacional

Paso**Hacer
predicciones
con el modelo**

Luego de entrenar un modelo en el aprendizaje computacional, llega el momento de desplegarlo, lo que significa que esté disponible para las predicciones bajo demanda.

Usted puede desplegar su modelo en su propia computadora, incluirlo como un componente de un complicado sistema de inteligencia artificial, publicarlo como una Interfaz de Programación de Aplicaciones (API) o incluirlo en una aplicación inteligente para plataformas web o móviles.

Actualmente, existen multitud de plataformas y librerías que permiten desplegar modelos de aprendizaje computacional de manera fácil.

Introducción al Aprendizaje Computacional

Plataformas



Google Cloud

Paperspace

colab

FLOYDHUB

PyTorch



Keras

TensorFlow



Introducción al Aprendizaje Computacional

Ejemplo clasificación de emails



Introducción al Aprendizaje Computacional / Aplicaciones

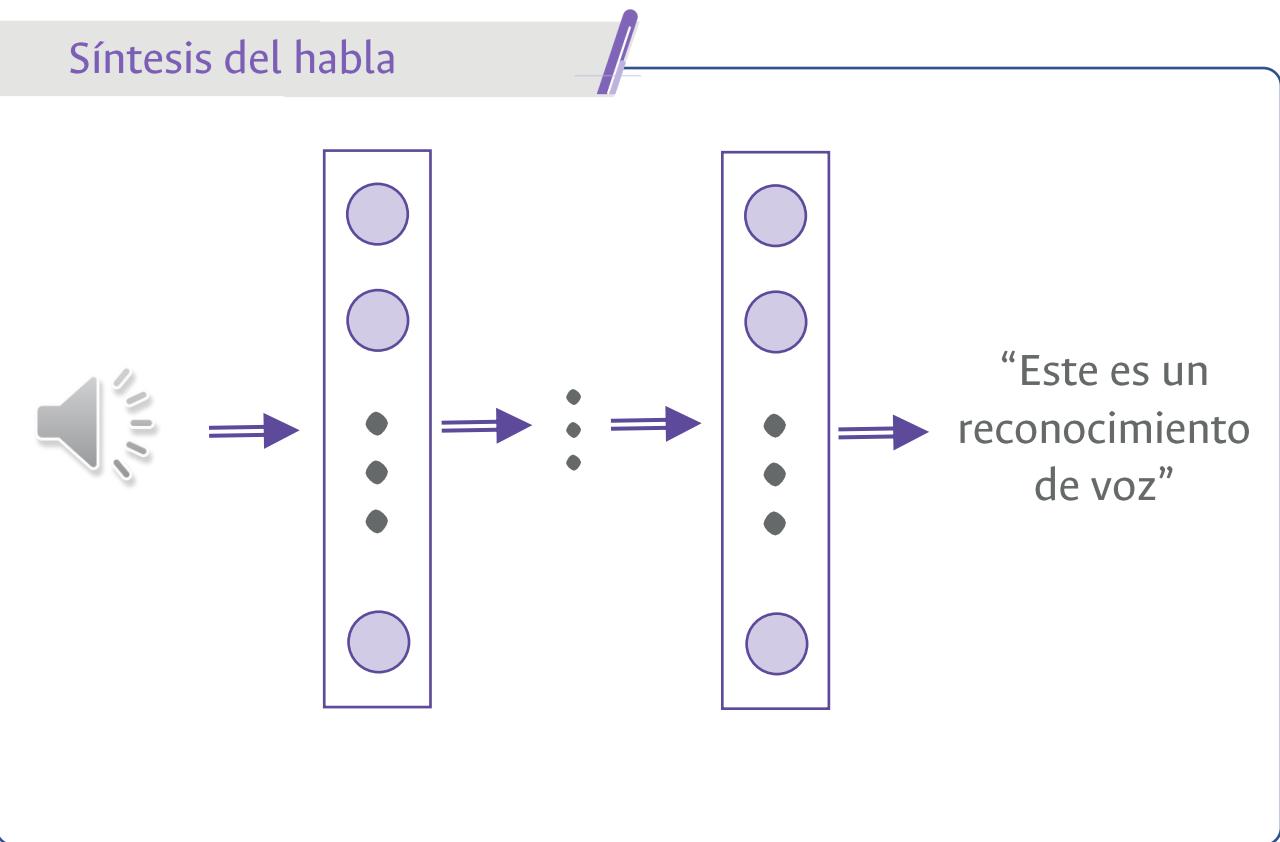


Introducción al Aprendizaje Computacional

Aplicaciones

- Visión por computador
- Reconocimiento del habla
- Síntesis de voz
- Vehículos autónomos
- Análisis de sentimientos
- Generación de texto
- Predicción de series de tiempo
- Identificación de correo no deseado

Síntesis del habla



Introducción al Aprendizaje Computacional

Deep learning

Scientists discover powerful antibiotic using AI

⌚ 21 February 2020

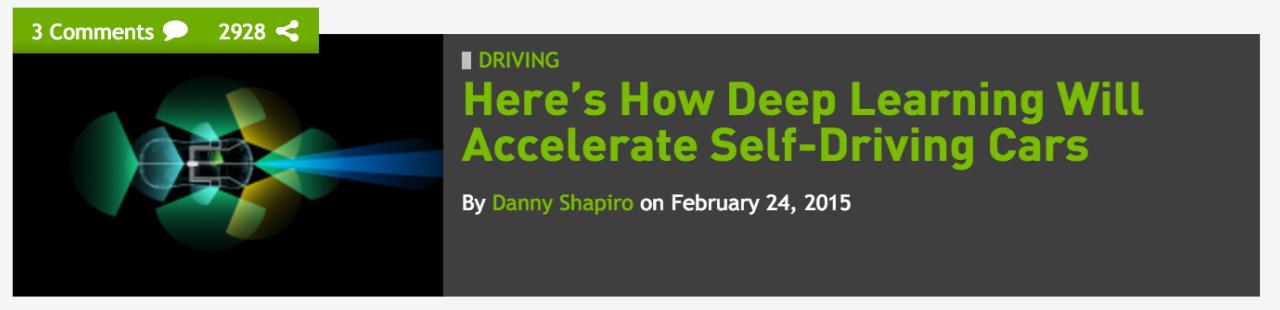
f 📲 🐦 📧 Share



Artificial intelligence diagnoses lung cancer

By James Gallagher
Health and science correspondent, BBC News

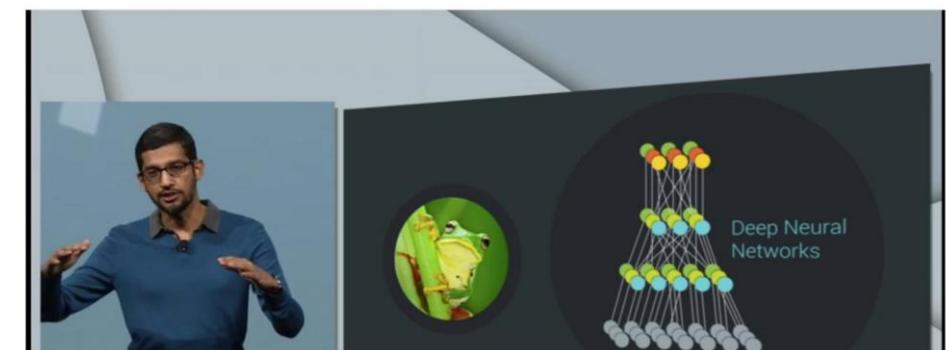
⌚ 20 May 2019



Google says its speech recognition technology now has only an 8% word error rate

JORDAN NOVET MAY 28, 2015 10:40 AM

TAGS: ARTIFICIAL INTELLIGENCE, DEEP LEARNING, GOOGLE, GOOGLE I/O 2015, SUNDAR PICHAI



Introducción al Aprendizaje Computacional

Ejemplo: Aplicaciones en Medicina

Artificial Intelligence in Medicine 64 (2015) 131–145

Contents lists available at ScienceDirect

Artificial Intelligence in Medicine

journal homepage: www.elsevier.com/locate/aim

CrossMark

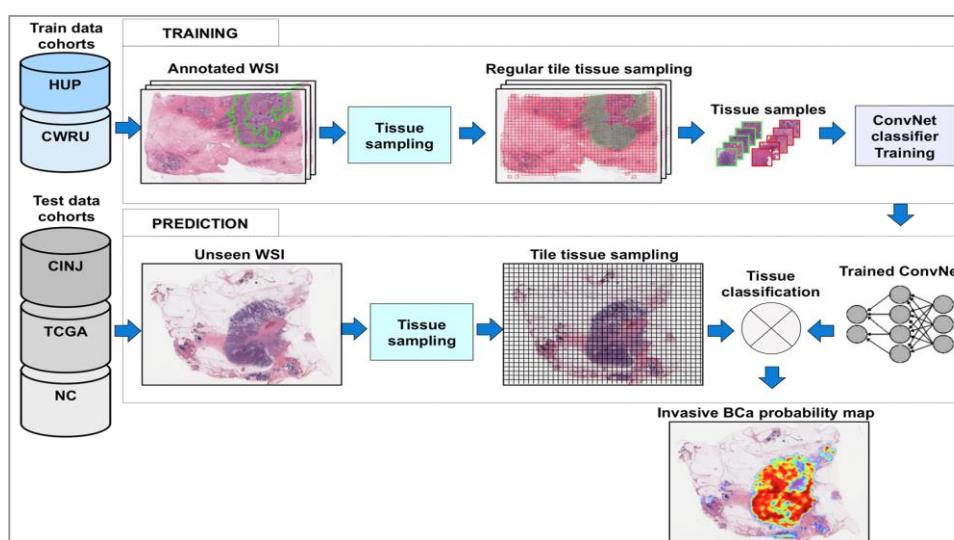
An unsupervised feature learning framework for basal cell carcinoma image analysis

John Arevalo ^a, Angel Cruz-Roa ^a, Viviana Arias ^b, Eduardo Romero ^c, Fabio A. González ^{a,*}

^a Machine Learning, Perception and Discovery Lab, Systems and Computer Engineering Department, Universidad Nacional de Colombia, Faculty of Engineering, Cra 30 No 45 03-Ciudad Universitaria, Building 453 Office 114, Bogotá DC, Colombia

^b Pathology Department, Universidad Nacional de Colombia, Faculty of Medicine, Cra 30 No 45 03-Ciudad Universitaria, Bogotá DC, Colombia

^c Computer Imaging & Medical Applications Laboratory, Universidad Nacional de Colombia, Faculty of Medicine, Cra 30 No 45 03-Ciudad Universitaria, Bogotá DC, Colombia



(12) **United States Patent**
Madabhushi et al.

(54) **AUTOMATIC DETECTION OF MITOSIS USING HANDCRAFTED AND CONVOLUTIONAL NEURAL NETWORK FEATURES**

(71) Applicant: **Case Western Reserve University**, Cleveland, OH (US)

(72) Inventors: **Anant Madabhushi**, Beachwood, OH (US); **Haibo Wang**, Cleveland Heights, OH (US); **Angel Cruz-Roa**, Bogota (CO); **Fabio Gonzalez**, Bogota (CO)

(73) Assignee: **Case Western Reserve University**, Cleveland, OH (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(10) **Patent No.:** US 9,430,829 B2
 (45) **Date of Patent:** Aug. 30, 2016

(58) **Field of Classification Search**
 None
 See application file for complete search history.

(56) **References Cited**
 U.S. PATENT DOCUMENTS

2007/0112701 A1*	5/2007	Chellapilla	G06K 9/6256
			706/15
2007/0140556 A1*	6/2007	Willamowski	G06K 9/0061
			382/167
2010/0002920 A1*	1/2010	Cosatto	G06K 9/00147
			382/128
2010/0172568 A1*	7/2010	Malon	G06K 9/00147
			382/133
2014/0139625 A1*	5/2014	Mathuis	G03H 1/0005
			348/40
2014/0314292 A1*	10/2014	Kamen	A61B 6/463
			382/131

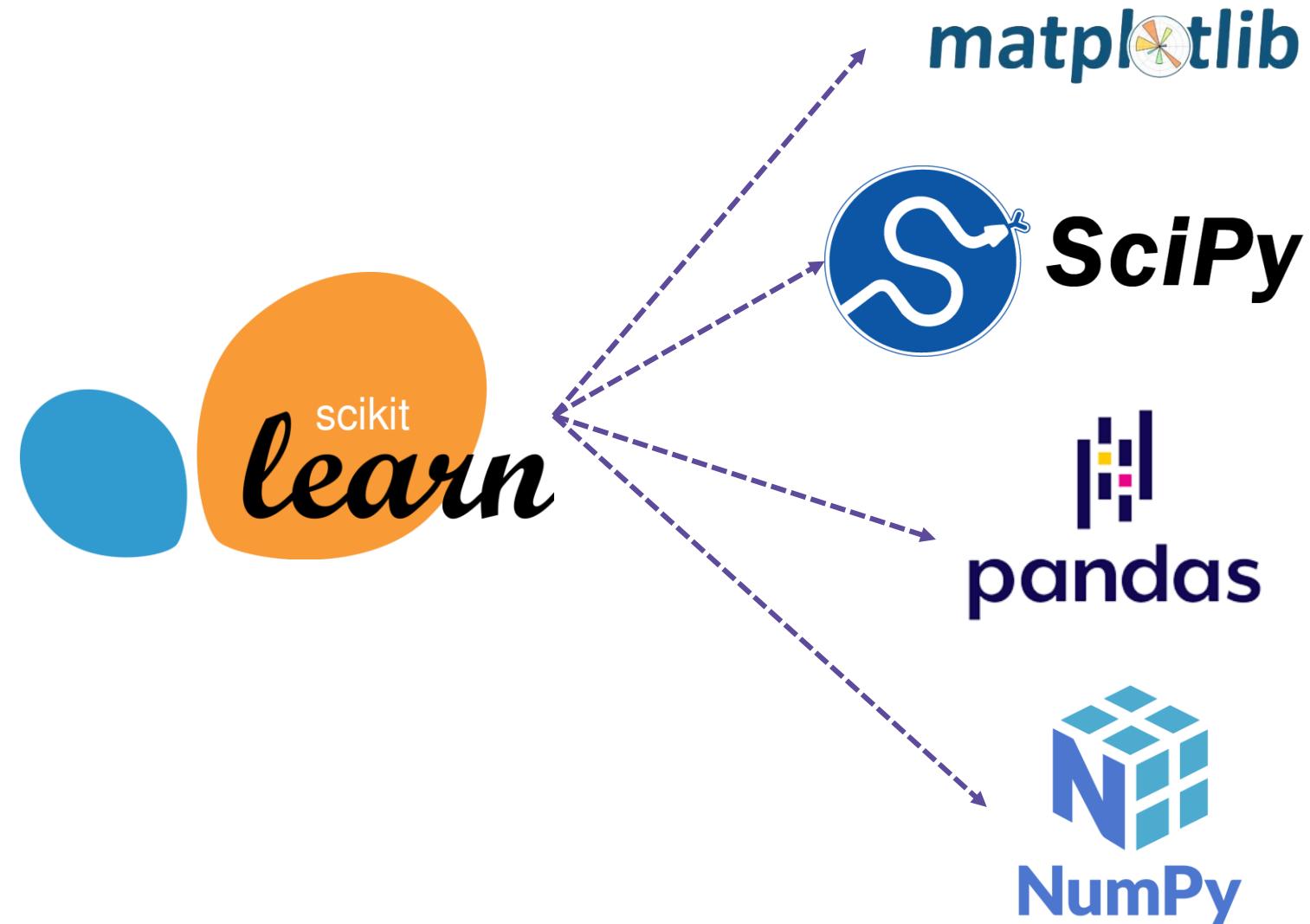
* cited by examiner

Scikit Learn



Scikit Learn Introducción

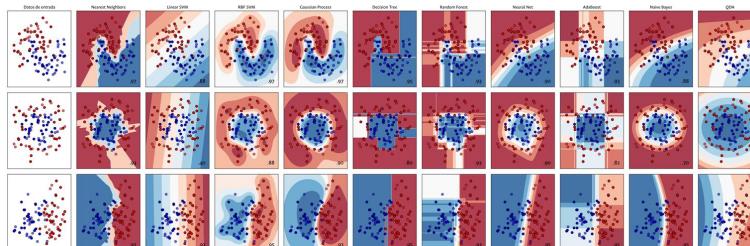
- Librería en Python que provee un conjunto de herramientas para minería y análisis de datos.
- Código abierto que puede ser usable comercialmente.
- Accesible a todos y reutilizable en muchos contextos.
- Construido sobre NumPy, SciPy y matplotlib.



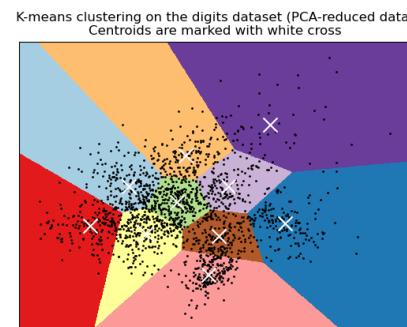
Scikit Learn Introducción

Scikit-learn cuenta con 6 módulos principales para el aprendizaje computacional

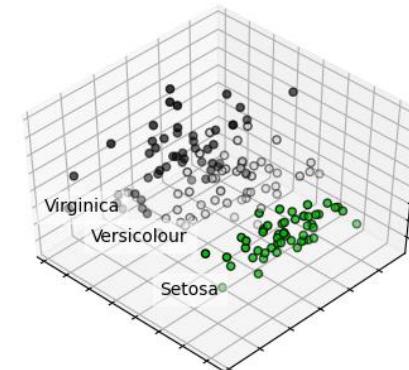
Clasificación



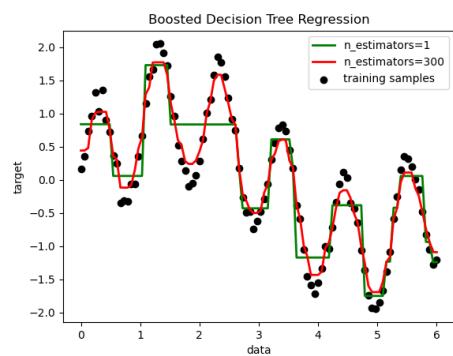
Agrupamiento



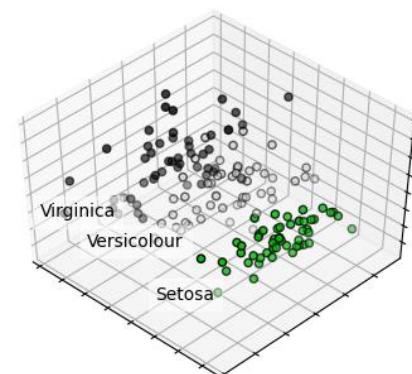
Selección de modelos



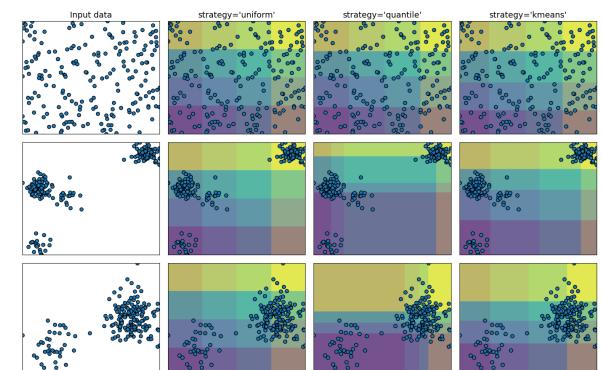
Regresión



Reducción de la dimensionalidad



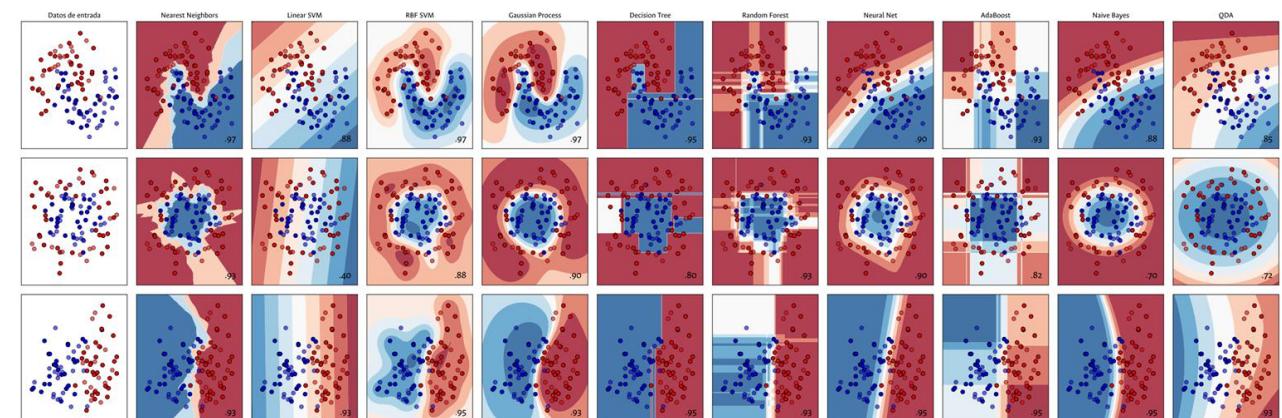
Preprocesamiento



Scikit Learn

 Clasificación

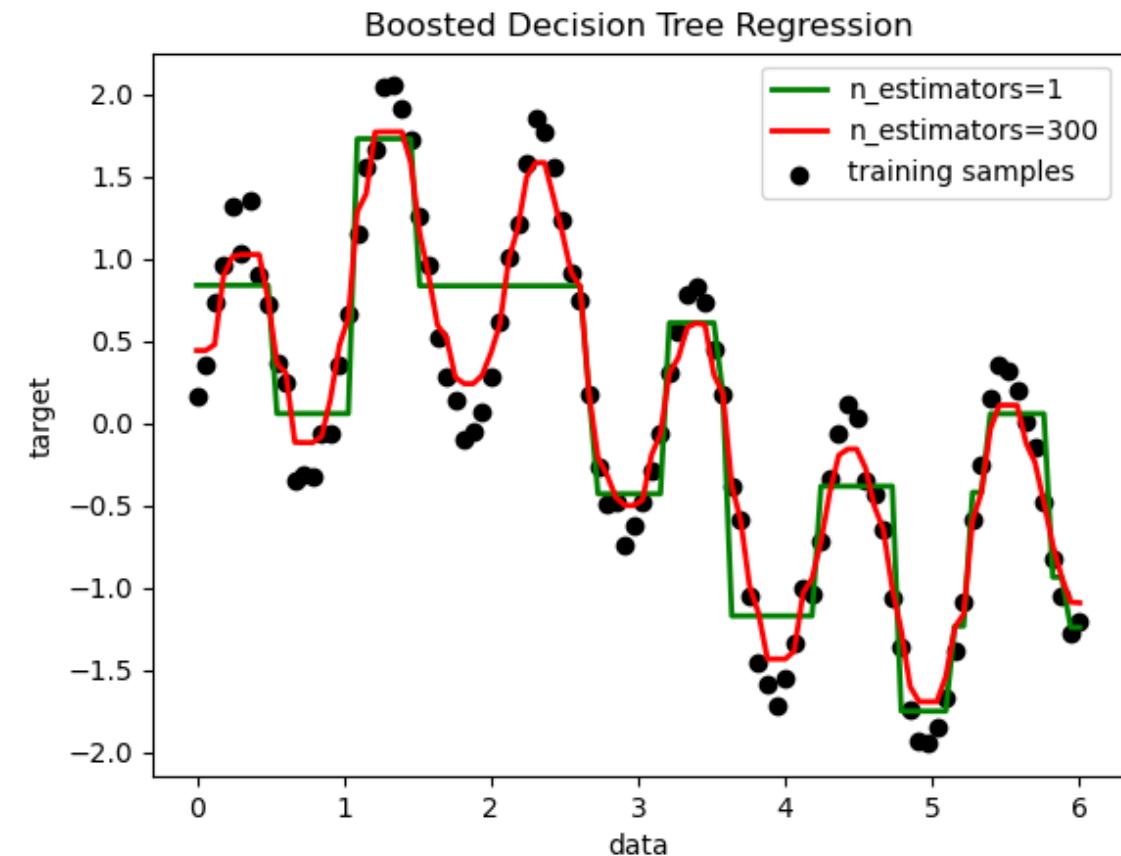
- Identificar la categoría a la que pertenece un objeto
- **Aplicaciones:** Detección de spam, reconocimiento de imágenes.
- **Algoritmos:** [SVM](#), [vecinos más cercanos](#), [bosque aleatorio](#), y más...



Scikit Learn

 Regresión

- Predecir un atributo de valor continuo asociado con un objeto.
- **Aplicaciones:** Precios de acciones, Farmacéutica, imputación de variables.
- **Algoritmos:** [SVR](#), [vecinos más cercanos](#), [bosque aleatorio](#), y [más...](#)

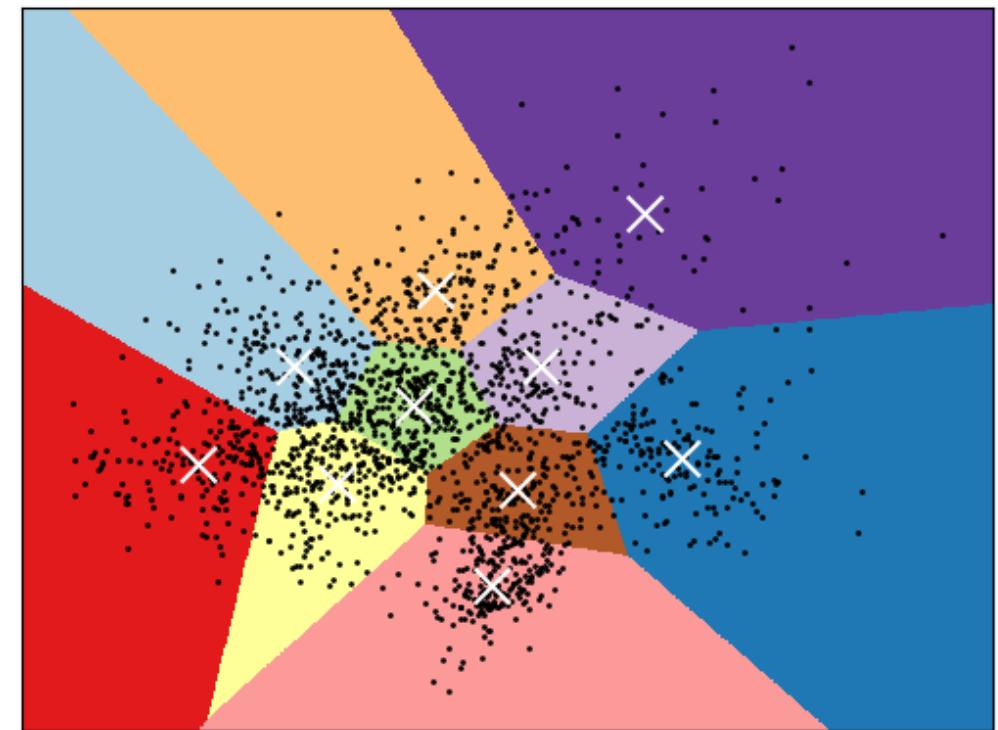


Scikit Learn

 Agrupamiento

- Agrupamiento automático de objetos similares en conjuntos.
- **Aplicaciones:** Segmentación de mercado, agrupamiento de documentos.
- Algoritmos: [k-Means](#), agrupamiento [espectral](#), [mean-shift](#), y [más...](#)

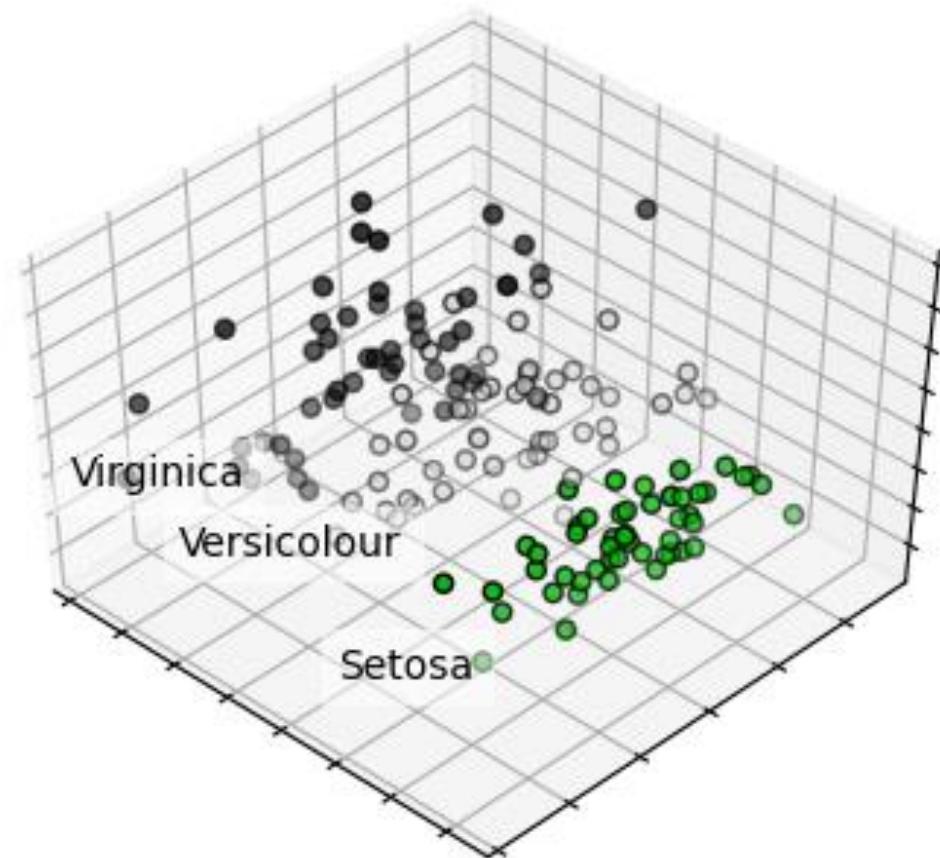
K-means clustering on the digits dataset (PCA-reduced data)
Centroids are marked with white cross



Scikit Learn

Reducción de la dimensionalidad

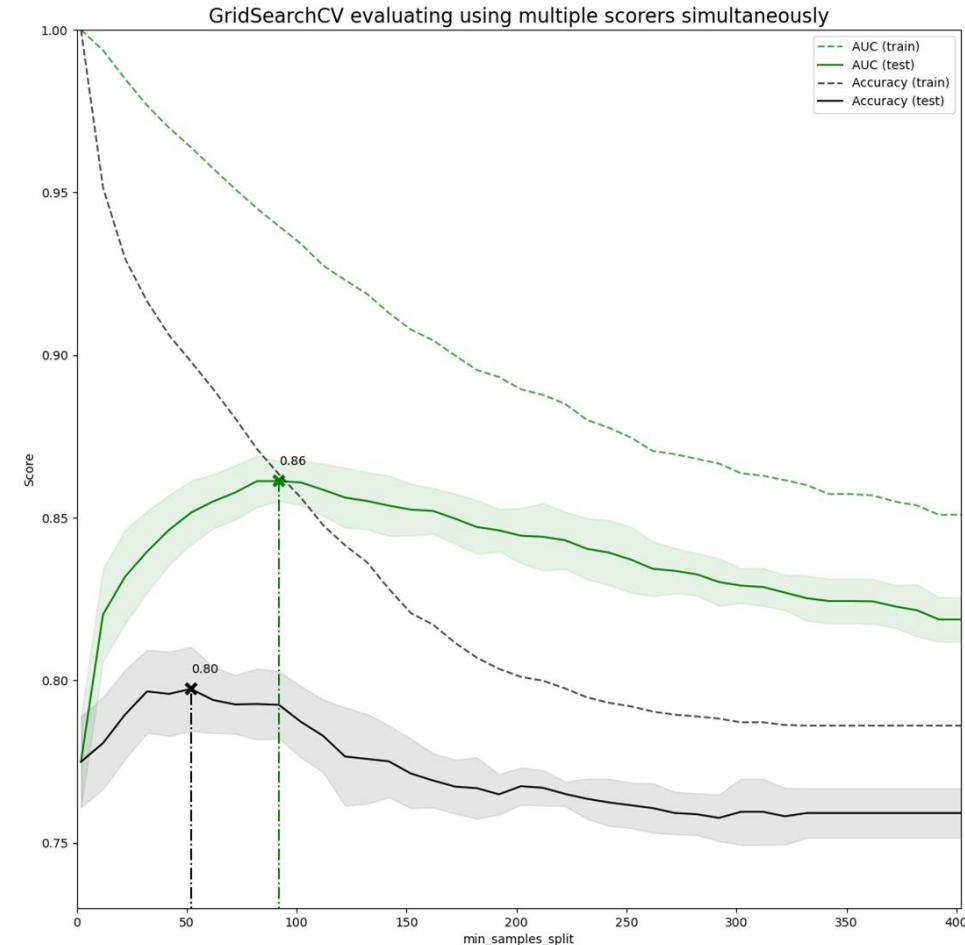
- Reducir el número de variables a considerar.
- **Aplicaciones:** Visualización, mejoras en la eficiencia.
- **Algoritmos:** k-Means, selección de características, factorización de matrices, y más...



Scikit Learn

Selección de modelos

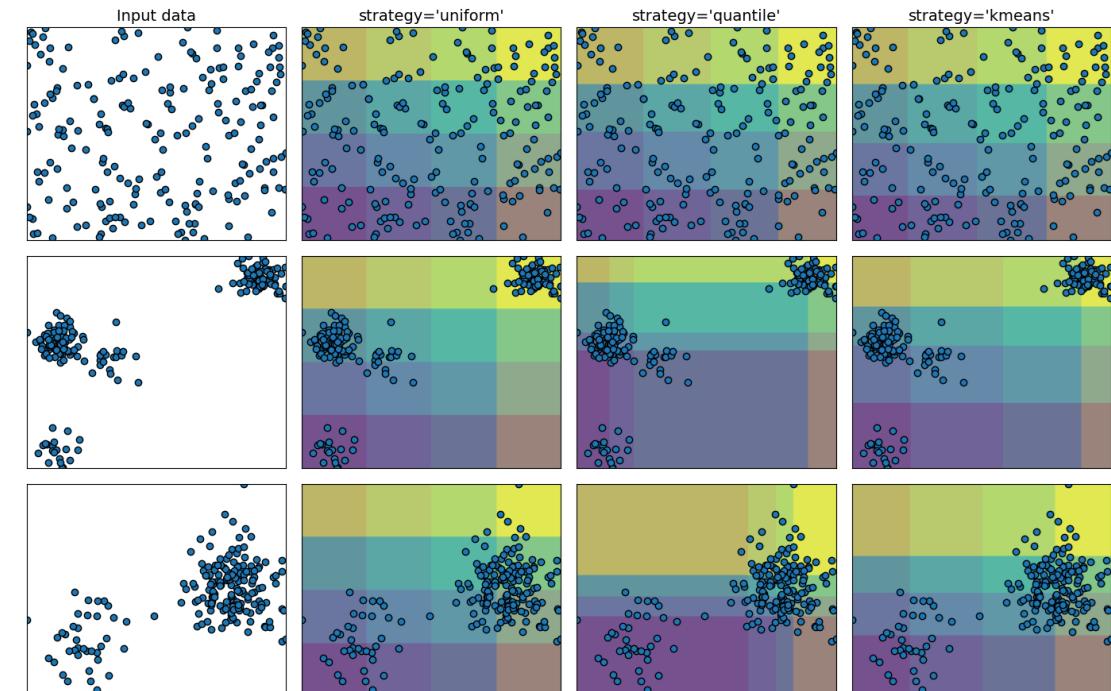
- Comparación, validación y selección de parámetros y modelos.
- Aplicaciones:** Mejor desempeño a través de la afinación de parámetros
- Algoritmos:** [grid search](#), [validación cruzada](#), [métricas](#), y más...



Scikit Learn

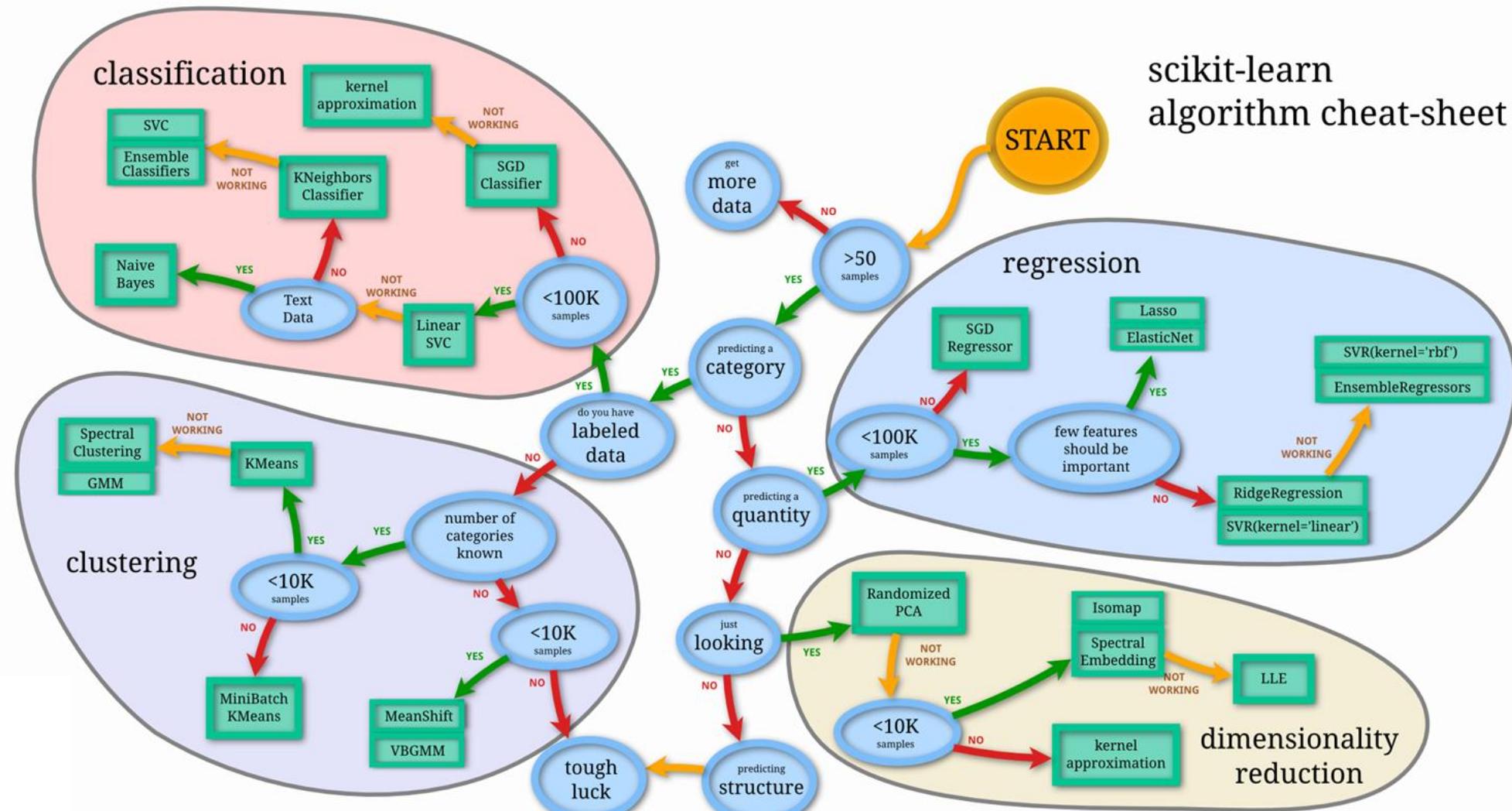
 Preprocesamiento

- Extracción de características y normalización
- **Aplicaciones:** Transformar datos de entrada tales como texto para ser usados en algoritmos de aprendizaje computacional.
- **Algoritmos:** preprocesamiento, extracción de características, y más...



Scikit Learn

Escogiendo el modelo correcto



Scikit Learn

 Taller



Despedida

Fabio Augusto Gonzalez, PhD.

<https://dis.unal.edu.co/~fgonza/>

fagonzalezo@unal.edu.co



UNIVERSIDAD
NACIONAL
DE COLOMBIA

Departamento de Ingeniería de Sistemas e Industrial
Facultad de Ingeniería
Universidad Nacional de Colombia
Sede Bogotá



Referencias

Scikit-learn

Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.



Recursos adicionales

Aprendizaje Computacional

Alpaydin, E. (2010). Introduction to Machine Learning. [Introducción al aprendizaje de máquinas]

https://kkpatel7.files.wordpress.com/2015/04/alppaydin_machinelearning_2010.pdf

Bibliografía

Mayo, M. (Mayo de 2018). Marcos para abordar el proceso de aprendizaje automático. Kdnuggets.

<https://www.kdnuggets.com/2018/05/general-approaches-machine-learning-process.html>

Mayo, M. (s.f.). El proceso de ciencia de datos, redescubierto. Kdnuggets. <https://www.kdnuggets.com/2016/03/data-science-process-rediscovered.html>

Google developers. (s.f.). Introducción a la estructura de problemas de aprendizaje automático.

<https://developers.google.com/machine-learning/problem-framing>



Derechos de imágenes

Creado por scikit-learn:

Imágenes sección scikit-learn.

Logo scikit-learn.

Aprendizaje computacional <https://pixabay.com/vectors/machine-learning-information-brain-5433370/>

https://www.freepik.es/vector-gratis/ordenador-portatil-pantalla-blanca-teclado_7222477.htm#page=1&query=computador&position=1

Aprendizaje computacional <https://pixabay.com/vectors/student-teenager-book-learning-147783/>

Tipos de <https://pixabay.com/vectors/thumb-positive-finger-excellent-1429327/> <https://pixabay.com/vectors/thumb-down-minus-failure-1429333/>

Scikit Learn <https://github.com/scikit-learn/scikit-learn/blob/master/doc/logos/scikit-learn-logo.svg>



Derechos de imágenes

Scikit-learn. (s.f). Imágenes sección scikit-learn. [Logo]. <https://scikit-learn.org/stable/>

Scikit-learn. (s.f). Logo scikit-learn. [Logo]. <https://scikit-learn.org/stable/>

Karmie. (s.f). Logo IBM Watson.[Logo]. <https://www.cleanpng.com/free/ibm-watson.html>

Amazon. (s.f). Logo AWS. [Logo].

https://upload.wikimedia.org/wikipedia/commons/thumb/9/93/Amazon_Web_Services_Logo.svg/1200px-Amazon_Web_Services_Logo.svg.png

Microsoft. (s.f). Logo Azure. [Logo]. <https://integration.team/wp-content/uploads/2019/07/azure.png>

Png egg. (s.f). Logo Google Cloud. [Logo]. <https://www.pngegg.com/es/png-ecwy>

Paperspace. (s.f). Logo Paperspace. [Logo]. <https://jobs.lever.co/paperspace/a99ba88f-dc65-402b-b2d3-6e13c9304679>

Google. (s.f). Logo Colaboratory. [Logo]. <https://colab.research.google.com/notebooks/intro.ipynb#>



Derechos de imágenes

Floydhub. (s.f). Floydhub. [Logo].

https://www.floydhub.com/static/media/floydhub_logo_large_transparent.cba00f0d.png

PyTorch (s.f.). PyTorch. [Logo]. https://upload.wikimedia.org/wikipedia/commons/9/96/Pytorch_logo.png

Keras (s.f.). Keras logo. [Logo]. <https://keras.io/img/logo.png>

Tensorflow (s.f) [Logo Tensorflow]NumPy. (s.f.). NumPy logo. [Logo]. <https://numpy.org/images/logos/numpy.svg>

Pandas. (s.f.). Pandas logo. [Logo]. https://pandas.pydata.org/static/img/pandas_white.svg

SciPy.org. (s.f.). SciPy logo. [Logo]. https://www.scipy.org/_static/logo.png

Matplotlib. (s.f.). matplotlib. [Logo]. https://matplotlib.org/_static/logo2_compressed.svg



Créditos

Facultad de
INGENIERÍA

Autores

Fabio Augusto González Osorio, PhD

Asistente docente

Miguel Ángel Ortiz Marín

Diseño instruccional

Claudia Patricia Rodríguez Sánchez

Diseño gráfico

Clara Valeria Suárez Caballero

Milton R. Pachón Pinzón

Diagramadora PPT

Daniela Duque

Fecha

2021-I

