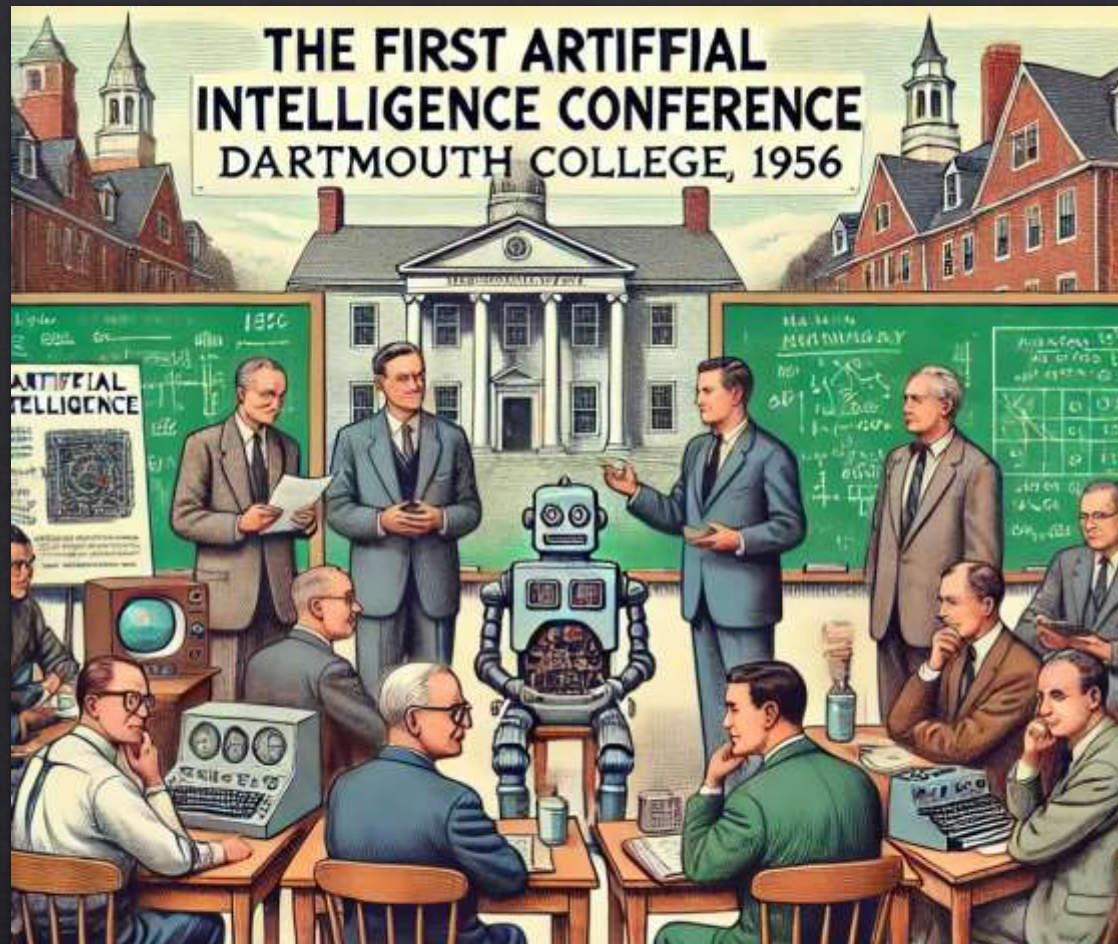


Introducción a Machine Learning



Fundamentos:

Autor: Estefanía Mendoza



Participantes Clave

- **John McCarthy:** Informático estadounidense, acuñó el término "inteligencia artificial".
- **Marvin Minsky:** Científico cognitivo y pionero en el campo de la IA.
- **Nathaniel Rochester:** Ingeniero de IBM que desarrolló uno de los primeros ordenadores.
- **Claude Shannon:** Matemático e ingeniero eléctrico conocido como el "padre de la teoría de la información".

Arthur Samuel

Arthur Samuel fue un pionero en el campo de la inteligencia artificial y la programación de computadoras.

Aportaciones Clave:

- ♦ - Acuñó el término 'Machine Learning' en la década de 1950.
- ♦ - Desarrolló programas de autoaprendizaje para jugar al ajedrez y damas.
- ♦ - Definió Machine Learning como 'el campo de estudio que da a las computadoras la capacidad de aprender sin ser explícitamente programadas'.
- ♦ **Importancia:**
- ♦ El trabajo de Samuel demostró que las computadoras podían aprender de la experiencia, sentando las bases para el desarrollo futuro de la IA.



Data Scientist: The Sexiest Job of the 21st Century

Harvard Business Review publicó un artículo en 2012 titulado "Data Scientist: The Sexiest Job of the 21st Century", en el cual describieron la ciencia de datos como la profesión más atractiva del siglo XXI



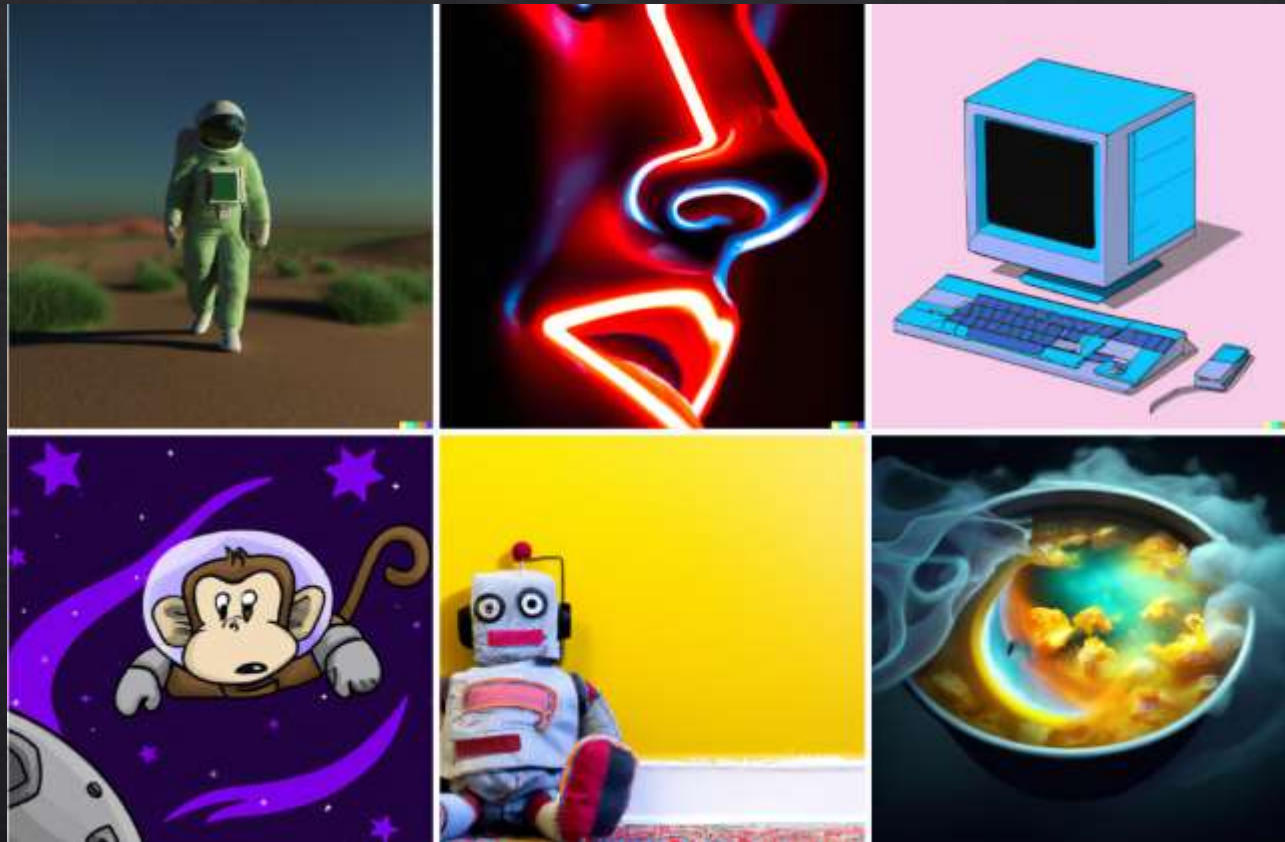
The Future of Jobs Report 2020

OCTOBER 2020





Este modelo de Machine Learning es capaz de crear imágenes realistas y también artísticas a partir de una descripción escrita



Perdida de empleo por tecnologías emergentes



Se van a destruir 85 millones de Puestos
Debido a que puedan ser automatizados

97 millones de trabajo se van a crear



Inteligencia artificial, big data y computación en la nube

Encuestas de empresas interesadas en implementar soluciones con IA

Comunicaciones 96%

Finanzas 90%

Salud 79%

Transporte y logística 88 %

Inteligencia Artificial

Big data y Analítica de datos

Industrias minera y metalúrgicas ,
educación y en el sector de transporte
todas estas con un 94%

1. Analítica y Científico de Datos
2. Especialista en IA y Machine Learning
3. Especialista en Big Data



8% a 14 %

Contenido del programa

1.1 Introducción y fundamentos del machine learning: introducción, jupyter, sklearn. (12O)

2.Regresión lineal y regresión logística: métricas de evaluación

3.Regularización, Clasificación y selección de modelos: paramétricos vs No paramétricos: KNN vs Gaussian, selección de modelos, overfitting, métricas, Árboles de decisión y máquinas de soporte vectorial, Boosting y selección de características . (26-02ON)

4.Aprendizaje no supervisado: Fundamentos de clústering y reducción de dimensionalidad . (09N)



COMPETENCIAS

Competencia general

Comprende y utiliza los algoritmos de aprendizaje automático para extraer e interpretar patrones a partir de grandes volúmenes de datos.



Competencias específicas:

- Comprender y crear modelos de regresión de aprendizaje automático supervisado a grandes volúmenes de datos.
- Comprender, crear y validar modelos de clasificación supervisada a grandes volúmenes de datos.
- Comprender, crear y validar modelos de clasificación no supervisada a grandes volúmenes de datos.

RESULTADOS DE APRENDIZAJE



- Identificar situaciones que requieran la creación de modelos de *machine learning*.
- Comprender y utilizar diferentes técnicas de *machine learning* acordes al tipo de problema que se quiere resolver.
- Usar **Python** como herramienta para la adquisición, carga, procesamiento y modelado de datos.

MATERIAL BÁSICO



Bagnato, J. (2017). *Aprende Machine Learning en Español Teoría + Práctica*.
<https://leanpub.com/aprendeml> ➔

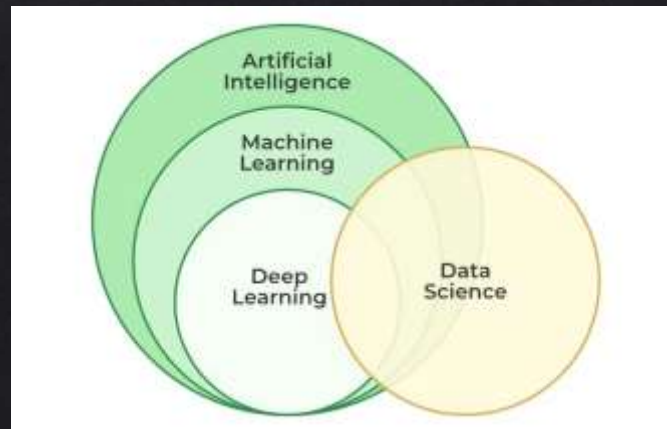
Guido, S., & Müller, A. (2017). *Introduction to Machine Learning with Python*. Sebastopol, CA, USA: O'Reilly Media, Inc.

ESQUEMA DE CALIFICACIÓN

- ◊ Guía 1: 15%
- ◊ Guía 2: 15%
- ◊ Guía 3: 15%
- ◊ Trabajo en clase: 20%
- ◊ Evaluaciones: 15%
- ◊ Encuentro Final: 20%

¿Qué es Machine Learning?

- ♦ Machine Learning (Aprendizaje Automático) es una rama de la inteligencia artificial (IA) que se centra en el desarrollo de algoritmos y modelos que permiten a las computadoras aprender y hacer predicciones o decisiones basadas en datos. En lugar de ser programadas explícitamente para realizar una tarea específica, las computadoras utilizan datos de entrenamiento para aprender patrones y relaciones en los datos, y luego aplicar ese conocimiento para hacer predicciones o tomar decisiones en nuevos datos.



Inteligencia Artificial (IA)

└─ Aprendizaje Automático (Machine Learning, ML)

└─ Aprendizaje Supervisado

└─ Regresión Basado en datos etiquetados

└─ Clasificación

└─ Aprendizaje No Supervisado

└─ Clustering Descubrimiento de patrones en datos

└─ Reducción de Dimensionalidad

└─ Aprendizaje por Refuerzo

└─ Juegos Toma de decisiones basadas en recompensar y castigos

└─ Control Robótico



DATOS

Michael Palmer dijo: “los datos son valiosos, pero si no están refinados, en realidad no se pueden usar. El petróleo debe transformarse en gas, plástico, productos químicos, etc. para crear una entidad valiosa que impulse una actividad rentable; por lo tanto, los datos deben desglosarse y analizarse para que tengan valor”.



Después de leer esto se podrá estar preguntando ¿Cómo refinar los datos?

Pareciera que la estadística es la respuesta, esa rama que acuña su nombre en 1749 cuando Gottfried Achenwall quería hablar del análisis de datos .

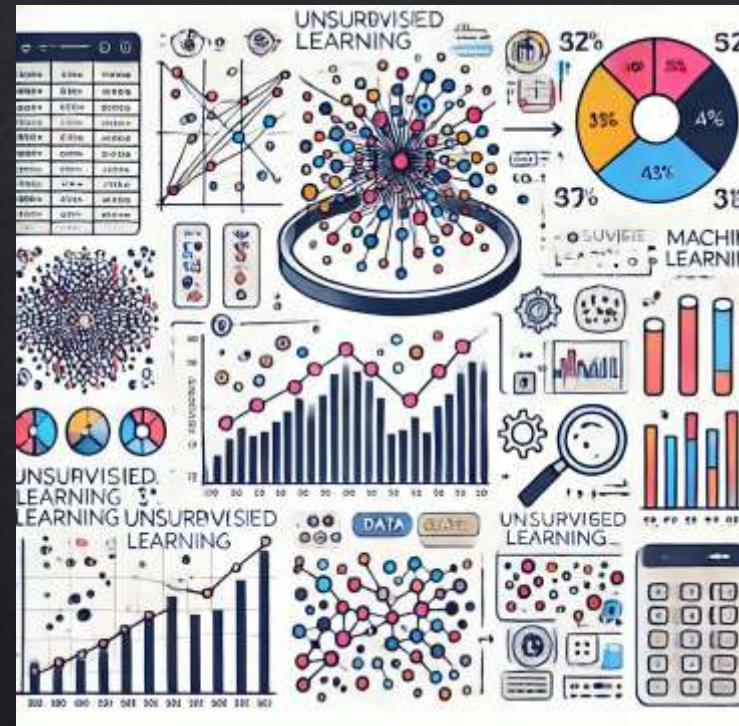
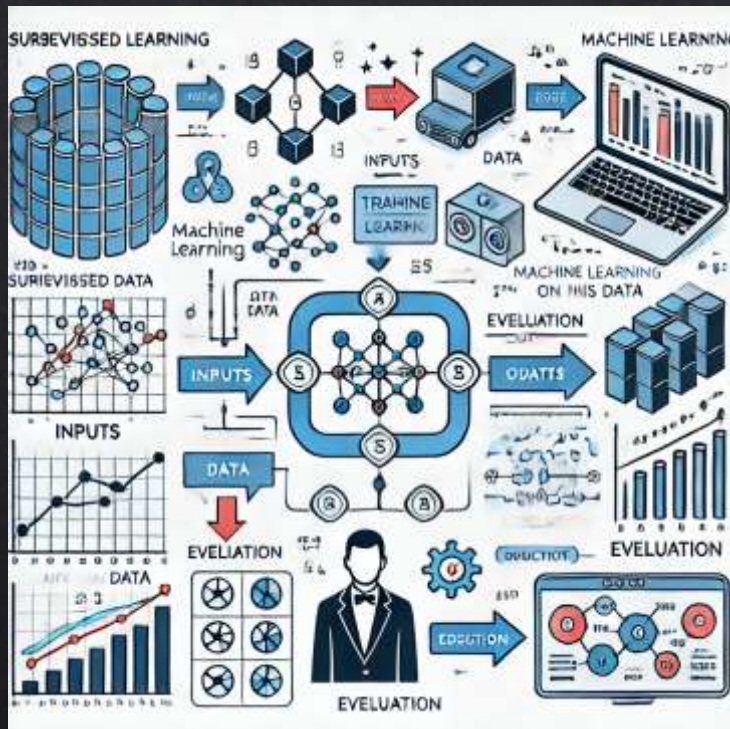
Restictiva y en algunos casos omitir información importante , además del trabajo de estadístico

SOLUCION :

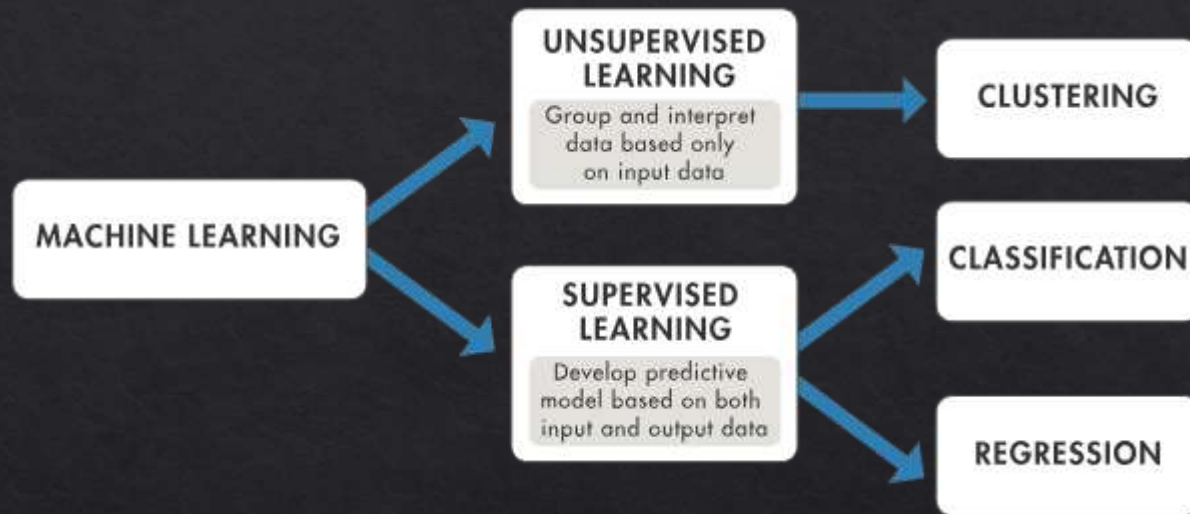
Podemos pensar en la posibilidad de que a partir de los datos la máquina pueda establecer patrones y aprenda de la información guardada para proveer lo mismo que tenemos con la estadística.

Podemos buscar estrategias para visualizar lo que encuentra la máquina y opciones, cuando se pueda, para entender los resultados que la máquina encontró.

¿Cómo enseñarle a una máquina?



Cómo funciona Machine Learning



Las técnicas de Machine Learning incluyen el aprendizaje supervisado y el aprendizaje no supervisado.

El aprendizaje supervisado

Aprendizaje supervisado hace referencia a esa colección de algoritmos que son capaces de predecir el valor de una variable objetivo cuando son entrenados por una base previamente etiquetada

Entonces, fue el proceso iterativo que programé el que permitió que la computación de la máquina determinará los parámetros que permite la decisión.

Correos Personales

Correos Spam

La máquina aprendió a determinar las etiquetas que yo le ponía a cada correo

Técnicas de clasificación: Estas técnicas predicen respuestas discretas; por ejemplo, si un email es genuino o spam, o si un tumor es maligno o benigno. Los modelos de clasificación clasifican los datos de entrada en categorías. Entre sus aplicaciones típicas se incluyen captura de imágenes médicas, reconocimiento de voz y puntuación crediticia.

Clasificación

Algoritmos de aprendizaje supervisados y semisupervisados para problemas de clasificación.

La clasificación es un tipo de machine learning supervisado en el que un modelo aprende a clasificar datos etiquetados. Para explorar modelos de clasificación de forma interactiva, consulte [Clasificación](#). Para explorar los datos de los predictores o las características con las etiquetas o respuestas, consulte [Comandos](#).

Para entrenar modelos de regresión, por ejemplo, regresión logística, árboles de decisión, consulte [Regresión](#).

Categorías

[App Classification Learner](#)

Entrene, valide y ajuste modelos de clasificación de forma interactiva.

[Árboles de clasificación](#)

Árboles de decisión binarios para aprendizaje multiclase.

[Análisis discriminante](#)

Análisis discriminante lineal y cuadrático regularizado.

[Naive Bayes](#)

Modelo Naive Bayes con predictores gaussianos, multinomiales o de kernel.

Técnicas de regresión: Estas técnicas predicen respuestas continuas; por ejemplo, cantidades físicas difíciles de medir, tales como estado de carga de una batería, carga eléctrica de la red o precios de activos financieros. Entre sus aplicaciones típicas se incluyen sensorización virtual, predicción de carga eléctrica y trading algorítmico.

Utilice técnicas de regresión si trabaja con un intervalo de datos o la respuesta es un número real, como los grados de temperatura o el tiempo hasta el fallo de una pieza electrónica. [Aquí](#) puede ver los algoritmos de regresión más comunes.

Toolbox

Categorías

App Regression Learner

Entrene, valide y ajuste modelos de regresión de forma interactiva

Regresión lineal

Modelos de regresión múltiple, escalonada y multivariante, y mucho más

Modelos lineales generalizados

Regresión logística, regresión multinomial, regresión de Poisson y mucho más

Regresión no lineal

Modelos de regresión de efectos mixtos y fijos no lineales

Regresión con máquinas de vectores de soporte

Máquinas de vectores de soporte para modelos de regresión

Regresión de procesos gaussianos

Modelos de regresión de procesos gaussianos (kriging)

Árboles de regresión

Árboles de decisión binarios para regresión

Ensembles de árboles de regresión

Bosques aleatorios, árboles de regresión potenciados y empaquetados

Modelo aditivo generalizado

Modelo interpretable compuesto por funciones de forma univariadas y bivariadas



El aprendizaje NO supervisado

mis datos no tienen etiquetas, busco que la máquina identifique patrones o datos anómalos a partir de ciertos criterios que programo en esos modelos. Los modelos más interesantes del aprendizaje no supervisado son los modelos de clusterización, usando una distancia definida identifica los datos cercanos entre sí y los agrupa de acuerdo con el criterio que se defina. Por ejemplo, imagina los siguientes datos binarios (ceros y unos) dibujados en

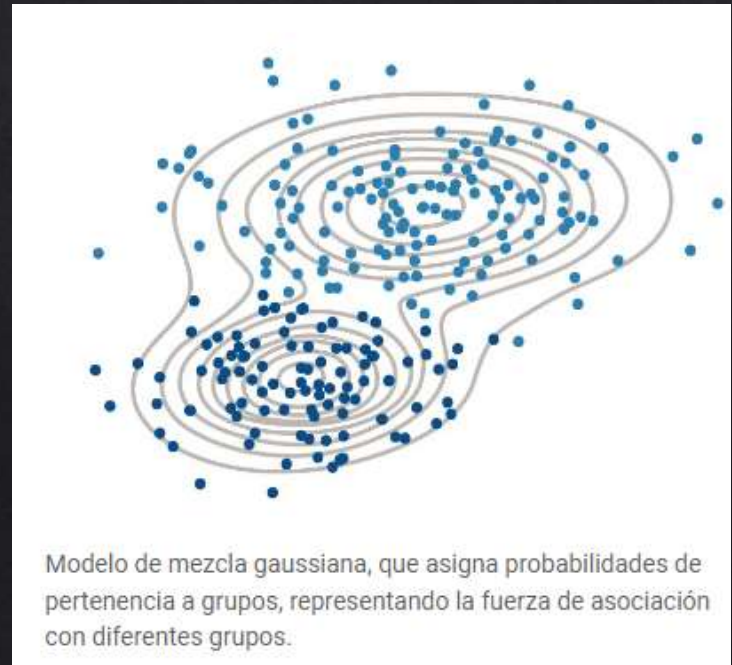
[0]	[0]	[0]		[1]	[1]	[1]	
[0]	[0]	[0]		[1]	[1]	[1]	
[0]	[0]	[0]		[1]	[1]	[1]	
[0]	[0]	[0]		[1]	[1]	[1]	

Aquí no hubo una etiqueta, el algoritmo descubre por sí mismo las diferencias en los datos y separa los grupos.

¿Qué es el análisis de clústeres?

El análisis de conglomerados implica la aplicación de algoritmos de agrupamiento con el objetivo de encontrar patrones ocultos o agrupaciones en un conjunto de datos. Por lo tanto, se utiliza con frecuencia en el análisis exploratorio de datos, pero también se utiliza para la detección de anomalías y el preprocesamiento para el aprendizaje supervisado.

Los algoritmos de agrupamiento forman agrupaciones de tal manera que los datos dentro de un grupo (o conglomerado) tienen una medida de similitud más alta que los datos de cualquier otro conglomerado. Se pueden utilizar varias medidas de similitud, incluidas las euclidianas, las probabilísticas, la distancia del coseno y la correlación. La mayoría de los métodos de aprendizaje no supervisado son una forma de análisis de conglomerados.



Ejemplo de análisis de conglomerados en MATLAB

Con el `imsegkmeans` comando (que utiliza el algoritmo k -means), MATLAB asignó tres grupos a la imagen original (tejido teñido con hematoxilina y eosina), lo que proporcionó una segmentación del tejido en tres clases (representadas como blanco, negro y gris). Pruébalo usted mismo y pruebe los enfoques de segmentación relacionados en [este ejemplo de código](#).

H&E image

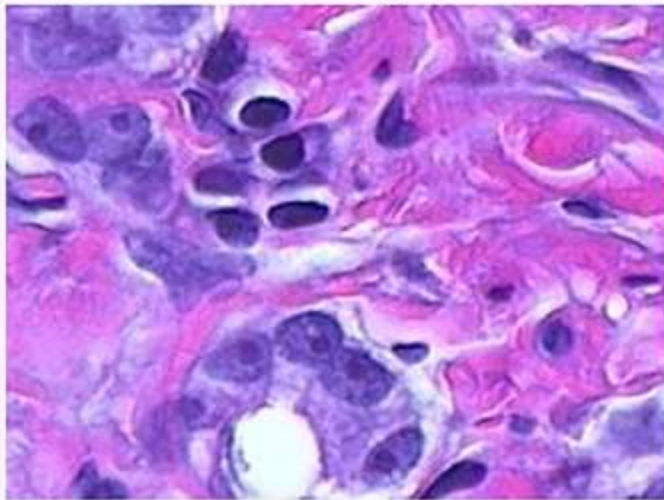
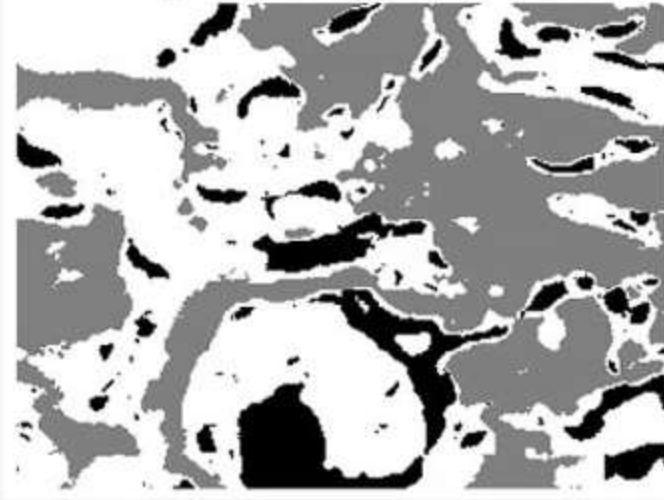


Image courtesy of Alan Partin, Johns Hopkins University

Image Labeled by Cluster Index



Machine Learning Summarized

