



Introducción al Machine learning

Clase sincrónica

Mapa general de la carrera

Revisión modular

SQL para el
análisis de
datos

Programación
con Python
para el análisis
de datos

Análisis
estadístico con
Python

Visualización y
comunicación
de insights

Machine
Learning*

Redes
neuronales*



Machine Learning

Unidad	Clases (sincrónico)	Autoaprendizaje (asincrónico)	Tutoría (sincrónico)
Introducción al Machine Learning	2 horas	Desde 6 horas	2 horas
Aprendizaje supervisado y no supervisado (Parte I)	2 horas	Desde 6 horas	2 horas
Aprendizaje supervisado y no supervisado (Parte II)	2 horas	Desde 6 horas	2 horas
Aprendizaje supervisado y no supervisado (Parte III)	2 horas	Desde 6 horas	2 horas
Aprendizaje supervisado y no supervisado (Parte IV)	2 horas	Desde 6 horas	2 horas
Aprendizaje supervisado y no supervisado (Parte V)	2 horas	Desde 6 horas	2 horas
Aplicando lo aprendido (Parte I)	2 horas	Desde 6 horas	2 horas
Aplicando lo aprendido (Parte II)	2 horas	Desde 6 horas	2 horas
<i>Prueba</i>	<i>0 horas</i>	<i>Desde 6 horas</i>	<i>0 horas</i>
<i>Receso</i>	<i>0 horas</i>	<i>0 horas</i>	<i>0 horas</i>

¿Qué aprenderemos en este módulo?

Al finalizar el módulo serás capaz implementar modelos de aprendizaje automático por medio de técnicas estadísticas, adecuando los diferentes algoritmos debidamente a la situación y requerimientos necesarios.



Implementar modelos de aprendizaje automático por medio de técnicas estadísticas, adecuando los diferentes algoritmos debidamente a la situación y requerimientos necesarios



- Unidad 1: Introducción al Machine Learning
- Unidad 2: Aprendizaje Supervisado y No Supervisado
(Parte I: No supervisado)
(Parte II: Clasificación)
(Parte III: Clasificación)
(Parte IV: Regresión)
(Parte V: Series de tiempo)
- Unidad 3: Aplicando lo aprendido
(Parte I: Preprocesamiento de datos)
(Parte II: Modelamiento)



¿Qué aprenderás en esta sesión?

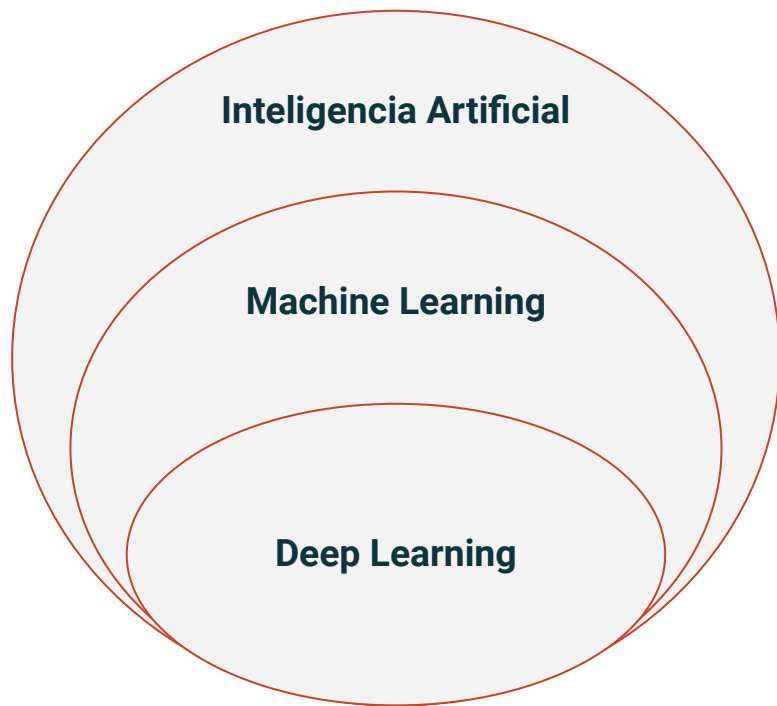
Identificar el conjunto de modelos que mejor se adapta a la resolución de una problemática de negocio que involucre aprendizaje automático.

¿Qué han escuchado del
concepto de Machine
Learning?



/*Conceptos Básicos de Machine Learning*/

¿Qué es el Machine Learning?



Inteligencia Artificial

Subdisciplina del campo de la informática que busca la creación de máquinas que puedan imitar un comportamiento “inteligente”.

Machine Learning

Busca dotar a las máquinas de capacidad de aprendizaje/generalización a partir de la experiencia (Datos).

Deep Learning

Rama del ML que utiliza redes neuronales profundas para el aprendizaje.

¿Cómo funciona el Machine Learning?

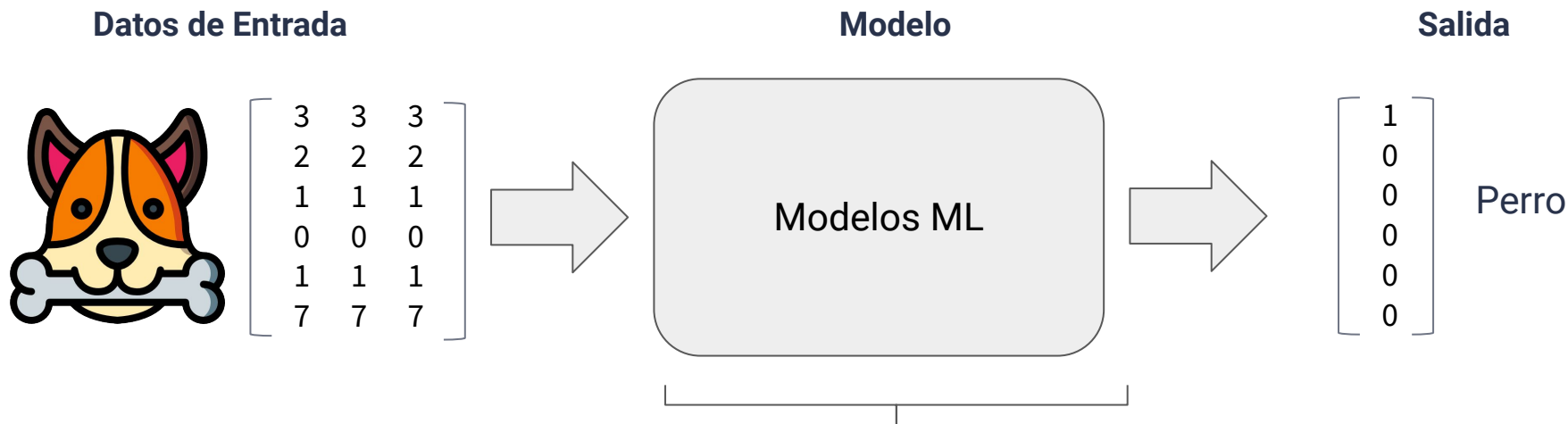
Para trabajar con técnicas de Machine Learning se utilizan entidades en el mundo real como: **clientes, productos, fotos, canciones, fenómenos, etc.**

El primer paso es poder representar estas entidades a partir de vectores numéricos, para que estos puedan ser comprendidos por una computadora.


$$\begin{bmatrix} 3 & 3 & 3 \\ 2 & 2 & 2 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \\ 7 & 7 & 7 \end{bmatrix}$$

$$\begin{bmatrix} x1 \\ x2 \\ x3 \\ x4 \\ \dots \\ xn \end{bmatrix}$$

¿Qué es un modelo de Machine Learning?



Representación de un problema o sistema que aprende a partir de los **datos** (relación entre los diferentes datos) utilizando un algoritmo de aprendizaje y se utiliza para hacer predicciones o tomar decisiones en nuevos datos.

`/*Tareas con Machine Learning*/`

Tareas con Machine Learning

Tipos de aprendizaje

1

Aprendizaje supervisado

busca descubrir la **relación** y/o **patrones** existente entre variables de entrada y de salida.

Aplica para set de **datos etiquetados**

2

Aprendizaje no supervisado

busca descubrir relaciones entre variables de entrada sin una guía del aprendizaje.

Aplica para set de **datos sin etiquetas**.

3

Aprendizaje reforzado

un **agente** aprende a **tomar decisiones** en un **entorno interactivo** para maximizar una recompensa acumulativa.

Aprendizaje Supervisado

Clasificación y regresión

Existen 2 tipos de aprendizaje supervisado dependiendo de la naturaleza de la variable objetivo.

Clasificación: Cuando la variable objetivo es una categoría.

Por ej, clasificación de imágenes, predicción de fuga, clasificación de correos.

Regresión: Cuando la variable objetivo es un valor numérico.

Por ej, predicción del precio de vivienda, predicción de demanda.

Entrada

Salida

2	4
5	10
7	14
10	?

Aprendizaje Supervisado

Clasificación y regresión

Existen 2 tipos de aprendizaje supervisado dependiendo de la naturaleza de la variable objetivo.

Clasificación: Cuando la variable objetivo es una categoría. Por ej, clasificación de imágenes, predicción de fuga, clasificación de correos.

Regresión: Cuando la variable objetivo es un valor numérico. Por ej, predicción del precio de vivienda, predicción de demanda.

Entrada

Salida

2	4
5	10
7	14
10	20

Ejercicio: ¿Regresión o clasificación?



Ejercicio

¿Clasificación o regresión?

Detección de Spam

	longitud	num_palabras	num_links
0	27	6	4
1	46	3	2
2	12	6	0
3	17	10	0
4	26	4	1
5	45	4	3
6	38	3	0
7	21	9	3
8	17	6	4

Variables explicativas

	spam
0	si
1	no
2	no
3	no
4	si
5	no
6	no
7	no
8	no

Variable Objetivo



Ejercicio

¿Clasificación o regresión?

Predicción precio de una casa

	LotArea	Street	GrLivArea	BsmtFullBath
0	8450	Pave	1710	1
1	9600	Pave	1262	0
2	11250	Pave	1786	1
3	9550	Pave	1717	1
4	14260	Pave	2198	1
5	14115	Pave	1362	1
6	10084	Pave	1694	1
7	10382	Pave	2090	1
8	6120	Pave	1774	0

Variables explicativas

	SalePrice
0	208500
1	181500
2	223500
3	140000
4	250000
5	143000
6	307000
7	200000
8	129900

Variable Objetivo



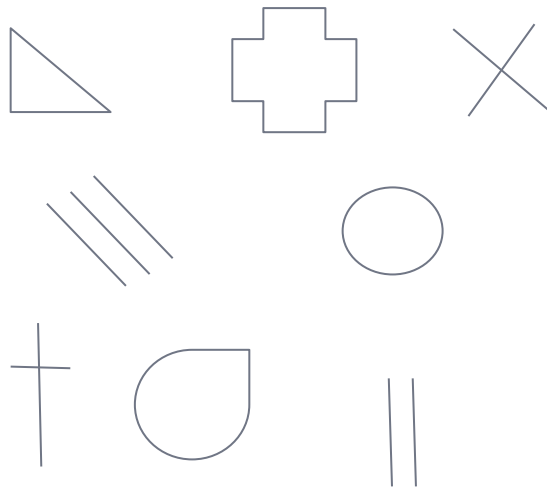
Aprendizaje No Supervisado

Técnicas

Reducción de dimensionalidad: Se utiliza para reducir la cantidad de dimensiones aprovechando la relación entre las diferentes variables.

Clustering: Se utiliza para agrupar conjunto de datos que se asemejan entre ellos. Por ejemplo, segmentación de clientes para descubrir segmentos de clientes similares entre sí.

Signos



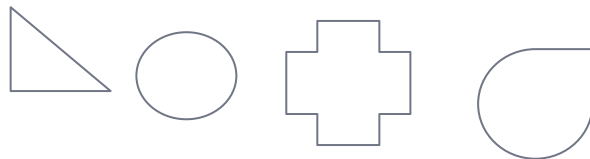
Aprendizaje No Supervisado

Técnicas

Reducción de dimensionalidad: Se utiliza para reducir la cantidad de dimensiones aprovechando la relación entre las diferentes variables.

Clustering: Se utiliza para agrupar conjunto de datos que se asemejan entre ellos. Por ejemplo, segmentación de clientes para descubrir segmentos de clientes similares entre sí.

Lenguaje 1



Lenguaje 2



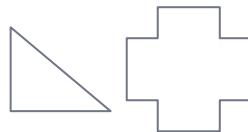
Aprendizaje No Supervisado

Técnicas

Reducción de dimensionalidad: Se utiliza para reducir la cantidad de dimensiones aprovechando la relación entre las diferentes variables.

Clustering: Se utiliza para agrupar conjunto de datos que se asemejan entre ellos. Por ejemplo, segmentación de clientes para descubrir segmentos de clientes similares entre sí.

Lenguaje 1



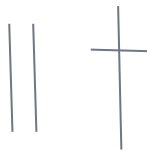
Lenguaje 3



Lenguaje 2



Lenguaje 4



Ejercicio: Detección de dígitos con clustering



Detección de dígitos con clustering

Utilicemos clusters

A continuación veremos cómo el uso de clusters puede ayudarnos a identificar patrones utilizando Python, y las bibliotecas que nos provee para esto.

Observa atentamente la presentación de tu profesor (la forma de llegar a estos resultados la comprenderás más adelante en el módulo)



Aprendizaje por Reforzamiento

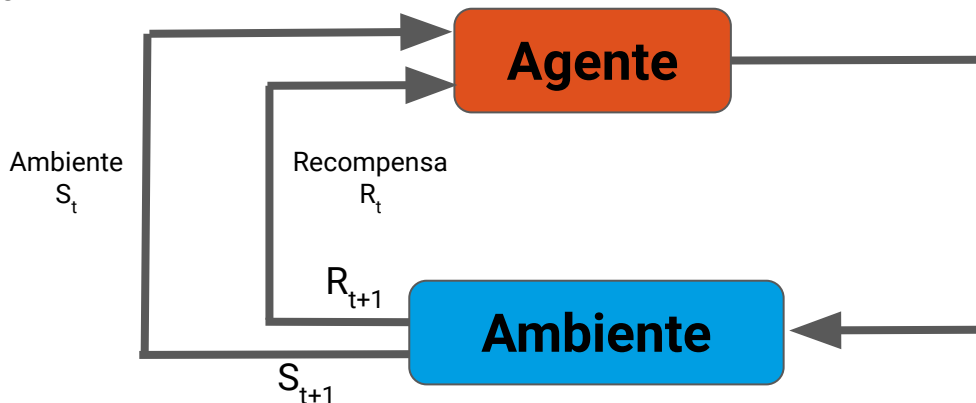
Elementos del aprendizaje por reforzamiento

Corresponde a un problema de optimización, en el cual un agente aprende a tomar decisiones en un entorno interactivo para maximizar una recompensa.

Agente: Puede realizar **acciones** dentro de un entorno, con el objetivo de obtener una recompensa.

Entorno: Corresponde al contexto en que está inserto el agente.

Recompensa: Es la respuesta que entrega el entorno al agente respecto a la acción que realizó en el estado que se encontraba.



Ejercicio: Comprendiendo las tareas de Machine Learning



Ejercicio

Clasifiquemos tareas de Machine Learning

1. Clasifique los siguientes ejemplos:
 - a. Predecir la demanda de los clientes a diferentes productos.
 - b. Analizar el sentimiento de las reseñas de los clientes.
 - c. Detección de transacciones fraudulentas en tiempo real.
 - d. Segmentar clientes para realizar ofertas dirigidas a los segmentos.
 - e. Predicción de fuga de clientes.
 - f. Diagnósticos con imágenes.
 - g. Visualización de múltiples variables en espacios más pequeños.
 - h. Predicción de falla en equipos mineros.
 - i. Predicción de la necesidad de camas críticas en el sector de salud.
 - j. Enseñar a un robot a caminar.
 - k. Predicción del precio de una propiedad



Ejercicio

Tipos de aprendizajes de máquinas

A partir del entendimiento de los diferentes tipos de aprendizaje que hay, entrega un ejemplo para cada uno (supervisado, no supervisado y por reforzamiento).

Comenta sobre sus áreas de expertise y experiencias varias.



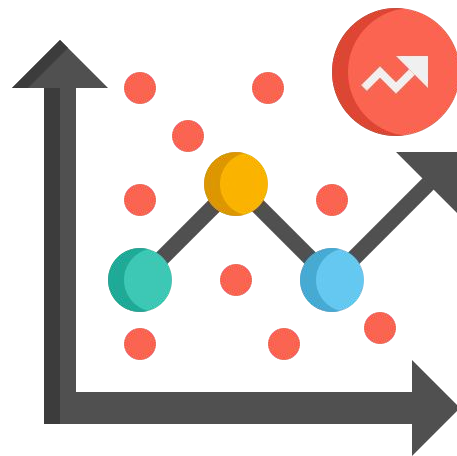
/*Retomando la Regresión Lineal*/

Retomando la regresión lineal

Scikit learn

La principal librería de machine learning para python es **scikit-learn**. En ella, encontraremos diferentes módulos como:

- Dataset de prueba
- Funciones para preparación de datos (esquemas de evaluación)
- Algoritmos de aprendizaje supervisado y no supervisado
- Métricas de evaluación
- Mucho más



Retomando la regresión lineal

Perspectiva del Machine Learning

Para realizar el ejemplo de la regresión lineal desde la perspectiva del Machine Learning vamos a proceder a aplicar los siguientes pasos:



División de los datos

Entrenamiento del
modelo

Predicción del modelo

Evaluación del modelo

Objetivo: Dotar al algoritmo de una buena capacidad de generalización en nuevos datos, para poder utilizar el algoritmo en los procesos pertinentes.

Machine Learning con Python

División de los datos

Dividimos el set de datos en entrenamiento y testeo.

Objetivo: Poder estimar los parámetros de algoritmo en el test de entrenamiento y testear cuál es el poder predictivo del algoritmo en datos nuevos, no vistos anteriormente.

Importante: Asegurar independencia entre ambos set de datos, para asegurar la generalización.

Atributos
entrenamiento

Vector objetivo
entrenamiento

Atributos
validación

Vector objetivo
validación

Machine Learning con Python

Entrenamiento

El algoritmo se entrena en el set de entrenamiento, donde ajusta sus parámetros para realizar predicciones o tomar decisiones más precisas.

$$y = \beta_0 + \beta_1 x_i + \epsilon_i$$

Atributos
entrenamiento

Vector objetivo
entrenamiento

Machine Learning con Python

Predicción

El algoritmo ya entrenado se utiliza para predecir en el set de testeo.

$$y = \beta_0 + \beta_1 x_i + \epsilon_i$$

Atributos
entrenamiento

Vector objetivo
entrenamiento

$$\hat{y} = \beta_0 + \beta_1 \mathbf{X}$$

Atributos
validación

Vector objetivo
validación

Machine Learning con Python

Evaluación del modelo

A partir de la predicción y el valor real de la variable objetivo del set de testeo se calculan métricas del desempeño del algoritmo.

$$MSE = \frac{1}{n} \sum_{i=1}^N (\mathbf{y}_i - \hat{\mathbf{y}}_i)^2$$

Importante: elegir métricas adecuadas para cada tipo de aprendizaje y cada caso de uso. Pero la idea siempre es la misma, tener una métrica que indica el grado de generalización del modelo.

{desafío}
latam_

$$y = \beta_0 + \beta_1 x_i + \epsilon_i$$

Atributos
entrenamiento

Vector objetivo
entrenamiento

$$\hat{y} = \beta_0 + \beta_1 \mathbf{X}$$

Atributos
validación

Vector objetivo
validación

¿Qué métricas podrían utilizarse para medir el desempeño de la regresión lineal?



Desafío - Introducción al Machine Learning



Desafío

"Introducción al Machine Learning"

- Descarga el archivo "Desafío".
- Tiempo de desarrollo asincrónico: desde 4 horas.
- Tipo de desafío: individual.

¡AHORA TE TOCA A TI! 💪



Ideas fuerza



Machine Learning es un subcampo de la IA que permite analizar datos, identificar patrones a partir de los datos.



Entender las diferentes **tareas** que se pueden resolver con ML (**Aprendizaje Supervisado, no supervisado y reforzamiento**) y escoge el indicado según sea el caso.



Conocer los usos y métodos de la librería **scikit learn** para la aplicación de **algoritmos de aprendizaje con python**

¿Qué contenidos de la clase
no te quedan totalmente
claros?



Recursos asincrónicos

¡No olvides revisarlos!

Para esta semana deberás revisar:

- Guía de estudio
- Desafío “Introducción al Machine Learning”





Próxima sesión...

- *Aprendizaje No supervisado*
- *Usos y técnicas de clusterización*
- *Evaluación de técnicas no supervisadas y consideraciones.*

{desafío}
latam_

*Academia de
talentos digitales*

