

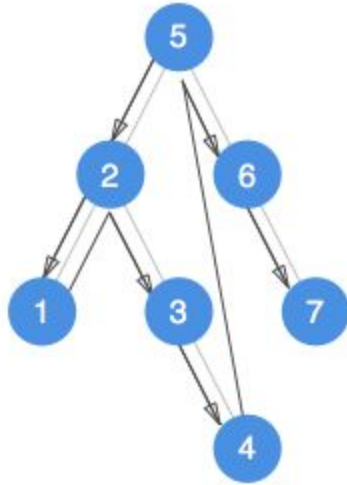
# Semana 13 – Machine Learning

## Temas para el día de hoy

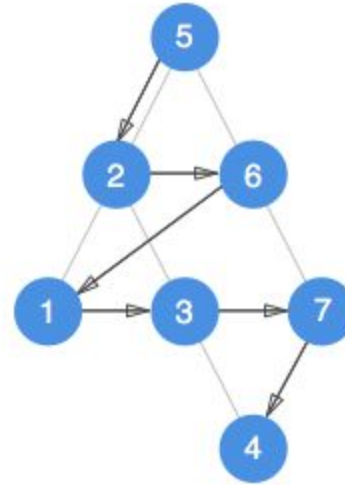
- ¿Qué es Machine learning?
- Tipos de Aprendizaje
- Pandas
- Clasificación vs. Regresión
- Regresión lineal y no lineal
- Regresión Logística

## Recordamos la sesión anterior

¿Qué es un algoritmo de búsqueda?



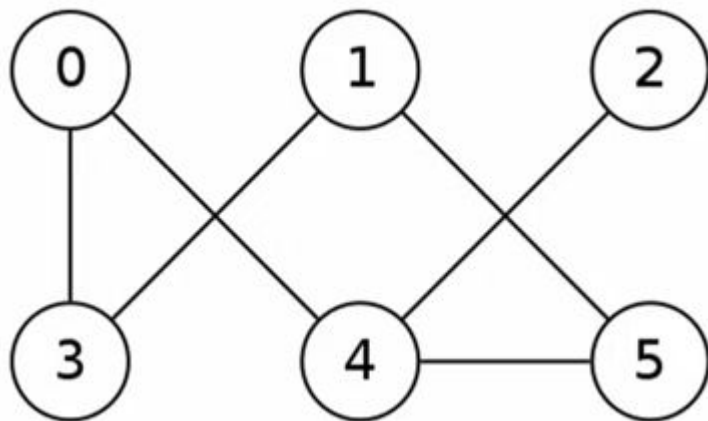
Depth-first traversal



Breadth-first traversal

En la modalidad virtual participamos a través del chat; para ello usamos la opción “Levantar la mano”.

## Búsqueda BFS DFS



**DFS:**031542

**BFS** 034125

## Logro de la semana



Al finalizar la sesión, el estudiante identifica los tipos de aprendizaje automático, y aplica los conceptos en ejercicios.



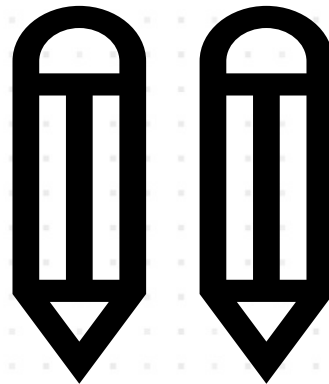
Universidad  
Tecnológica  
del Perú

# Motivación

## Actividad inicial

Respondamos las siguientes preguntas:

- ¿Qué es una machine learning?
- Mencione un ejemplo donde se utiliza machine learning



En la modalidad virtual participamos a través del chat, para ello usamos la opción “Levantar la mano”

## ¿Cuál de las 2 opciones es SPAM?

**A**

Estimado cliente de  
BigBanking,

Hemos detectado una  
actividad sospechosa en su  
cuenta.

Por favor, haga clic en el  
siguiente [enlace](#) y  
proporcione sus datos  
bancarios.

Lo mejor,  
XXX

**B**

Estimado cliente,

Su cambio de contraseña  
ha sido aceptado. Gracias  
por utilizar los servicios de  
BigBanking.

Lo mejor,  
XXX



## ¿Cuál de las 2 opciones es SPAM?

### Spam

Estimado cliente de  
BigBanking,

Hemos detectado una  
actividad sospechosa en su  
cuenta.

Por favor, haga clic en el  
siguiente [enlace](#) y  
proporcione sus datos  
bancarios.

Lo mejor,  
XXX

### Normal

Estimado cliente,

Su cambio de contraseña  
ha sido aceptado. Gracias  
por utilizar los servicios de  
BigBanking.

Lo mejor,  
XXX

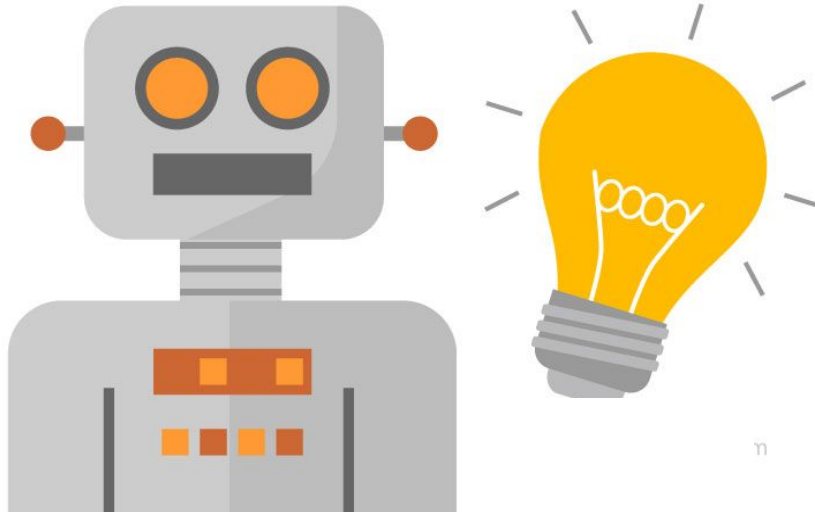


Universidad  
Tecnológica  
del Perú

# Machine Learning

## Machine learning

Es la ciencia de usar algoritmos para identificar patrones con el fin de resolver un problema de interés.



## Utilidad de machine learning

- **Seguridad informática**, diagnóstico de ataques, prevención de fraude online, detección de anomalías, etc.
- **Reconocimiento de imágenes o patrones** (facial, dactilar, objetos, voz, etc.)
- **Conducción autónoma**, mediante algoritmos deep learning: identificación de imágenes en tiempo real, detección de obstáculos y señales de tráfico, prevención de accidentes...
- **Salud**: evaluación automática de pruebas diagnósticas, robótica médica, etc.
- **Análisis de mercado de valores** (predicciones financieras, evolución de mercados, etc.)
- **Motores de recomendación**

## Pandas

Biblioteca de software escrita como extensión de NumPy para manipulación y análisis de datos para el lenguaje de programación Python. En particular, ofrece **estructuras de datos** y operaciones para manipular tablas **numéricas y series temporales**.



# Pandas

```
#Importar librería Pandas
import pandas as pd
```

```
[ ] #Crear una serie de 5 elementos numéricos
series = pd.Series( [5, 10, 15, 20, 25] )
series
```

```
0    5
1   10
2   15
3   20
4   25
dtype: int64
```

```
[ ] #Conocer el tipo de dato
type(series)
```

```
pandas.core.series.Series
```

```
[ ] #Obtener el índice 3
series[3]
```

```
20
```

```
[ ] #Crear una serie de caracteres con 3 elementos
cad = pd.Series( ['U', 'T', 'P'] )
cad
```

```
0    U
1    T
2    P
dtype: object
```

```
[ ] # Crear una DataFrames
lst = ['Hola', 'mundo', 'robótico']
df = pd.DataFrame(lst)
print(df)
```

```
0      0
0    Hola
1    mundo
2  robótico
```

```
[ ] # Estructura de datos con cabecera
data = { 'Nombre': ['Juan', 'Ana', 'Jose', 'Arturo' ],
        'Edad' : [25, 18, 23, 27],
        'Pais': ['MX', 'CO', 'BR', 'PE'] }
df = pd.DataFrame(data)
df
```

	Nombre	Edad	Pais
0	Juan	25	MX
1	Ana	18	CO
2	Jose	23	BR
3	Arturo	27	PE

```
[ ] #Imprimir dos columnas de estructura de datos
print(df[['Nombre', 'Pais']])
```

```
Nombre Pais
0    Juan  MX
1     Ana  CO
2    Jose  BR
3  Arturo  PE
```

# Pandas

```
[ ] #Cargar una estructura de datos y mostrar los 5 primeros elementos
data = pd.read_csv('canciones_2022.csv')
data.head(5)
```

	id	name	artists	danceability	energy
0	6DCZcSspjsKoFjzrWoCd	God's Plan	Drake	0.754	0.449
1	3ee8Jmje8o58CHK66QrVC	SAD!	XXXTENTACION	0.740	0.613
2	0e7ipj03S05BNilyu5bRz	rockstar (feat. 21 Savage)	Post Malone	0.587	0.535
3	3swc6WTsr7rl9DqQKQA55	Psycho (feat. Ty Dolla Sign)	Post Malone	0.739	0.559
4	2G7V7zsVDxg1yRsu7Ew9R	In My Feelings	Drake	0.835	0.626

```
[ ] #Imprimir un dato específico
artista = data.artists
artista[2]
```

'Post Malone'

```
[ ] #Mostrar la información una fila en específico
info = data.iloc[2]
info
```

```
id          0e7ipj03S05BNilyu5bRz
name        rockstar (feat. 21 Savage)
artists     Post Malone
```

```
[ ] #Mostrar los últimos registros de la estructura
data.tail()
```

	id	name	artists	danceability	energy	key	loudness
95	630sXRhlcfwr2e4RdNgK	Rewrite The Stars	Zac Efron	0.684	0.619	10.0	-7.005
96	2xmrfQpmS2jExTikLoA	I Miss You (feat. Julia Michaels)	Clean Bandit	0.638	0.658	3.0	-6.318
97	5WvAo7DNuPRmk4APhdPzi	No Brainer	DJ Khaled	0.552	0.760	0.0	-4.706
98	1j4kHkqpqZRBwEOA4CN4Y	Dusk Till Dawn - Radio Edit	ZAYN	0.258	0.437	11.0	-6.593
99	3EPXxR3ImUwfayaurPi3c	Be Alright	Dean Lewis	0.553	0.586	11.0	-6.319

```
[ ] # Conocer el tamaño del Datasets
data.shape
```

(100, 16)

```
[ ] #Conocer las columnas del Datasets
data.columns
```

```
Index(['id', 'name', 'artists', 'danceability', 'energy', 'key', 'loudness',
      'mode', 'speechiness', 'acousticness', 'instrumentalness', 'liveness',
      'valence', 'tempo', 'duration_ms', 'time_signature'],
      dtype='object', name='columns')
```

## Pandas

Resolvamos las actividades propuestas

Crear la siguiente matriz en pandas

	animal	favorite_food	cuteness_level
0	panda	bamboo	100
1	koala	eucalyptus	100



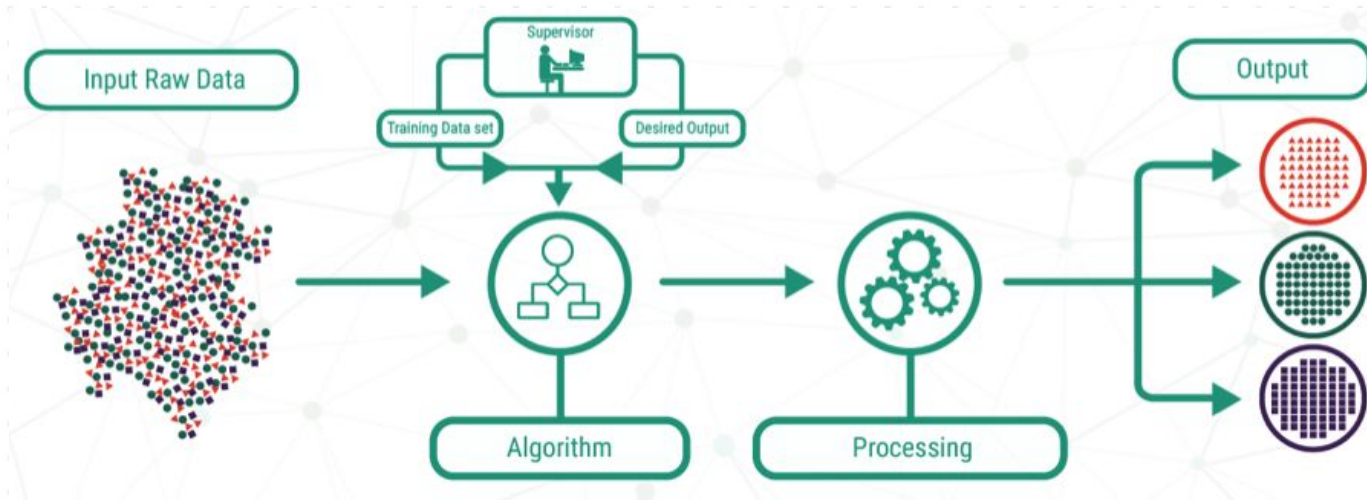
# Tipos de machine learning

Los tipos de implementación de machine Learning pueden clasificarse en tres categorías:



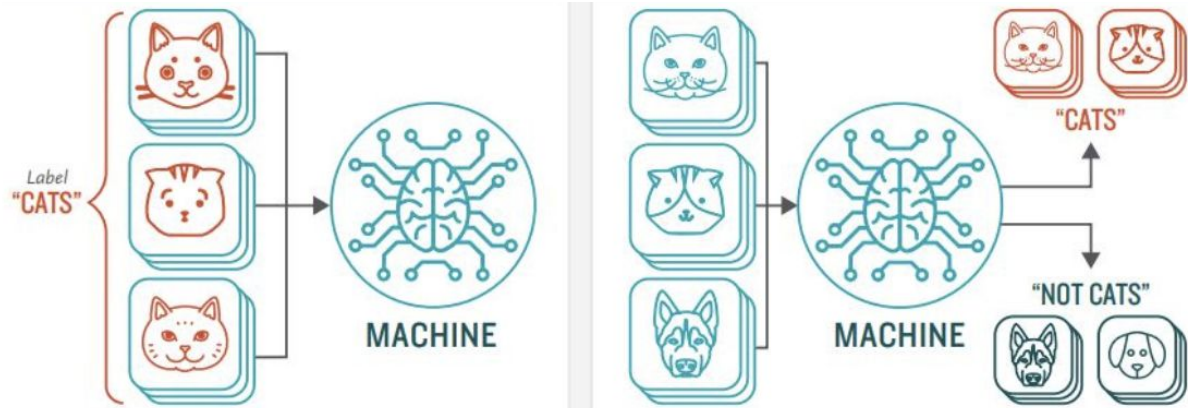
## Aprendizaje supervisado

Los algoritmos trabajan con datos “etiquetados” (labeled data), intentando encontrar una función que, dadas las variables de entrada (input data), les asigne la etiqueta de salida adecuada.



## Ejemplo aprendizaje supervisado

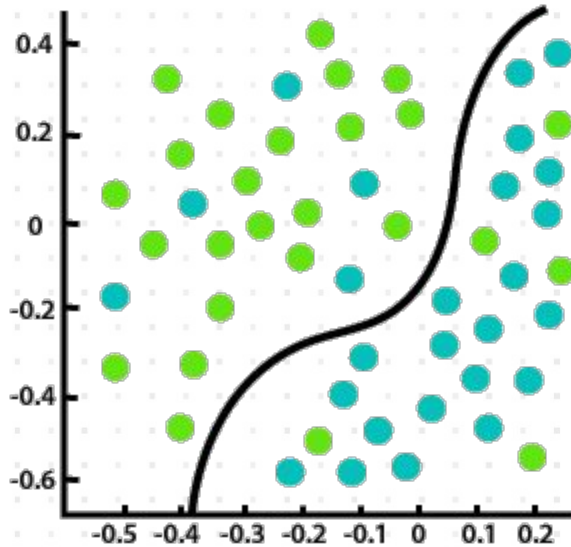
Para saber si en una imagen tenemos un **gato o un perro**, entrenamos a nuestro modelo con miles de imágenes y le decimos cuáles son de perros y cuáles son de gatos (**labels**). Después de muchos ejemplos, dada una nueva imagen (**sin labels**), nuestro modelo podrá determinar si se trata de un gato o un perro. A este problema se le llama **clasificación**.



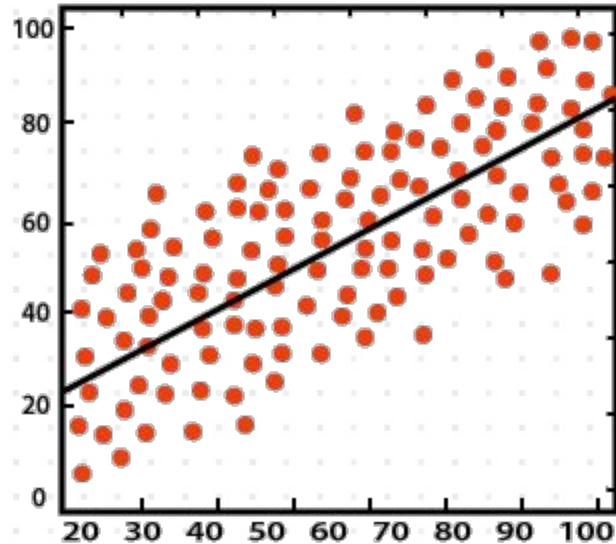
## Los algoritmos más usados en aprendizaje supervisado

- Árboles de decisión.
- Clasificación de Naïve Bayes.
- Regresión por mínimos cuadrados.
- Regresión Logística.
- Support Vector Machines (SVM).
- Métodos “Ensemble” (Conjuntos de clasificadores).

## Técnicas de Aprendizaje supervisado



**Clasificación**

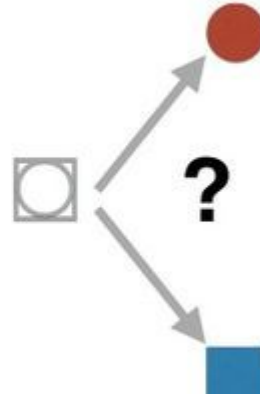
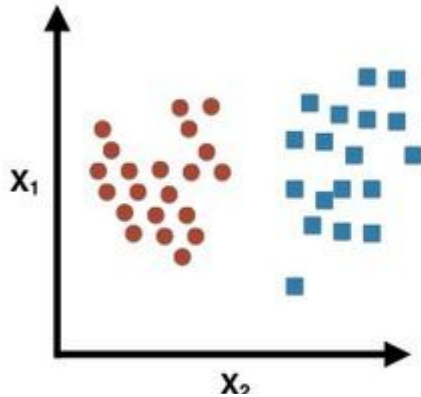


**Regresión**

## Clasificación

Es un sistema de clasificación que predice una categoría.

Ejemplo: Los correos se “categorizan” como “spam” o como “legítimos”.



# Técnicas de Clasificación

Logistic Regression



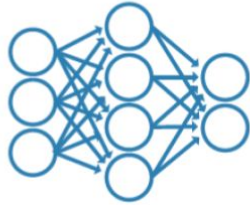
k Nearest Neighbor (kNN)



Support Vector Machine (SVM)



Neural Network



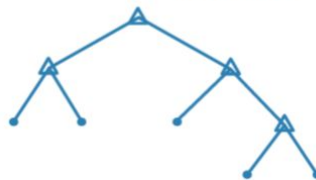
Naïve Bayes



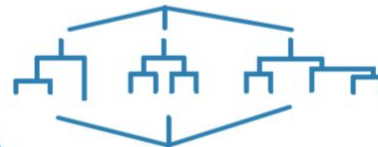
Discriminant Analysis



Decision Tree



Bagged and Boosted Decision Trees

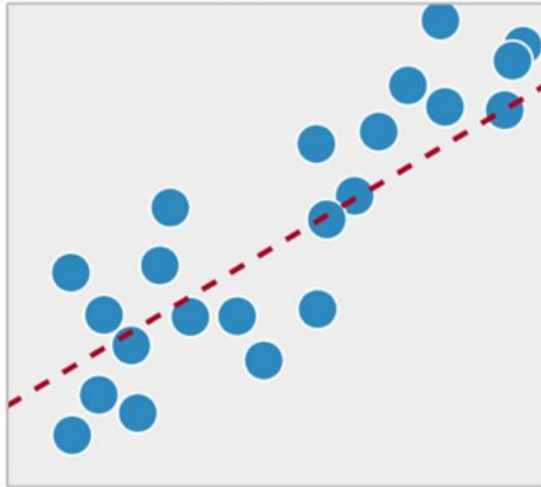


Common Classification Algorithms



## Regresión

Es una técnica de modelado predictivo que evalúa la relación entre las variables dependientes (es decir, la variable objetivo) y las independientes.





## Regresión lineal simple

La regresión lineal es la forma más básica de algoritmos de regresión en el aprendizaje automático. El modelo consta de un solo parámetro y una variable dependiente.

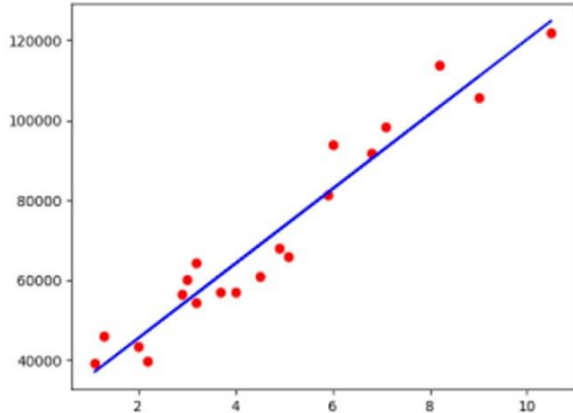
$$y = mx + c + e$$

Donde:

m= es la pendiente de la línea,

c= es una intersección

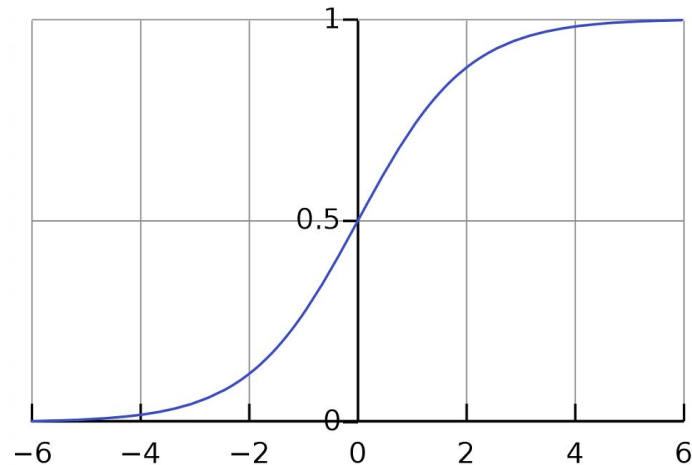
e=representa el error en el modelo.



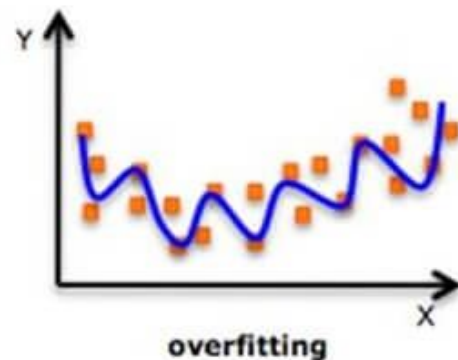
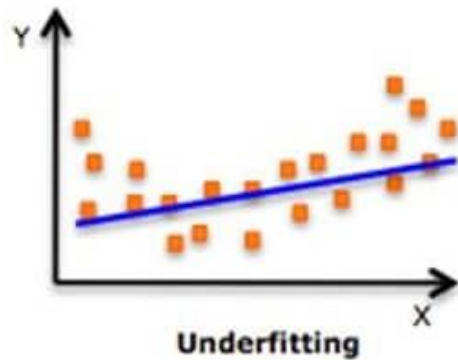
## Regresión logística

Es un tipo de técnica de regresión cuando la variable dependiente es discreta. Ejemplo: 0 o 1, verdadero o falso, etc. Esto significa que la variable objetivo solo puede tener dos valores, y una función sigmoidea muestra la relación entre la variable objetivo y la variable independiente.

$$f(x) = \frac{1}{1 + e^{-x}}$$

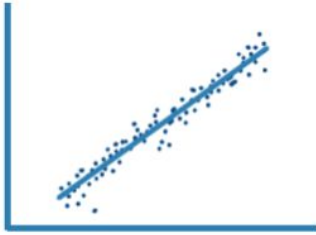


## Problemas de entrenamiento

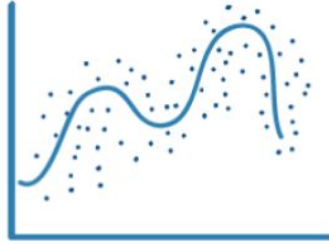


# Técnicas de regresión

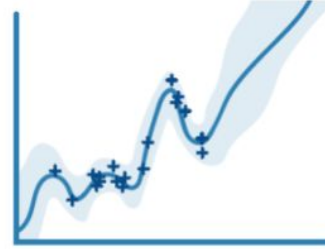
Linear Regression



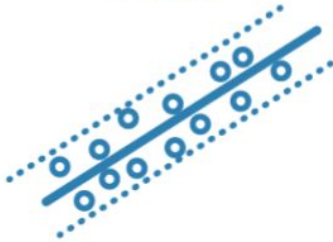
Nonlinear Regression



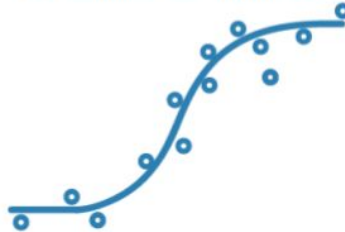
Gaussian Process Regression



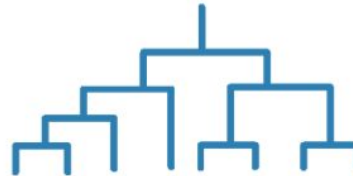
SVM Regression



Generalized Linear Models

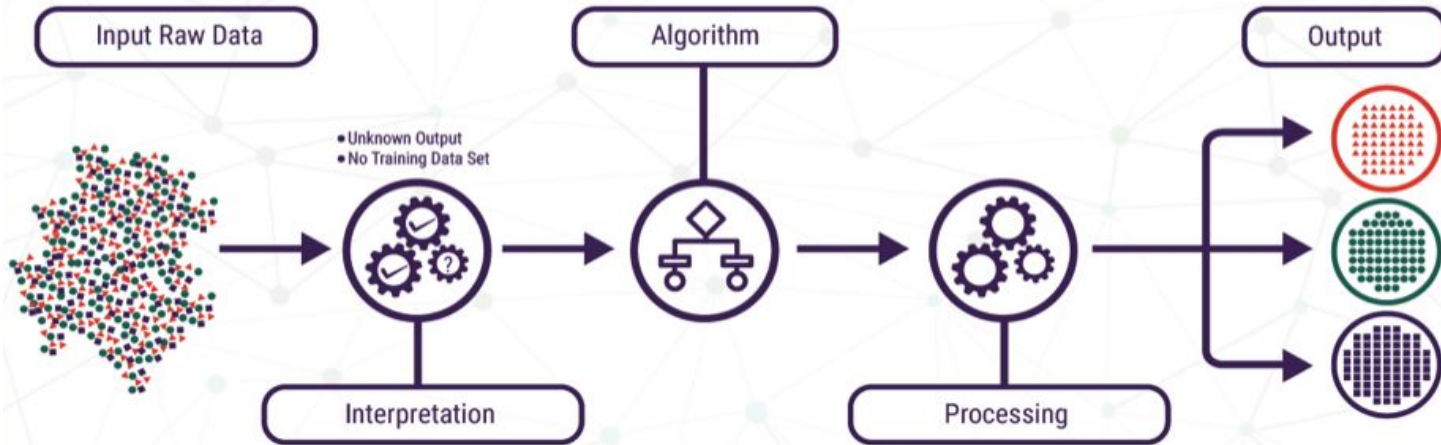


Regression Tree



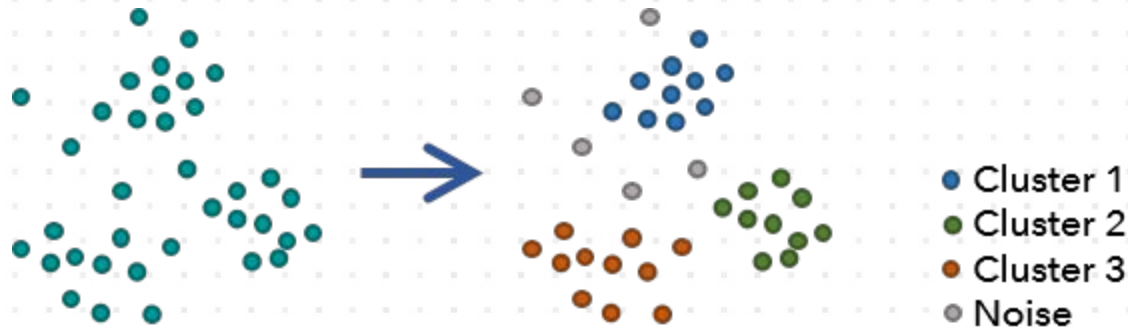
## Aprendizaje no supervisado

Tiene lugar cuando no se dispone de datos “etiquetados” para el entrenamiento. Por tanto, solo podemos describir la estructura de los datos, para intentar encontrar algún tipo de organización que simplifique el análisis.



## Ejemplo de aprendizaje no supervisado

Las tareas de clustering, buscan agrupamientos basados en similitudes, pero nada garantiza que estas tengan algún significado o utilidad. En ocasiones, al explorar los datos sin un objetivo definido, se pueden encontrar correlaciones espurias, curiosas, pero poco prácticas.



## **Algoritmo más habituales en aprendizaje no supervisado:**

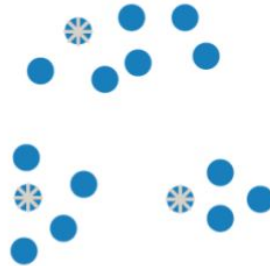
1. Algoritmos de clustering
2. Análisis de componentes principales
3. Descomposición en valores singulares (singular value decomposition)
4. Análisis de componentes principales (Independent Component Analysis)

# Técnicas de cluster

k-Means

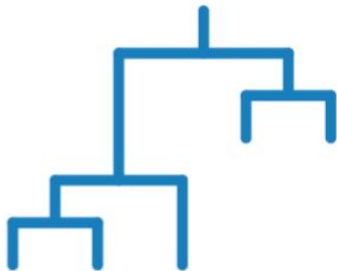


k-Medoids

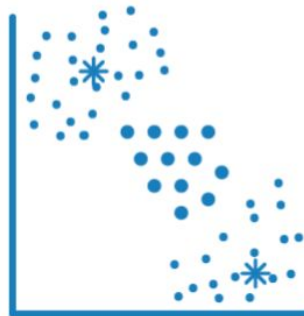


*Common  
Clustering  
Algorithms*

Hierarchical Clustering



Fuzzy c-Means



Gaussian Mixture Model





## Aprendizaje de refuerzo

Este tipo aprendizaje se basa en mejorar la respuesta del modelo usando un proceso de retroalimentación. El algoritmo aprende observando el mundo que le rodea.



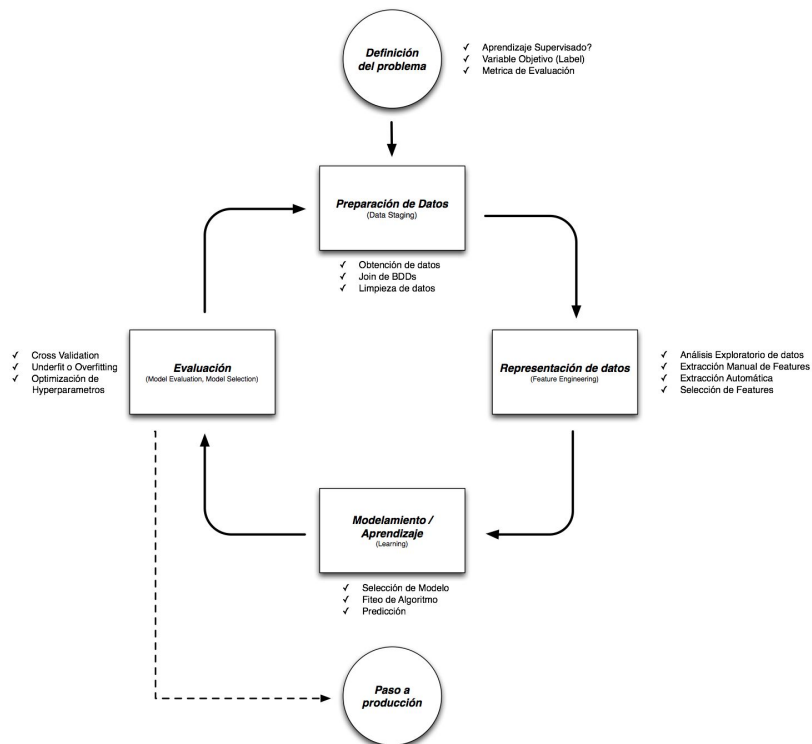
## Recordemos lo aprendido

Respondamos la siguiente pregunta:

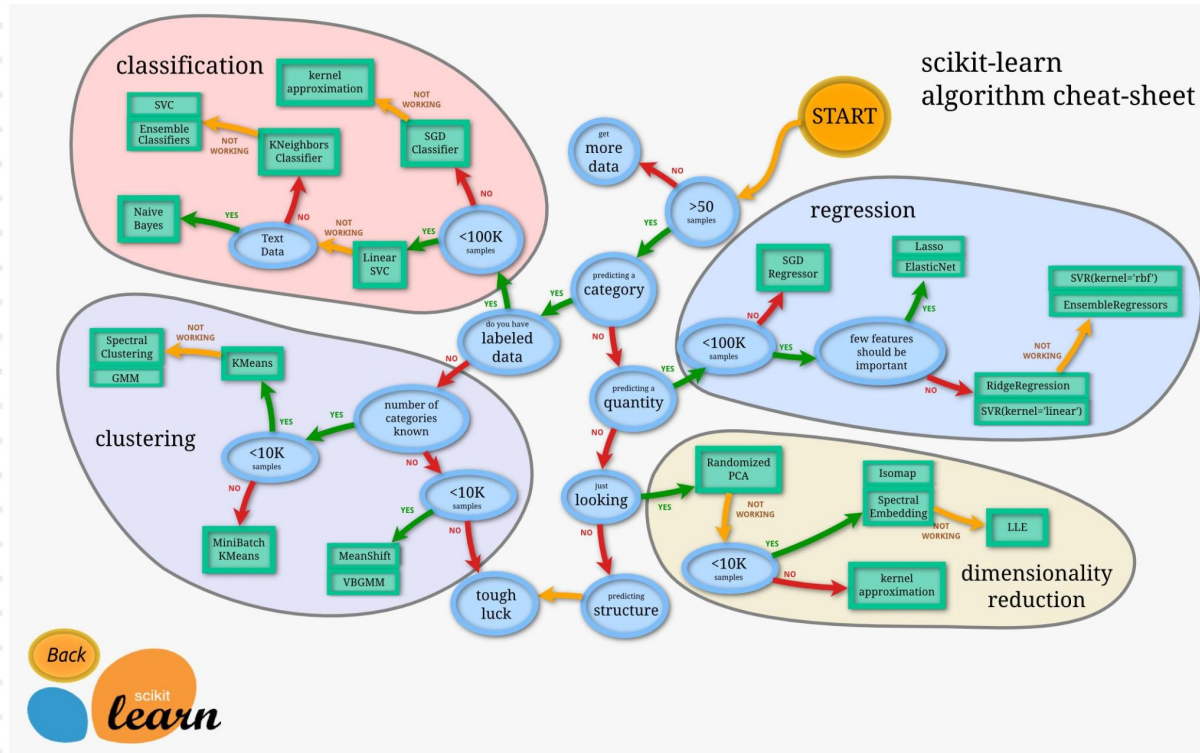
- ¿Cuáles son los tipos de aprendizaje automático”?



# El ciclo de trabajo del Machine Learning



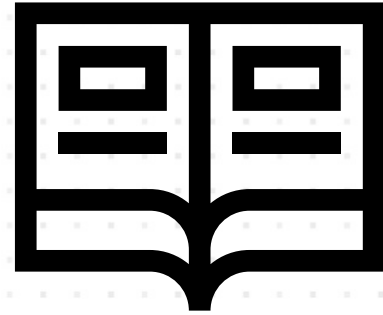
# Algoritmos de Machine Learning



# Pongamos en práctica lo aprendido

Resolvamos las actividades propuestas en

Plantear un proyecto de inteligencia artificial que pueda solucionar los problemas cotidianos.



# Conclusiones

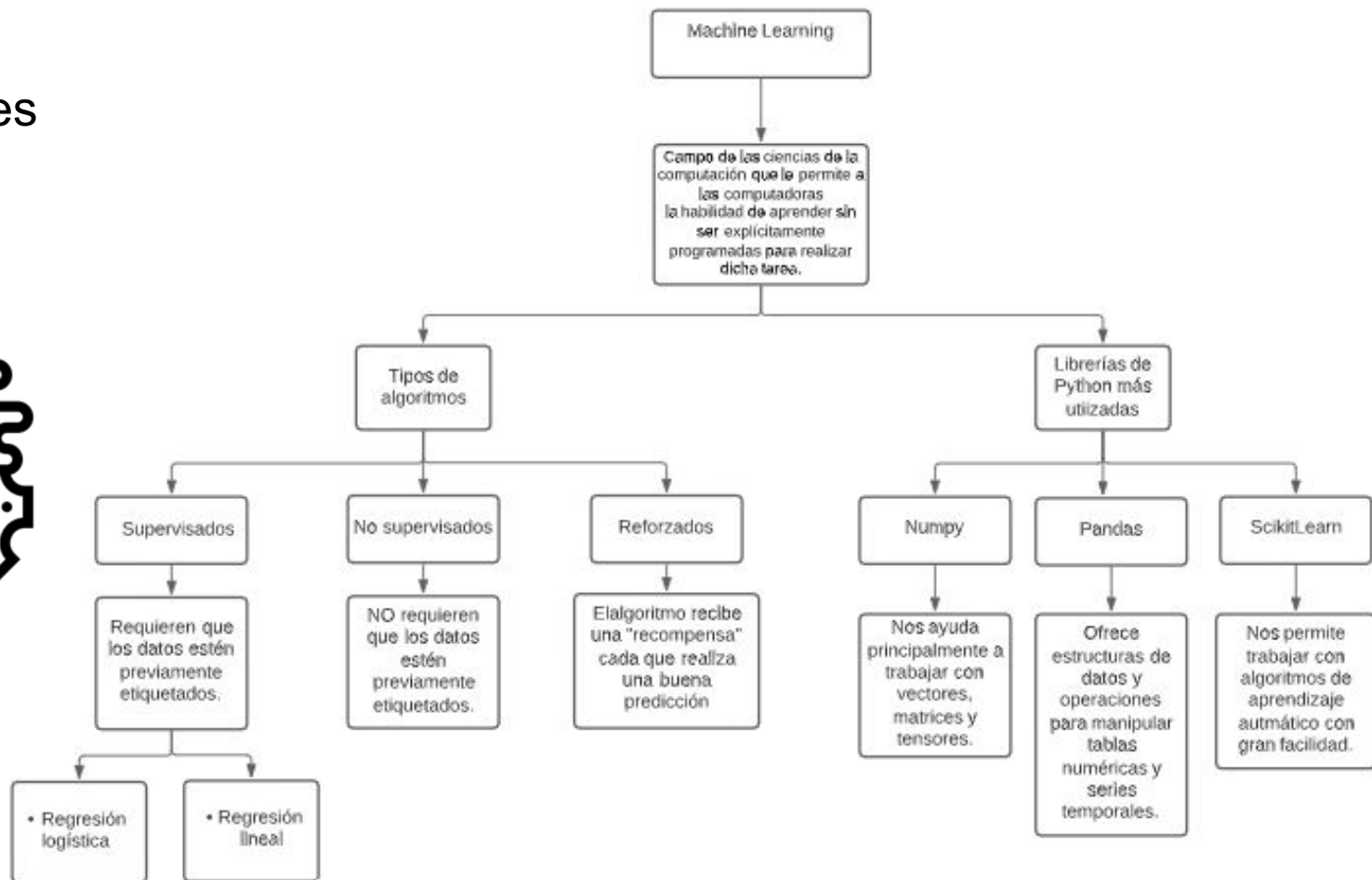
## Conclusiones



¿Qué aprendiste en esta sesión? Te invitamos a compartir tus conclusiones en clase.



## Conclusiones



# ¿Preguntas?