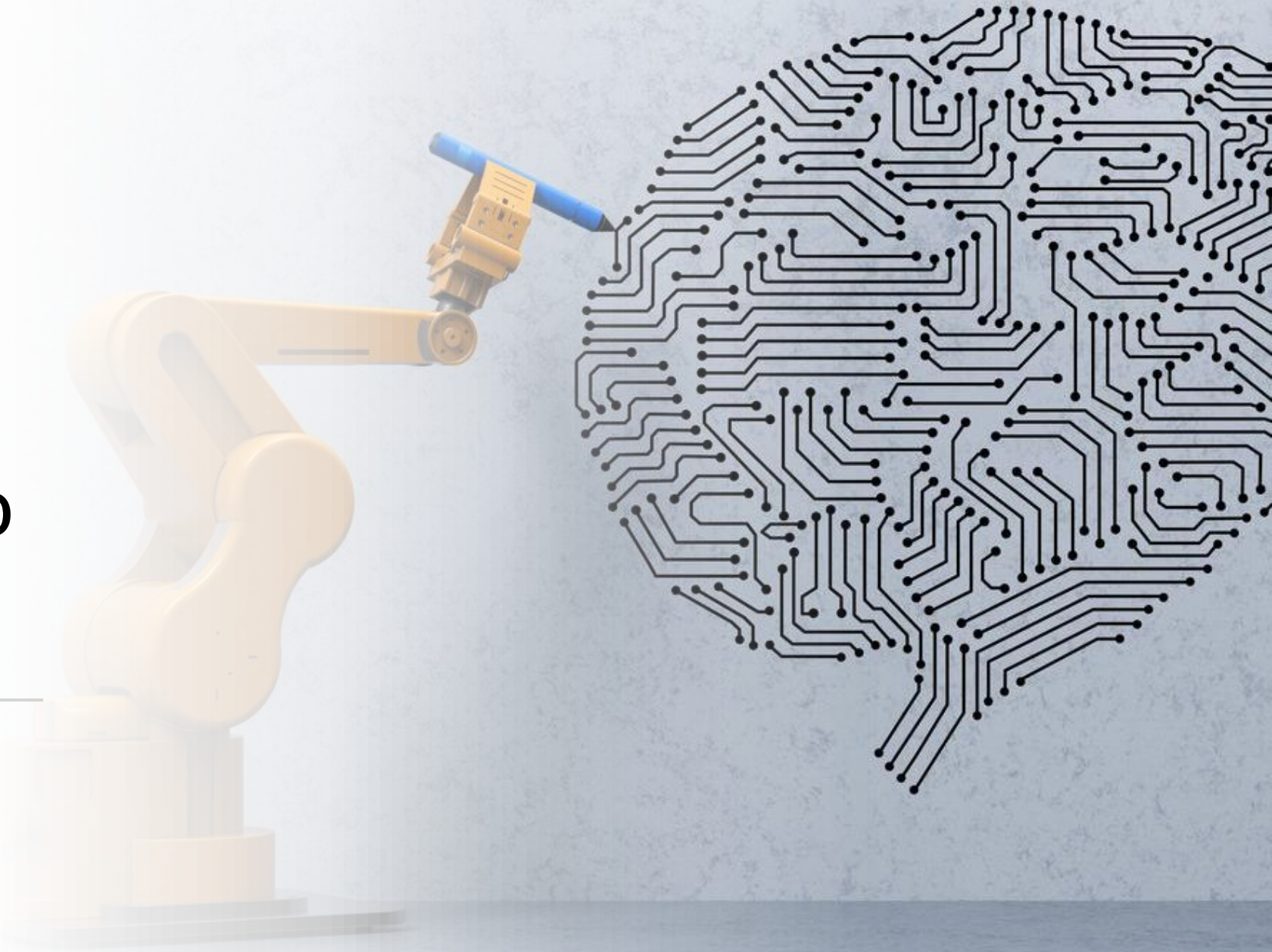




Reconocimiento de Patrones

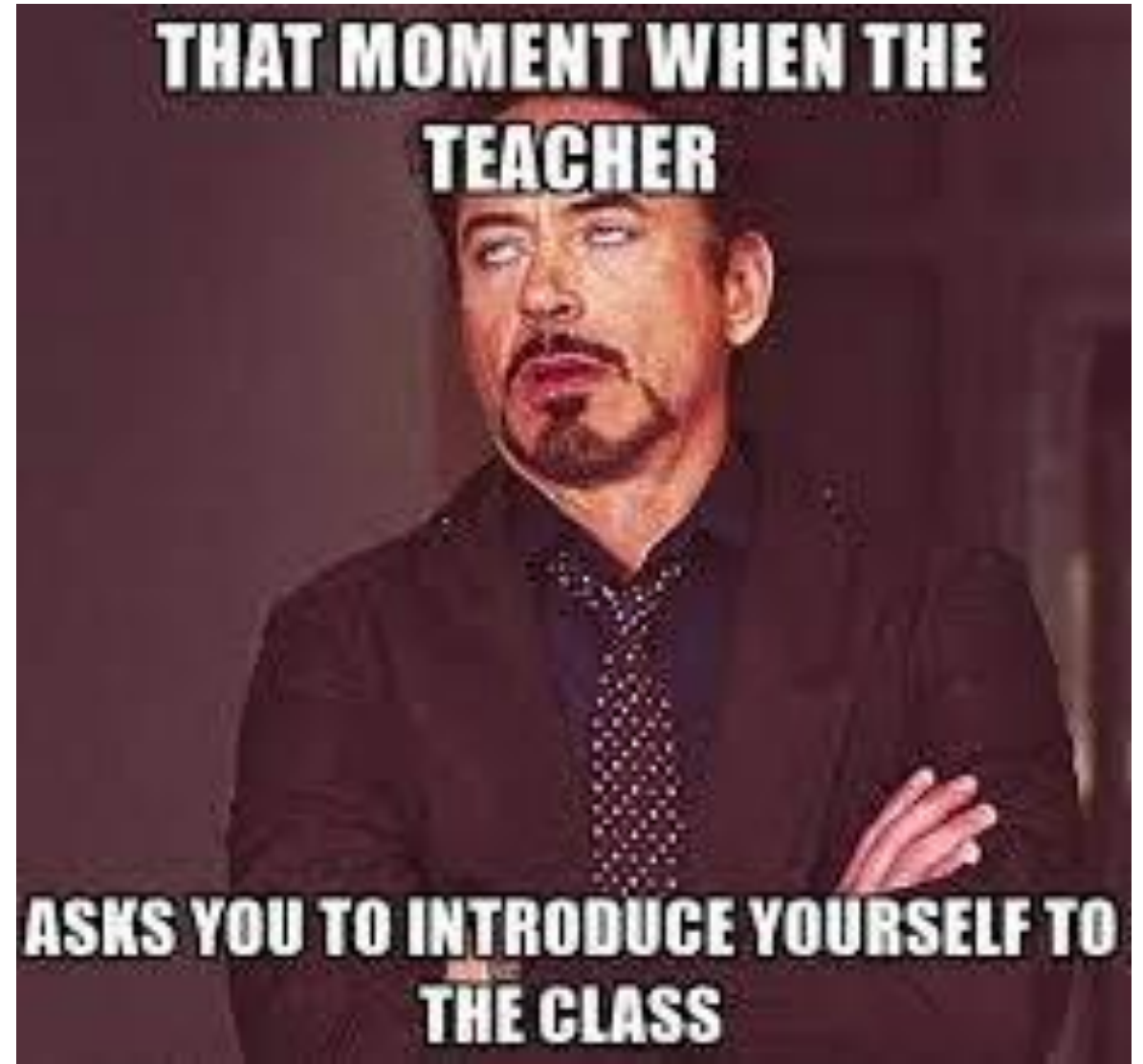
2027-1

Rubén Omar Azuara Domínguez



Presentaciones

- ¿Quiénes son?
- ¿Cuál carrera están estudiando y en qué semestre están?
- ¿Qué esperan del curso?
- ¿En que esperan trabajar?





Temario

1. Conceptos básicos de reconocimiento de patrones
2. Funciones de decisión.
3. Clasificación de patrones por medio de funciones de distancia.
4. Clasificación de patrones por medio de funciones de similitud.
5. Reconocimiento de patrones sintácticos.

Temario

1. Conceptos básicos de reconocimiento de patrones.
2. Extracción de características.
3. Selección y reducción de dimensionalidad.
4. Evaluación de modelos.
5. Funciones de decisión.
6. Clasificación de patrones por medio de funciones de distancia.
7. Clasificación de patrones por medio de funciones de similitud.
8. Clasificadores de Aprendizaje Automático.
9. Aprendizaje Profundo.
10. Patrones Estructurales y Procesamiento de Lenguaje Natural.

Información de curso



Horario

Lunes y Viernes de 07:00 a
08:30 hrs.



Lugar

Salón S227 conjunto sur



Contacto

ruben.azuara@ingenieria.
unam.edu



Datos útiles

15 semanas

28 sesiones

Días inhábiles: 2 y 16 de
noviembre

Evaluación



Tareas

15%



Practicas/Examen 45%



Proyecto

40%

¡Importante!

- El que una calificación mayor a .5 suba a la inmediata superior queda a criterio del profesor con base en el desempeño del alumno durante el semestre.

Tareas

- Los trabajos son individuales.
- Anotar nombre y numero de cuenta en la parte superior.
- Es posible entregar tareas posterior a la fecha de entrega con una penalización del 10% por cada día de retraso (hasta llegar a 0).

Prácticas

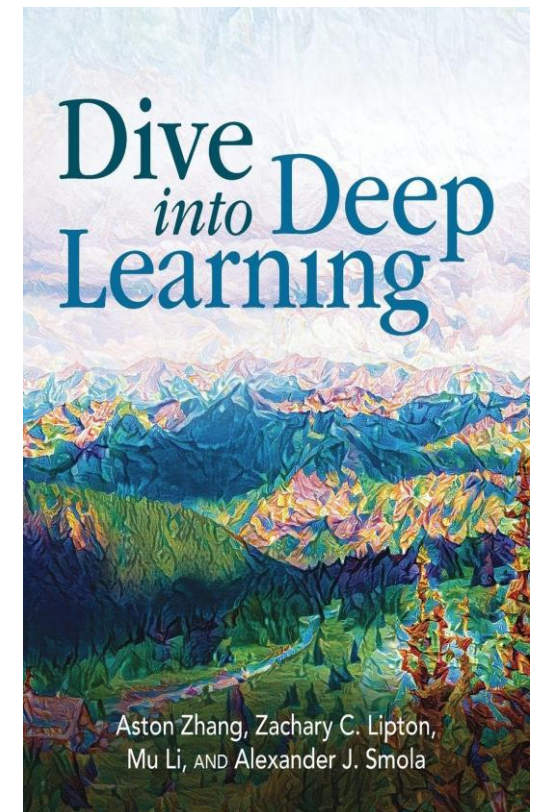
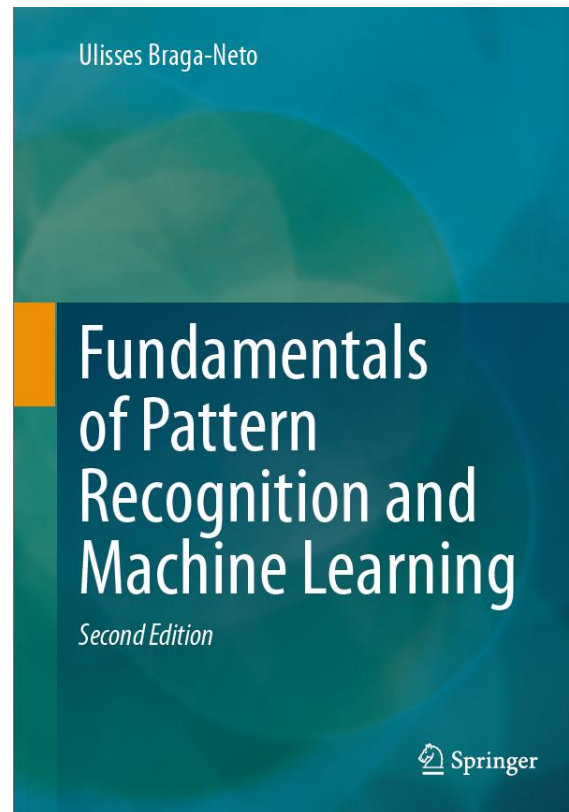
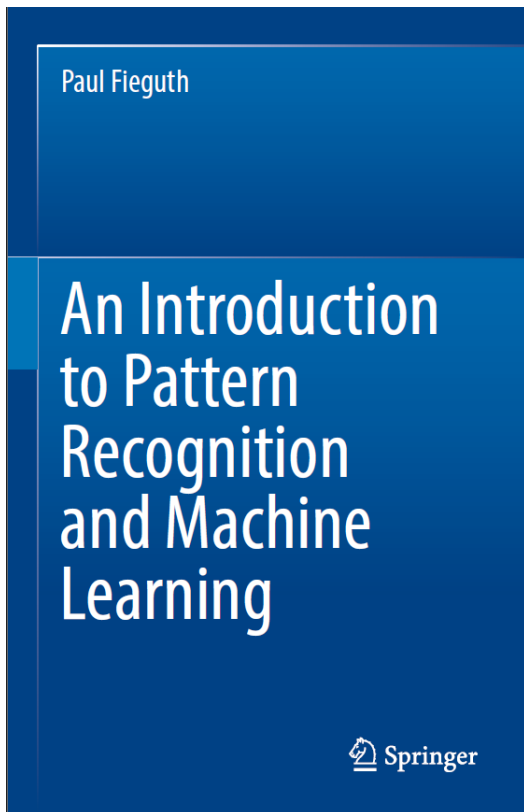
- Las prácticas son por equipos.
- Se evaluarán 3 componentes:
 - Implementación: código reproducible.
 - Reporte escrito del trabajo en formato IEEE.
 - Presentación oral de todos los integrantes del equipo.
- No se aceptan prácticas posteriores a la fecha de entrega.

Proyecto Final

- Aplicación de alguna de las opciones propuestas de Reconocimiento de Patrones.
- Se evaluarán los mismos 3 componentes solicitados en las prácticas.
- Fecha de entrega tentativa: 30 de noviembre
- Se darán más detalles posteriormente.

Todos los integrantes del equipo deberán subir el reporte así como los archivos de código correspondientes al classroom.

Recursos



Otros comentarios



El lenguaje de programación que se utilizará para este curso es Python.



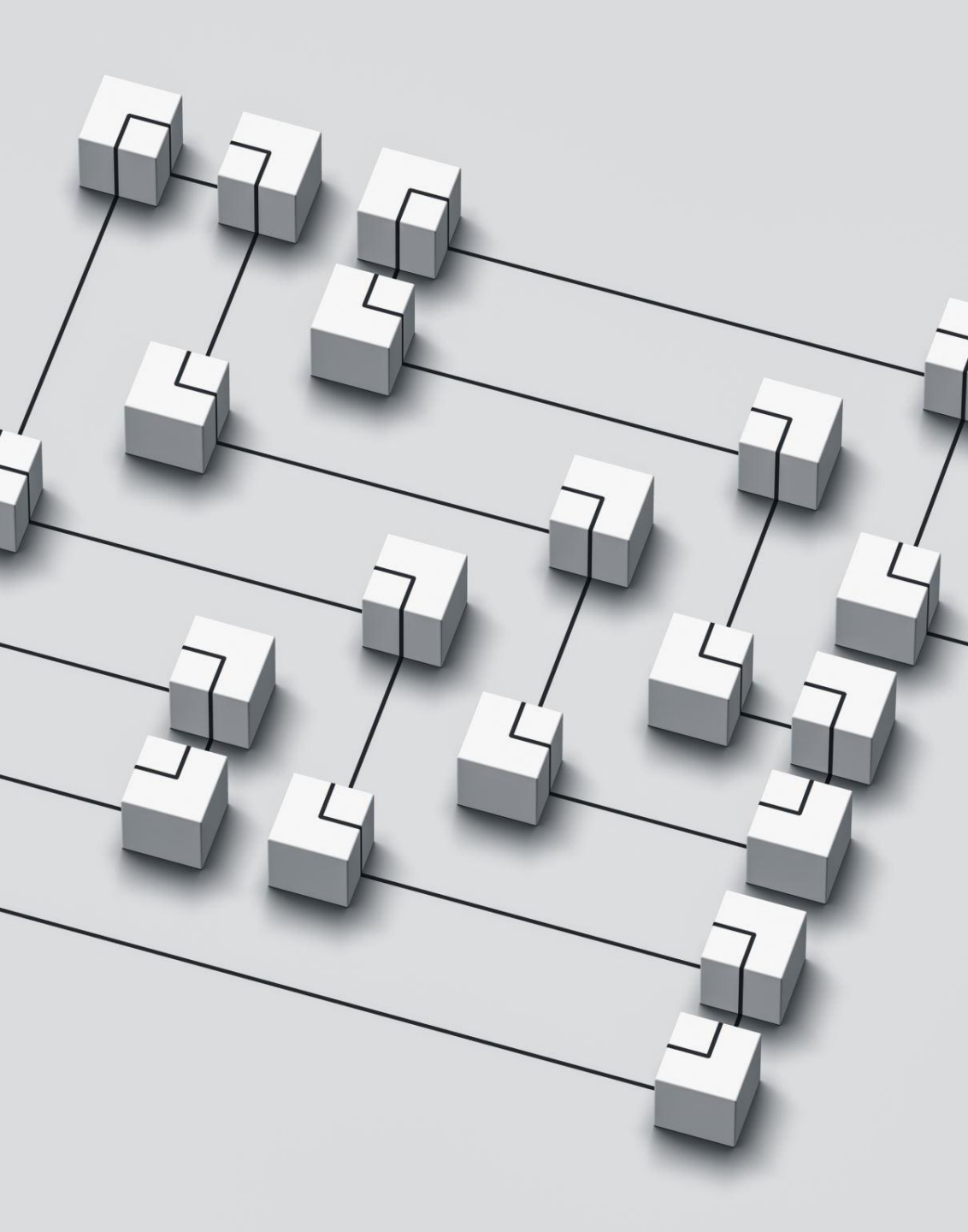
Los oyentes son bienvenidos y serán tratados como estudiantes regulares.



Se utilizará Google Classroom para dar seguimiento al curso: tz6dpily



Y más importante... la idea es que ¡aprendan los contenidos!



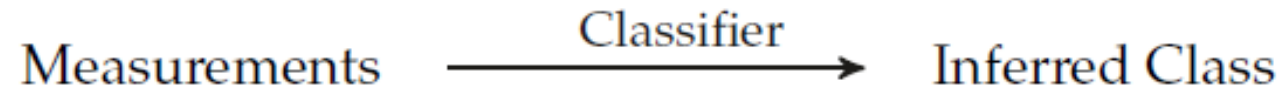
¡Empezamos!

1. Conceptos básicos de reconocimiento de patrones.

¿Qué es el reconocimiento de patrones?

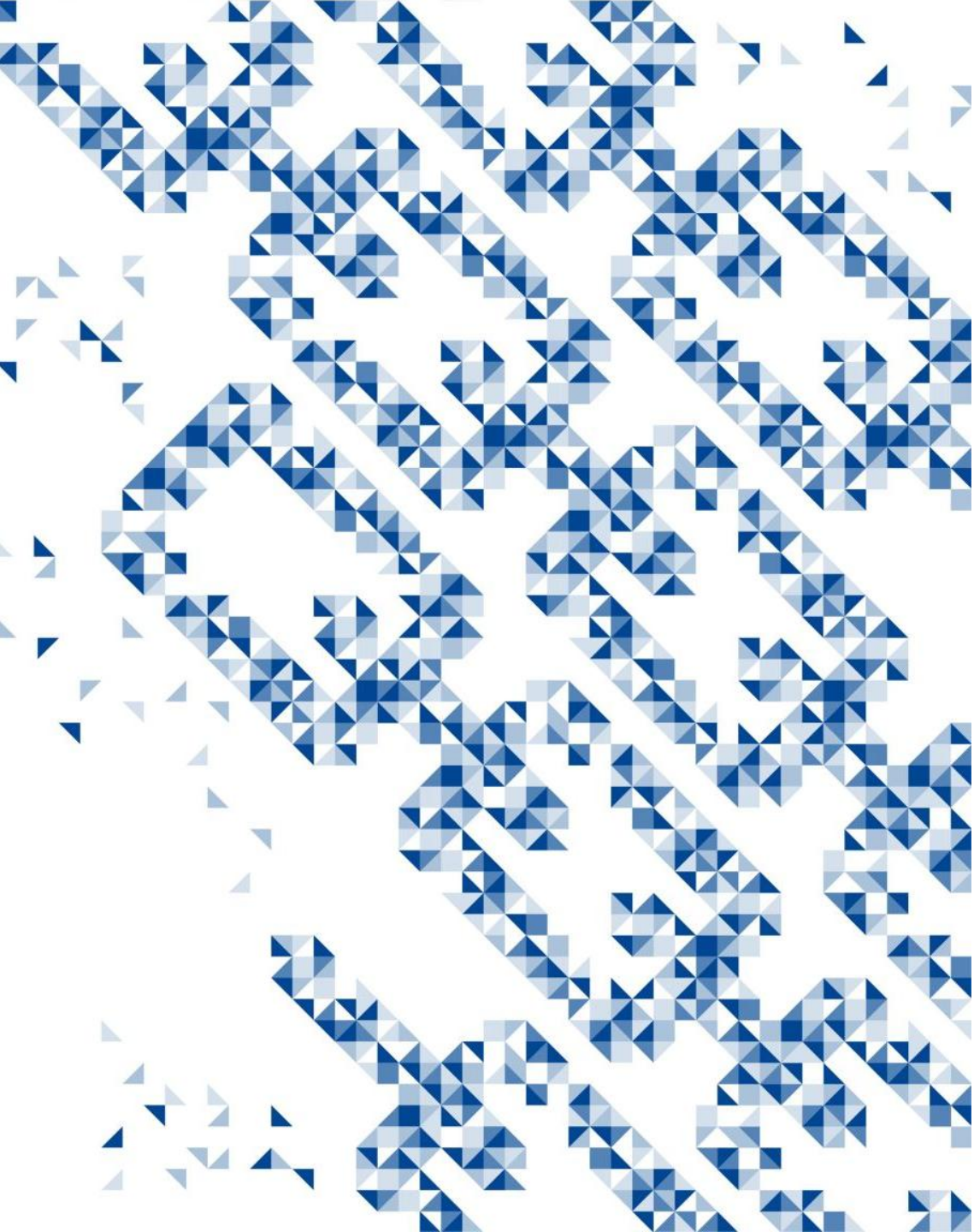
- Durante mucho tiempo se ha buscado realizar inferencias de alguna cantidad desconocida con base mediciones relacionadas.
- Si cantidad desconocida es...
 - Continua -> teoría de la estimación
 - Discreta -> reconocimiento de patrones

El reconocimiento de patrones se enfoca en determinar una identidad particular (la **clase** del patrón) basándose en información medida, en donde el proceso que selecciona la clase se conoce como clasificador.



La tarea central del reconocimiento de patrones consta de 2 partes

1. Extracción de características.
2. Clasificación.



- El reconocimiento de patrones es un proceso mediante el cual las entradas se miden, analizan y luego se clasifican como pertenecientes a una de un conjunto de clases.
- Categorización de la información de entrada en clases identificables a través de la extracción de características o atributos significativos de los datos con respecto al fondo o detalles irrelevantes.

- Este proceso se lleva a cabo de forma cotidiana en nuestras vidas...



- ¿Qué está frente a ti cuando caminas?



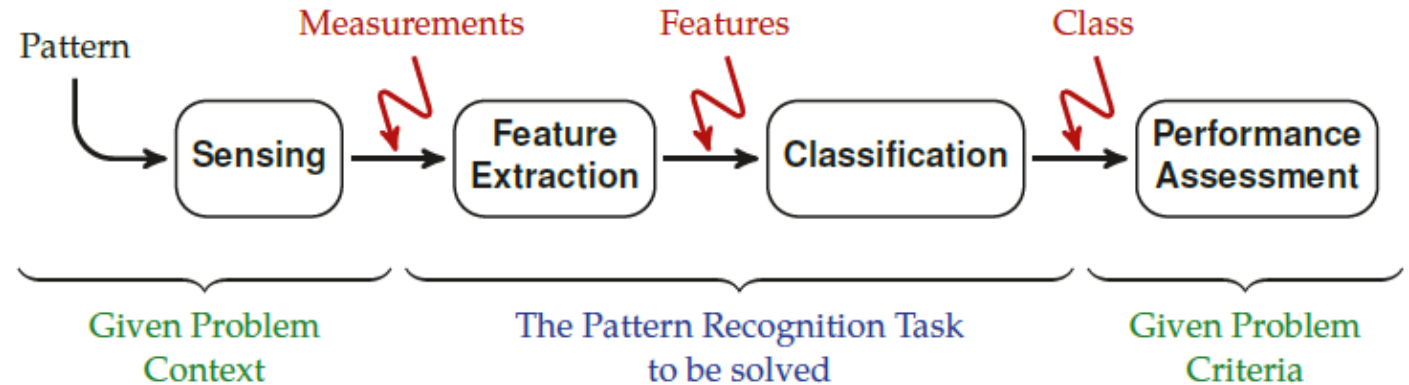
- ¿Es seguro cruzar la calle?

- 
- ¿Qué música estas escuchando? Conocida o desconocida, genero, nombre del compositor o grupo.
 - Oliste algo en tu casa... La comida ya está lista, algo de la estufa se esta quemando, algo del refrigerador se descompuso.
 - Desde el punto de vista de la experiencia humana el reconocimiento de patrones es el proceso en que alguna forma de entrada sensorial es detectada, analizada y reconocida (clasificada) ya sea consciente o inconscientemente.

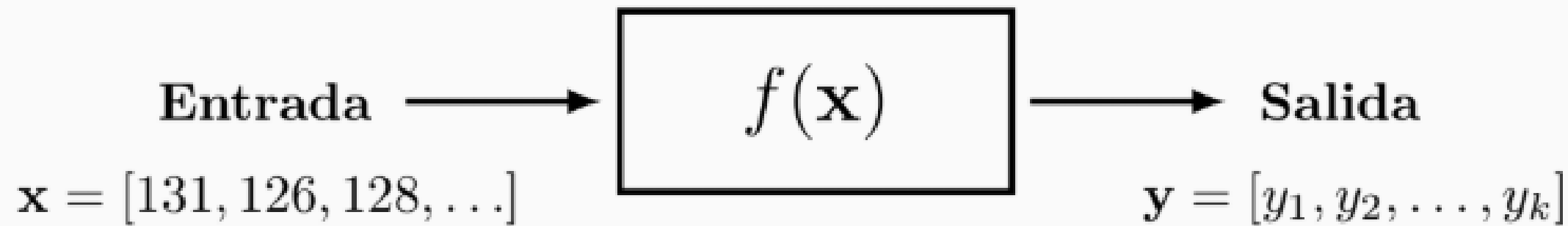
Pero antes de seguir avanzado hay que definir algunos términos:

- Patrón: es cualquier objeto, proceso o evento que puede ser descrito y distinguido de otros.
- Clase: una categoría o etiqueta a la que pertenece un patrón. Un conjunto de patrones con propiedades similares.
- Característica: es una medición individual que describe un aspecto del patrón.
- Clasificación: el proceso de asignar un patrón a una de varias clases predefinidas basándose en sus características.
- Entrenamiento: El proceso mediante el cual un sistema de reconocimiento de patrones aprende a partir de ejemplos conocidos (datos de entrenamiento) para poder clasificar nuevos patrones.

Una estructura conceptual para el reconocimiento de patrones es:



Un patrón es detectado dando lugar a mediciones, de las cuales se extraen características y se entregan a un clasificador, que se encarga de seleccionar la clase asociada con el patrón detectado.



$$y = \begin{cases} 0 & \text{Gato} \\ 1 & \text{Perro} \end{cases}$$

Problemas fundamentales en el diseño de un sistema de reconocimiento de patrones

1. Referente a la representación de la información que puede medirse de los objetos a ser reconocidos.

Cada cantidad medida describe una característica del patrón. El vector de características contiene toda la información medida disponible sobre los patrones.

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix}$$

Pero...

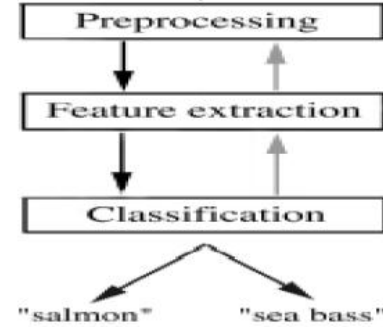
Ejemplo 1

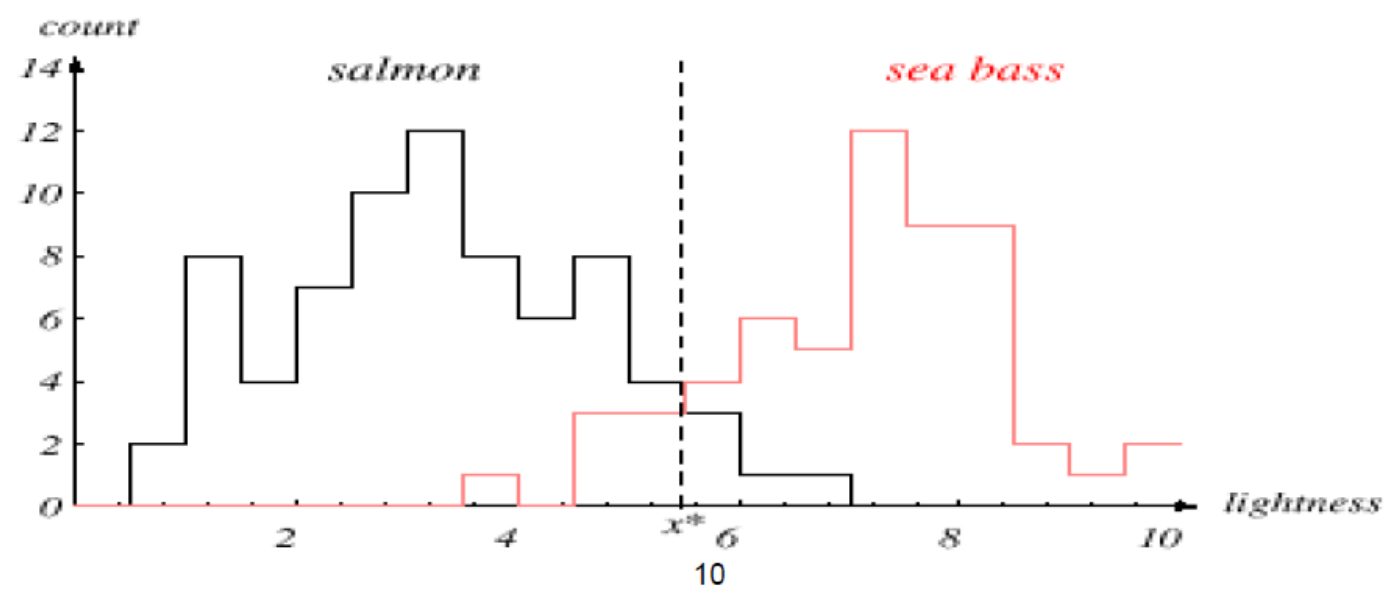
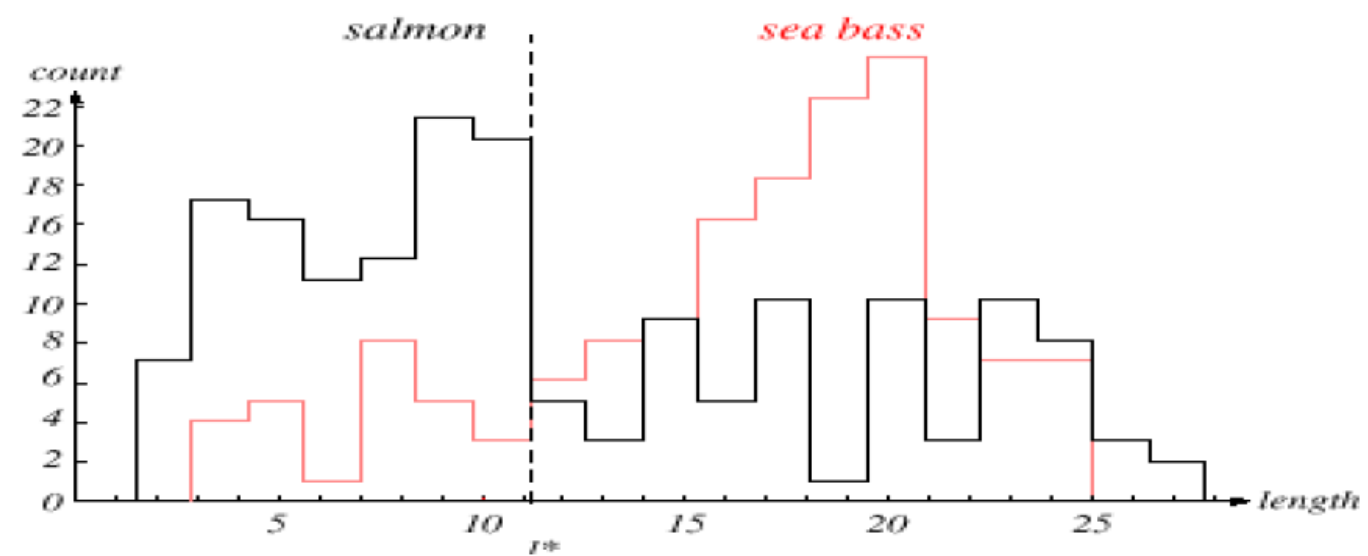
Weight (kg)	Class label
50	Unhealthy
60	Healthy
70	Healthy
80	Unhealthy

¿Tenemos suficiente información?

Ejemplo 2

¿Salmón o robalo?





2. Referente a la extracción de atributos o rasgos característicos de los datos de entrada recibidos y reducción de dimensionalidad.

¿Qué pasa si nuestro conjunto de datos tiene muchas características?

Ejemplo 1. Selección.

Breast Cancer Wisconsin (Diagnostic)

Conjunto de datos para la detección de cáncer de pecho. 569 instancias donde: 357 benignos, 212 malignos.

- a) radius (mean of distances from center to points on the perimeter)
- b) texture (standard deviation of gray-scale values)
- c) perimeter
- d) area
- e) smoothness (local variation in radius lengths)
- f) compactness ($\text{perimeter}^2 / \text{area} - 1.0$)
- g) concavity (severity of concave portions of the contour)
- h) concave points (number of concave portions of the contour)
- i) symmetry
- j) fractal dimension ("coastline approximation" - 1)

```
Index(['id', 'diagnosis', 'radius_mean', 'texture_mean', 'perimeter_mean',
      'area_mean', 'smoothness_mean', 'compactness_mean', 'concavity_mean',
      'concave points_mean', 'symmetry_mean', 'fractal_dimension_mean',
      'radius_se', 'texture_se', 'perimeter_se', 'area_se', 'smoothness_se',
      'compactness_se', 'concavity_se', 'concave points_se', 'symmetry_se',
      'fractal_dimension_se', 'radius_worst', 'texture_worst',
      'perimeter_worst', 'area_worst', 'smoothness_worst',
      'compactness_worst', 'concavity_worst', 'concave points_worst',
      'symmetry_worst', 'fractal_dimension_worst', 'Unnamed: 32'],
      dtype='object')
(569, 33)
```

$$x = [x_1, x_2, x_3, \dots, x_{30}]$$

¿Realmente todas las características nos aportan información relevante?

Característica	Correlación
Área	0.82
Textura	0.79
Concavidad	0.85
Simetría	0.12
Fractal	0.05

$$x = [x_{\text{area}}, x_{\text{textura}}, x_{\text{concavidad}}]$$

Ejemplo 2.

¿Y si tenemos un vector de características muchísimo más grande?

Tamaño: 3000 x 1800

Al aplanarlo en un vector $x \in R^{5,400,000}$

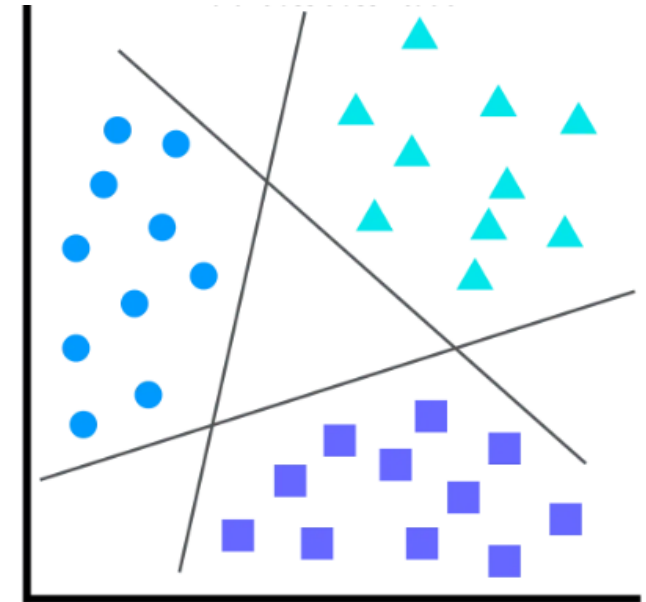
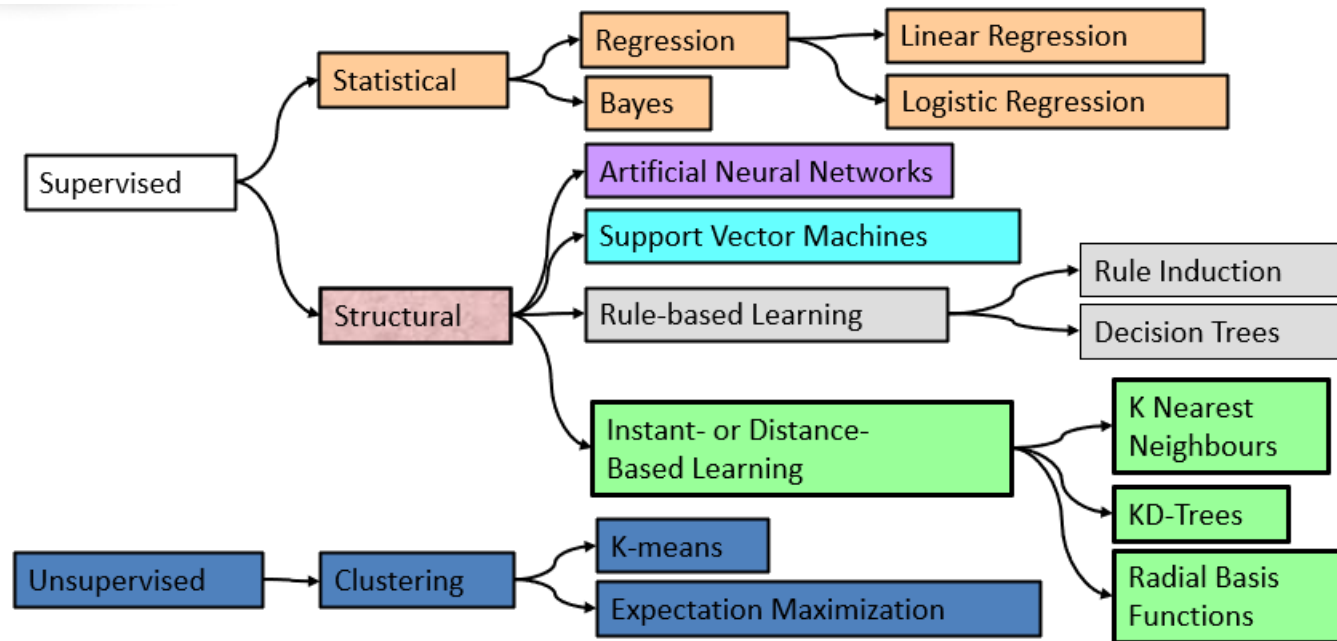
Reducción de dimensionalidad

El vector pasa de 5,400,000 $\rightarrow$ 500

$$z \in R^{500}$$

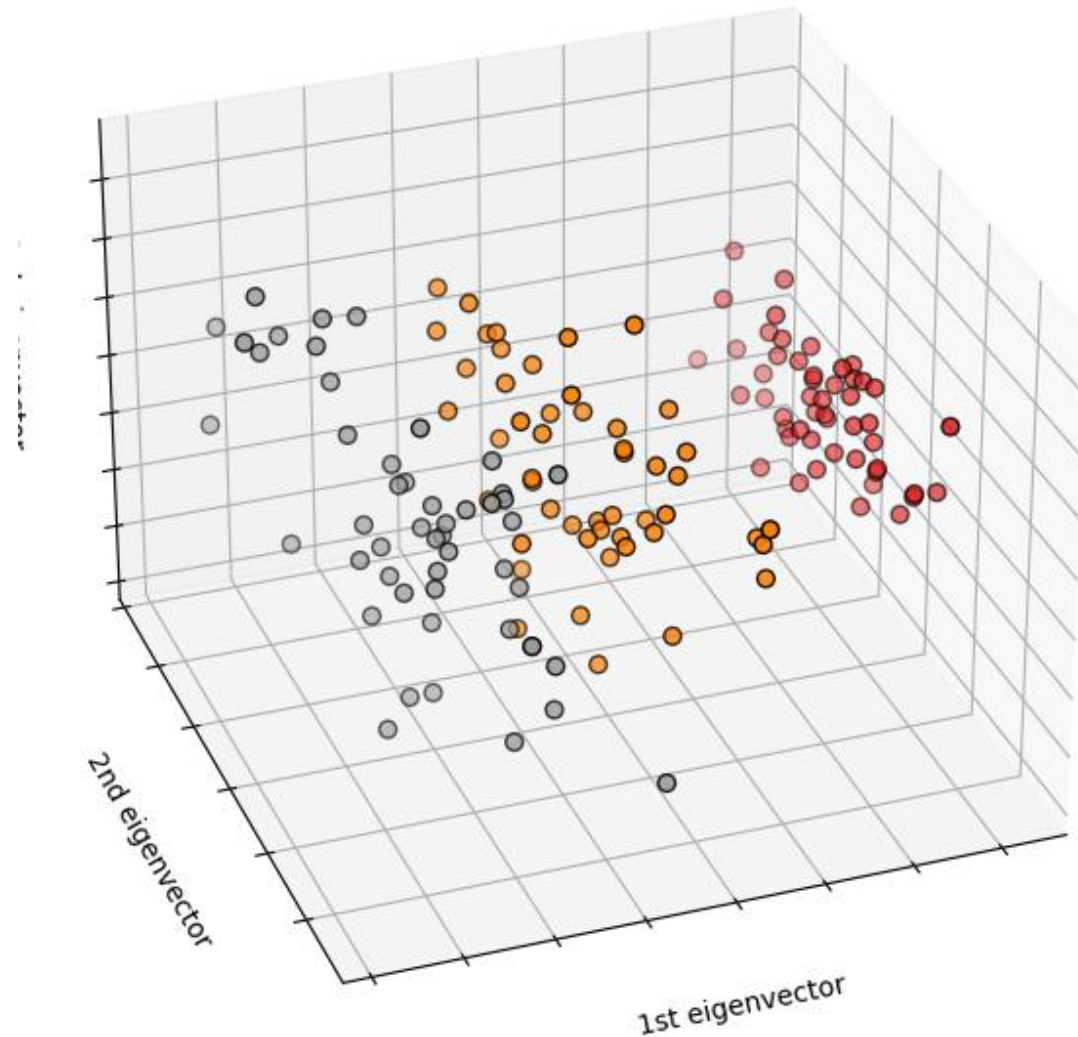


3. Determinación de procedimientos de decisión óptimos, necesarios para la clasificación e identificación.



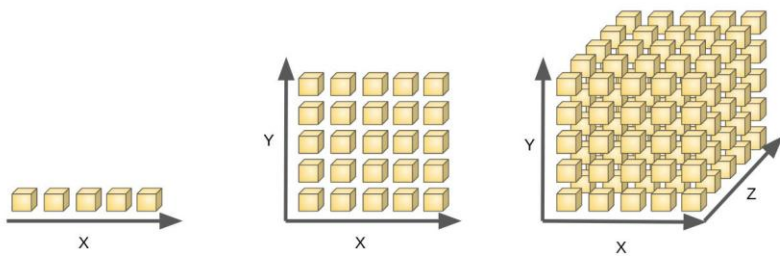
La maldición de la dimensionalidad

- En el contexto del análisis de datos y el aprendizaje automático, las dimensiones se refieren a las características o atributos de los datos.



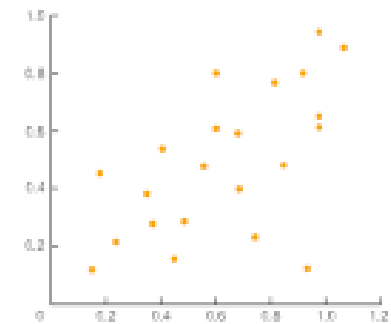
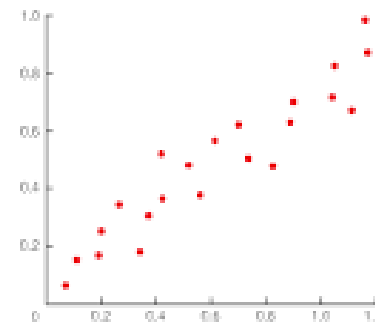
¿Cómo se produce la maldición de la dimensionalidad?

- A medida que añadimos más dimensiones a nuestro conjunto de datos, el volumen del espacio aumenta exponencialmente.

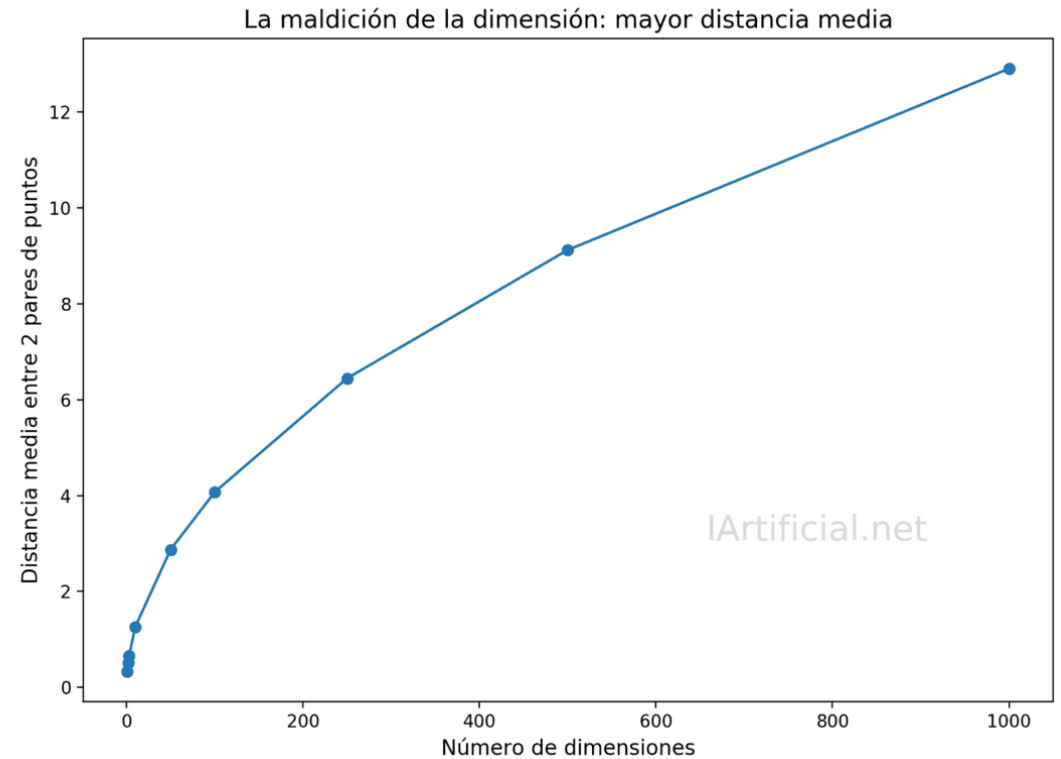


Esto provoca:

- Dispersión de datos
- Mayor computación
- Sobreajuste
- Las distancias pierden sentido
- Degradación del rendimiento
- Retos de visualización



-
- La maldición de la dimensionalidad se manifiesta de dos formas:
 1. La distancia media entre los datos aumenta con el número de dimensiones
 2. La variabilidad de la distancia disminuye exponencialmente con el número de dimensiones (éste es el verdadero problema).



¿Cómo se resuelve?

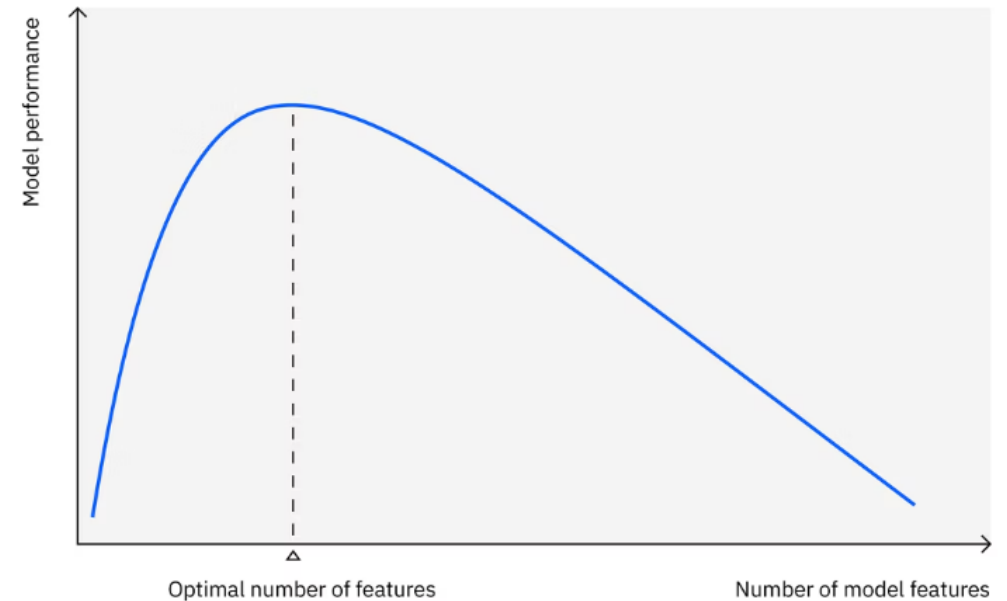
- Reducción de la dimensionalidad.

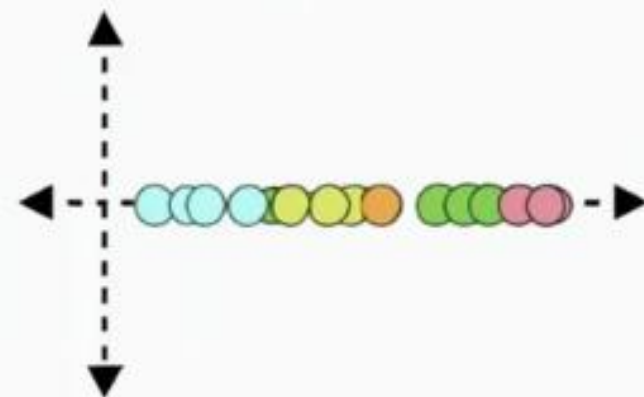
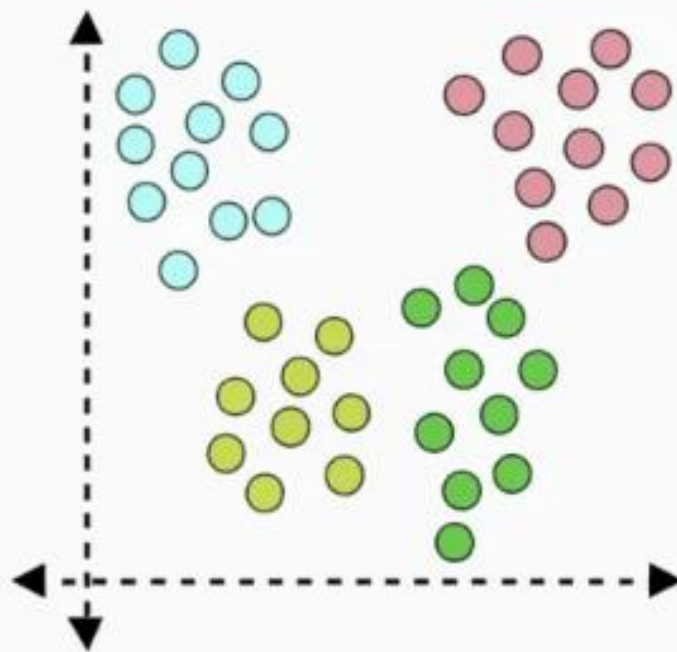
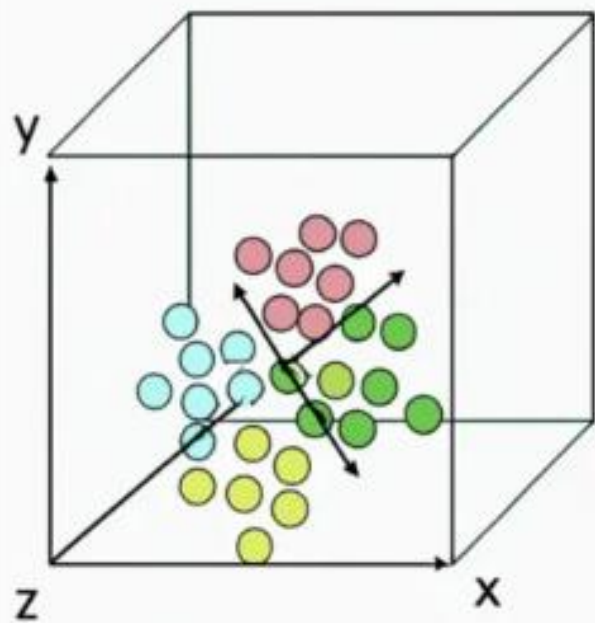
Este proceso busca reducir el numero de variables aleatorias de forma que se obtenga un conjunto de variables principales.

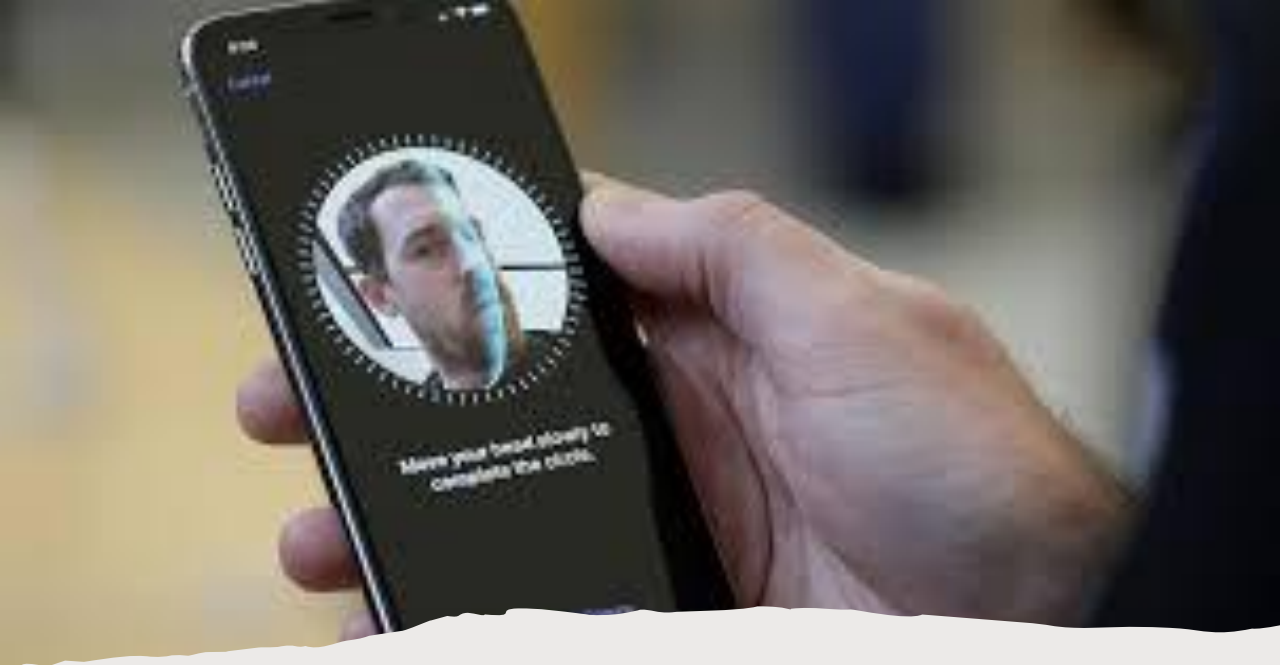
Esto permite:

- Conservar la información más importante de los datos.
- Descartar características redundantes o menos importantes.

Pero, al eliminar características esto inevitablemente implica perdida de la información.

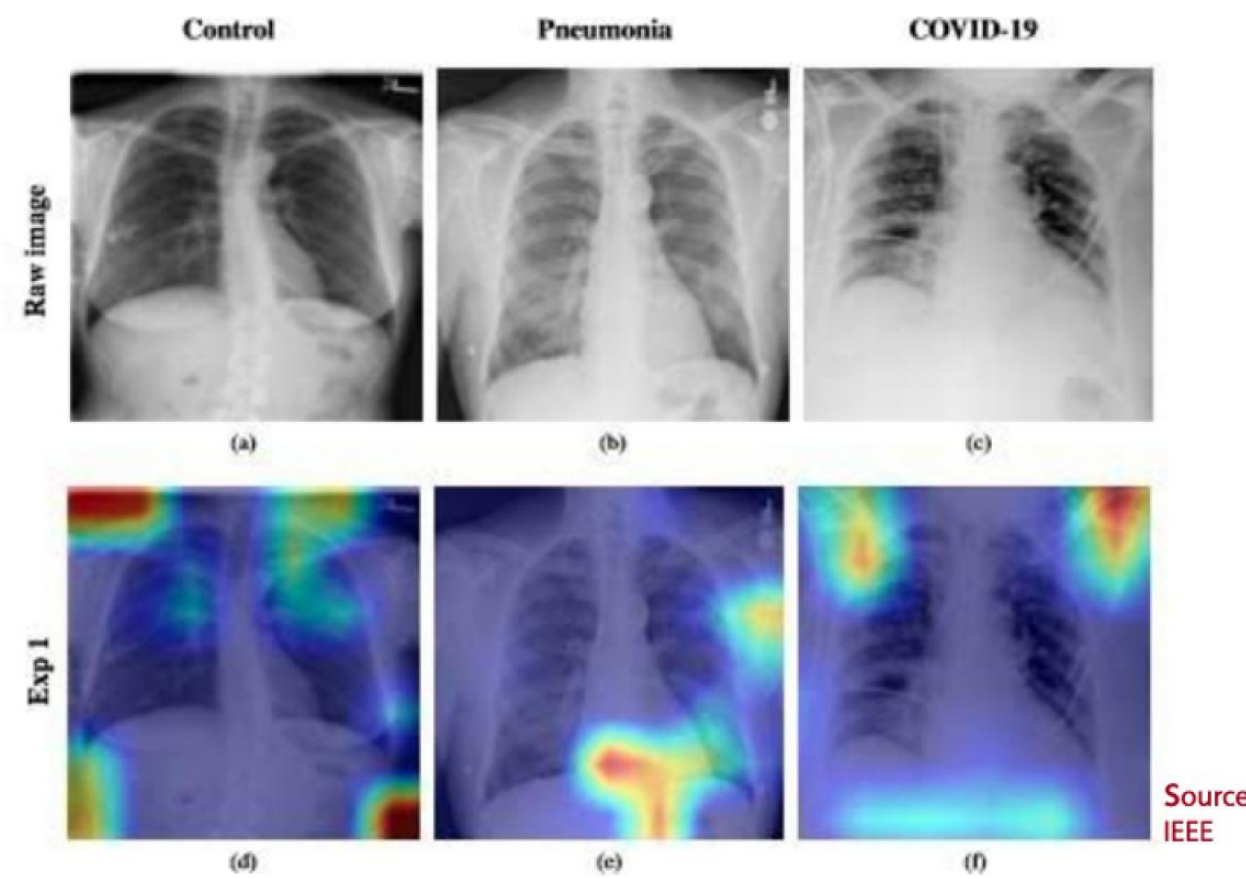
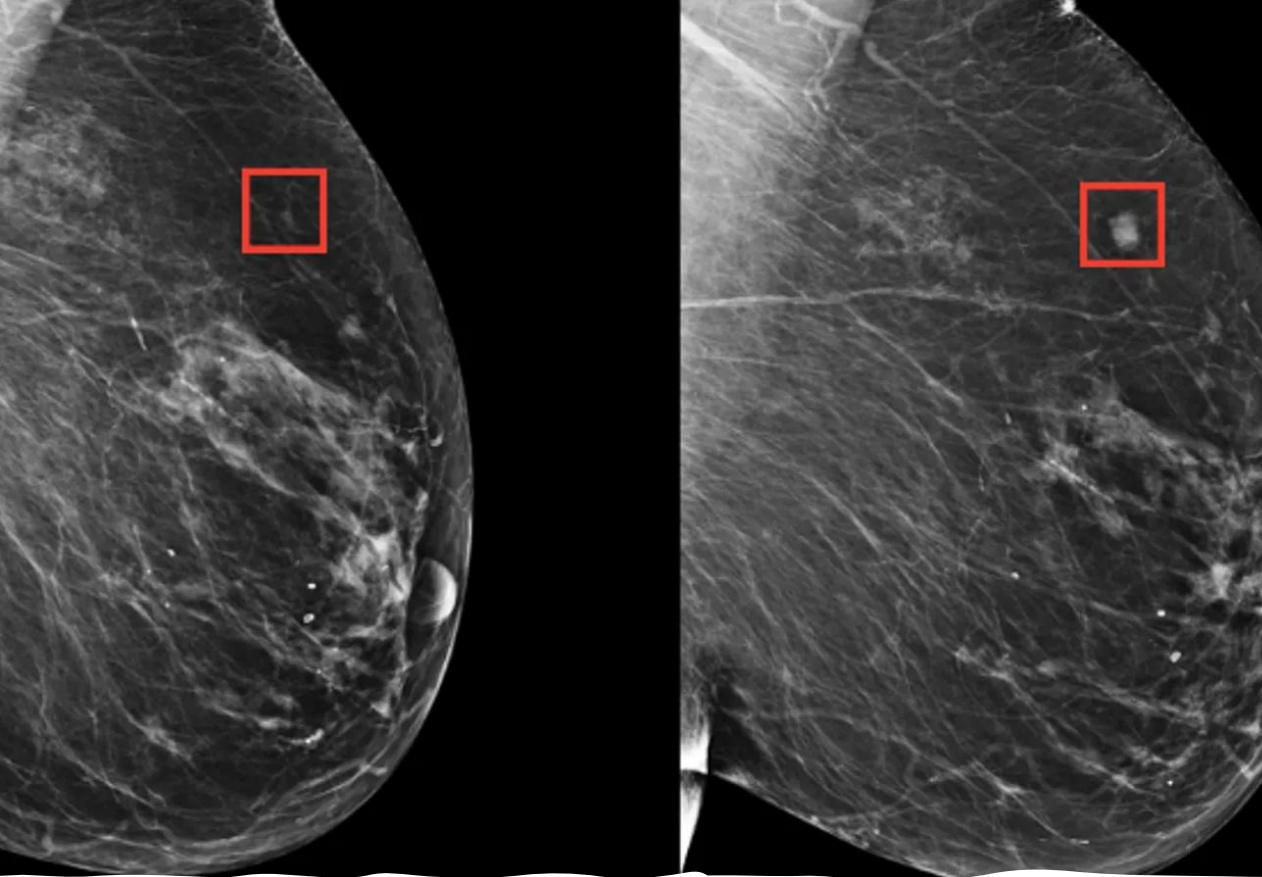




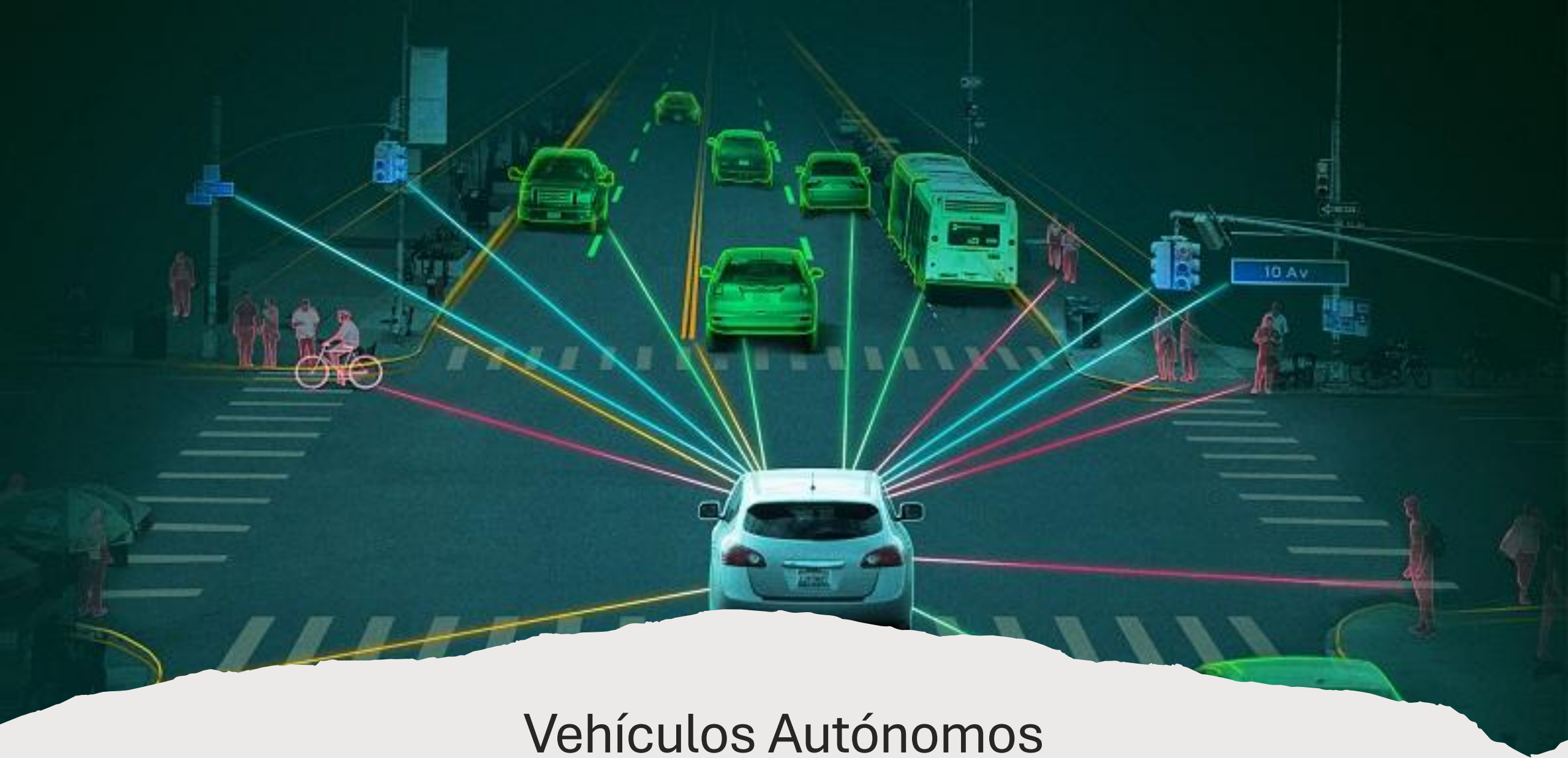


Algunos ejemplos de sistemas de RP...

- **Visión Computacional**
Reconocimiento Facial



Diagnostico Médico por Imagen



Vehículos Autónomos



- **Procesamiento del Lenguaje Natural**

Asistentes Virtuales

Traducción Automática





- **Biometría**

Reconocimiento de huellas dactilares

Reconocimiento de iris



Para la próxima clase...

- Extracción de características.

