CIS

Introducción al Machine Learning
Temas básicos en Álgebra Lineal

APRENDIZAJE DE MÁQUINA
FERNANDO TESHIMA

UNIVERSIDAD
DE LIMA

Asistencia

Martes: Presencial

Jueves: Virtual por Zoom

Ambas clases serán grabadas

Evaluaciones

- EC1 (50%)
  - Semana 5 (40%): Trabajo de investigación
    - Componente grupal (70%)
    - Componente individual (30%)
  - Semana 7 (50%): Examen
  - Continuo (10%): Participación

- EC2 (50%)
  - Semana 12 (30%): Trabajo de investigación
    - Componente grupal (70%)
    - Componente individual (30%)
  - Semana 14 (50%): Examen
  - Semana 15 (20%): Exposición
    - Componente grupal (30%)
    - Componente individual (70%)

<table>
<tr><td colspan="6">Nota del Curso
(100%)</td></tr>
<tr><td colspan="3">EC1
(50%)</td><td colspan="3">EC2
(50%)</td></tr>
<tr><td>Trabajo Inv Grup
(20%)</td><td>Examen
(25%)</td><td>Particip
ación
(5%)</td><td>Trabajo Inv Grup
(15%)</td><td>Examen
(25%)</td><td>Expo
(10%)</td></tr>
</table>

Datos de Contacto

fteshima@ulima.edu.pe

mensajes en BlackBoard

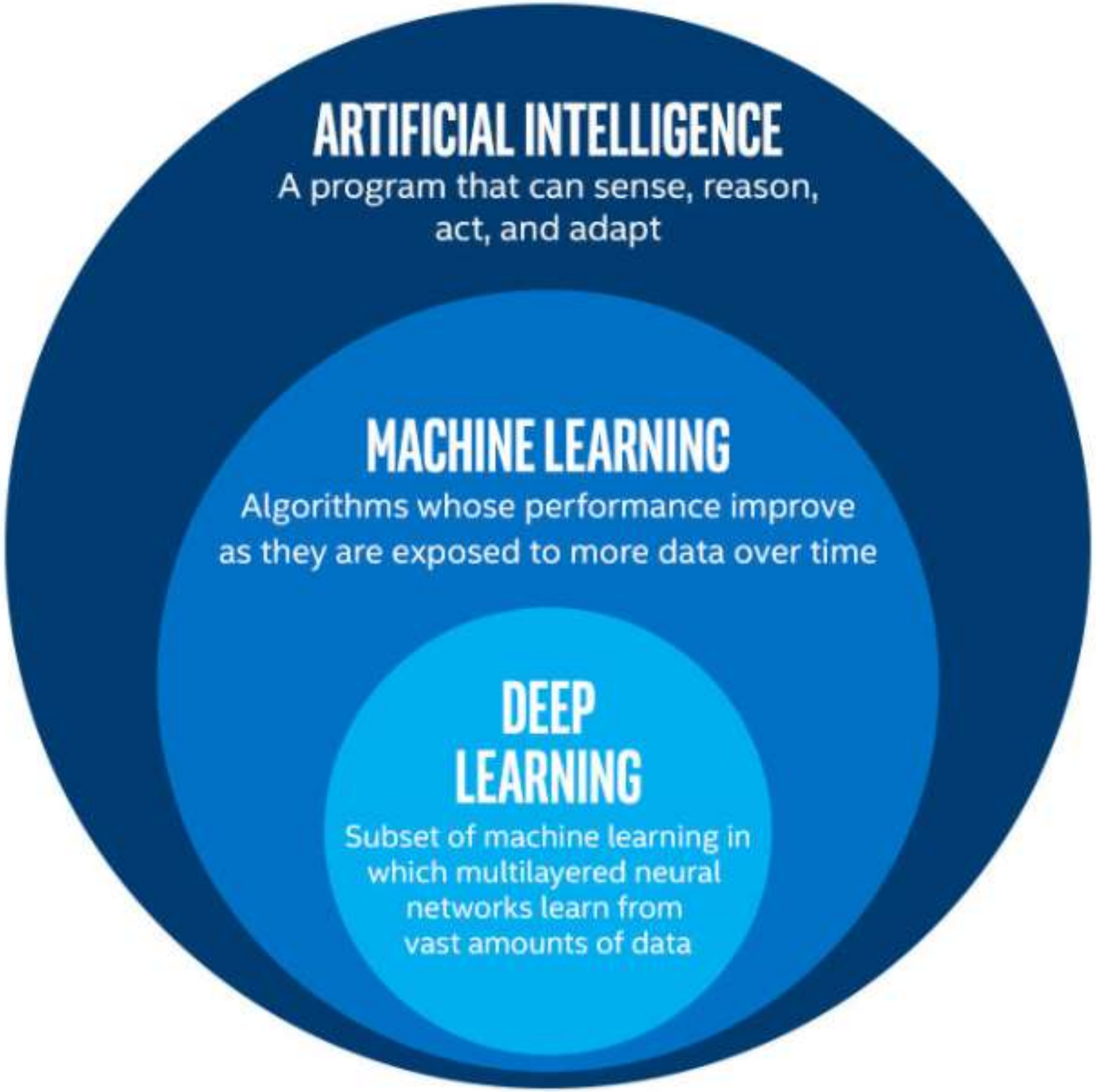
martes 3pm-4pm

Aprendizaje de Máquina (Machine Learning)

MSc. Fernando Teshima

Definición Formal

- Según Mitchell (1997) “Un programa de computador se dice que aprende de la experiencia **E** con respecto a una clase de tarea **T** y con una medida de rendimiento **P**, si su rendimiento en la tarea **T**, medido por **P**, mejora con la experiencia **E**”



Fuente: https://towardsdatascience.com/cousins-of-artificial-intelligence-dda4edc27b55

Aplicaciones Prácticas

Agrupar clientes por el comportamiento de compra

Gente que compra X producto también compra producto Y

Análisis de lenguaje (acentos, pronunciación)

Detección de transacciones fraudulentas

Tipos de ML

Supervisado

Predicción y etiquetado

No Supervisado

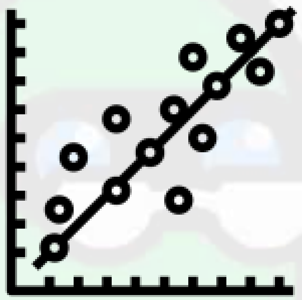
Identifica clústeres

Reforzado

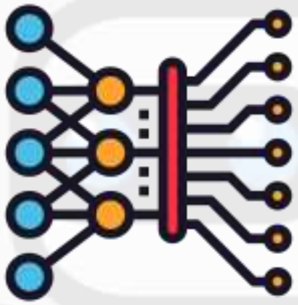
Aprende de errores

Aprendizaje Supervisado

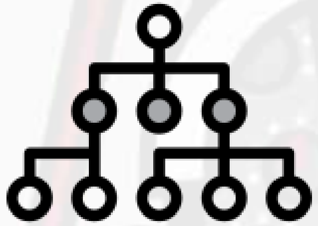
- Algoritmos que aprenden, en base a ejemplos, una función que relaciona entradas a salidas.



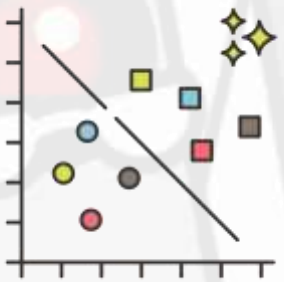
Regresión



Redes Neuronales

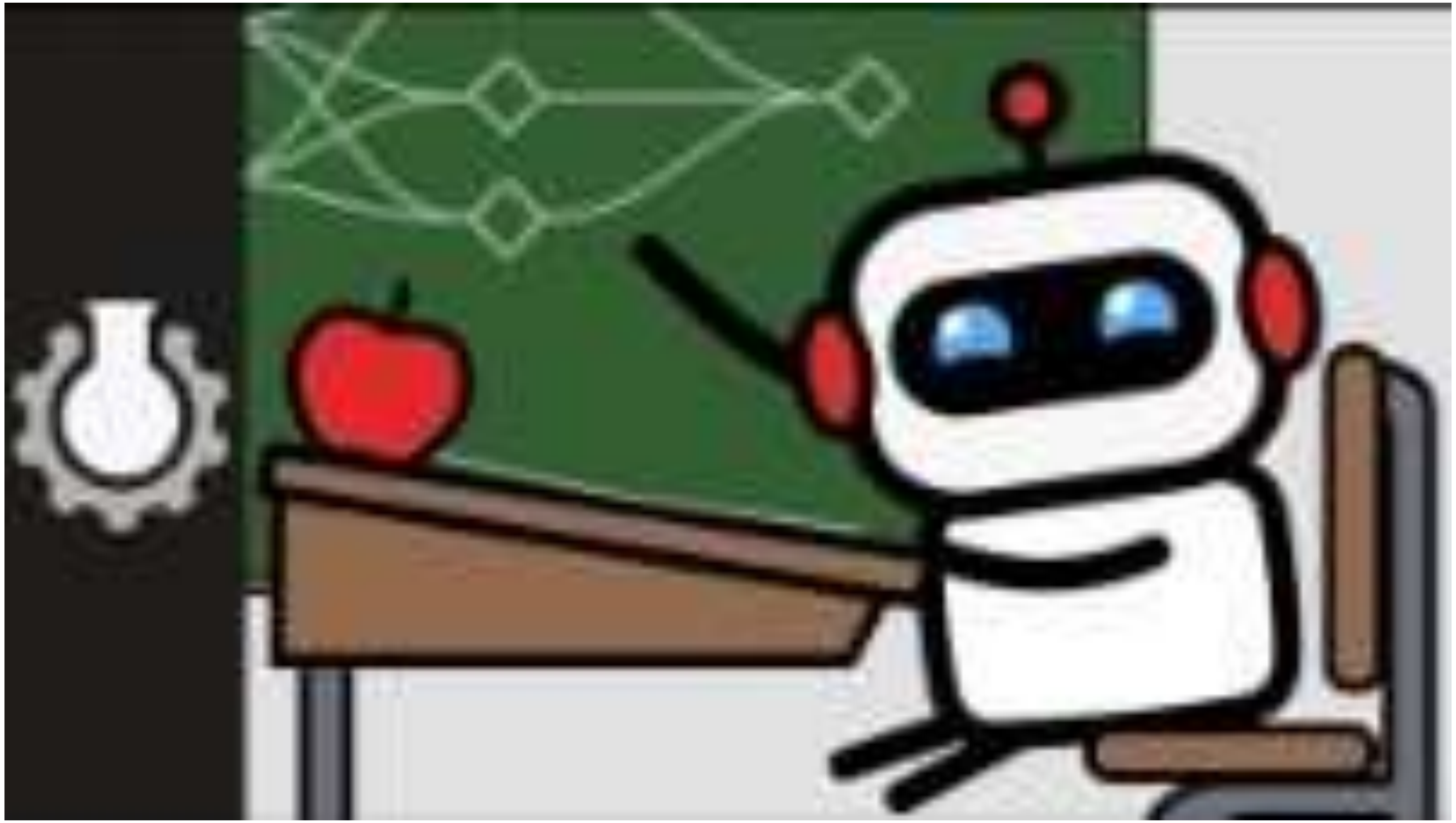


Árboles de Decisión

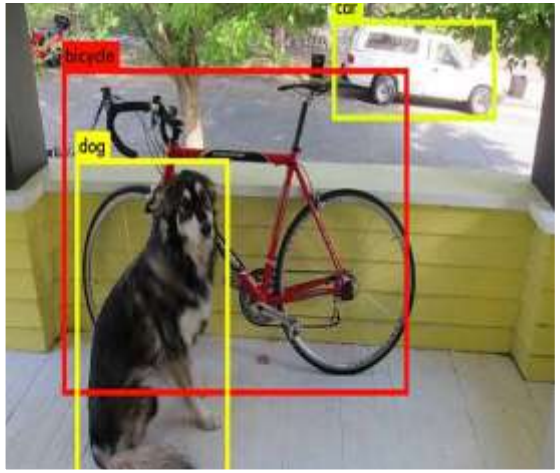


SVM

Aprendizaje Supervisado



Aplicaciones



Reconocimiento de Objetos

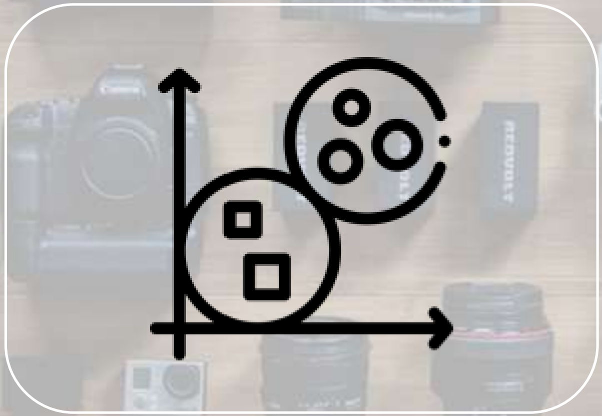
Tiempo de viaje

Filtros spam

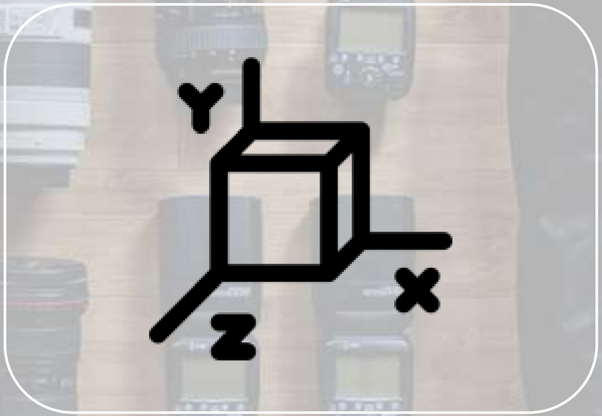
¿Más ejemplos?

Aprendizaje No Supervisado

- No existe etiqueta de salida
- El objetivo es descubrir relaciones entre la data existente

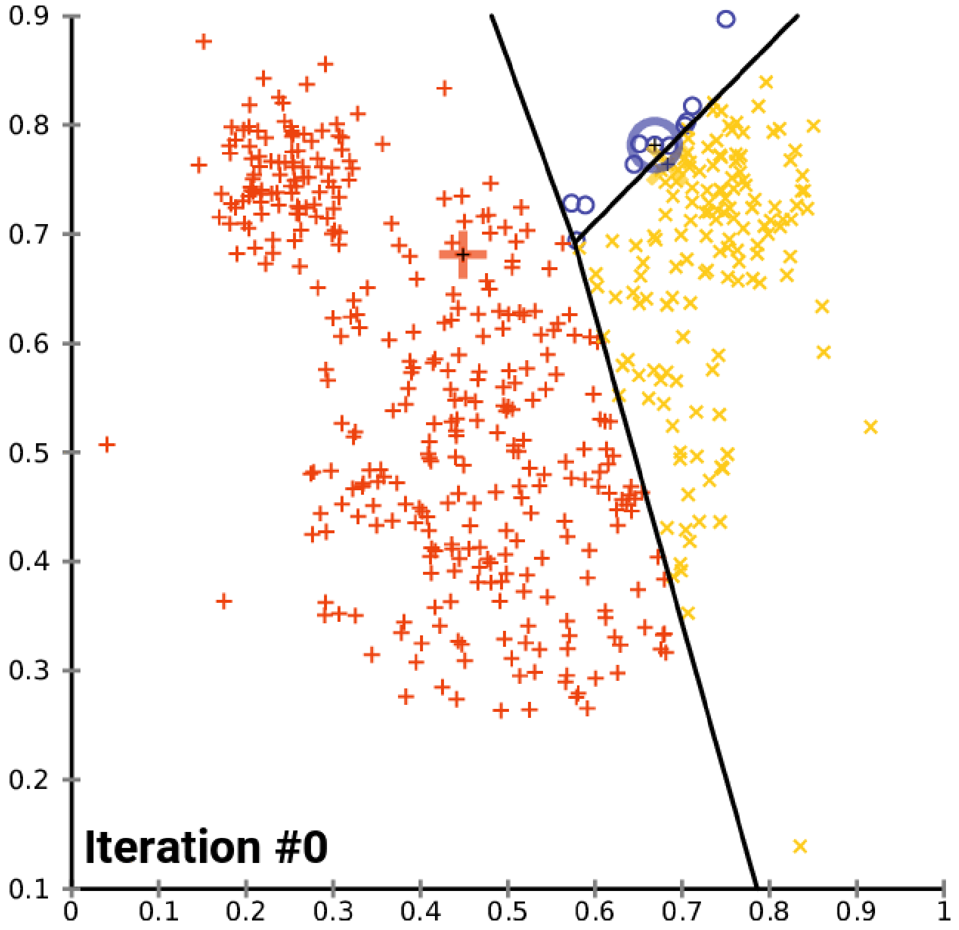


Clasificación

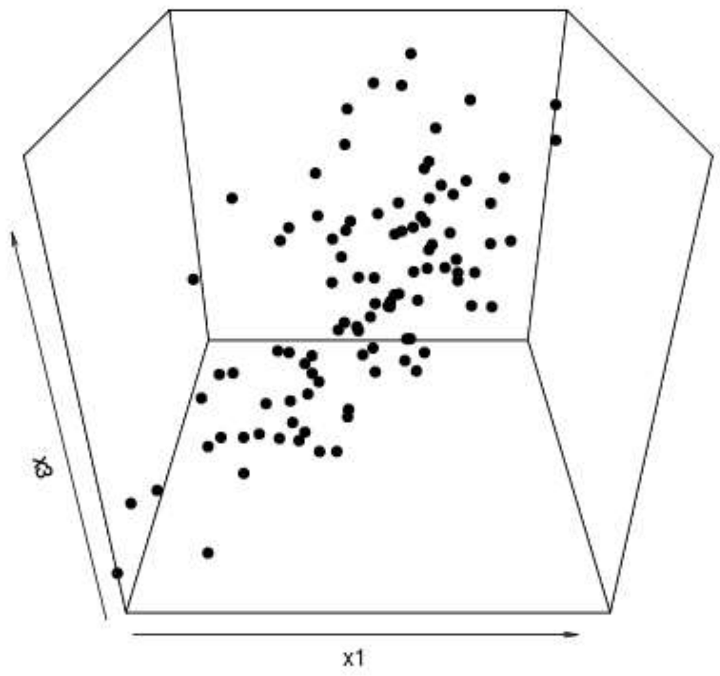


Reducción de dimensiones

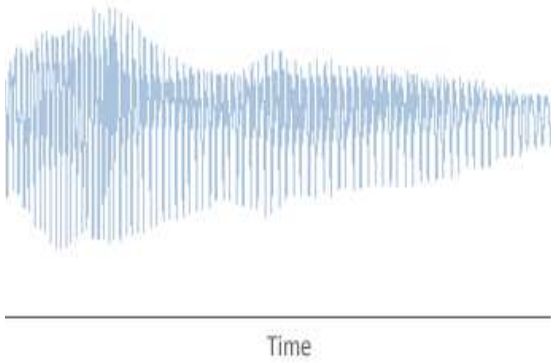
Aprendizaje No Supervisado



Aprendizaje No Supervisado

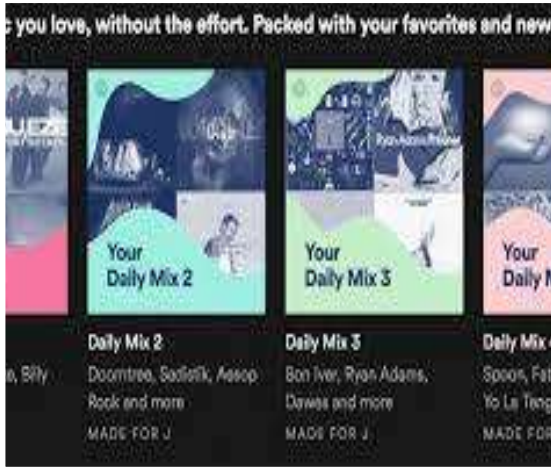


Aplicaciones



Detección de Acentos

Segmentación de clientes



Sistemas de Recomendación

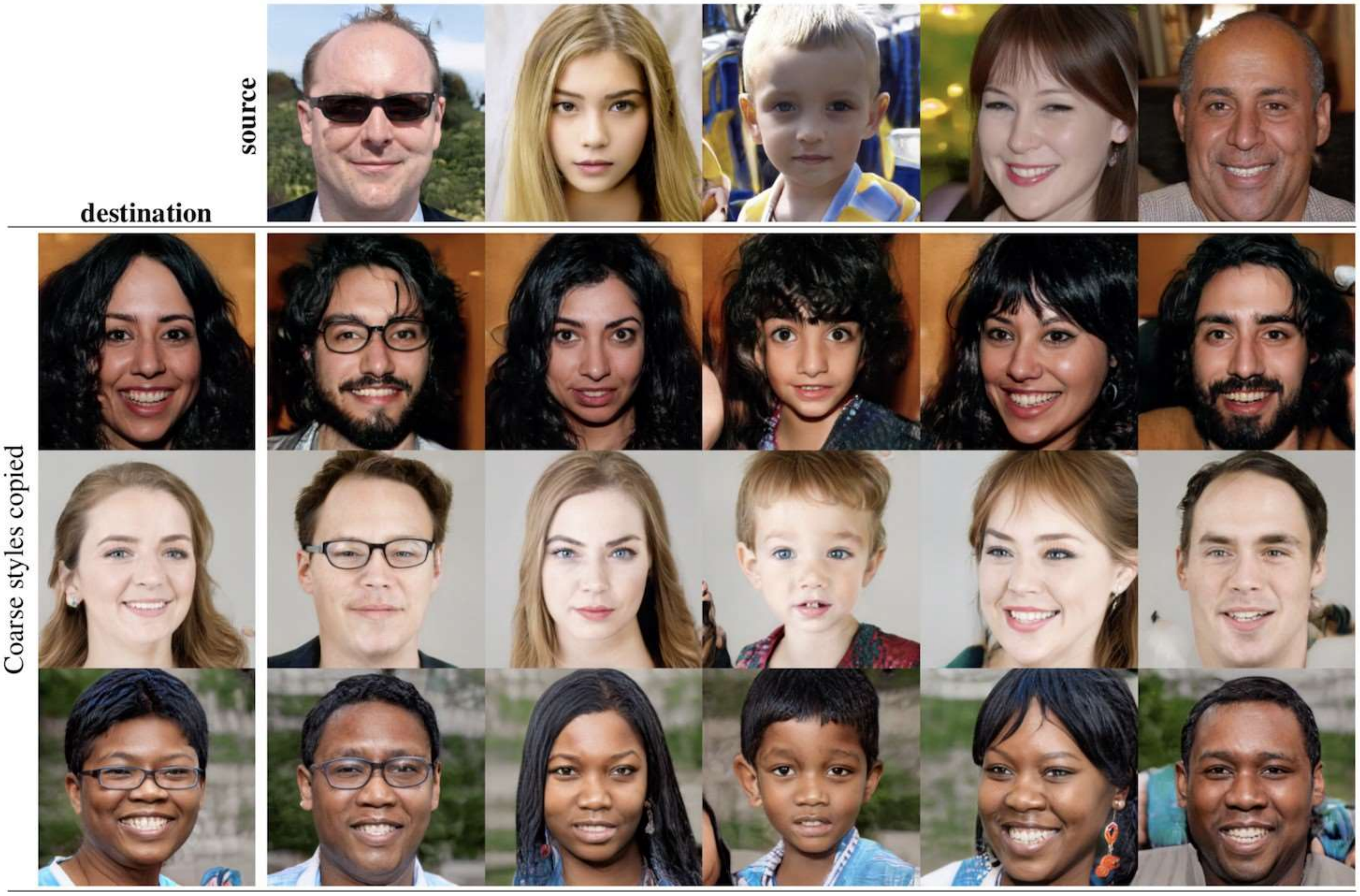
¿Más ejemplos?

Aprendizaje Reforzado

El algoritmo interactúa con el entorno mediante acciones

El algoritmo recibe una recompensa que depende de las acciones

El objetivo del algoritmo es maximizar la recompensa acumulada

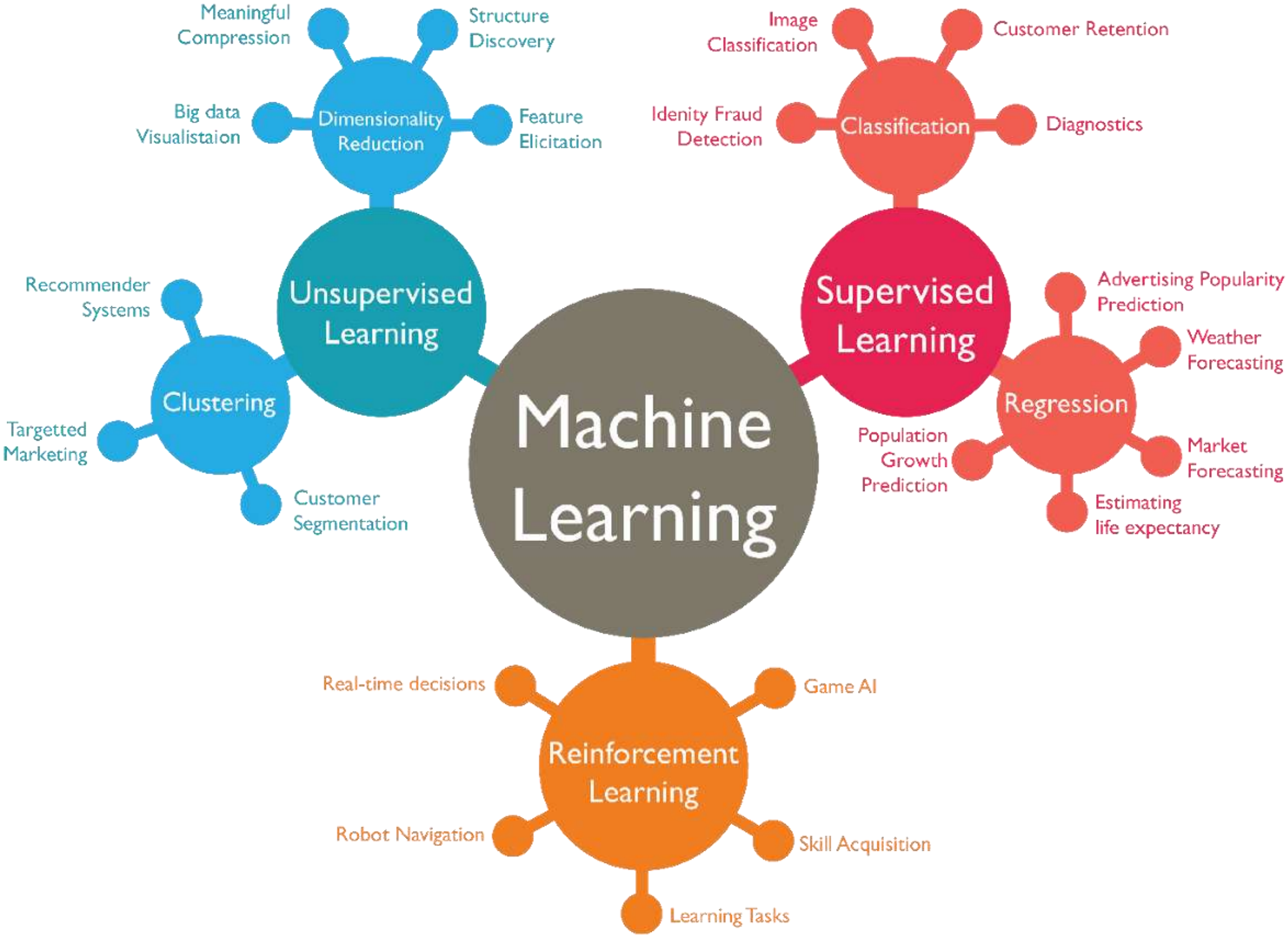


Fuente: https://arxiv.org/abs/1812.04948

Aprendizaje Reforzado

Trabajemos

Buscar 3 ejemplos adicionales de los 3 tipos de ML que hay



Fuente: https://7wdata.be/visualization/types-of-machine-learning-algorithms-2/

Repasemos

- Predecir el precio de una casa dado el área y número de habitaciones
- Indicar si una imagen es un perro o un gato
- Encontrar segmentos de clientes
- Tengo muchos atributos y quiero ver cuáles son los más relevantes

Ventajas

- Identifica patrones complejos
- Manejo de datos complejos
- Automatización
- Mejora continua

Desventajas

- Data
- Tiempo
- Interpretabilidad
- Susceptible a errores

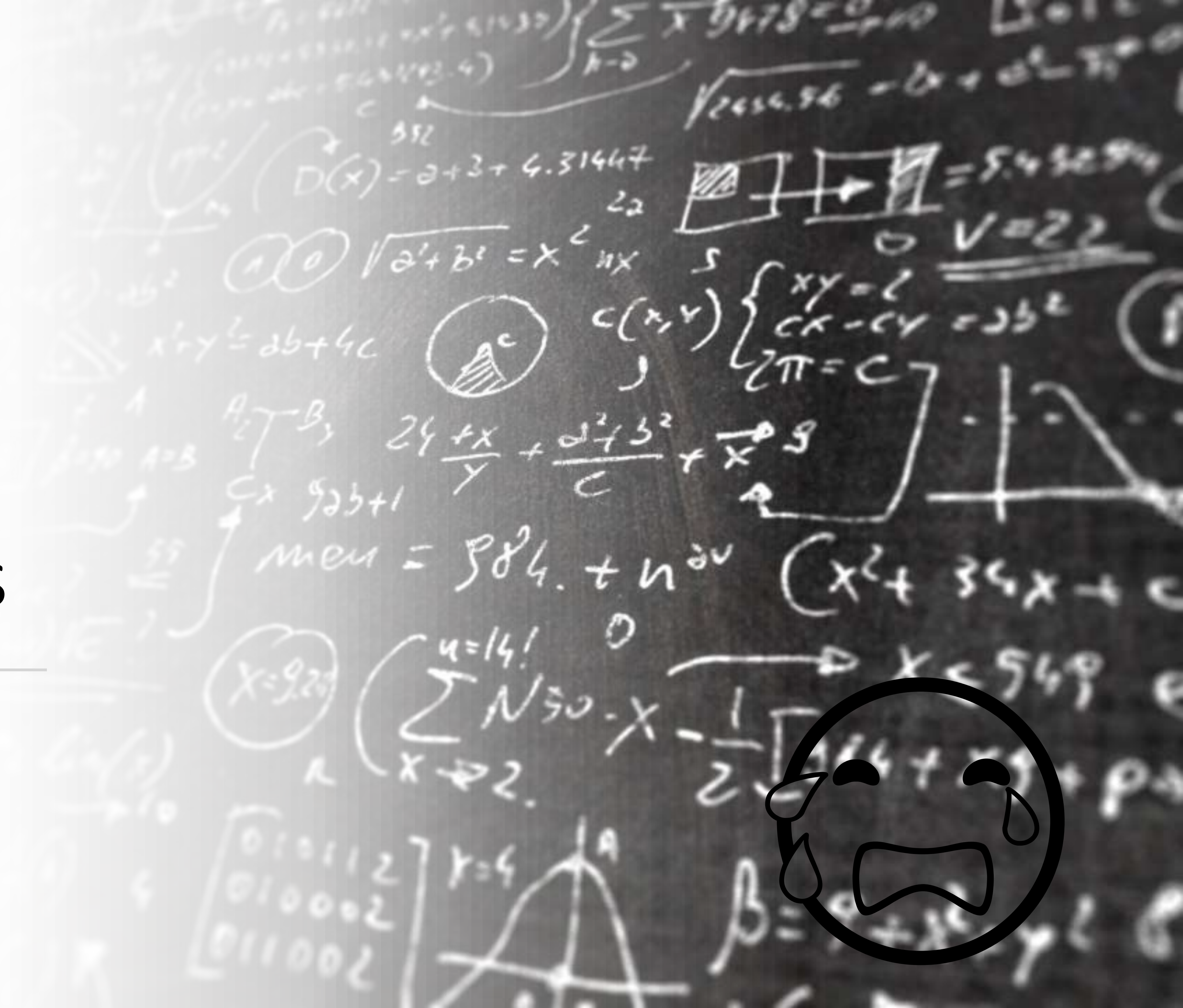
Notas

ML No es la solución para todos los problemas

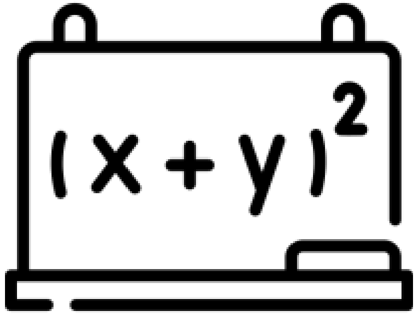
Problemas bien definidos

Data, Data, Data!

Las matemáticas son inevitables



Las matemáticas son inevitables



Álgebra Lineal

- Matrices
- Vectores

Estadística

Cálculo

- Derivadas

Representación de datos

Caso: Precio de Casas

12 atributos

5000 casas

x1	y	x2	x3	x4
sqft	price	City	bedrooms	baths
3392	339000	Dublin	3	2.1
4100	899900	pleasanton	4	3
3200	448641	Clayton	5	4
1436	239999	Moraga	4	3
1944	377500	Antioch	3	2
1500	299900	Danville	3	2.5
1700	265000	El Dorado Hills	4	3
2507	449000	Shingle Springs	4	3
1580	439950	McKinleyville	3	2
1500	699888	Marina	4	2
2705	1250000	Roseville	3	2
1715	439000	Rocklin	4	3

Representación de datos

- x_i son los atributos
- y variable de salida

12 atributos

5000 casas

x_1	y	x_2	x_3	x_4
sqft	price	City	bedrooms	baths
3392	339000	Dublin	3	2.1
4100	899900	pleasanton	4	3
3200	448641	Clayton	5	4
1436	239999	Moraga	4	3
1944	377500	Antioch	3	2
1500	299900	Danville	3	2.5
1700	265000	El Dorado Hills	4	3
2507	449000	Shingle Springs	4	3
1580	439950	McKinleyville	3	2
1500	699888	Marina	4	2
2705	1250000	Roseville	3	2
1715	439000	Rocklin	4	3

Representación de datos

- $x^{(i)}$ variables de entrada o *input feature*
- $y^{(i)}$ función de salida, *target variable*
- x_j^i dato de entrenamiento, donde *i* es el *feature* y *j* es un registro en particular.

x1	y	x2	x3	x4
sqft	price	City	bedrooms	baths
3392	339000	Dublin	3	2.1
4100	899900	pleasanton	4	3
3200	448641	Clayton	5	4
1436	239999	Moraga	4	3
1944	377500	Antioch	3	2
1500	299900	Danville	3	2.5
1700	265000	El Dorado Hills	4	3
2507	449000	Shingle Springs	4	3
1580	439950	McKinleyville	3	2
1500	699888	Marina	4	2
2705	1250000	Roseville	3	2
1715	439000	Rocklin	4	3

Representación de datos

a) Una variable del tipo x_{ij} indicaría el valor del *feature* i para la observación j; es decir:

$$i = 1,2,..,p$$
$$j = 1,2,\ldots,n$$

Donde n es el total de casas (5000) y p es la cantidad de attributos (12)

En forma de una matriz $\boldsymbol{X}_{np}$

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \ldots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{np} \end{bmatrix}$$

Representación de datos

b) Cuando se requiere **filas**: $x_1, x_2, \dots, x_n$

$$\boldsymbol{x}_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{bmatrix}$$

Es un vector de longitud p

c) A veces se requiere **columnas**: $x_1, x_2, \dots, x_p$

$$\boldsymbol{x}_j = \begin{bmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{nj} \end{bmatrix}$$

Es un vector de longitud n

Representación de datos

d) Vector de vectores

$$\boldsymbol{X} = [\boldsymbol{x}_1 \quad \boldsymbol{x}_2 \quad \dots \quad \boldsymbol{x}_p]$$

$$\boldsymbol{X} = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{bmatrix}$$

Resumiendo

	sqft	City	bedrooms	baths	price
x_i	$\boxed{3392}$	$\boxed{\text{Dublin}}$	$\boxed{3}$	$\boxed{2.1}$	$\boxed{339000}$
	4100	pleasanton	4	3	899900
	3200	Clayton	5	4	448641
	1436	Moraga	4	3	239999
	1944	Antioch	3	2	377500
	1500	Danville	3	2.5	299900
	1700	El Dorado Hills	4	3	265000
	2507	Shingle Springs	4	3	449000
	1580	McKinleyville	3	2	439950
	1500	Marina	4	2	699888
	2705	Roseville	3	2	1250000
	1715	Rocklin	4	3	439000

Resumiendo

b)

$$x_i = \begin{array}{l} \texttt{3392} \\ \texttt{Dublin} \\ \texttt{3} \\ \texttt{2.1} \end{array}$$

c)

$$x_j = \begin{array}{l} \texttt{3392} \\ \texttt{4100} \\ \texttt{3200} \\ \texttt{1436} \\ \texttt{1944} \\ \texttt{1500} \\ \texttt{1700} \\ \texttt{2507} \\ \texttt{1580} \\ \texttt{1500} \\ \texttt{2705} \\ \texttt{1715} \end{array}$$

d)

$$\boldsymbol{X} = \begin{array}{cccc} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_p \\ \texttt{sqft} & \texttt{City} & \texttt{bedrooms} & \texttt{baths} \\ \begin{bmatrix} \texttt{3392} \\ \texttt{4100} \\ \texttt{3200} \\ \texttt{1436} \\ \texttt{1944} \\ \texttt{1500} \end{bmatrix} & \begin{bmatrix} \texttt{Dublin} \\ \texttt{pleasanton} \\ \texttt{Clayton} \\ \texttt{Moraga} \\ \texttt{Antioch} \\ \texttt{Danville} \end{bmatrix} & \begin{array}{c} \texttt{3} \\ \texttt{4} \\ \texttt{5} \\ \texttt{4} \\ \texttt{3} \\ \texttt{3} \end{array} & \begin{bmatrix} \texttt{2.1} \\ \texttt{3} \\ \texttt{4} \\ \texttt{3} \\ \texttt{2} \\ \texttt{2.5} \end{bmatrix} \end{array}$$

$$\boldsymbol{X} = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{bmatrix} \begin{array}{l} \texttt{3392} \quad \texttt{4100} \quad \texttt{3200} \quad \texttt{1436} \quad \cdots \\ \texttt{Dublin pleasanton} \quad \cdots \\ \\ \texttt{2.1} \quad \texttt{3} \quad \texttt{4} \end{array}$$

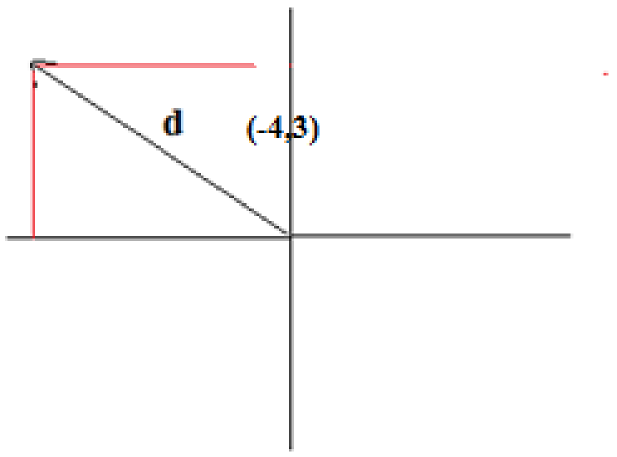
Vectores

Video Recomendado

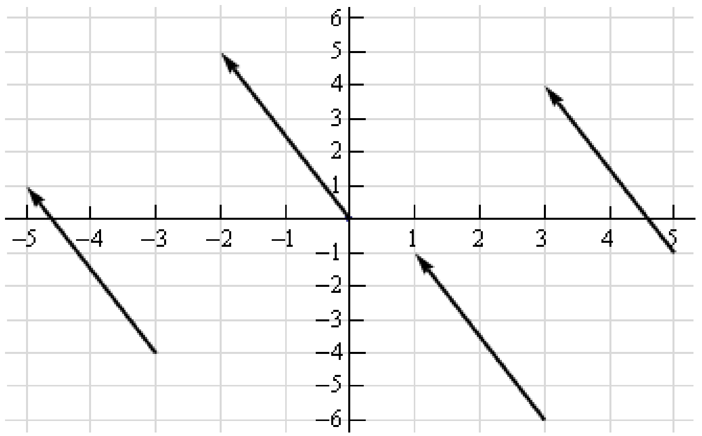
- https://www.youtube.com/watch?v=wiuEEkP_XuM

Vectores

- Magnitud
- Dirección



$$d = \sqrt{3^2 + (-4)^2} = 5$$

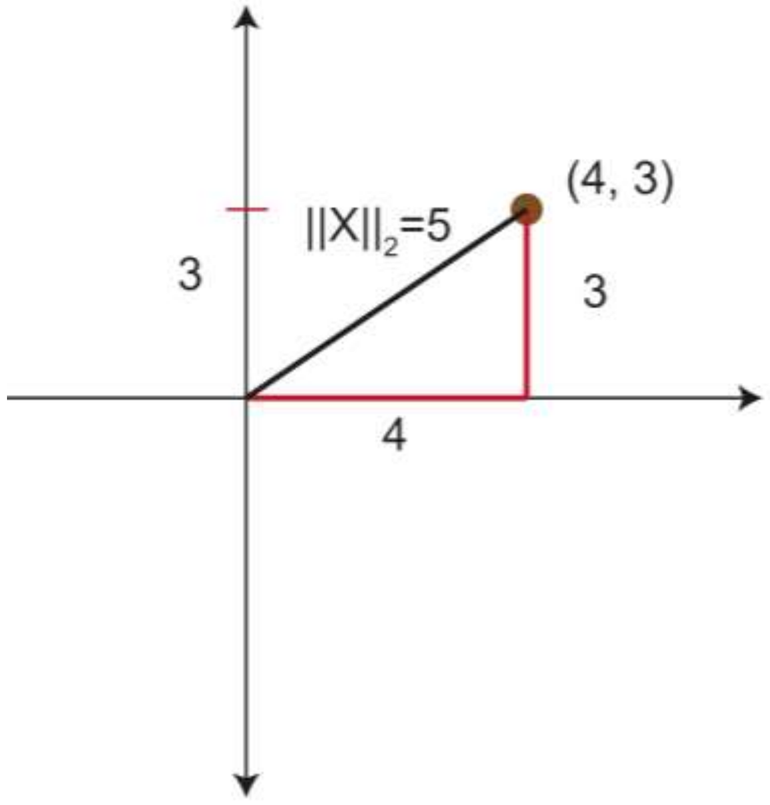


(gráfica tomada de http://tutorial.math.lamar.edu/Classes/CalcII/Vectors_Basics.aspx)

Norma o longitud de un vector

Si tenemos un vector $\vec{v} = \langle x_1, x_2, x_3 \rangle$ entonces su longitud estará dada por:

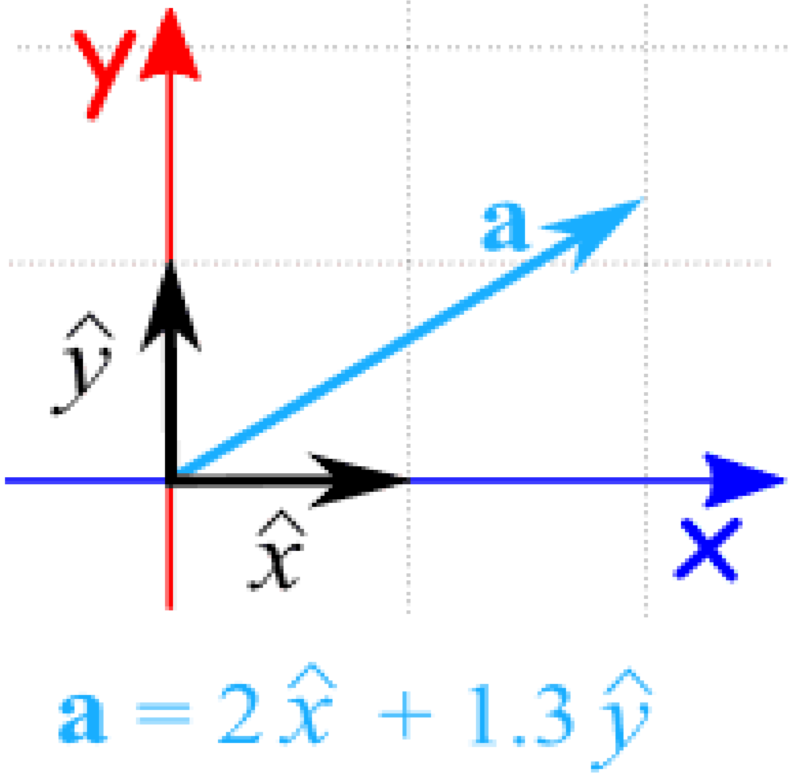
$$||\vec{v}|| = \sqrt{v_0^2 + v_1^2 + v_3^2}$$



Vector unitario

- Es cualquier vector cuya magnitud es igual a 1, es decir ||**x**||=1.

- **Notación acerca de los vectores:** Algunos textos suelen representar a los vectores o matrices mediante el siguiente símbolo $\vec{v}$ o **v**



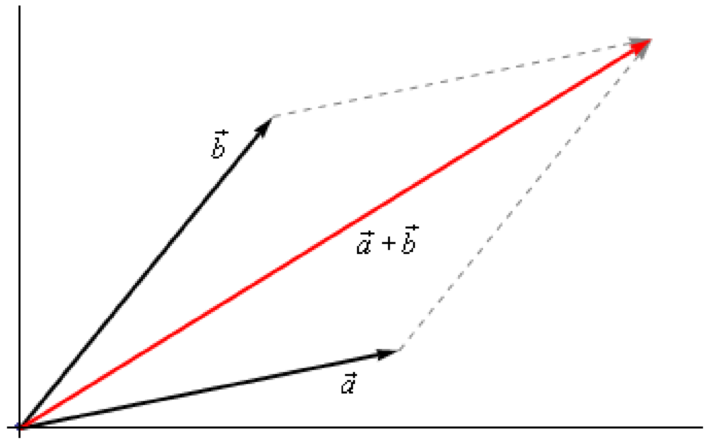
Suma de Vectores

Si se tienen dos vectores:

x=$\begin{pmatrix} x_0 \\ x_1 \\ x_2 \end{pmatrix}$, **y**=$\begin{pmatrix} y_0 \\ y_1 \\ y_2 \end{pmatrix}$, entonces su suma estará dada por:

x + **y** =$\begin{pmatrix} x_0 + y_0 \\ x_1 + y_1 \\ x_2 + y_2 \end{pmatrix}$

De forma geométrica tendríamos:



(gráfica tomada de http://tutorial.math.lamar.edu/Classes/CalcII/Vectors_Basics.aspx)

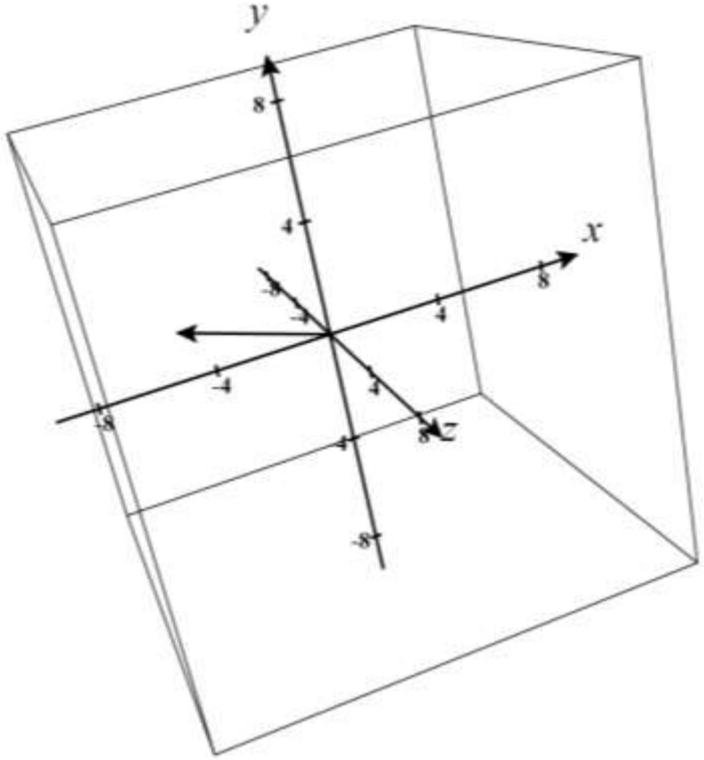
Multiplicación escalar

Se puede multiplicar un vector **x** por un número cualquiera u denominado escalar, es decir:

$$u * \vec{x}$$

Multiplicación escalar

- Encontrar un vector unitario que apunte en la misma dirección de $\vec{\boldsymbol{w}} = \begin{pmatrix} -5 \\ 2 \\ 1 \end{pmatrix}$



Multiplicación escalar

- **Solución**: Lo que necesitamos es hallar un vector cualquiera que sea paralelo a $\vec{\boldsymbol{w}}$**.**
- Esto se puede obtener al multiplicar nuestro vector original por un escalar
- $\vec{\boldsymbol{u}} = \vec{\boldsymbol{w}}$, donde u es un vector unitario
- $\vec{\boldsymbol{u}} = \alpha\vec{\boldsymbol{w}}$**,** podemos multiplicar por un escalar, el cual puede tomar la siguiente forma:

$$\vec{\boldsymbol{u}} = \frac{1}{||\boldsymbol{w}||}\vec{\boldsymbol{w}}$$

Concepto básico
usado en SVM

Multiplicación escalar

- $u = \begin{pmatrix} -5 \\ 2 \\ 1 \end{pmatrix} \cdot \frac{1}{\sqrt{(-5)^2+2^2+1^2}}$
- $= \begin{pmatrix} -5 \\ 2 \\ 1 \end{pmatrix} \cdot \frac{1}{\sqrt{30}} = \begin{pmatrix} \frac{-5}{\sqrt{30}} \\ \frac{2}{\sqrt{30}} \\ \frac{1}{\sqrt{30}} \end{pmatrix}$
- $\vec{\boldsymbol{u}} = \sqrt{\left(\frac{-5}{\sqrt{30}}\right)^2 + \left(\frac{2}{\sqrt{30}}\right)^2 + \left(\frac{1}{\sqrt{30}}\right)^2} = 1$

$$\vec{\boldsymbol{u}} = \frac{1}{||\boldsymbol{w}||}\vec{\boldsymbol{w}}$$

Transpuesta de un vector

- La transpuesta de un vector columna **v** consiste en convertirlo en un vector fila con los mismos componentes

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix}, entonces\ \boldsymbol{x^T} = (x_1 x_2 \quad \dots \ \ x_n)$$

Importante: Esta sección y las siguientes se utiliza en todo lo que es manipulación de datos.

Producto punto

- Si tenemos dos vectores $\boldsymbol{x} = (x_1, x_2, x_3), \boldsymbol{y} = (y_1, y_2, y_3)$, entonces su producto punto o dot product estará dado por:

$$\boldsymbol{x}.\boldsymbol{y} = \sum_{i=1}^{n} x_i y_i$$

$$\boldsymbol{x}.\boldsymbol{y} = x_1 y_1 + x_2 y_2 + x_3 y_3$$

Algunos Ejercicios

- $\vec{a} = [1, 2, 8]$
- $b = [-5, 10, 0]$
- $c = [7, -1, 11]$

- Calcular:
  - La norma
  - El vector unitario de cada uno

  - $\vec{a} + \vec{b}$
  - $\vec{a} - \vec{b} + \vec{c}$
  - $-7\vec{a} + 1.3\vec{b} - 3.7\vec{c}$

Referencias

1. Paul´s Online Math Notes, disponible en: http://tutorial.math.lamar.edu/Classes/CalcII/VectorArithmetic.aspx **(estos apuntes son especialmente útiles en caso no recuerde temas de matemática básica o de cálculo en todos los niveles)**
2. Poole, D. (2003). Linear algebra: A modern introduction. Australia: Brooks/Cole-Thomson Learning.
3. Strogatz, S. H. (1994). Nonlinear dynamics and Chaos: With applications to physics, biology, chemistry, and engineering. Reading, MA: Addison-Wesley Pub. **(este autor hace una explicación bastante interesante de la utilidad del cálculo de los eigenvalues y eigenvectors para casos de Dinámicas no lineales)**
4. Williams, G., & Williams, G. (1984). Linear algebra with applications. Boston: Allyn and Bacon.
5. Tan, P., Steinbach, M., & Kumar, V. (2005). Introduction to data mining. Boston: Pearson Addison Wesley.
6. Müller, Andreas & Guido, Sara (2017). Introduction to Machine Learning with Python.
7. Hastie, T., Tibshirani, R.,, Friedman, J. (2001). The Elements of Statistical Learning. New York, NY, USA: Springer New York Inc.
8. James, G., Witten, D., Hastie, T.,, Tibshirani, R. (2013). An Introduction to Statistical Learning -- with Applications in R (Vol. 103). New York: Springer. ISBN: 978-1-4614-7137-0