

Curso de Machine Learning

Hayde Martínez



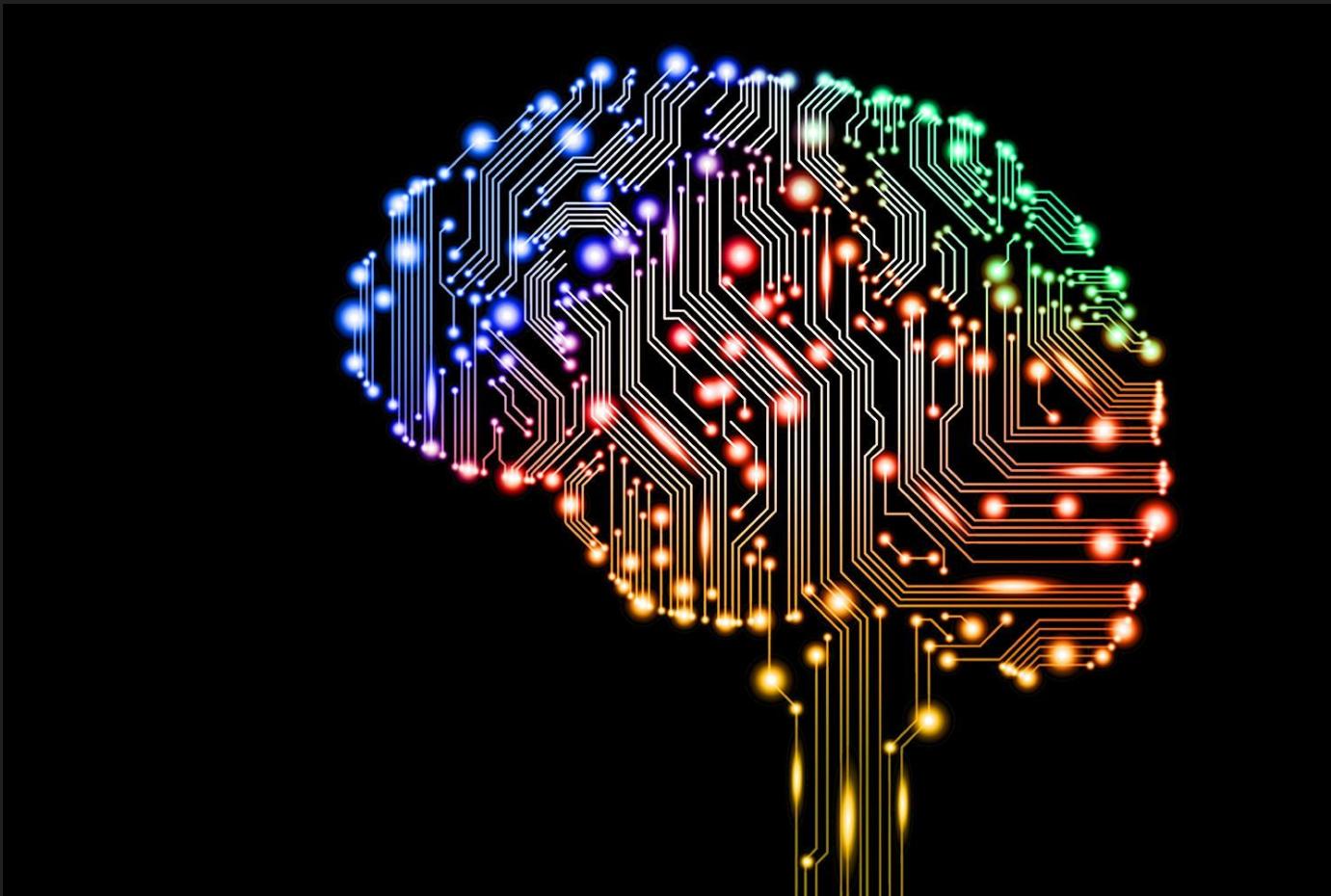




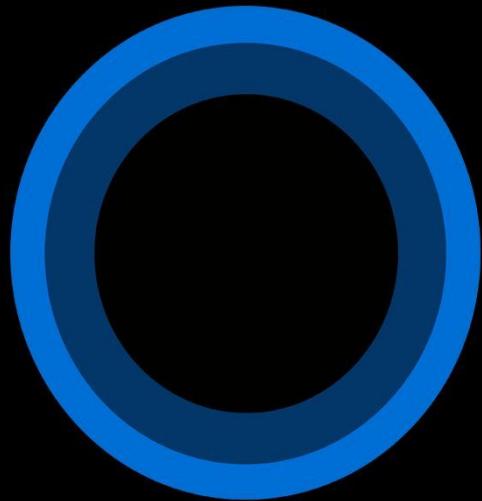
Temporally Recurrent Optimal Learning

TROL(L)

Inteligencia Artificial

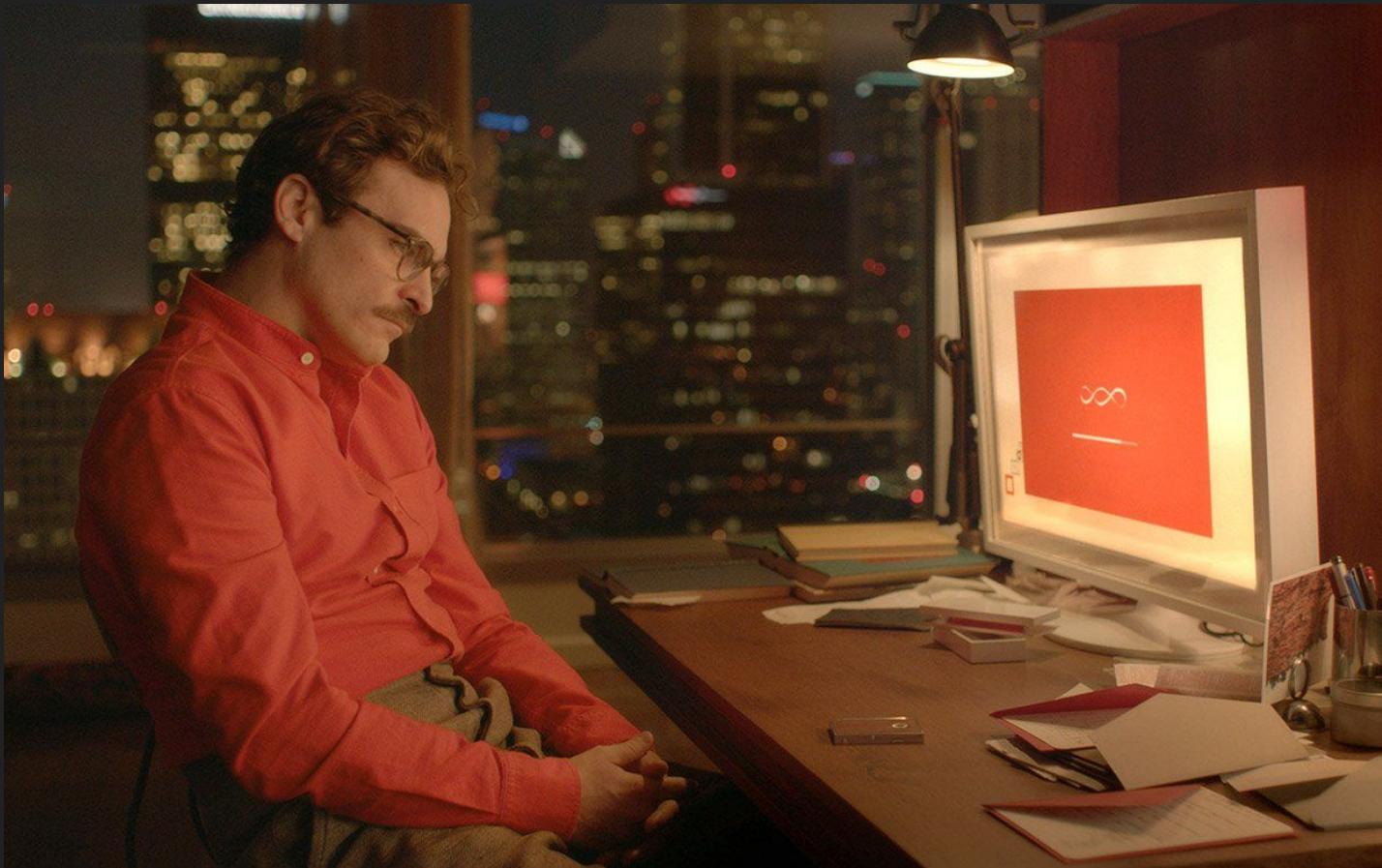


Inteligencia Artificial Débil

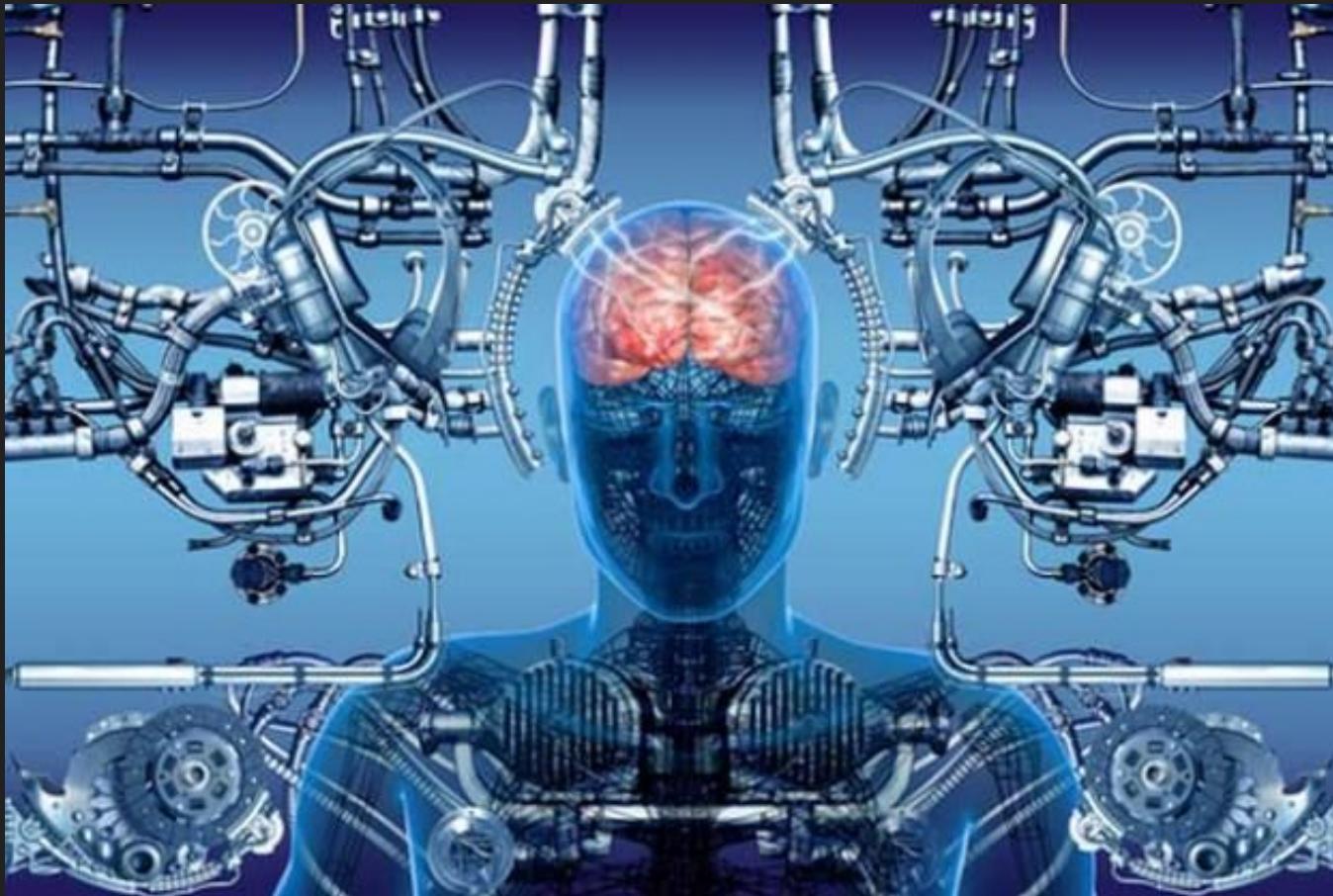


Hi. I'm Cortana.
Ask me a question!

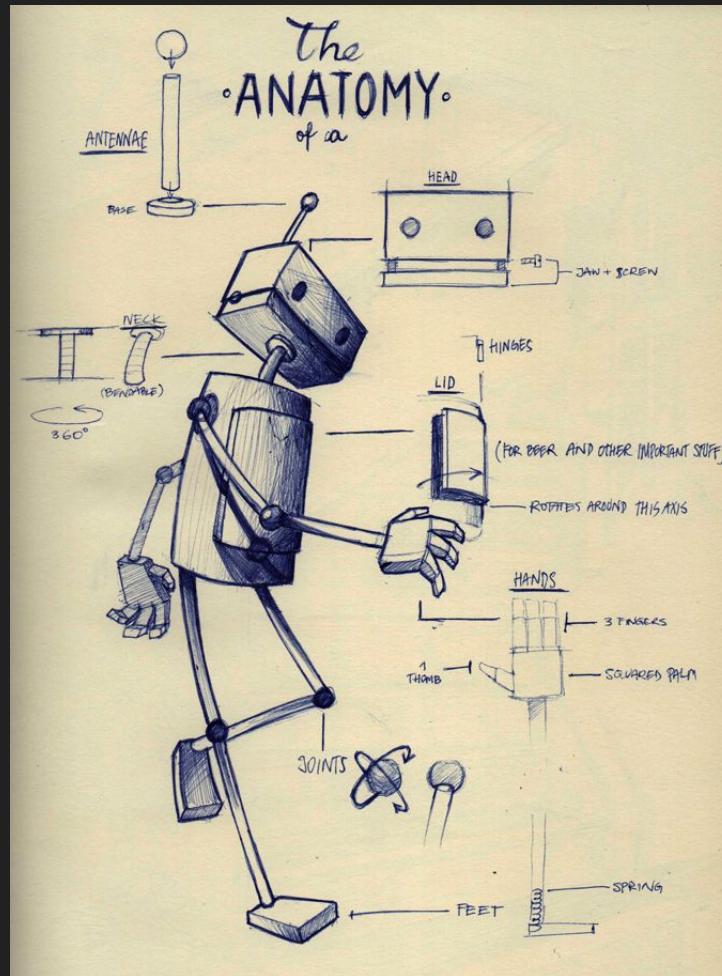
Inteligencia Artificial Fuerte



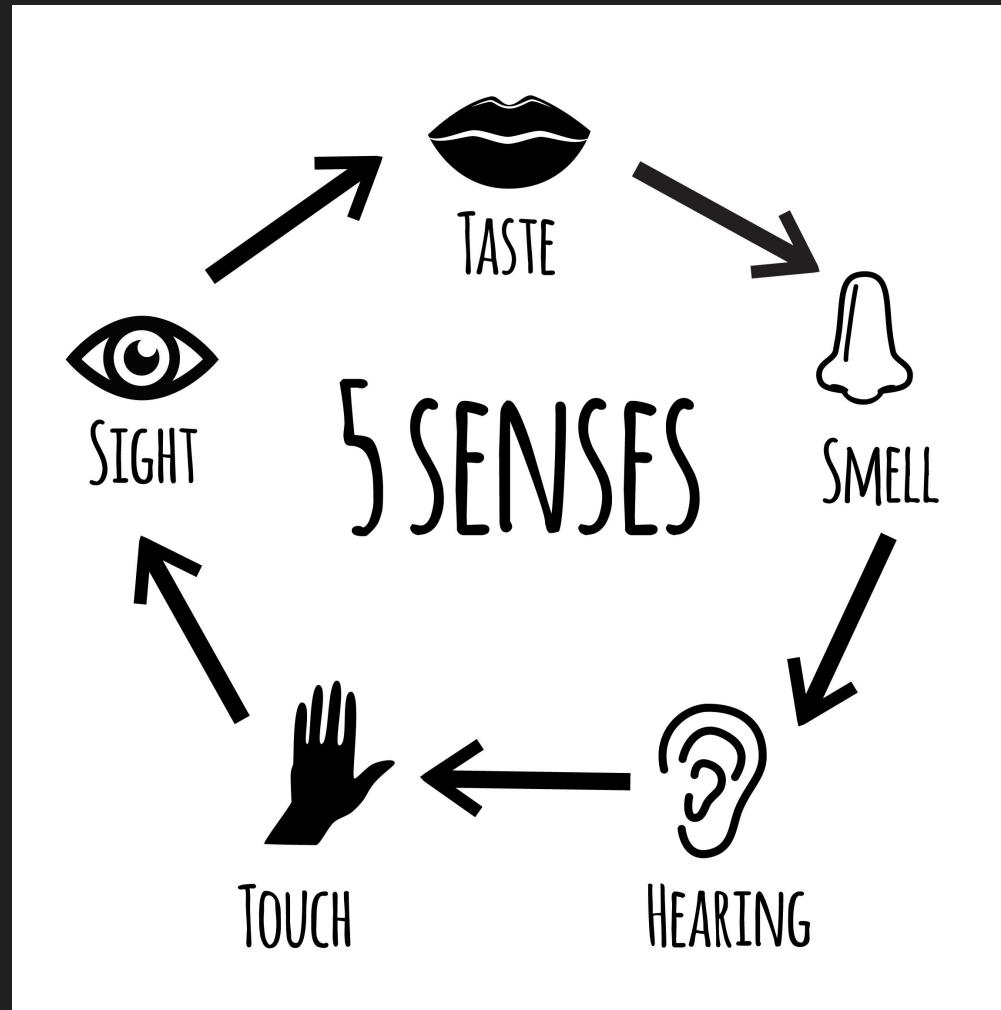
Super Inteligencia Artificial



La Anatomía de la Inteligencia Artificial



Percepción



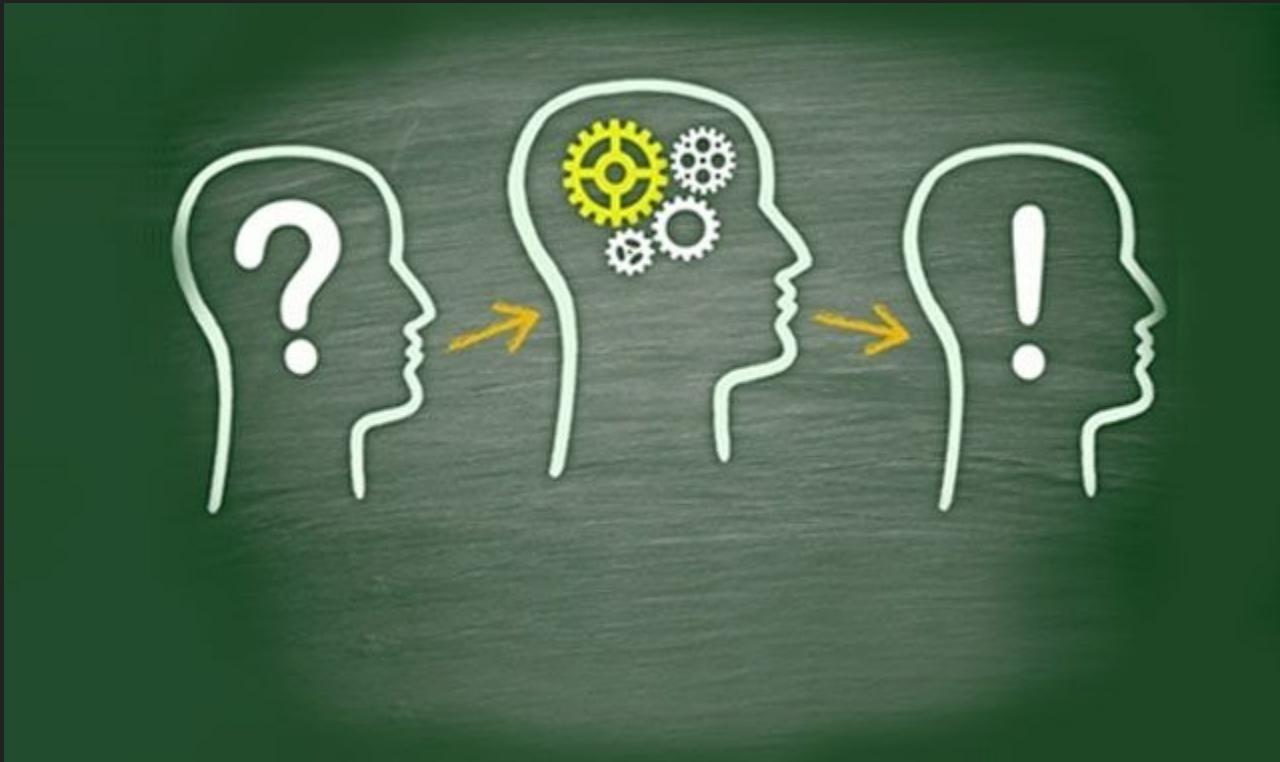
Procesamiento Natural del Lenguaje



Representación del Conocimiento



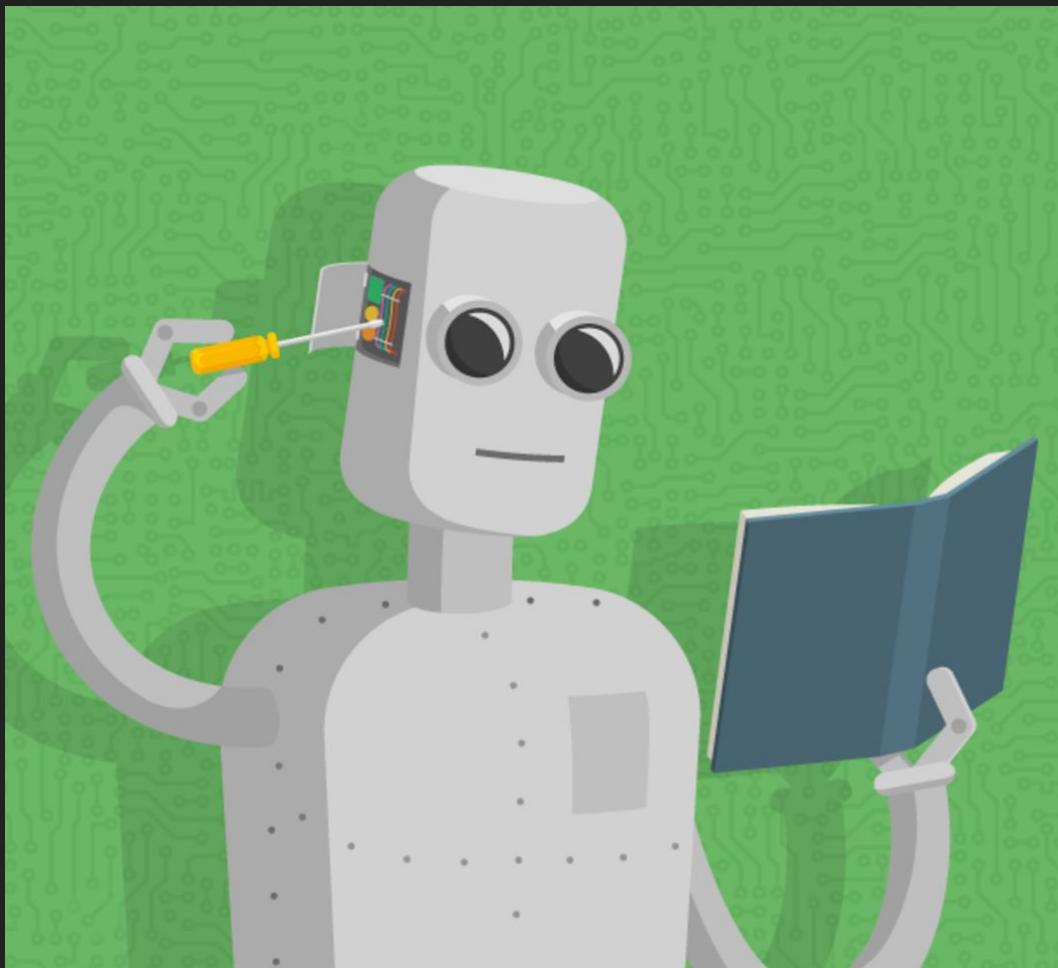
Razonamiento



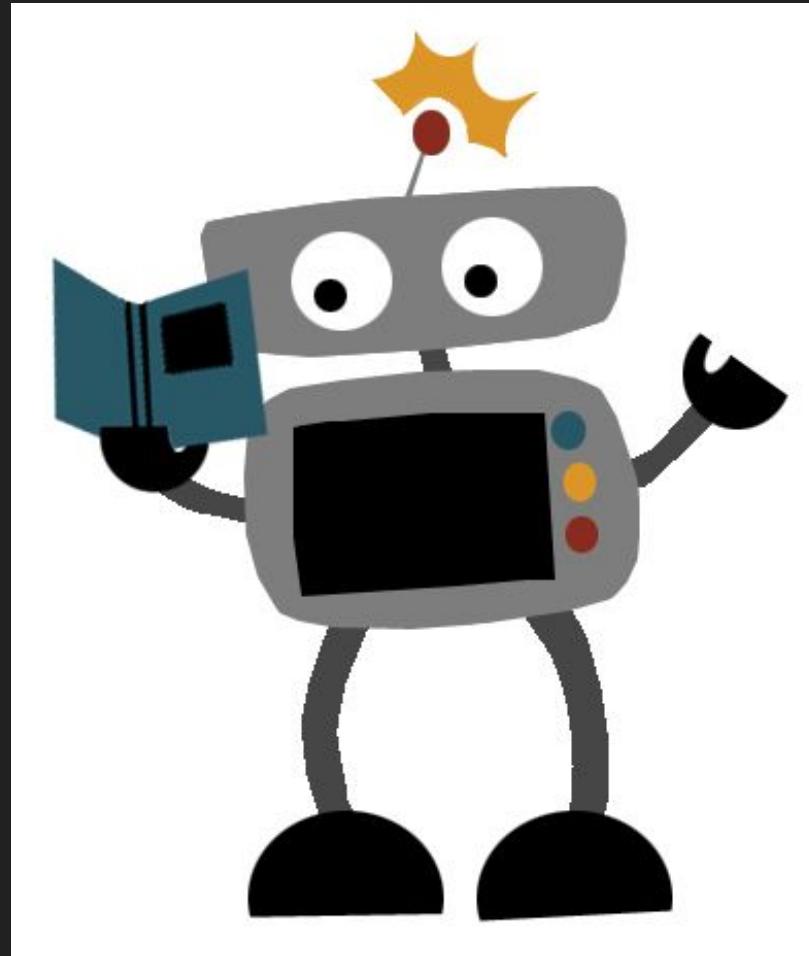
Planeación y Navegación



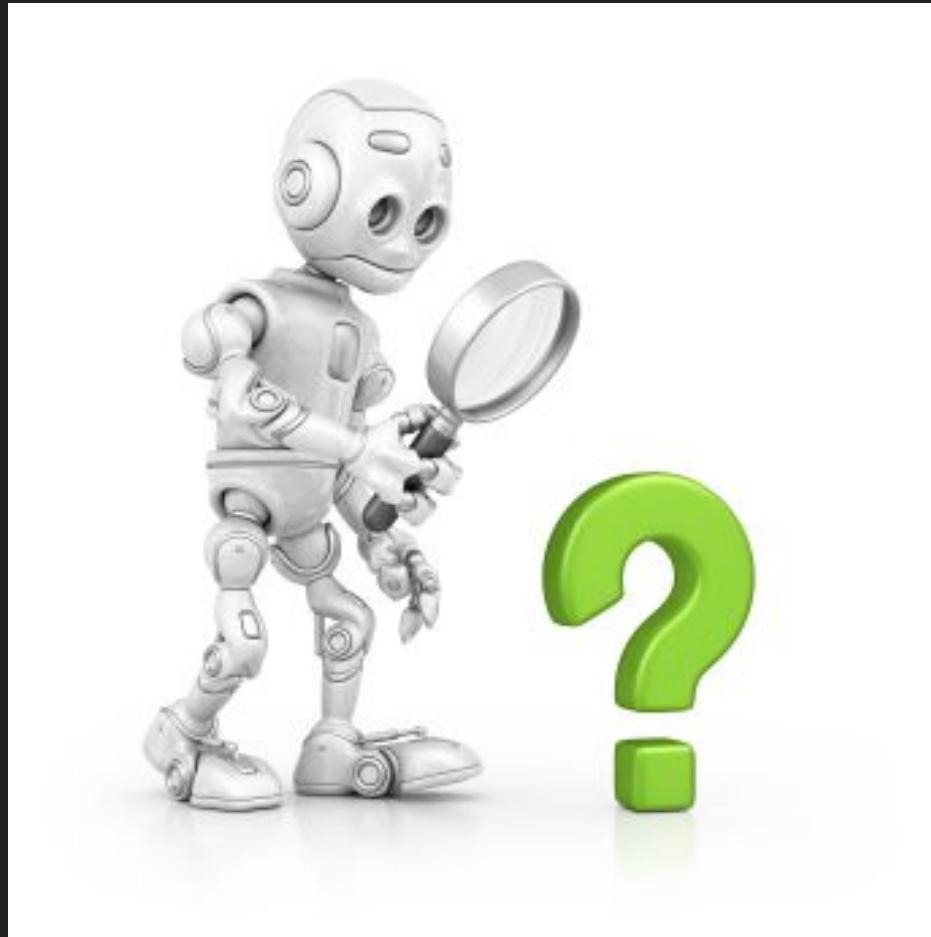
Machine Learning



Aprendizaje supervisado



Aprendizaje sin supervisión



Revisando la Probabilidad y Estadística



Media

Es la suma de las medidas divididas entre el número total de datos.

$$\bar{x} = \frac{\sum x_i}{n}$$

Datos = 1, 8, 7, 9, 5 n = 5 Sumatoria = 30 Resultado = 30/5 = 6

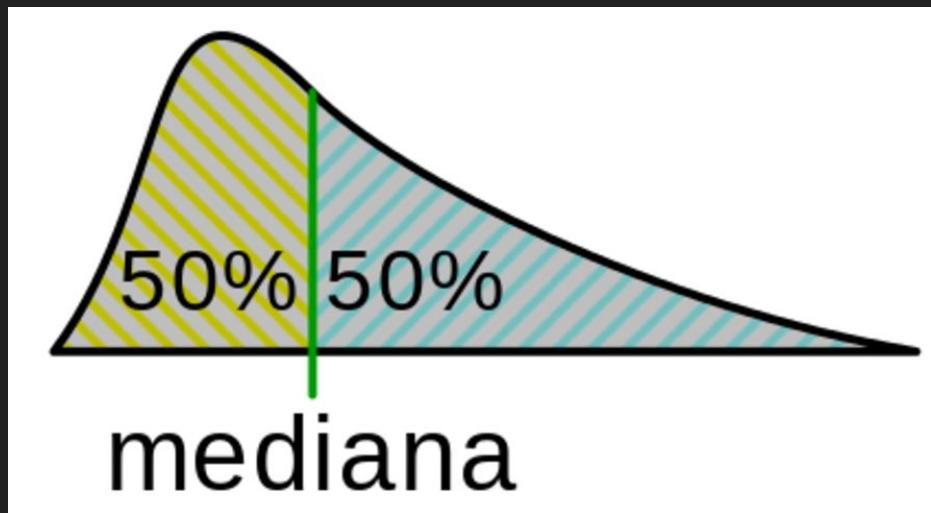
Moda

Es el valor con mayor frecuencia en una distribución de datos.

Datos = 1, 8, 8, 8, 5 n = 5 Moda = 8

Mediana

Es el valor que ocupa el lugar central de todos los datos cuando están agrupados de menor a mayor.



Varianza

Es una medida de dispersión definida como la esperanza del cuadrado de la desviación de dicha variable respecto a su media. O en pocas palabras, la media de los residuos al cuadrado.

$$\sigma_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

Varianza

Pasos:

1. Calcular la suma de los datos
2. Calcular la media
3. Restar la media a cada dato
4. Elevar al cuadrado cada resultado
5. Sumar todos los resultado y sacar la media de esta sumatoria

Varianza

Ejemplo:

En un partido de baloncesto, se tiene la siguiente anotación en los jugadores de un equipo:

0,2,4,5,8,10,10,15,38. Calcular la varianza de las puntuaciones de los jugadores del equipo.

1. **Sumatoria** = $0+2+4+5+8+10+10+15+38 / 9 = 92$
2. **Media** = $92 / 9 = 10.22$
3. **Restar Media** =
 $(0-10.22)^2+(2-10.22)^2+(4-10.22)^2+(5-10.22)^2+(8-10.22)^2+(10-10.22)^2+(10-10.22)^2+(15-10.22)^2+(38-10.22)^2 / 9$
4. **Elevar al cuadrado** = $10.22^2+8.22^2+6.22^2+5.22^2+2.22^2+0.22^2+4.78^2+27.78^2 / 9$
5. **Sumatoria de los resultados y media** =
 $104.4484+67.5684+38.6884+27.2484+4.9284+0.0484+22.8484+771.72849 = 1037.5556 / 9 = 115.28$

Varianza = 115.28

Desviación Estándar

Se define como la raíz cuadrada de la varianza de la variable.

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

Desviación Estándar

Pasos:

1. Calcular la suma de los datos
2. Calcular la media
3. Restar la media a cada dato
4. Elevar al cuadrado cada resultado
5. Sumar todos los resultado y sacar la media de esta sumatoria
6. Raíz cuadrada del resultado

Desviación Estándar

Ejemplo:

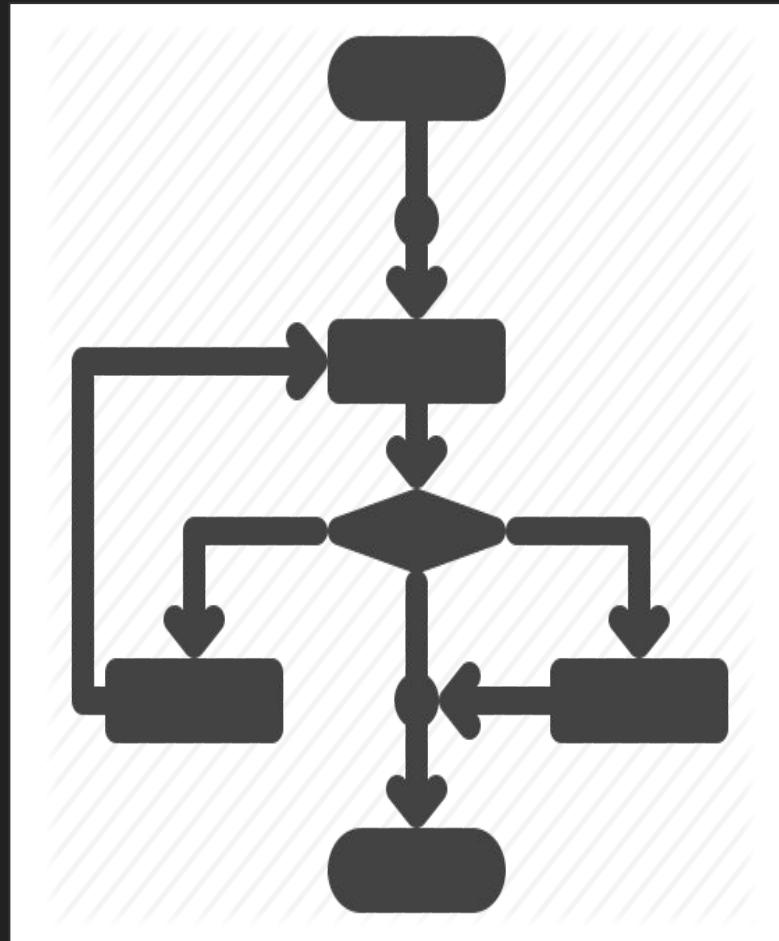
En un partido de baloncesto, se tiene la siguiente anotación en los jugadores de un equipo:

0,2,4,5,8,10,10,15,38. Calcular la varianza de las puntuaciones de los jugadores del equipo.

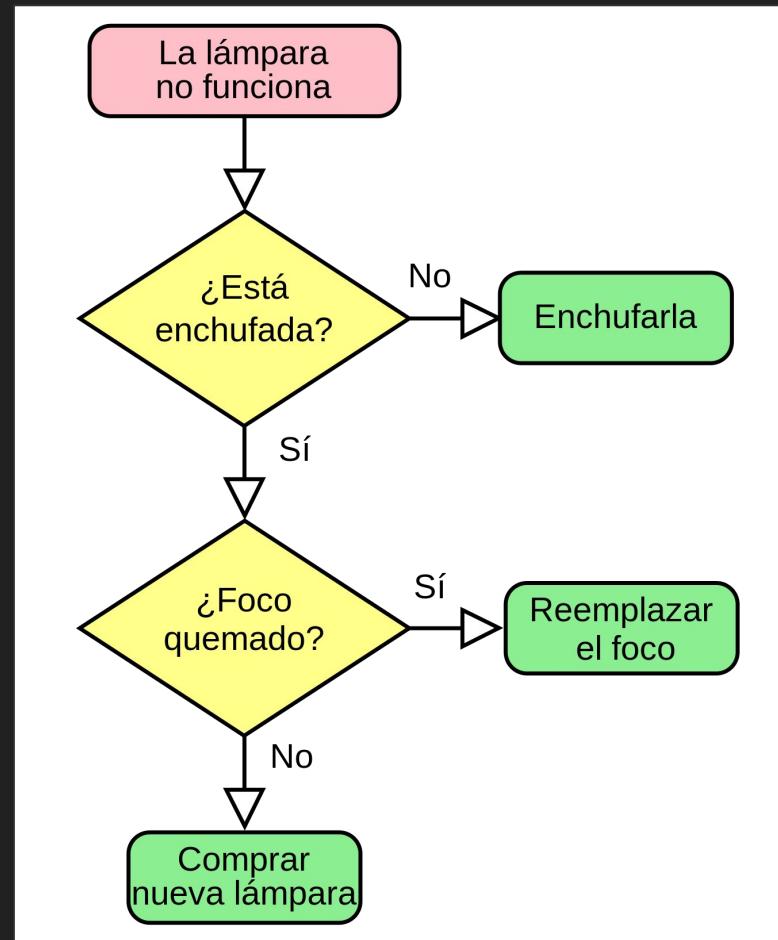
1. **Sumatoria** = $0+2+4+5+8+10+10+15+38 / 9 = 92$
2. **Media** = $92 / 9 = 10.22$
3. **Restar Media** =
 $(0-10.22)^2+(2-10.22)^2+(4-10.22)^2+(5-10.22)^2+(8-10.22)^2+(10-10.22)^2+(10-10.22)^2+(15-10.22)^2+(38-10.22)^2 / 9$
4. **Elevar al cuadrado** = $10.22^2+8.22^2+6.22^2+5.22^2+2.22^2+0.22^2+4.78^2+27.78^2 / 9$
5. **Sumatoria de los resultados y media** =
 $104.4484+67.5684+38.6884+27.2484+4.9284+0.0484+22.8484+771.72849 = 1037.5556 / 9 = 115.28$
6. **Raíz cuadrada** de 115.28 = 10.73

Desviación Estandar = 10.73

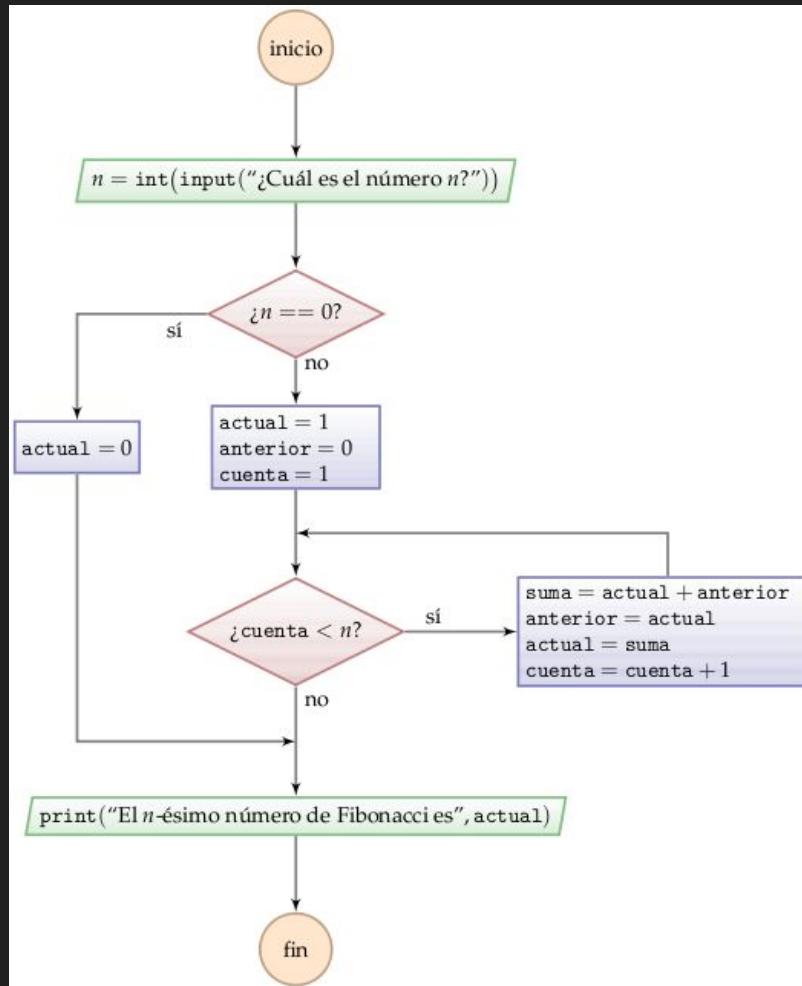
Algoritmos



Ejemplo

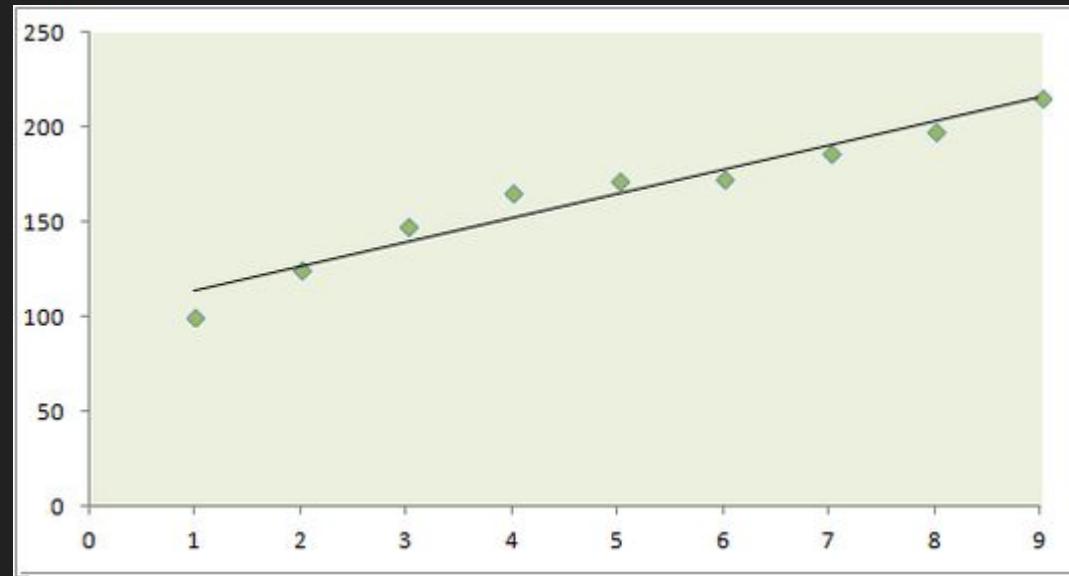


Ejemplo



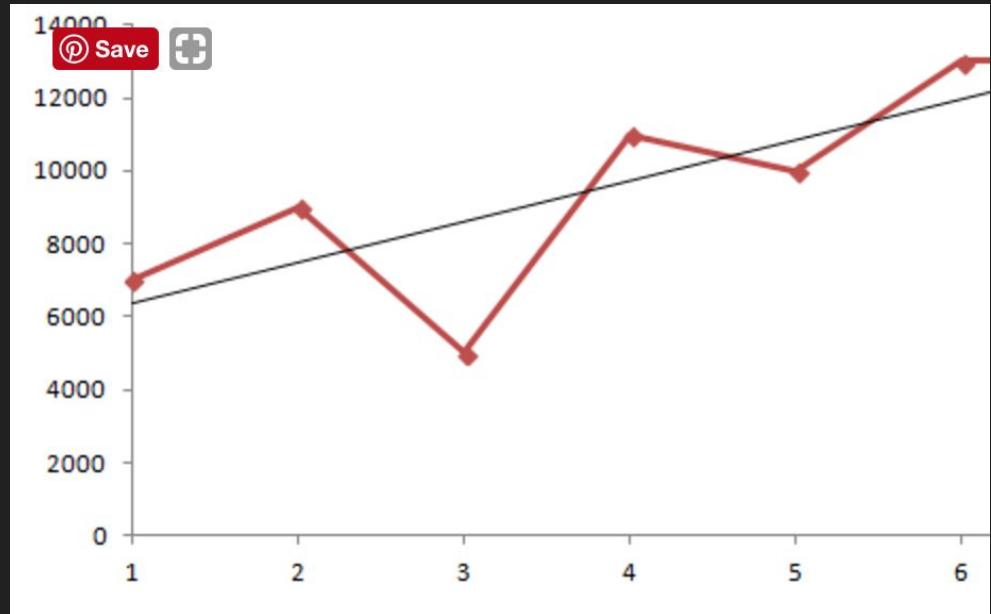
Regresión Lineal

$$\hat{X}_t = a + bt$$



Regresión Lineal

Mes	Ventas
1 Enero	7000
2 Febrero	9000
3 Marzo	5000
4 Abril	11000
5 Mayo	10000
6 Junio	13000



Regresión Lineal

Calcular la pendiente de los valores:

$$b = \frac{n \sum_{i=1}^n X_i t_i - \sum_{i=1}^n X_i \sum_{i=1}^n t_i}{n \sum_{i=1}^n t_i^2 - [\sum_{i=1}^n t_i]^2}$$

$$[(7000 * 1) + (9000 * 2) + (5000 * 3) + (11000 * 4) + (10000 * 5) + (13000 * 6)] = 212000$$

$$(7000 + 9000 + 5000 + 11000 + 10000 + 13000) = 55000$$

$$(1 + 2 + 3 + 4 + 5 + 6) = 21$$

$$[(1^2) + (2^2) + (3^2) + (4^2) + (5^2) + (6^2)] = 91$$

$$(1 + 2 + 3 + 4 + 5 + 6)^2 = 441$$

Sustituimos:

$$b = [[6*(212000)] - [(5500) * (21)]] / [[6*(91)]-(441)]$$

b = 1114.28 pendiente

Regresión Lineal

Calcular la intersección de la linea:

$$a = \bar{X} - b\bar{t}$$

$$(7000 + 9000 + 5000 + 11000 + 10000 + 13000) / 6 = 9166.67$$

$$(1 + 2 + 3 + 4 + 5 + 6) / 6 = 3.5$$

Sustituimos:

$$9166.67 - [(1114.28)*(3.5)] = 5266.68$$

Intersección de la línea = 5266.68

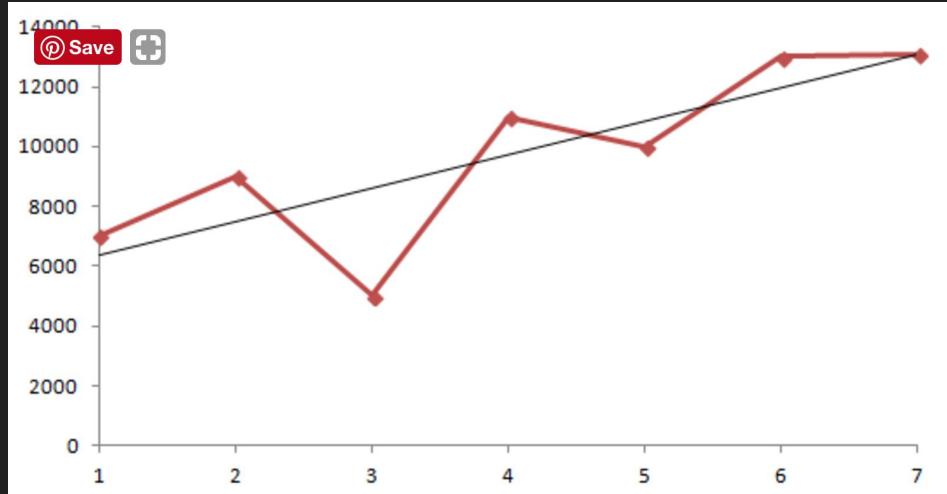
Regresión Lineal

$$\hat{X}_t = a + bt$$

Por último calculamos el pronóstico para el mes 7:

Pronóstico = Intersección + [pendiente * periodo de tiempo]

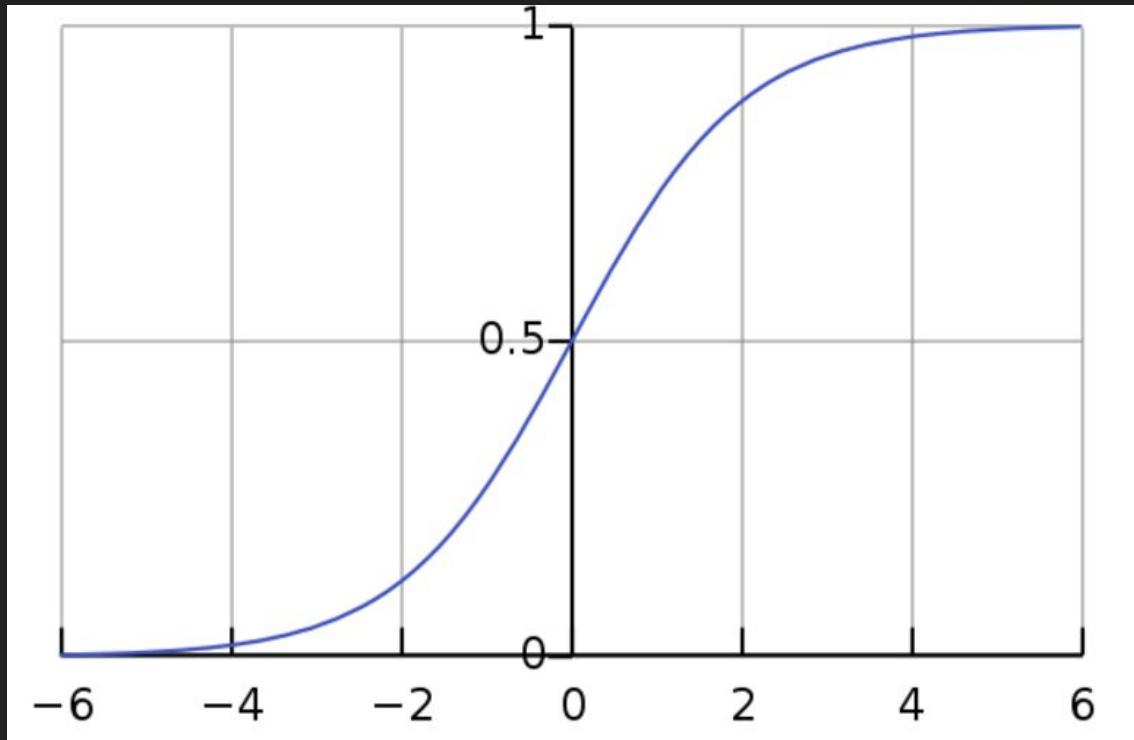
Ventas mes 7 = $5266.68 + [(1114.28) * (7)] = 13067$



Mes	Ventas
1 Enero	7000
2 Febrero	9000
3 Marzo	5000
4 Abril	11000
5 Mayo	10000
6 Junio	13000
7 Julio	13067

7 Julio 13067

Regresión Logística



Regresión Logística

Clasificación

Ejemplos:

Correos Electrónicos: Correo No Deseado / Correo Deseado

Diagnóstico: Enfermo / Sano

Tumor: Maligno / Benigno

$$y \in \{0, 1\}$$

0: Clase Negativa

1: Clase Positiva

Regresión Logística

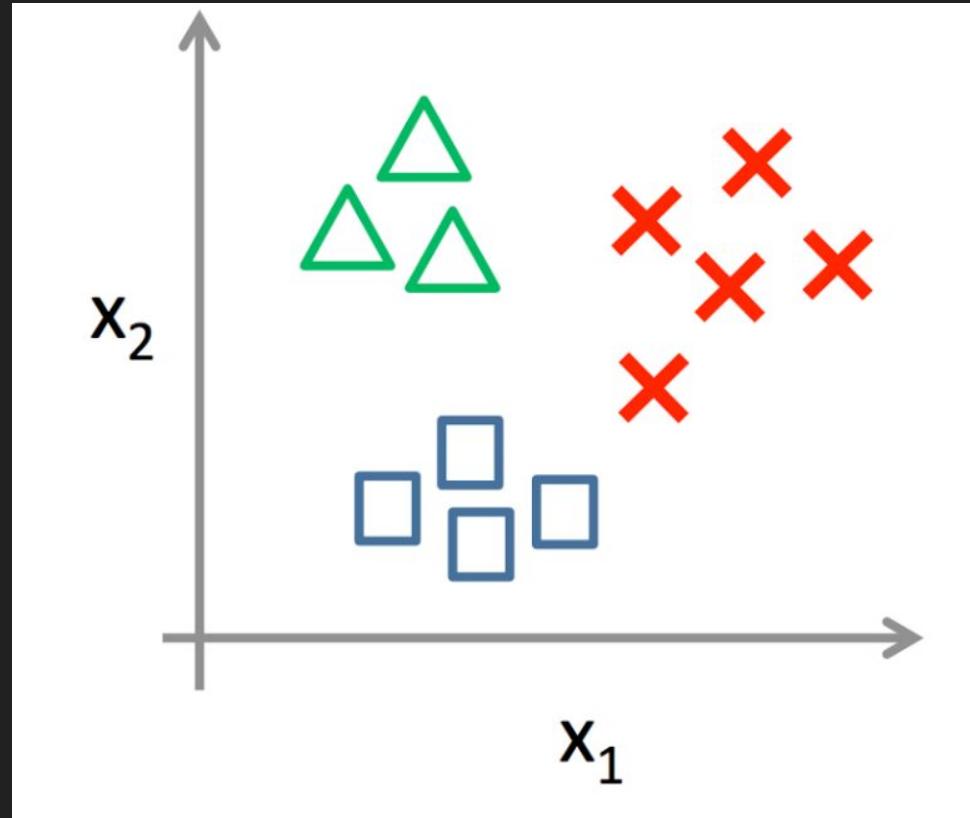
El clasificador está en 0.5:

Si $h_{\theta}(x) \geq 0.5$ entonces $y = 1$

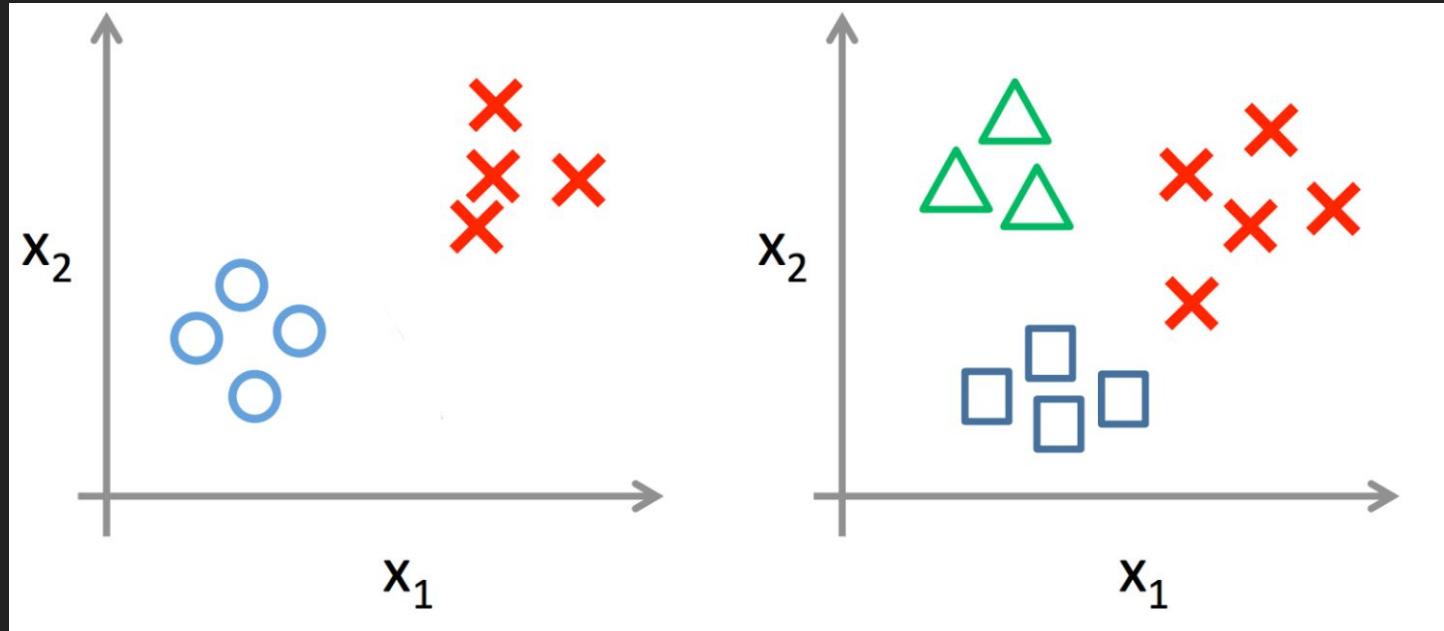
Si $h_{\theta}(x) < 0.5$ entonces $y = 0$

$$0 \leq h_{\theta}(x) \leq 1$$

Regresión Logística Multinomial



Regresión Logística Multinomial



Regresión Logística Multinomial

Clasificación

Ejemplos:

Clima: Soleado, Nublado, Lluvia.

Diagnóstico Médico: Saludable, Gripe, Enfermedad Estomacal

Clasificación de Correos: Trabajo, Familia, Viajes, Compras

Usando Ejemplo Clima:

Soleado = y_1 , Nublado = y_2 , Lluvia = y_3

$$h_{\theta}^{(i)}(x) \quad P(y = i|x; \theta) \quad (i = 1, 2, 3)$$

Regresión Logística Multinomial

Otros Casos de Uso:

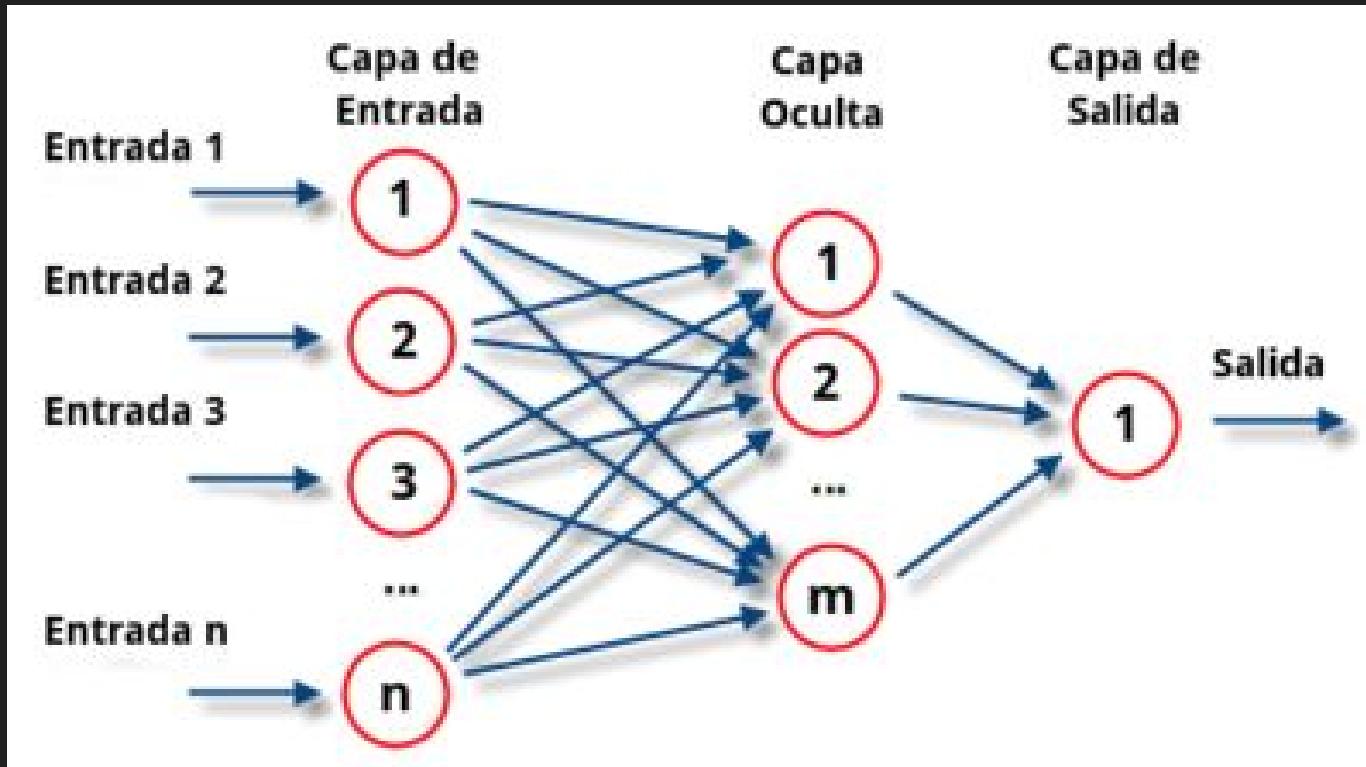
Determinar tipo de Sangre basado en resultados de estudios de laboratorio.

Qué candidato tendrá el voto de una persona teniendo un cuenta determinadas características demográficas.

Redes Neuronales



Redes Neuronales



Redes Neuronales

Compuerta lógica
AND

INPUT		OUTPUT
A	B	A AND B
0	0	0
0	1	0
1	0	0
1	1	1

$$\Theta^{(1)} = [-30 \quad 20 \quad 20]$$

$$h_{\Theta}(x) = g(-30 + 20x_1 + 20x_2)$$

$x_1 = 0$ and $x_2 = 0$ then $g(-30) \approx 0$

$x_1 = 0$ and $x_2 = 1$ then $g(-10) \approx 0$

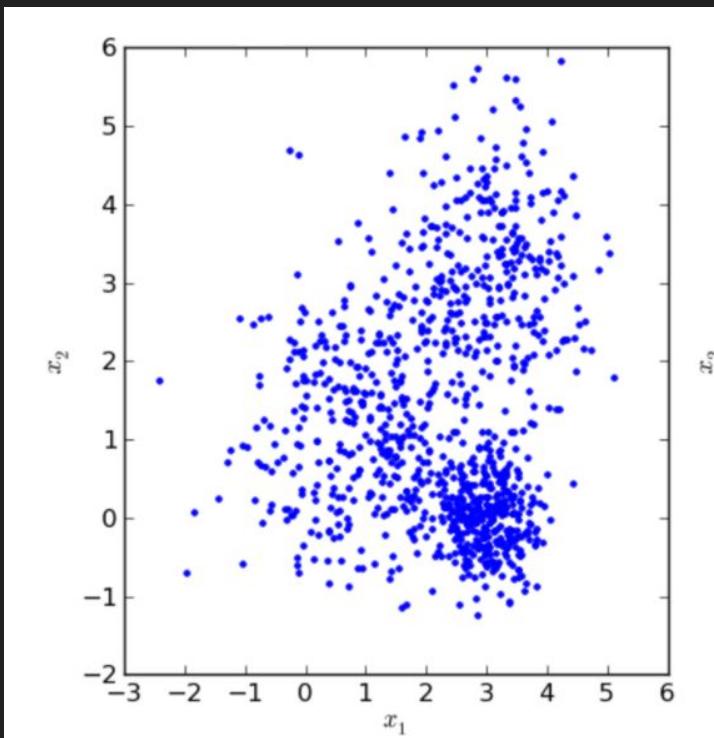
$x_1 = 1$ and $x_2 = 0$ then $g(-10) \approx 0$

$x_1 = 1$ and $x_2 = 1$ then $g(10) \approx 1$

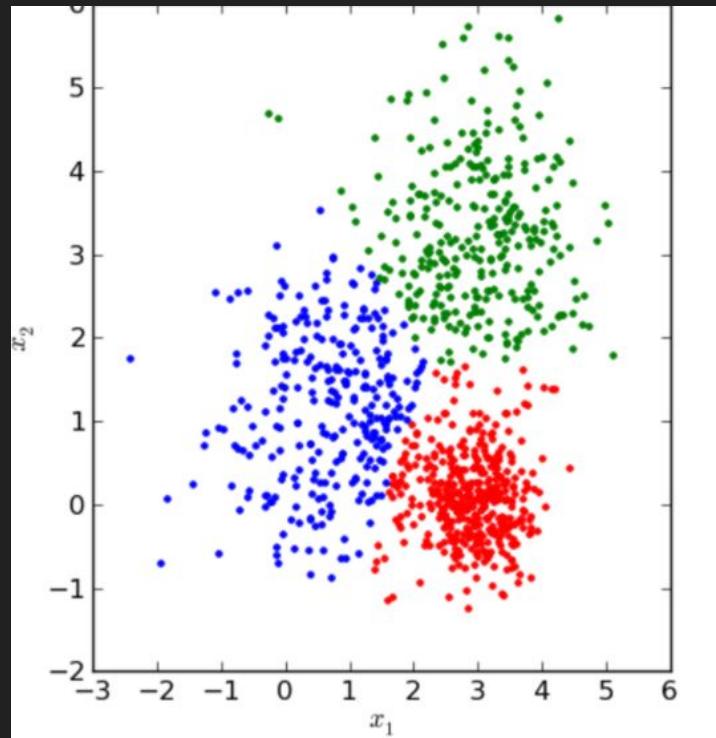
Output = [0, 0, 0, 1]

Algoritmo K-means

Datos originales sin agrupar



Datos Agrupados



Algoritmo K-means

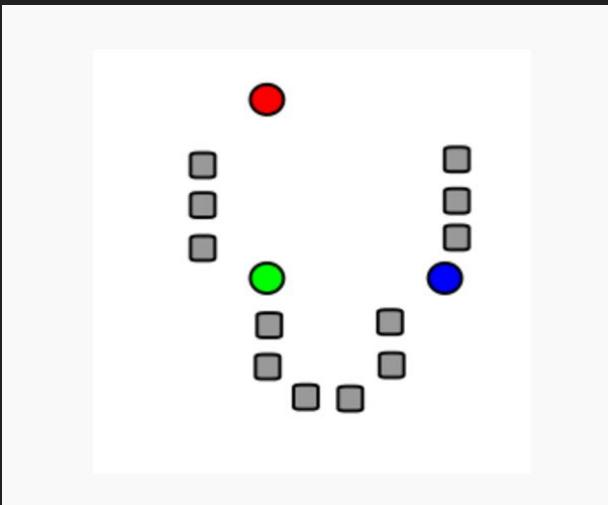
Procedimiento:

Dado un conjunto inicial de k centros, el algoritmo continúa iterando entre 2 pasos:

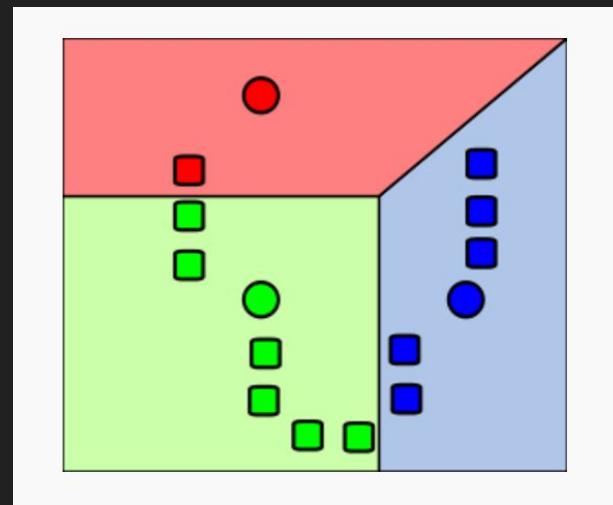
- **Paso de asignación:** Asigna cada dato al grupo con la media más cercana, es decir la partición de las observaciones generadas por los centros.
- **Paso de actualización:** Calcula nuevos centros como el centro de las observaciones en el grupo o partición.

El algoritmo se considera que **converge** cuando los centros ya no cambian.

Algoritmo K-means

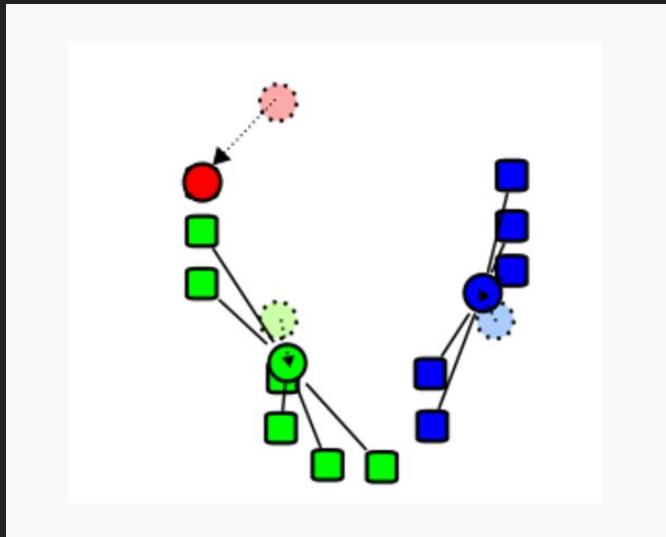


Paso 1: k centros iniciales, en este caso $k = 3$, son generados aleatoriamente en un conjunto de datos.

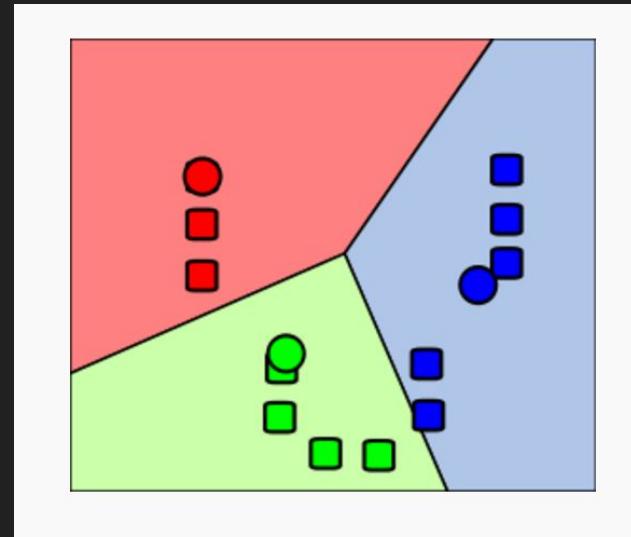


Paso 2: K grupos son generados, en este caso 3 grupos, asociando el dato con la media más cercana.

Algoritmo K-means

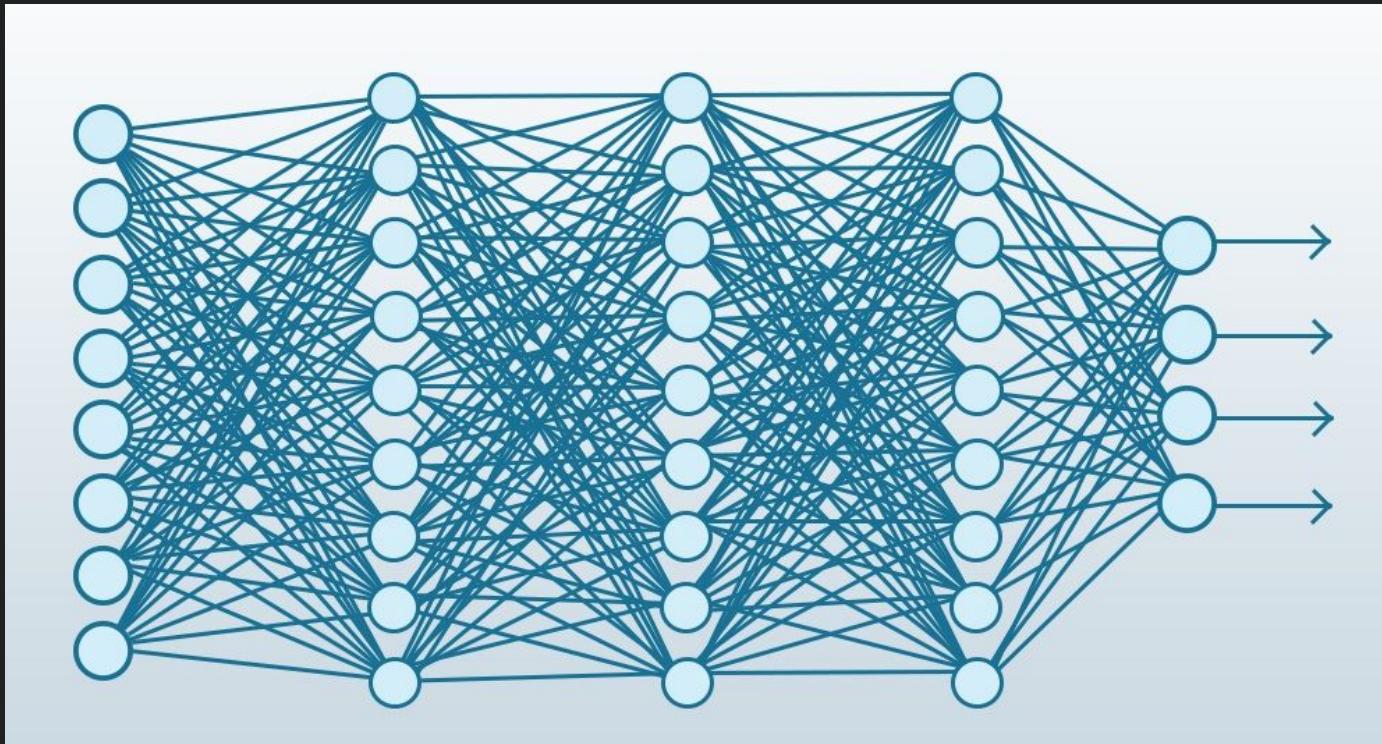


Paso 3: El centro de cada uno de los k grupos se recalcula basado en la media de la partición antes asignada.



Paso 4: Paso 2 y 3 se repiten hasta que los centros dejan de recalcularse.

Deep Learning (Aprendizaje Profundo)

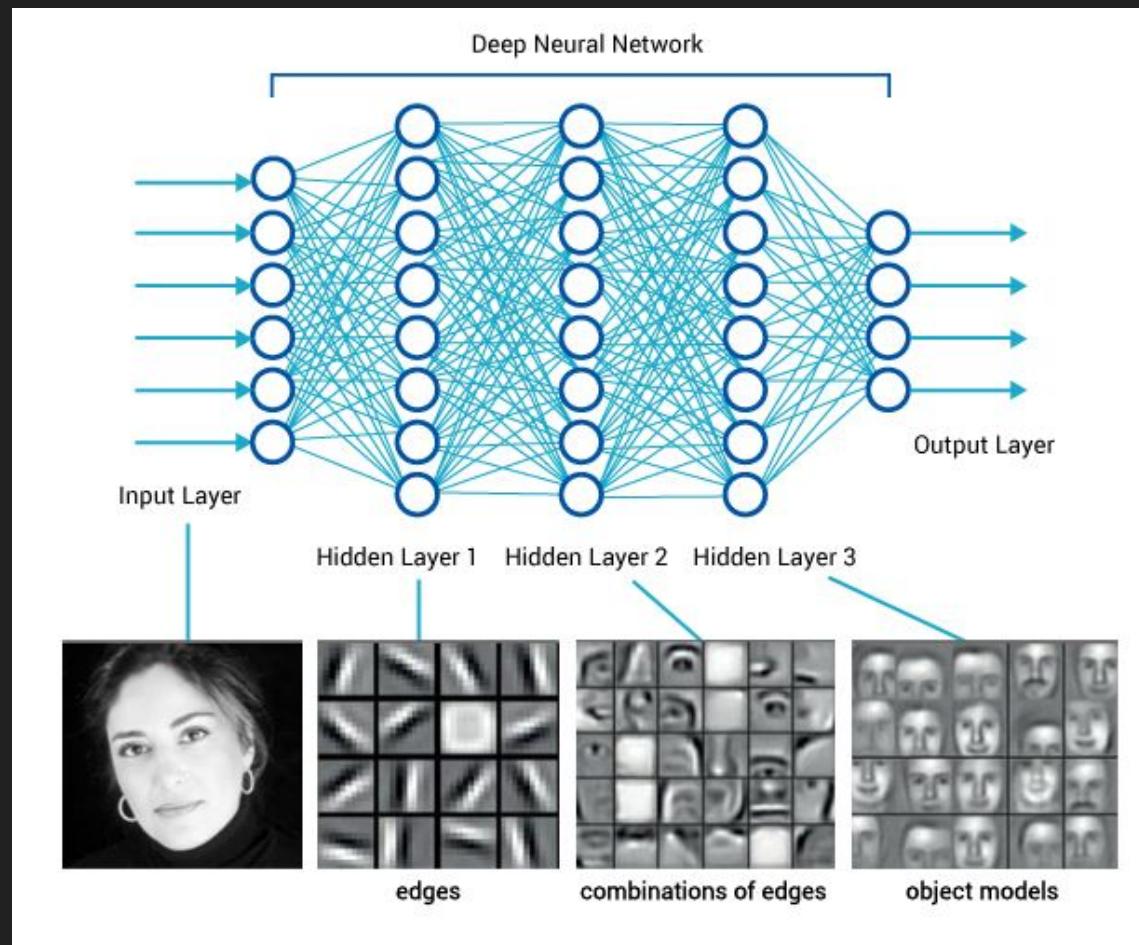


Deep Learning (Aprendizaje Profundo)

Concepto:

- Se utiliza una cascada de capas con unidades de procesamiento no lineal para extraer y transformar variables.
- Cada capa utiliza la salida de la capa anterior como entrada.
- Se puede utilizar con algoritmos de aprendizaje supervisado o de aprendizaje sin supervisión.
- Las características de más alto nivel se derivan de las características de nivel inferior, formando una representación jerárquica.

Deep Learning (Aprendizaje Profundo)



Deep Learning (Aprendizaje Profundo)

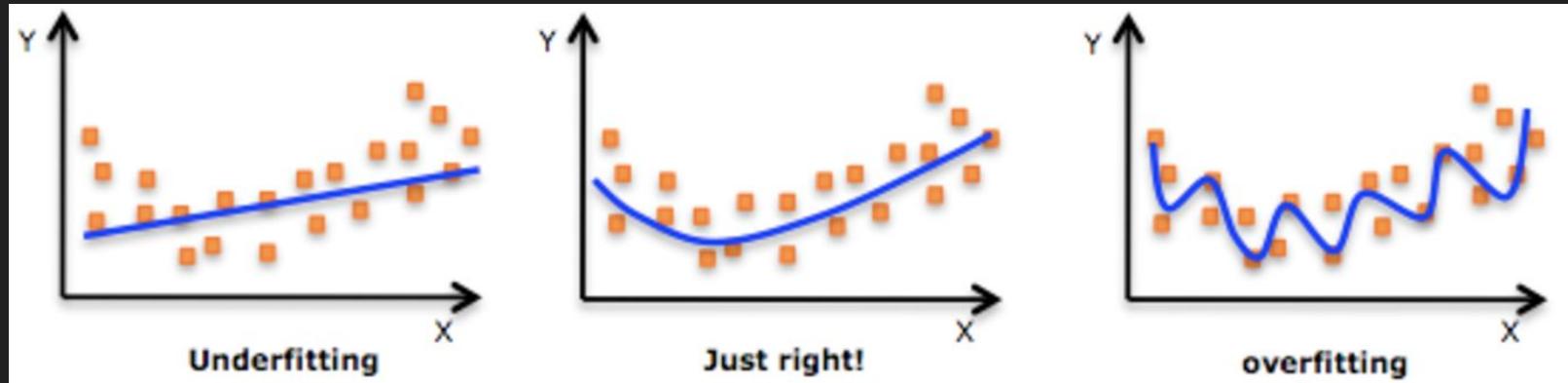
Algunos **algoritmos** de Deep Learning:

- Deep Neural Networks
- Convolutional Neural Networks
- Deep Belief Networks

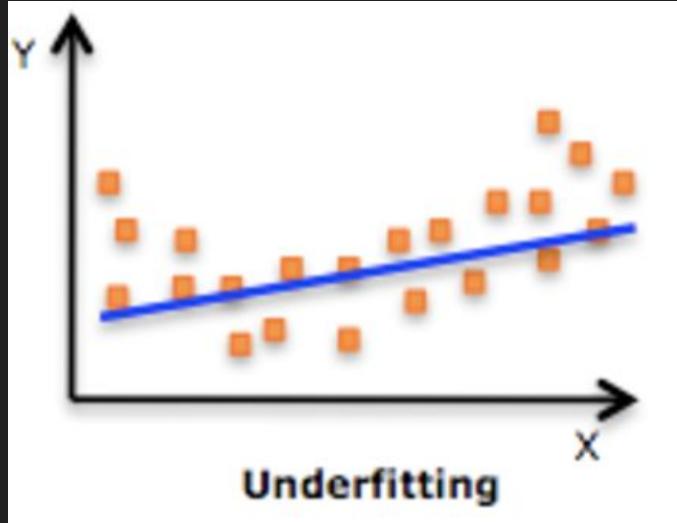
Algunas **áreas** donde se usa Deep Learning:

- Computer Vision
- Reconocimiento del habla
- Reconocimiento de audio

Problemas frecuentes (Underfitting vs. Overfitting)

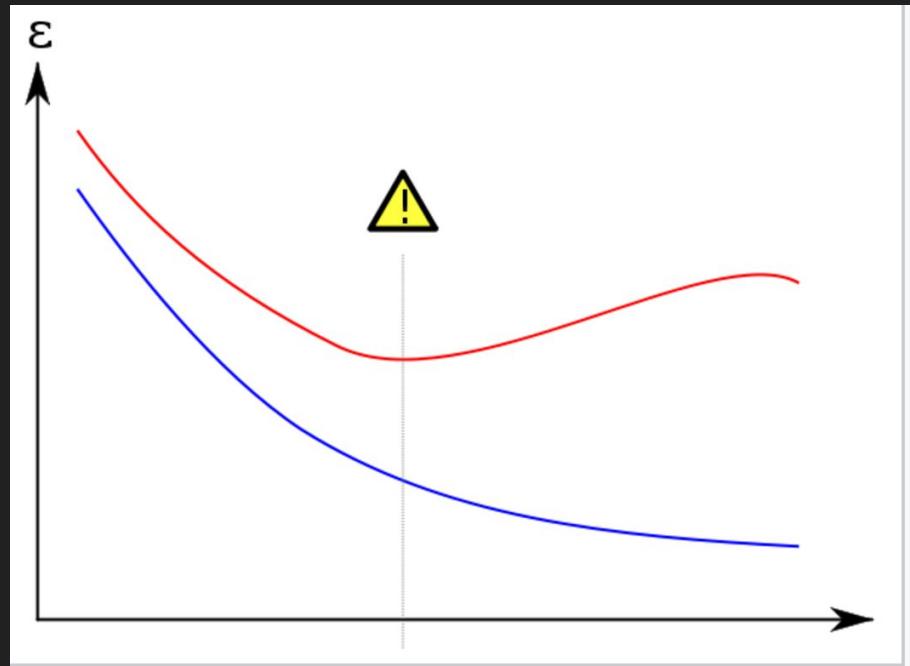
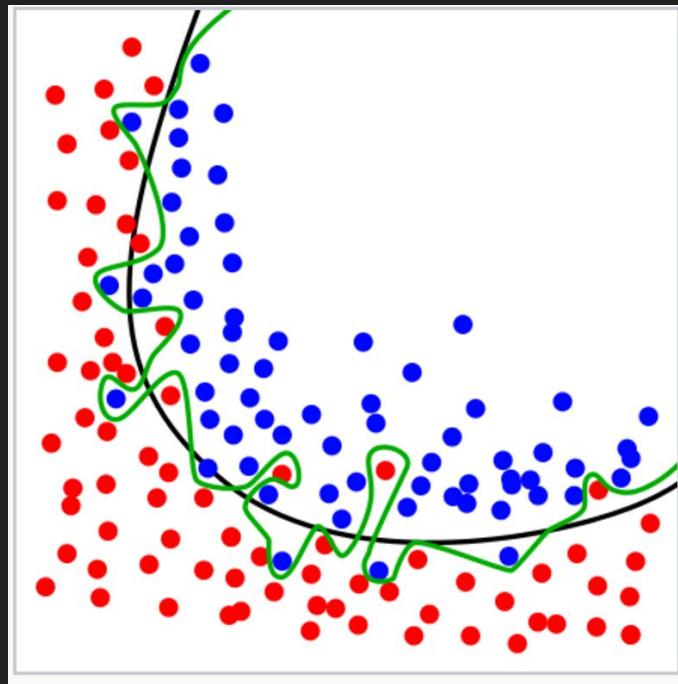


Underfitting (Subajustando)



Se presenta cuando un modelo no puede capturar la tendencia de los datos. Es generalmente el resultado de un modelo extremadamente simple

Overfitting (Sobrejustando)



Se presenta cuando el algoritmo está perfectamente adaptado a los datos con los que lo entrenamos, pero si tratamos de predecir datos nuevos, lo más probable es que nos de error.

Consejos para resolver los problemas

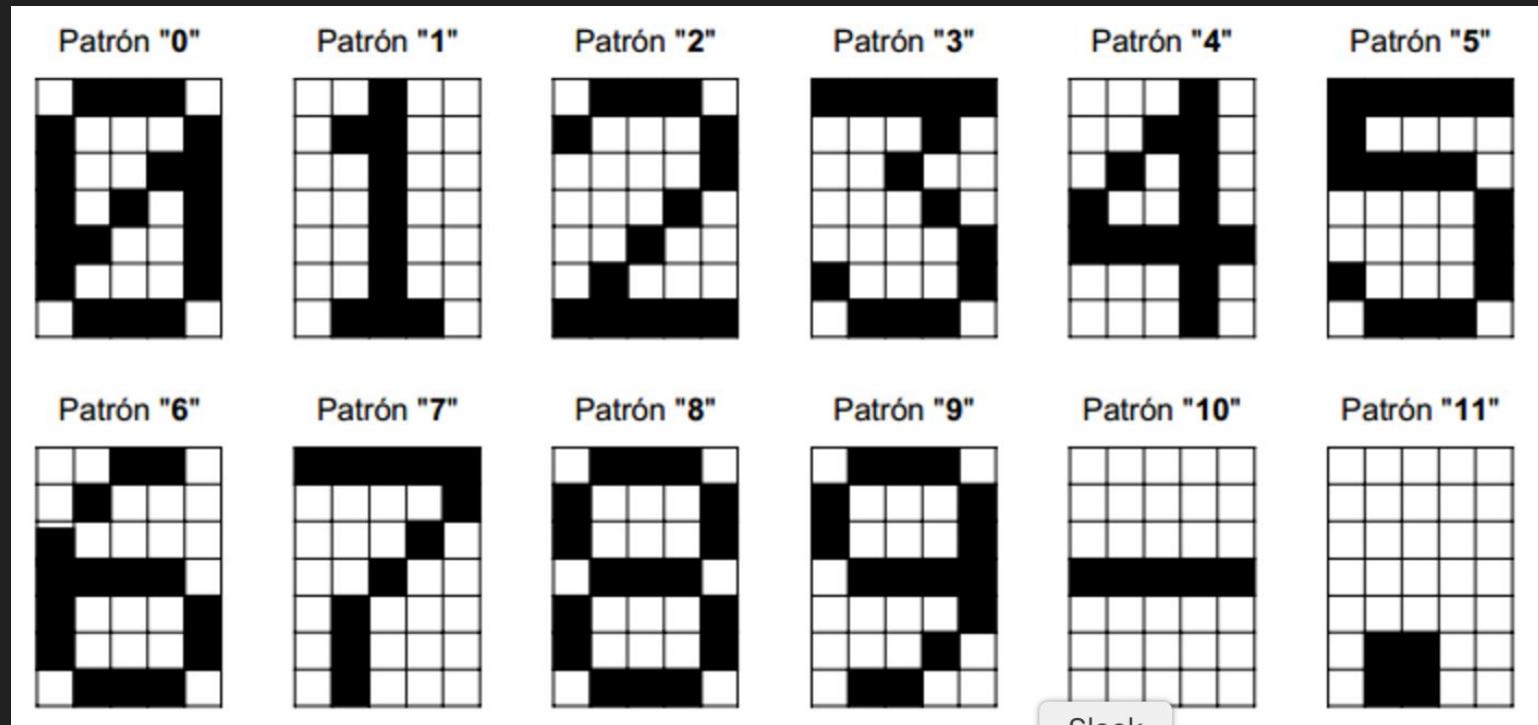
Para resolver **underfitting**:

- Se recomienda tratar de agregar más features a tu dataset o dicho de otra forma agregar más columnas con condiciones a tu modelo.

Para resolver **overfitting**:

- Se recomienda disminuir la cantidad de features a tu dataset o dicho de otra manera quitar columnas de condiciones a tu modelo.
- También se recomienda agregar más ejemplos para el entrenamiento o dicho de otra manera agregar renglones con ejemplos de datos para tu modelo.

Construyendo un Modelo de ML.



Python para ML



TensorFlow para ML



MNIST Dataset



Consejos para aplicar Machine Learning

Google

facebook

amazon

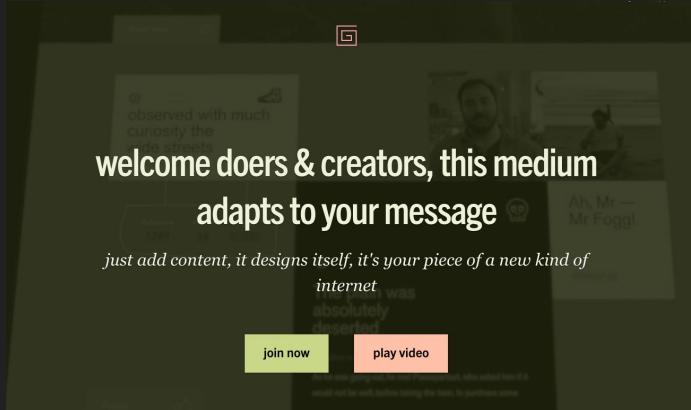


Hi, I'm Amy

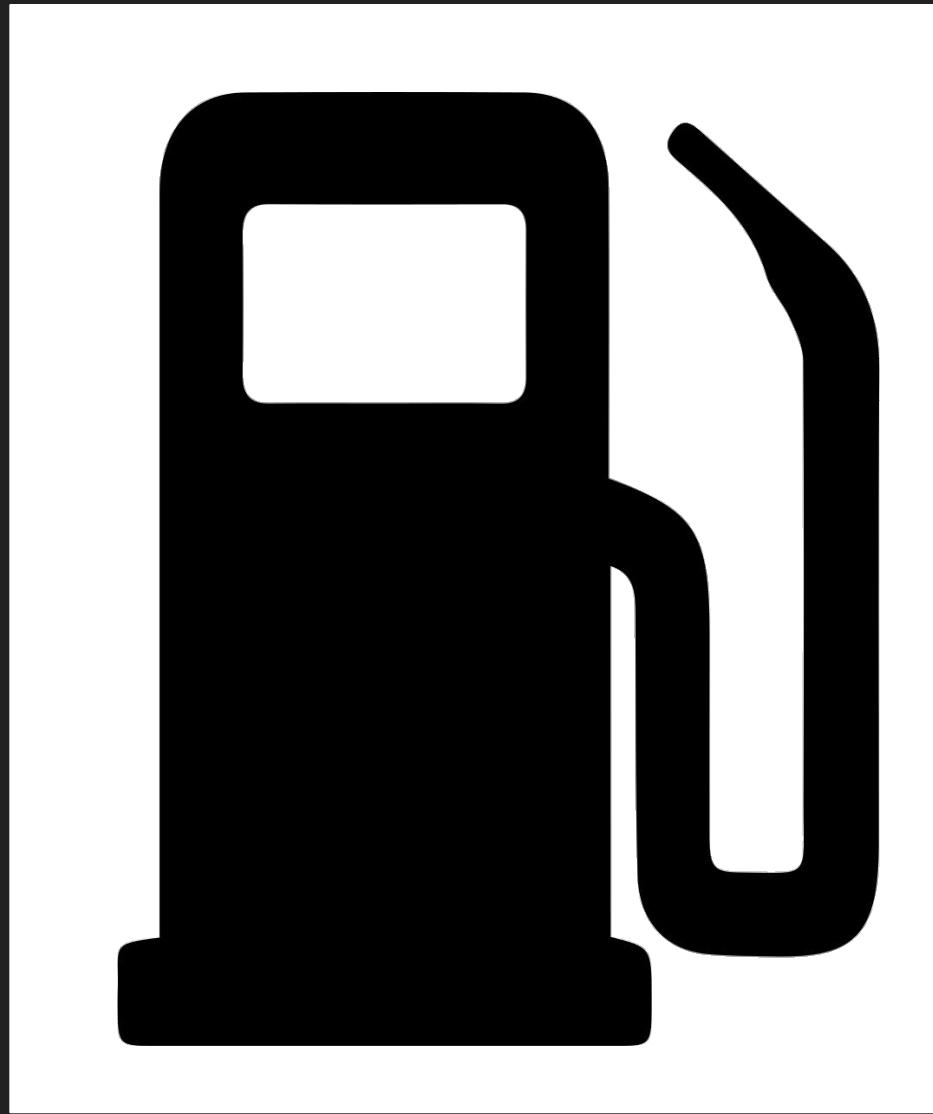
Your AI powered personal assistant for scheduling meetings. You interact with me as you would to any other person – and I'll do all the tedious email ping pong that comes along with scheduling a meeting.

No sign-in, no password, no download, all you do is:

[Cc: amy@x.ai](mailto:CC: amy@x.ai)



**El secreto del
cambio es
enfocar tu
energía no en
luchar contra lo
viejo sino en
construir lo
nuevo.**



Google

facebook

amazon



Lo que **destaca** a una
empresa de machine learning
aplicado con **éxito** es
generalmente **la novedad,**
calidad y/o cantidad de
datos a los que tienen acceso.

LISTEN UP

Shortage of data causing Samsung troubles in launching Bixby Voice for the U.S.

BY ANDREW MARTONIK • Wednesday, Jul 5, 2017 at 1:41 pm EDT

23 Comments

“El retraso se debe a una gran cantidad de problemas con el servicio, pero es generado por la falta de datos de uso que es lo que se necesita para que los sistemas de machine learning puedan trabajar a su máximo potencial.”