

ACÀMICA

¡Bienvenidos/as a Data Science!



Agenda

¿Cómo anduvieron?

Actividad: Data Science en mi vida

Explicación: Redes Neuronales

Hands-On

Break

Explicación: Descenso por gradiente

Hands-On

Resumen: Descenso por gradiente

Cierre



¿Dónde estamos?



¿Cómo anduvieron?



A partir de la semana que viene:
Data Science en mi vida



Data Science en mi vida

En 10/15 minutos, tendrán que contar a sus compañeros y equipo docente lo siguiente:

- a) En qué problemas estoy aplicando lo aprendido en DS y cómo lo estoy encarando.
- b) Contar algún tema que me interese o que proyecto aplicar relacionado con lo que vimos.

Data Science en mi vida: ¿Cómo preparo mi presentación?



LA GUÍA COMPLETA PARA NO ENTRAR EN PÁNICO



La regla de las 3 Ps:

1

PREPARAR

2

PRACTICAR

3

PRESENTAR

La regla de las 3 Ps:

1

PREPARAR

2

PRACTICAR

3

PRESENTAR

1

PREPARAR

Sugerimos que armen un esqueleto de presentación (Word, Excel, etc.) con lo que van a presentar teniendo en cuenta los siguientes puntos:

¿A quién?

¿Qué?

¿Cómo?

¿Tiempo?

1

PREPARAR

¿A quién?

¿Qué?

¿Cómo?

¿Tiempo?

Le presentan a la clase y al equipo docente.

¿Qué les puede servir a ellos/as de su experiencia?



1

PREPARAR

¿A quién?

¿Qué?

¿Cómo?

¿Tiempo?

Viaje

¿Con qué obstáculos te encontraste?

¿Qué decisiones tomaste?

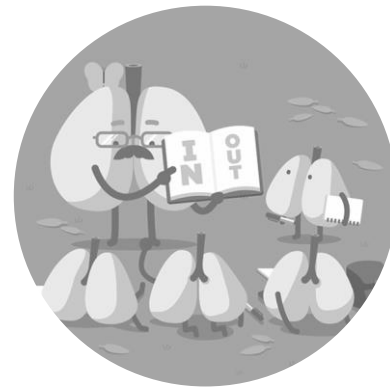
¿Qué quisiste lograr?

Puerto de llegada

¿Cuáles son los resultados que alcanzaste?

¿Qué te gustaría mejorar?

¿Qué aprendiste?



1

PREPARAR

¿A quién?

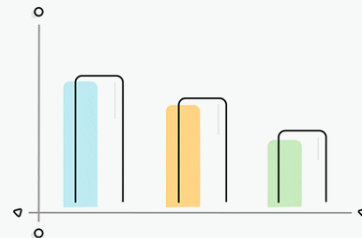
¿Qué?

¿Cómo?

¿Tiempo?

*TIP: usa fotos, gráficos, diagramas...en fin, elementos visuales.

¡El 65% de las personas son pensadores puramente visuales!



1

PREPARAR

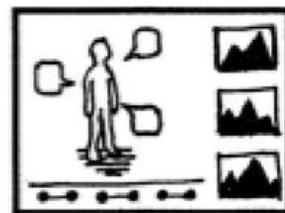
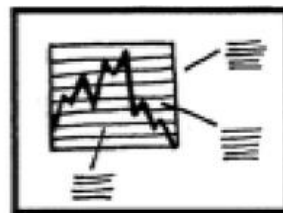
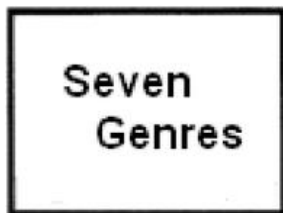
¿A quién?

¿Qué?

¿Cómo?

¿Tiempo?

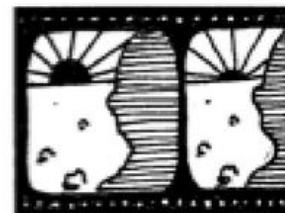
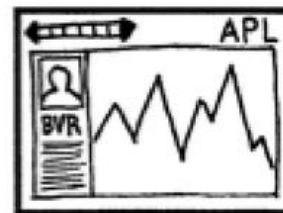
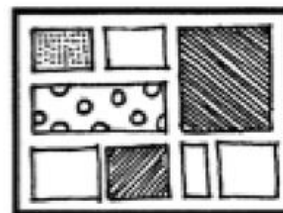
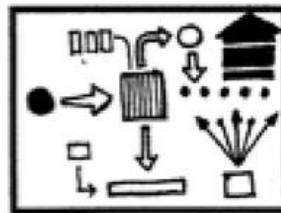
Los 7 géneros de Edward Segel



Magazine Style

Annotated Chart

Partitioned Poster



Flow Chart

Comic Strip

Slide Show

Film/Video/Animation

1

PREPARAR

¿A quién?

¿Qué?

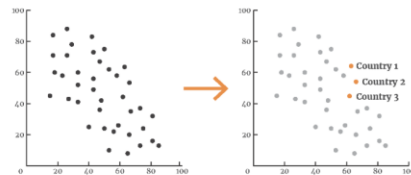
¿Cómo?

¿Tiempo?

Core Principles of Data Visualization

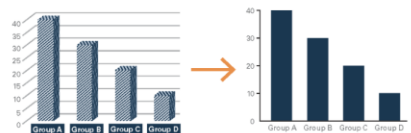
Show the data

People read graphs in a research report, article, or blog to understand the story being told. The data is the most important part of the graph and should be presented in the clearest way possible. But that does not mean that all of the data must be shown—indeed, many graphs show too much.



Reduce the clutter

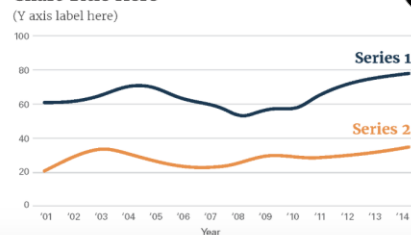
Chart clutter, those unnecessary or distracting visual elements, will tend to reduce effectiveness. Clutter comes in the form of dark or heavy gridlines; unnecessary tick marks, labels, or text; unnecessary icons or pictures; ornamental shading and gradients; and unnecessary dimensions. Too often graphs use textured or filled gradients.



Integrate the text and the graph

Standard research reports often suffer from the **slideshow effect**, in which the writer narrates the text elements that appear in the graph. A better model is one in which visualizations are constructed to complement the text and at the same time to contain enough information to stand alone. As a simple example, legends that define or explain a line, bar, or point are often placed far from the content of the graph—off to the right or below the graph. Integrated legends—right below the title, directly on the chart, or at the end of a line—are more accessible.

Chart Title Here



1

PREPARAR

¿A quién?

¿Qué?

¿Cómo?

¿Tiempo?

Ejemplos:

<http://drones.pitchinteractive.com/>

<https://www.nytimes.com/interactive/2017/06/14/world/europe/migrant-rescue-efforts-deadly.html>

<https://projects.propublica.org/houston-cypress/>

<https://projects.propublica.org/graphics/harvey-maps>

<https://podio.com/site/creative-routines>

<https://www.bloomberg.com/graphics/2015-whats-warming-the-world/>

<https://www.axios.com/wall-street-asset-management-climate-change-activism-19fe3b28-eab2-4b1c-951b-1c1c2ebe8f33.html>

1

PREPARAR

¿A quién?

¿Qué?

¿Cómo?

¿Tiempo?

¿Presento

- **primero el resultado final** y luego el proceso o
- cuento el proceso **cronológicamente** hasta el resultado al final?

1

PREPARAR

¿A quién?

¿Qué?

¿Cómo?

¿Tiempo?

ESTRUCTURA VISUAL

*apoya la
historia*

1

PREPARAR

¿A quién?

¿Qué?

¿Cómo?

¿Tiempo?

**ESTRUCTURA
VISUAL**

*apoya la
historia*

MENSAJE

*cuenta la
historia*

1

PREPARAR

¿A quién?

¿Qué?

¿Cómo?

¿Tiempo?

ESTRUCTURA VISUAL

*apoya la
historia*

MENSAJE

*cuenta la
historia*

INTERACTIVIDAD

*hazla dinámica y
atrapante*



1

PREPARAR

¿A quién?

¿Qué?

¿Cómo?

¿Tiempo?

Cada uno/a tendrá
10 minutos como máximo para exponer
(luego podremos hacerle preguntas y/o
darle feedback).

La regla de las 3 Ps:

1

PREPARAR

2

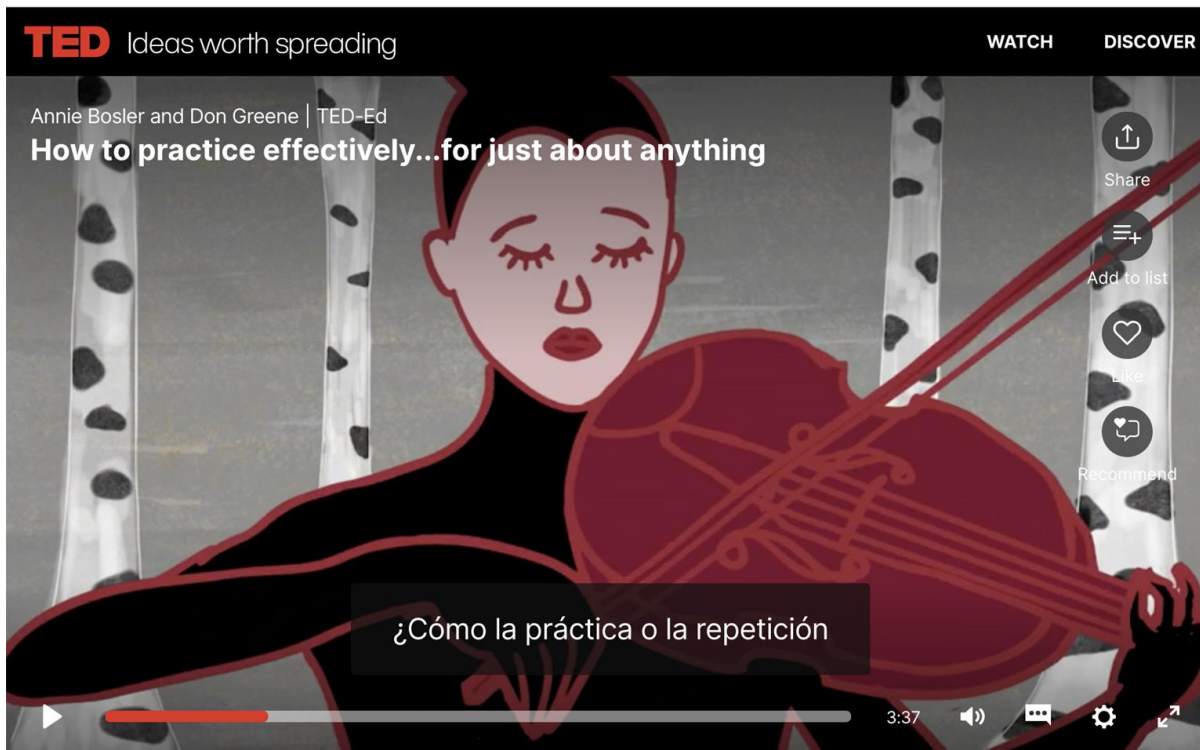
PRACTICAR

3

PRESENTAR

2

PRACTICAR



La regla de las 3 Ps:

1

PREPARAR

2

PRACTICAR

3

PRESENTAR

3

PRESENTAR

Emoción

Actitud

Intención

Si demuestras entusiasmo por tu trabajo (sin importar a lo que hayas llegado), vas a contagiar al público.

Conecta con tu entorno. Aunque te dé vértigo, miedo y timidez, ¡mira a los ojos!

***TIP: decir en voz alta “estoy muy nervioso/a” te ayudará a relajarte.** 😊

3

PRESENTAR

Emoción

Actitud

Intención

Si pretendes diseñar tu presentación con una determinada emoción, ¡acompañala con tu **cuerpo, gestos y cara!**



3

PRESENTAR

Emoción

Actitud

Intención

¿Buscas feedback del público?
¡Pídelo al terminar!

Redes neuronales



Redes neuronales

Esperamos que aprendan los siguientes conceptos:

- Perceptrón, Funciones de Activación
- Forward Propagation
- Backpropagation
- Descenso por gradiente (Gradient Descent)
- Redes Neuronales Profundas
- Regularización
- Redes Neuronales Convolucionales (CNN, si hay tiempo)
- Entornos de desarrollo: Keras, Tensor Flow
- Y muchos que probablemente nos estemos olvidando.

Redes neuronales

Las redes neuronales son particularmente buenas para trabajar con “Unstructured Data”: Imágenes/videos, Audio, Texto (NLP), juegos, etc.

Imágenes

- Clasificación de Imágenes en Facebook y Google
- [DeepFakes](#)
- Visión por computadora para vehículos autónomos

Redes neuronales

Las redes neuronales son particularmente buenas para trabajar con “Unstructured Data”: Imágenes/videos, Audio, Texto (NLP), juegos, etc.

Imágenes

- Clasificación de Imágenes en Facebook y Google
- [DeepFakes](#)
- Visión por computadora para vehículos autónomos

Texto

- [Google Translate](#) - Google Neural Machine Translation
- [BERT/ELMO](#)
- [GPT-2](#)
- [GPT-2 Explicada](#)
- Para jugar: <https://talktotransformer.com/>
- Canal de YouTube: [Two Minute Papers](#)

Redes neuronales

Las redes neuronales son particularmente buenas para trabajar con “Unstructured Data”: Imágenes/videos, Audio, Texto (NLP), juegos, etc.

Imágenes

- Clasificación de Imágenes en Facebook y Google
- [DeepFakes](#)
- Visión por computadora para vehículos autónomos

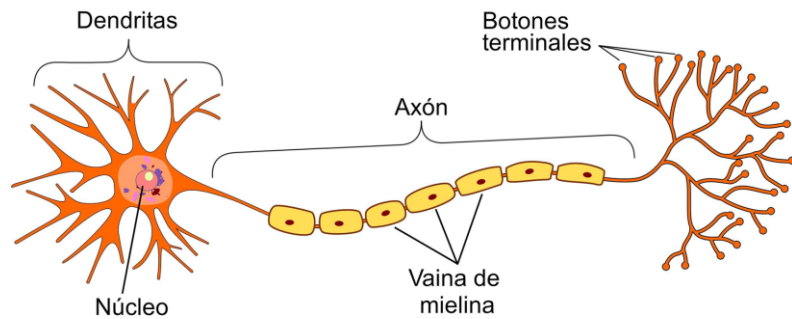
Texto

- [Google Translate](#) - Google Neural Machine Translation
- [BERT/ELMO](#)
- [GPT-2](#)
- [GPT-2 Explicada](#)
- Para jugar: <https://talktotransformer.com/>
- Canal de YouTube: [Two Minute Papers](#)

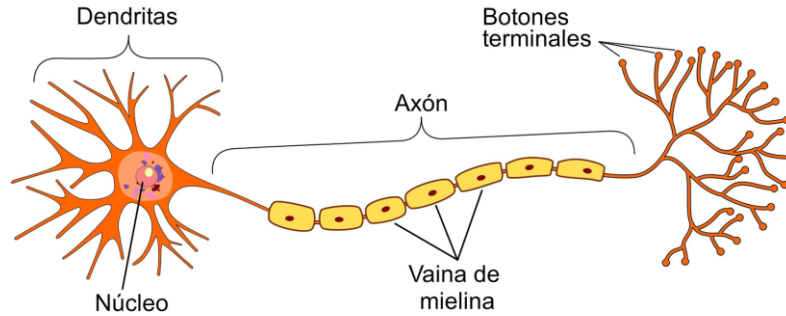
Juegos

- AlphaGo
- [Deep Reinforcement Learning](#)

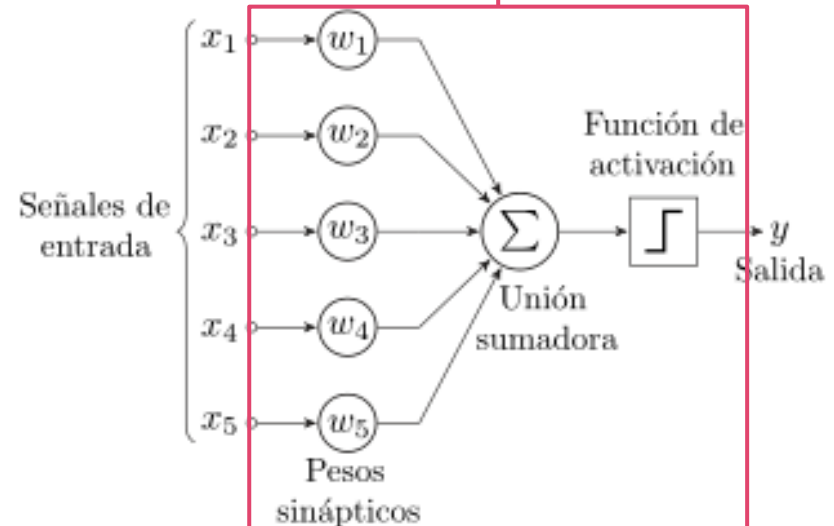
Redes neuronales



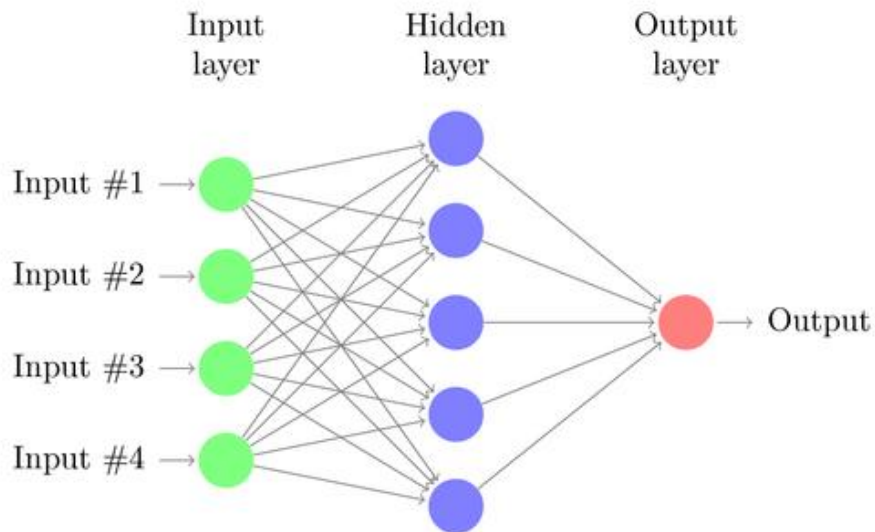
Redes neuronales



Neurona básica:
Perceptrón Simple

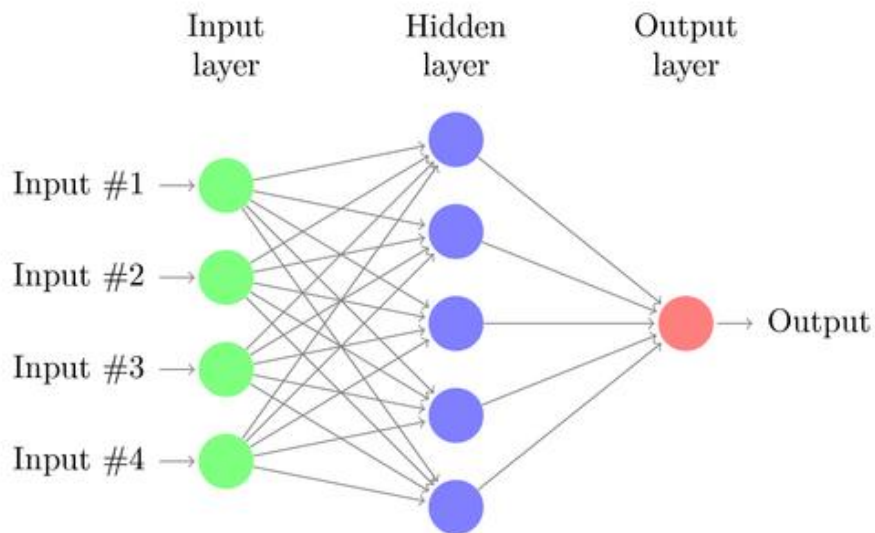


Redes neuronales



- Cada neurona tiene sus propios pesos/parámetros. En aplicaciones comunes suelen ser desde miles a millones de parámetros para toda la red.

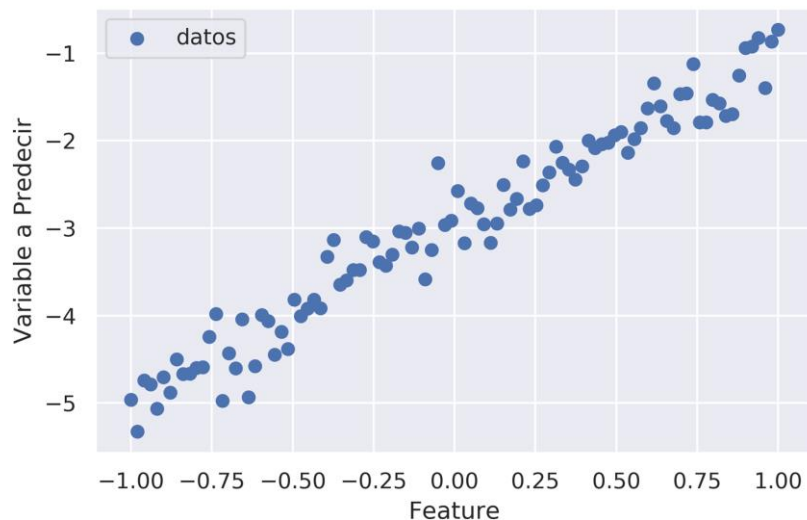
Redes neuronales



- Cada neurona tiene sus propios pesos/parámetros. En aplicaciones comunes suelen ser desde miles a millones de parámetros para toda la red.

- **Deep Learning** es encontrar esos pesos de manera eficiente, bajo la condición de realizar correctamente una tarea objetivo.

Redes neuronales • Preliminares



Buscamos $Y = mX + b$ que mejor ajuste a los datos

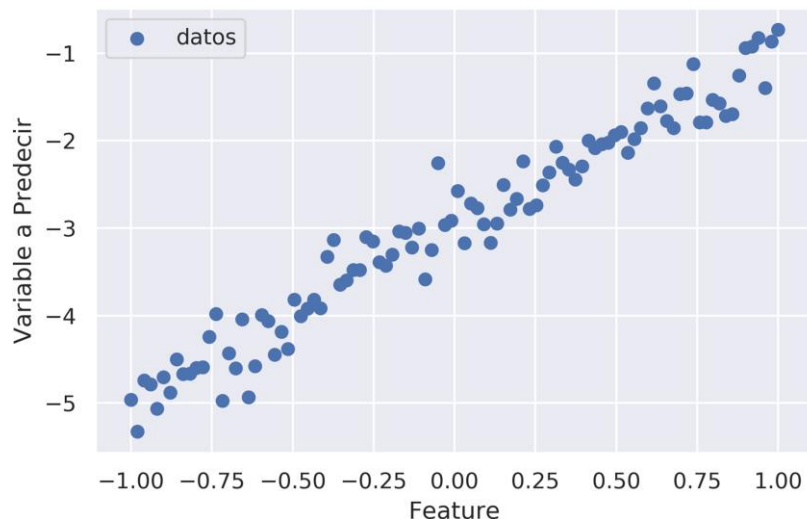
m: pendiente

b: ordenada al origen

¿Qué podemos hacer?



Redes neuronales • Preliminares



Buscamos $Y = mX + b$ que mejor ajuste a los datos

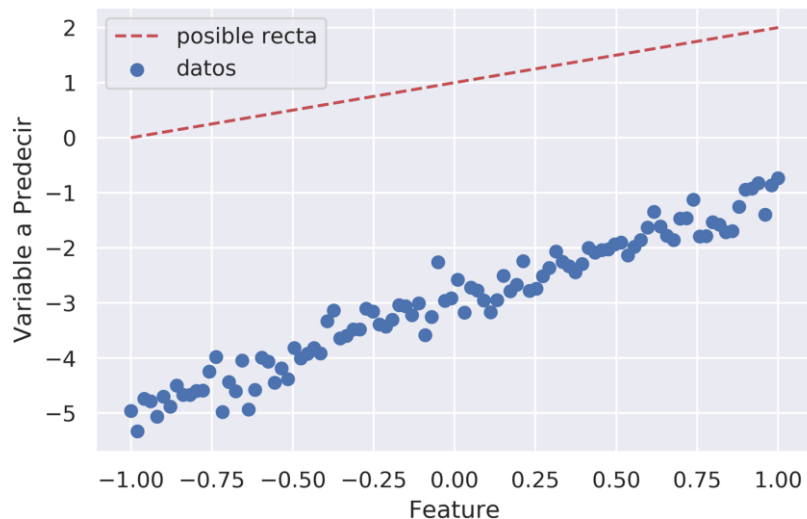
m: pendiente

b: ordenada al origen

1 - Tenemos que definir **qué es mejor ajuste a los datos**.

2 - Podemos **probar con distintos valores de m y b** , y quedarnos con el que mejor ajusta.

Redes neuronales • Preliminares



$$m = 1, b = 1$$

Buscamos $Y = mX + b$ que mejor ajuste a los datos

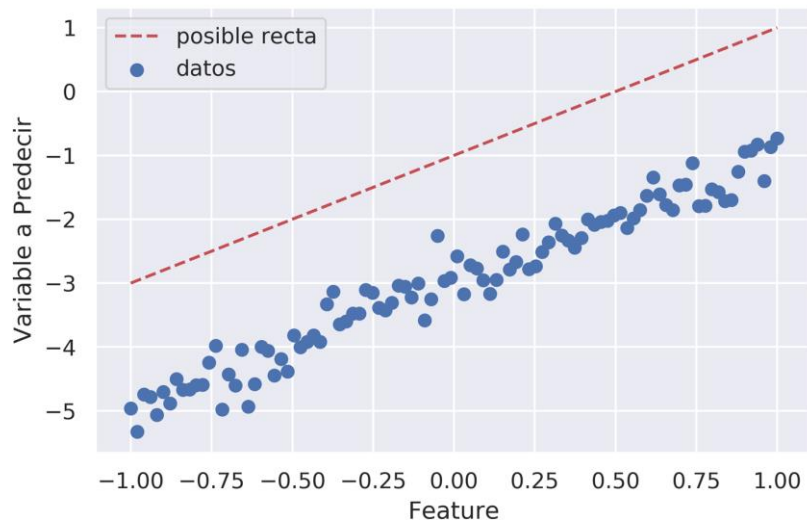
m: pendiente

b: ordenada al origen

1 - Tenemos que definir **qué es mejor ajuste a los datos**.

2 - Podemos **probar con distintos valores de m y b** , y quedarnos con el que mejor ajusta.

Redes neuronales • Preliminares



$$m = 2, b = -1$$

Buscamos $Y = mX + b$ que mejor ajuste a los datos

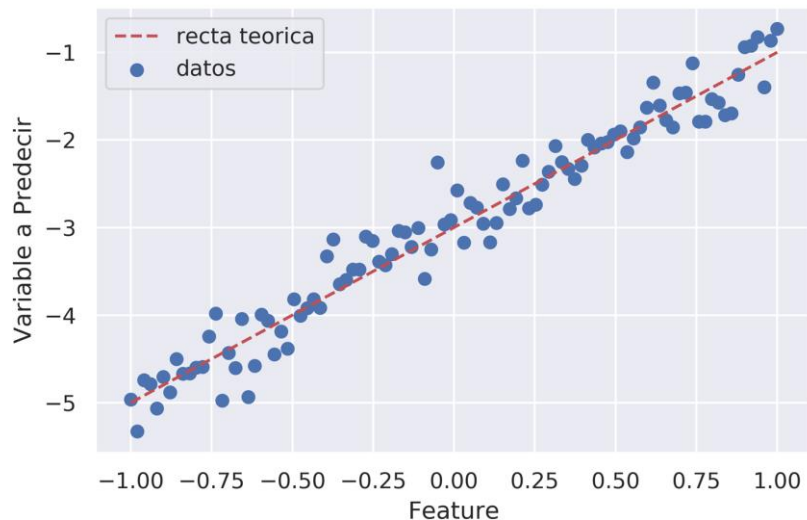
m: pendiente

b: ordenada al origen

1 - Tenemos que definir **qué es mejor ajuste a los datos**.

2 - Podemos **probar con distintos valores de m y b** , y quedarnos con el que mejor ajusta.

Redes neuronales • Preliminares



$$m = 2, b = -3$$

Buscamos $Y = mX + b$ que mejor ajuste a los datos

m: pendiente

b: ordenada al origen

1 - Tenemos que definir **qué es mejor ajuste a los datos**.

2 - Podemos **probar con distintos valores de m y b** , y quedarnos con el que mejor ajusta.

Redes neuronales • Preliminares

Entonces:

Si tenemos un modelo paramétrico

($y = mx + b$, m y b parámetros),

podemos hacer una búsqueda por fuerza bruta para encontrar el m y b .

¿Cómo guiamos la búsqueda?



Redes neuronales • Preliminares

Guiamos la búsqueda con una **función de costo**.

Nota 1: es muy similar a GridSearch, pero en este caso es para buscar los parámetros del modelo, NO sus hiperparámetros. Ejemplo de hiperparámetro: grado del polinomio.

Nota 2: Esta clase de problemas se conocen como problemas de optimización

Actividad: Hands-on training



DS_Encuentro_29_RN_parte1.ipynb

Este notebook es demostrativo,
pero recomendamos trabajarlo
con un/a compañero/a.



A close-up photograph of a white ceramic cup filled with a latte. The surface of the milk is decorated with intricate latte art, featuring a central heart shape surrounded by concentric, wavy lines. The cup is placed on a matching white saucer. In the background, a blurred white napkin and a silver spoon are visible on a dark, textured surface. The overall lighting is soft and focused on the cup.

¡BREAK!



Descenso por gradiente



Preliminares...
¿qué problemas tiene?



Preliminares • Problemas

- Si la grilla es muy gruesa podemos perder el mínimo.
- El costo computacional crece exponencialmente con las dimensiones.

Preliminares...

¿se puede hacer mejor?



Preliminares • Posibles soluciones

Opción 1

Formas analíticas (fórmulas matemáticas) que calculen exactamente el mínimo de la función de costo y a qué parámetros corresponde ese mínimo. Pocos casos pueden resolverse así.

Ejemplo: cuadrados mínimo para la regresión lineal.

Preliminares • Posibles soluciones

Opción 1

Formas analíticas (fórmulas matemáticas) que calculen exactamente el mínimo de la función de costo y a qué parámetros corresponde ese mínimo. Pocos casos pueden resolverse así.

Ejemplo: cuadrados mínimo para la regresión lineal.

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

Preliminares • Posibles soluciones

Opción 1

Formas analíticas (fórmulas matemáticas) que calculen exactamente el mínimo de la función de costo y a qué parámetros corresponde ese mínimo. Pocos casos pueden resolverse así.

Ejemplo: cuadrados mínimo para la regresión lineal.

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

m y b ,
como vector

Datos, en
forma matricial

Preliminares • Posibles soluciones

Opción 1

Formas analíticas (fórmulas matemáticas) que calculen exactamente el mínimo de la función de costo y a qué parámetros corresponde ese mínimo. Pocos casos pueden resolverse así.

Ejemplo: cuadrados mínimo para la regresión lineal.

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

m y b ,
como vector

Datos, en
forma matricial

Opción 2

Mezclar los dos mundos, es decir, combinar la búsqueda en un espacio de parámetros con una guía que nos diga dónde buscar.

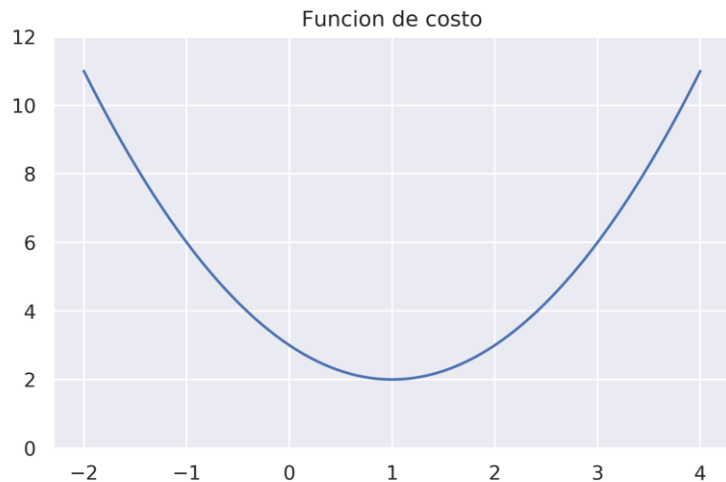
Ejemplo: descenso por gradiente.

Descenso por gradiente.



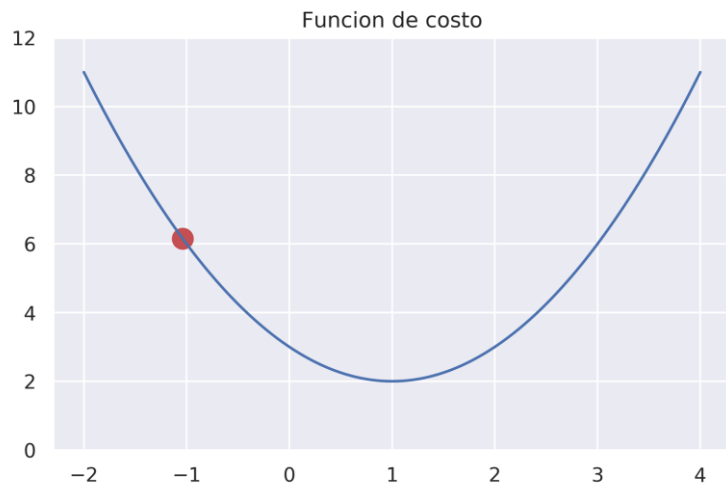
Descenso por gradiente

Queremos explorar el mínimo, pero no hicimos una exploración exhaustiva de la función de costo:



Descenso por gradiente

Queremos explorar el mínimo, pero no hicimos una exploración exhaustiva de la función de costo:

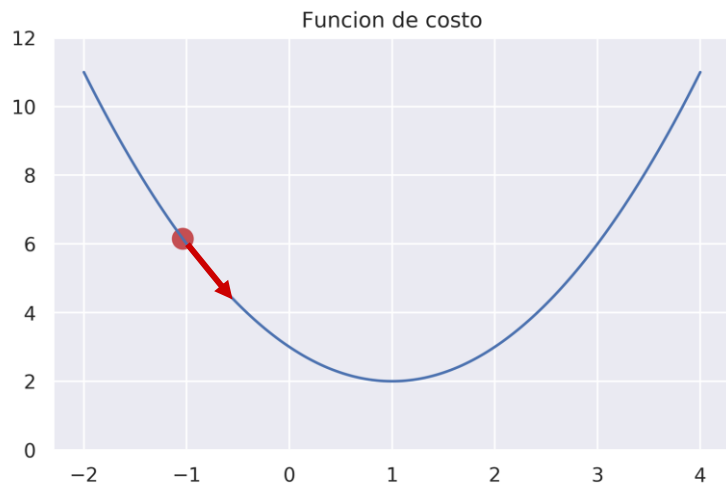


Pasos

1. Calculamos el costo para ciertos valores al azar de los parámetros.

Descenso por gradiente

Queremos explorar el mínimo, pero no hicimos una exploración exhaustiva de la función de costo:

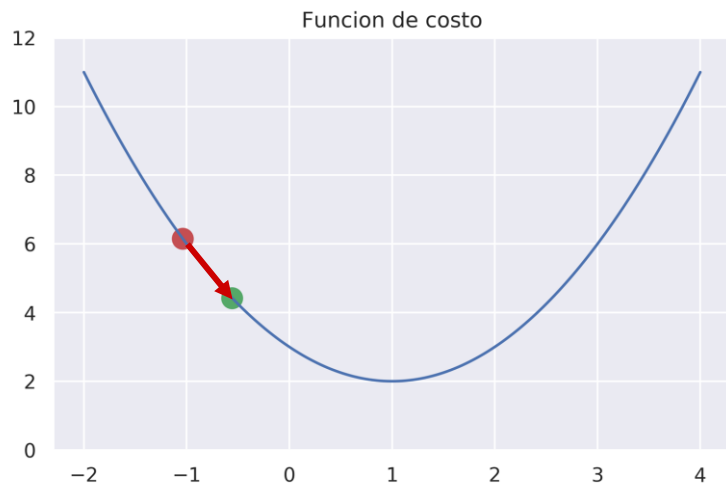


Pasos

1. Calculamos el costo para ciertos valores al azar de los parámetros.
2. Repetimos hasta converger
 - a. Nos fijamos la dirección de decrecimiento en ese punto. Técnicamente, derivamos o calculamos el gradiente.

Descenso por gradiente

Queremos explorar el mínimo, pero no hicimos una exploración exhaustiva de la función de costo:

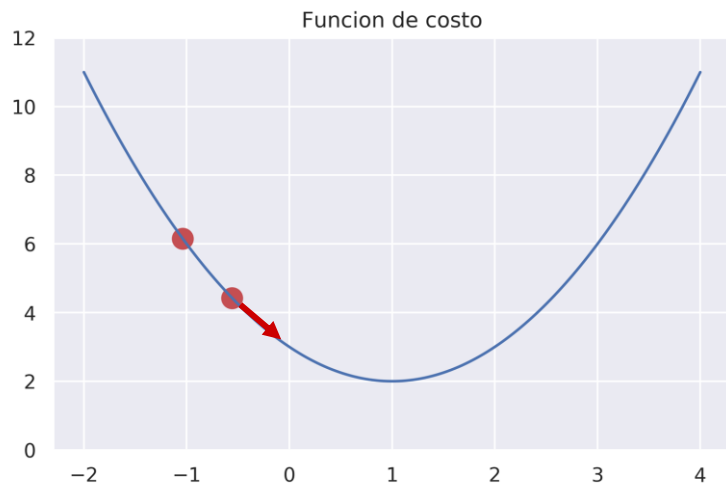


Pasos

1. Calculamos el costo para ciertos valores al azar de los parámetros.
2. Repetimos hasta converger
 - a. Nos fijamos la dirección de decrecimiento en ese punto. Técnicamente, derivamos o calculamos el gradiente.
 - b. Actualizamos los valores de los parámetros.

Descenso por gradiente

Queremos explorar el mínimo, pero no hicimos una exploración exhaustiva de la función de costo:

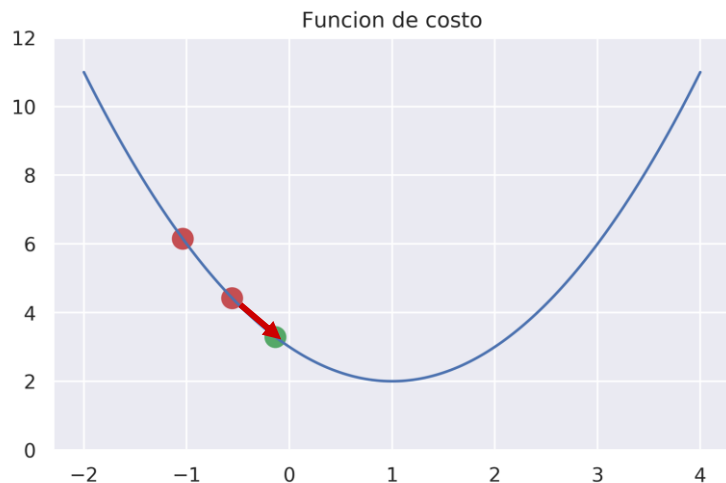


Pasos

1. Calculamos el costo para ciertos valores al azar de los parámetros.
2. Repetimos hasta converger
 - a. Nos fijamos la dirección de decrecimiento en ese punto. **Técnicamente**, derivamos o calculamos el gradiente.
 - b. Actualizamos los valores de los parámetros.

Descenso por gradiente

Queremos explorar el mínimo, pero no hicimos una exploración exhaustiva de la función de costo:



Pasos

1. Calculamos el costo para ciertos valores al azar de los parámetros.
2. Repetimos hasta converger
 - a. Nos fijamos la dirección de decrecimiento en ese punto. **Técnicamente**, derivamos o calculamos el gradiente.
 - b. Actualizamos los valores de los parámetros.

Descenso por gradiente

Queremos explorar el mínimo, pero no hicimos una exploración exhaustiva de la función de costo:

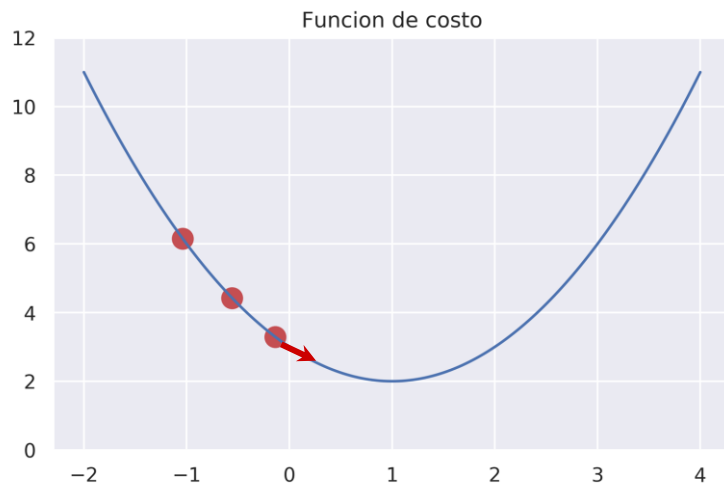


Pasos

1. Calculamos el costo para ciertos valores al azar de los parámetros.
2. Repetimos hasta converger
 - a. Nos fijamos la dirección de decrecimiento en ese punto. **Técnicamente**, derivamos o calculamos el gradiente.
 - b. Actualizamos los valores de los parámetros.

Descenso por gradiente

Queremos explorar el mínimo, pero no hicimos una exploración exhaustiva de la función de costo:

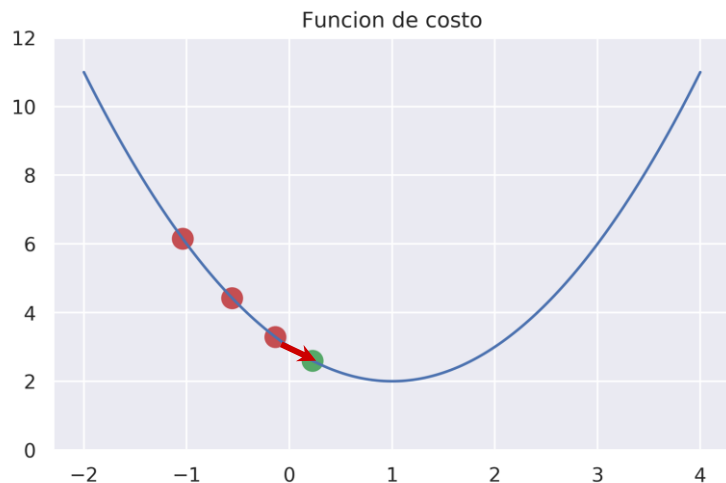


Pasos

1. Calculamos el costo para ciertos valores al azar de los parámetros.
2. Repetimos hasta converger
 - a. Nos fijamos la dirección de decrecimiento en ese punto. **Técnicamente**, derivamos o calculamos el gradiente.
 - b. Actualizamos los valores de los parámetros.

Descenso por gradiente

Queremos explorar el mínimo, pero no hicimos una exploración exhaustiva de la función de costo:

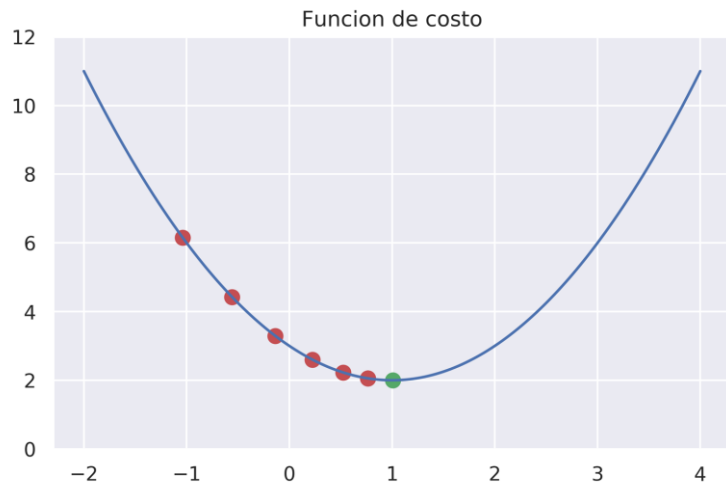


Pasos

1. Calculamos el costo para ciertos valores al azar de los parámetros.
2. Repetimos hasta converger
 - a. Nos fijamos la dirección de decrecimiento en ese punto. **Técnicamente**, derivamos o calculamos el gradiente.
 - b. Actualizamos los valores de los parámetros.

Descenso por gradiente

Queremos explorar el mínimo, pero no hicimos una exploración exhaustiva de la función de costo:



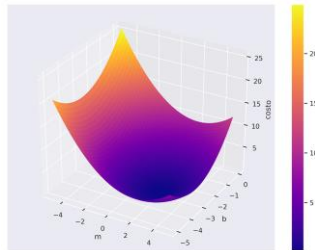
Pasos

1. Calculamos el costo para ciertos valores al azar de los parámetros.
2. Repetimos hasta converger
 - a. Nos fijamos la dirección de decrecimiento en ese punto. **Técnicamente**, derivamos o calculamos el gradiente.
 - b. Actualizamos los valores de los parámetros.

Actividad: Hands-on training



¡En 2 o más dimensiones es similar!



DS_Encuentro_29_DPG.ipynb



Este notebook es demostrativo,
pero recomendamos trabajarlo
con un/a compañero/a.

Descenso por gradiente: Resumen



Descenso por gradiente • Resumen

1. Necesitamos una función de costo/pérdida. La función de costo depende del problema (clasificación, regresión, etc).
2. La función de costo es una función de los parámetros de la red (bueno, también de los datos que tengo, pero ignoremos eso por ahora).
3. Los mejores parámetros de la red son aquellos que minimicen la función de costo.
4. Como explorar todo ese espacio de parámetros exhaustivamente (simil *grid search*) es imposible, necesitamos una técnica que lo haga eficientemente. Esa técnica es Descenso por Gradiente.

Descenso por gradiente • Resumen

1. Necesitamos una función de costo/pérdida. La función de costo depende del problema (clasificación, regresión, etc).
2. La función de costo es una función de los parámetros de la red (bueno, también de los datos que tengo, pero ignoremos eso por ahora).
3. Los mejores parámetros de la red son aquellos que minimicen la función de costo.
4. Como explorar todo ese espacio de parámetros exhaustivamente (simil *grid search*) es imposible, necesitamos una técnica que lo haga eficientemente. Esa técnica es Descenso por Gradiente.

Mucha de la jerga en redes neuronales refieren a técnicas para optimizar esta búsqueda

Para la próxima

1. Ver los videos de la plataforma de “Redes Neuronales”.
2. Trabajar en notebooks atrasados.

ACÀMICA