Escuela Profesional de
Ciencia de la Computación

ICC Fase 1

Inteligencia Artificial

Dr. Edward Hinojosa Cárdenas

Universidad Nacional de San Agustín de Arequipa

2021A/Semestre Impar

Índice

ehinojosa@unsa.edu.pe

Aprendizaje de Máquina

- Machine Learning
- Un agente está aprendiendo si mejora su desempeño después de hacer observaciones sobre el mundo.
- El aprendizaje puede ir desde lo trivial, como apuntar una lista de compras, hasta lo profundo, como cuando Albert Einstein infirió una nueva teoría del universo.
- Cuando el agente es una computadora, lo llamamos aprendizaje automático, es decir, una computadora observa algunos datos, construye un modelo basado en los datos y usa el modelo como una hipótesis sobre el mundo y una pieza de software que puede resolver problemas.

ehinojosa@unsa.edu.pe

Aprendizaje de Máquina

- ¿Por qué queremos que una máquina aprenda?
- ¿Por qué no programarlo de la manera correcta para empezar?
- Existen dos razones principales:

ehinojosa@unsa.edu.pe

Aprendizaje de Máquina

- Primero, los diseñadores no pueden anticipar todas las posibles situaciones futuras. Por ejemplo, un robot diseñado para navegar por laberintos debe aprender el diseño de cada nuevo laberinto que encuentra; Un programa para predecir los precios del mercado de valores debe aprender a adaptarse cuando las condiciones cambian de auge a caída.
- Segundo, en ciertas situaciones los diseñadores no tienen idea de cómo programar ellos mismos una solución. La mayoría de las personas son buenas para reconocer los rostros de los miembros de la familia, pero lo hacen de manera subconsciente, por lo que incluso los mejores programadores no saben cómo programar una computadora para realizar esa tarea, excepto mediante el uso de algoritmos de aprendizaje automático.

ehinojosa@unsa.edu.pe

Tipos de Aprendizaje

- Cualquier componente de un programa de agente se puede mejorar mediante el aprendizaje automático.
- Las mejoras y las técnicas utilizadas para realizarlas dependen de estos factores:
  - Qué componente se va a mejorar.
  - Qué conocimiento previo tiene el agente, que influye en el modelo que construye.
  - Qué datos y comentarios sobre esos datos están disponibles.

ehinojosa@unsa.edu.pe

Tipos de Aprendizaje

- Por ejemplo, un agente de vehículos autónomos que aprende observando a un conductor humano.
- Cada vez que el conductor frena, el agente puede aprender una regla de condición-acción sobre cuándo frenar.
- Al observar varias imágenes que contienen buses, puede aprender a reconocerlos.
- Intentar acciones y observar los resultados, por ejemplo, frenar con fuerza en una carretera mojada, puede aprender los efectos de sus acciones.
- Cuando recibe quejas de pasajeros que han sido profundamente conmocionados durante el viaje, puede aprender un componente útil de su función de utilidad general.

ehinojosa@unsa.edu.pe

Tipos de Aprendizaje

- El uso de aprendizaje automático se ha convertido en una parte estándar de la Ingeniería de Software.
- Cada vez que crea un sistema de software, incluso si no es considerado coomo un agente inteligente, los componentes del sistema pueden mejorarse potencialmente con el uso de aprendizaje automático.
- Por ejemplo, el software para analizar imágenes de galaxias bajo lentes gravitacionales se aceleró en un factor de 10 millones con un modelo de aprendizaje automático (Hezaveh, 2017).

Tipos de Aprendizaje

- Otro ejemplo, el uso de energía para enfriar centros de datos se redujo en un 40 % con otro modelo de aprendizaje automático (Gao, 2014).
- Por último, el ganador del premio Turing, David Patterson, y el jefe de inteligencia artificial de Google, Jeff Dean, declararon el comienzo de una "Edad de Oro" para la arquitectura informática debido al aprendizaje automático (2018).

ehinojosa@unsa.edu.pe

Tipos de Aprendizaje

- Nosotros vamos a asumir que el agente comienza de cero y aprende de los datos.
- Vamos a utilizar la inducción, lo cual consiste de pasar de un conjunto específico de observaciones a un regla general. Por ejemplo, a partir de las observaciones de que el sol salió todos los días en el pasado, inducimos que el sol saldrá mañana.

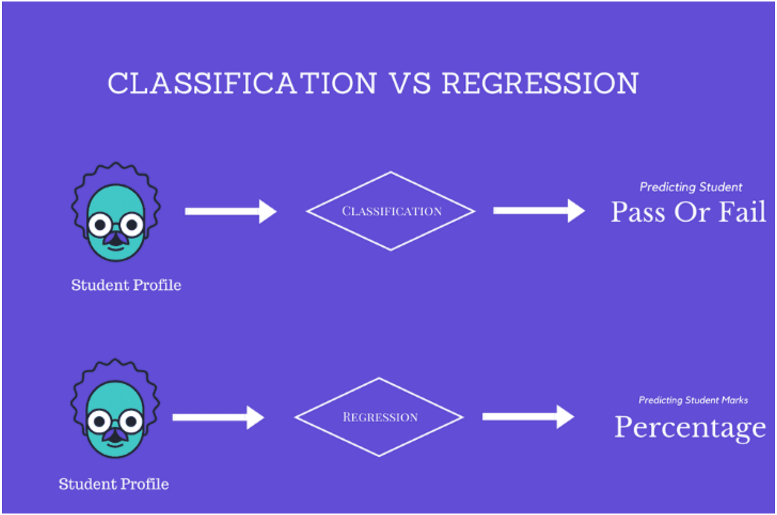
ehinojosa@unsa.edu.pe

Tipos de Aprendizaje

- Nos vamos a centrar en problemas que la entrada sea una representación factorizada (es decir, un vector de valores de atributos, vector de atributos o vector de características), valores atómicos y relacionales.
- Cuando el resultado es uno de un conjunto finito de valores (como soleado / nublado / lluvioso o verdadero / falso), el problema de aprendizaje se denomina clasificación.
- Cuando el resultado es un número (como la temperatura de mañana, medida como un número entero o un número real), el problema de aprendizaje se denomina regresión (aproximación de funciones, predicción numerica).

ehinojosa@unsa.edu.pe

Tipos de Aprendizaje



Tipos de Aprendizaje

- Las formas o tipos de aprendizaje dependen de los datos del problema:
  - Aprendizaje Supervisado - Supervised learning
  - Aprendizaje No Supervisado - Unsupervised learning
  - Aprendizaje por Refuerzo - Reinforcement learning

ehinojosa@unsa.edu.pe

Aprendizaje Supervisado

- El agente observa los datos (entrada-salida) y aprende una función que mapea la entrada a la salida. Por ejemplo, las entradas podrían ser imágenes de cámara, cada una acompañada de una salida que diga "automóvil" o "peatón'. Una salida como esta se llama etiqueta. El agente aprende una función que, cuando se le da una nueva imagen, predice la etiqueta apropiada.
- En el caso de las acciones de frenado, una entrada es el estado actual (velocidad y dirección del automóvil, estado de la carretera) y una salida es la distancia que tomó para detenerse.

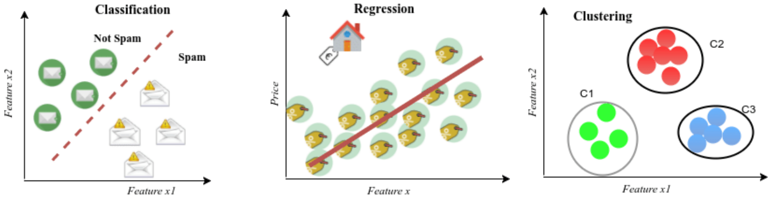
ehinojosa@unsa.edu.pe

Aprendizaje No Supervisado

- En el aprendizaje no supervisado, el agente aprende patrones en los datos que contienen una entrada pero sin ninguna salida explícita.
- La tarea de aprendizaje no supervisada más común es la agrupación: detectar grupos de ejemplos de entrada potencialmente útiles.
- Por ejemplo, cuando se muestran millones de imágenes tomadas de Internet, un sistema de visión computacional puede identificar un gran grupo de imágenes similares que contengan un gato.

ehinojosa@unsa.edu.pe

Aprendizaje No Supervisado

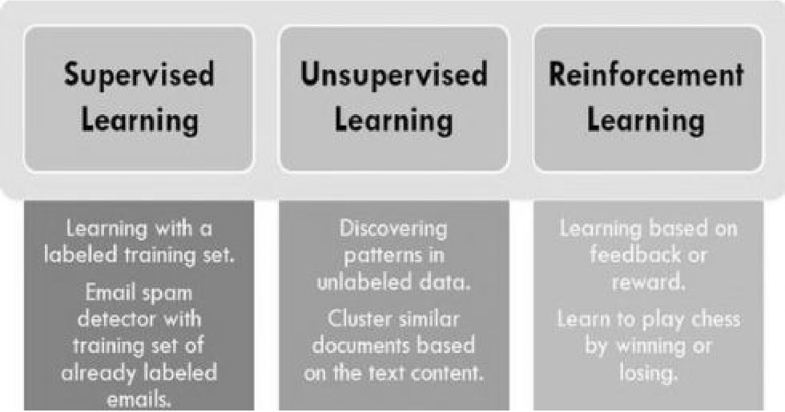


Aprendizaje por Refuerzo

- El agente aprende de una serie de refuerzos: recompensas y castigos.
- Por ejemplo, al final de una partida de ajedrez, se le dice al agente que ha ganado (una recompensa) o perdido (un castigo).
- Depende del agente decidir cuál de las acciones previas al refuerzo fueron las que tuvieron mayor influencia y modificar sus acciones con el objetivo a tener mayor recompensas en el futuro.

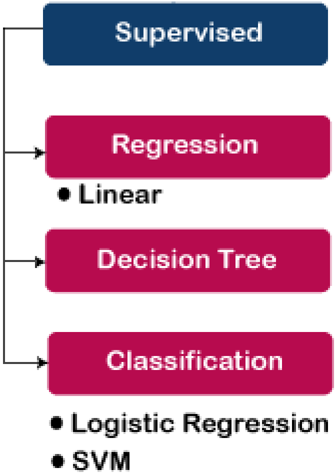
ehinojosa@unsa.edu.pe

Tipos de Aprendizaje



ehinojosa@unsa.edu.pe

Aprendizaje Supervisado



Regresión Lineal

Regresión Lineal

- La regresión lineal es uno de los algoritmos de aprendizaje automático más populares.
- Es un método estadístico que se utiliza para el análisis predictivo.
- La regresión lineal hace predicciones para variables continuas/reales o numéricas como ventas, salario, edad, precio del producto, etc.

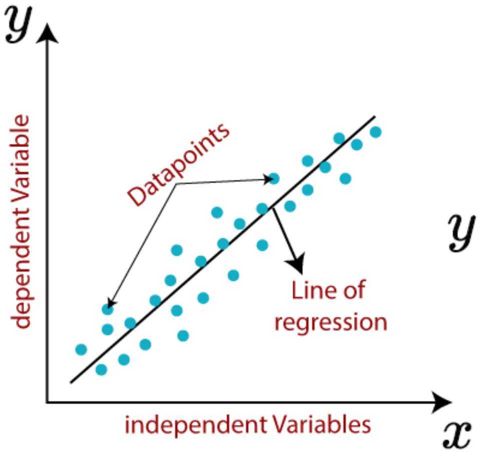
ehinojosa@unsa.edu.pe

Regresión Lineal

- El algoritmo de regresión lineal muestra una relación lineal entre una variable dependiente (y) y una o más variables independientes (x).
- Dado que la regresión lineal muestra la relación lineal, lo que significa que encuentra cómo cambia el valor de la variable dependiente de acuerdo con el valor de la variable independiente.
- El modelo de regresión lineal proporciona una línea recta inclinada que representa la relación entre las variables.

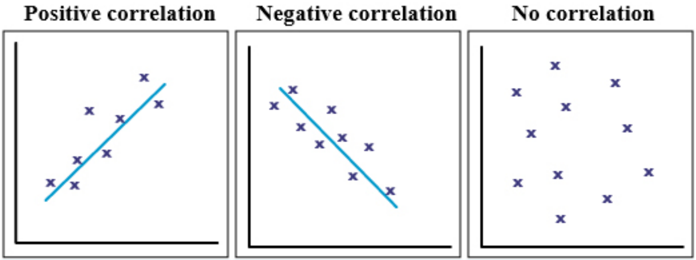
ehinojosa@unsa.edu.pe

Regresión Lineal



$$y = w_1 x + w_0$$

Regresión Lineal



The points lie close to a straight line, which has a positive gradient.

This shows that as one variable **increases** the other **increases**.

The points lie close to a straight line, which has a negative gradient.

This shows that as one variable **increases**, the other **decreases**.

There is no pattern to the points.

This shows that there is **no connection** between the two variables.

Método de Mínimos Cuadrados

- Es un método no iterativo que ajusta el modelo de modo que la suma de cuadros de las diferencias de los valores observados y previsto sea minimizada.

$$w_1 = \frac{N(\sum x_j y_j) - (\sum x_j)(\sum y_j)}{N(\sum x_j^2) - (\sum x_j)^2}$$

$$w_0 = \frac{\sum y_j - w_1(\sum x_j)}{N}$$

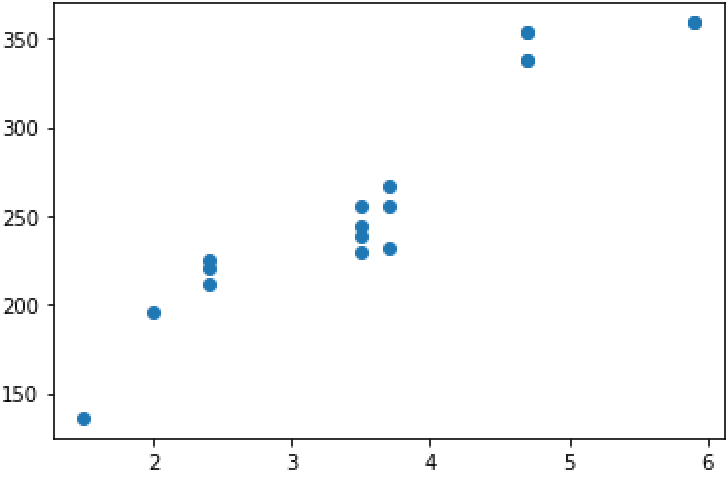
ehinojosa@unsa.edu.pe

Método de Mínimos Cuadrados

x (motor)	y (Emisión de CO_2)
2.0	196
2.4	221
1.5	136
3.5	255
3.5	244
3.5	230
3.7	232
3.7	255
3.7	267
2.4	212
2.4	225
3.5	239
5.9	359
5.9	359
4.7	338
4.7	354
4.7	338
4.7	354

ehinojosa@unsa.edu.pe

Método de Mínimos Cuadrados



ehinojosa@unsa.edu.pe

Método de Mínimos Cuadrados

x	y	$x * y$	x^2
2	196	392	4
2.4	221	530.4	5.76
1.5	136	204	2.25
3.5	255	892.5	12.25
3.5	244	854	12.25
3.5	230	805	12.25
3.7	232	858.4	13.69
3.7	255	943.5	13.69
3.7	267	987.9	13.69
2.4	212	508.8	5.76
2.4	225	540	5.76
3.5	239	836.5	12.25
5.9	359	2118.1	34.81
5.9	359	2118.1	34.81
4.7	338	1588.6	22.09
4.7	354	1663.8	22.09
4.7	338	1588.6	22.09
4.7	354	1663.8	22.09
66.4	**4814**	**19094**	**271.58**

ehinojosa@unsa.edu.pe

Método de Mínimos Cuadrados

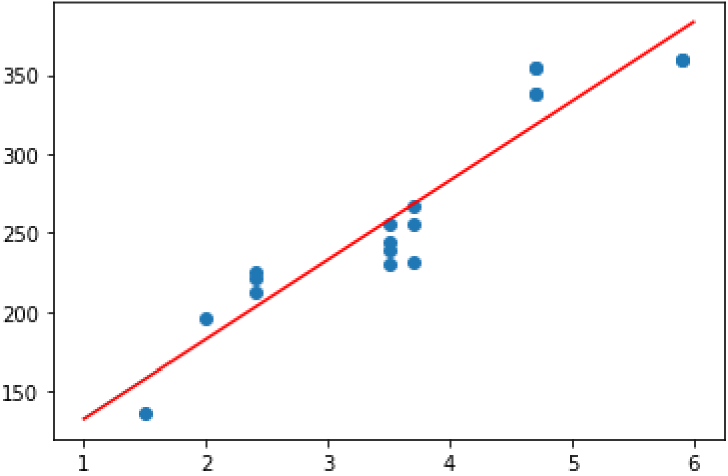
$$w_1 = \frac{18(19094) - (66{,}4)(4814)}{18(271{,}58) - (66{,}4)^2} \approx 50{,}14$$

$$w_0 = \frac{4814 - 50{,}14(66{,}4)}{18} \approx 82{,}47$$

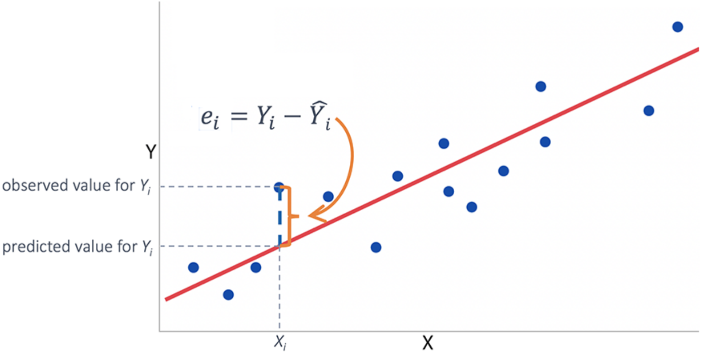
$$y = 50{,}14x + 82{,}47$$

ehinojosa@unsa.edu.pe

Método de Mínimos Cuadrados



Gradiente Descendiente



Gradiente Descendiente

- El descenso de gradiente se utiliza para minimizar el MSE (Minimum Square Error - Error Cuadrático Medio) calculando el gradiente de la función de costo.
- Utiliza el descenso de la gradiente para actualizar los coeficientes de la línea al reducir la función de costo.
- Se realiza mediante una selección aleatoria de valores de coeficiente y luego se actualizan iterativamente los valores para alcanzar la función de costo mínimo.

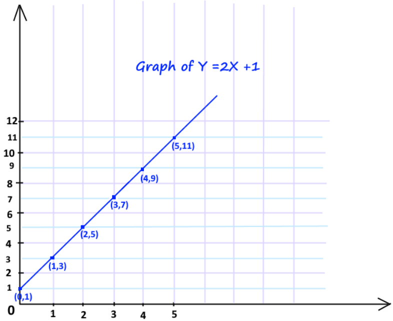
ehinojosa@unsa.edu.pe

Gradiente Descendiente - Derivada

- La derivada de una función describe la razón de cambio instantáneo de la función en un cierto punto.

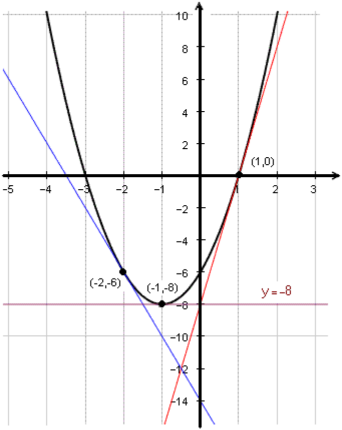
$$f(y) = 2x + 1$$

$$f'(y) = 2$$



Gradiente Descendiente - Derivada

- La derivada es el resultado de un límite y representa la pendiente de la recta tangente a la gráfica de la función en un punto.



Gradiente Descendiente - Derivada

Differentiation Rules	
Constant Rule	$\frac{d}{dx}[c] = 0$
Power Rule	$\frac{d}{dx}x^n = nx^{n-1}$
Chain Rule	$\frac{d}{dx}\left[f(g(x))\right] = f'(g(x))g'(x)$

ehinojosa@unsa.edu.pe

Gradiente Descendiente - Derivada

The Chain Rule

$$\frac{d}{dx}\left[f(g(x))\right] = \overbrace{f'(g(x))} \cdot \overbrace{g'(x)}$$

$$y' = f'(\underbrace{u}) \cdot u'$$

Example: $y = (x^2+1)^3 \Rightarrow y = (\overset{g(x)}{x^2+1})^3$

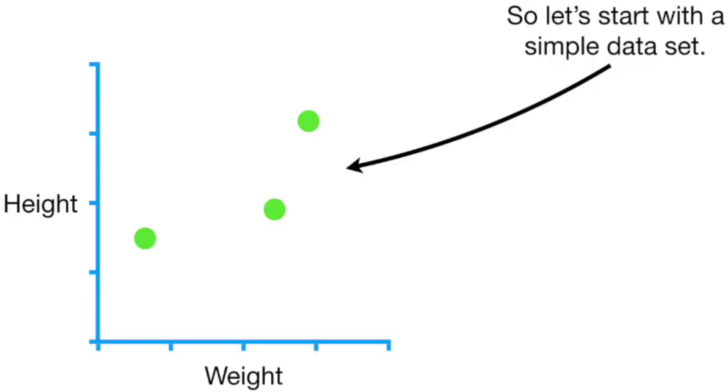
$$y = (t)^3$$

$$y = 3(t)^2 \cdot (2x)$$

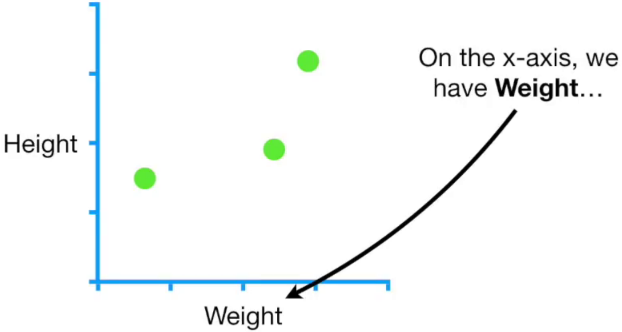
$$y = 3(x^2+1)^2 \cdot (2x$$

ehinojosa@unsa.edu.pe

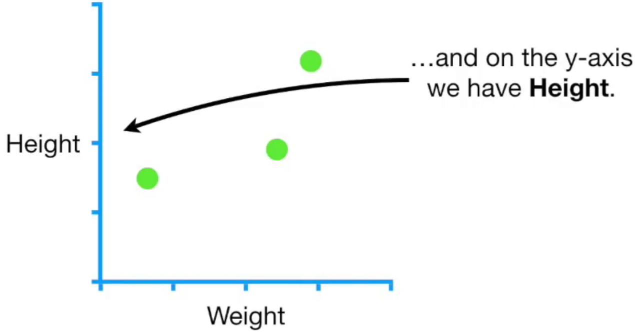
Gradiente Descendiente



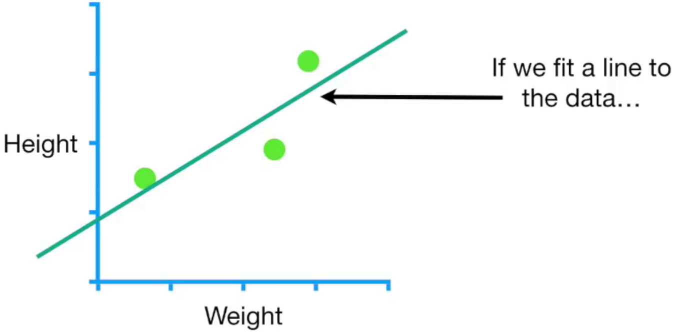
Gradiente Descendiente



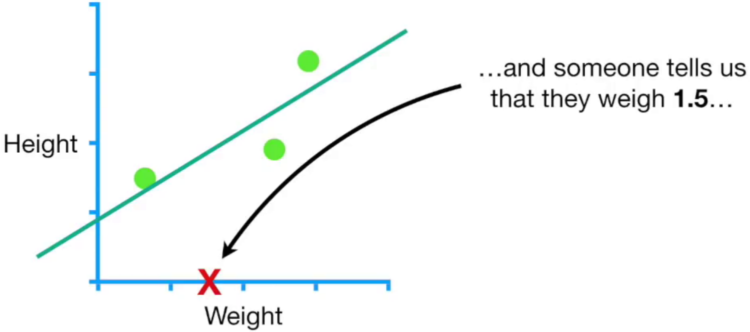
Gradiente Descendiente



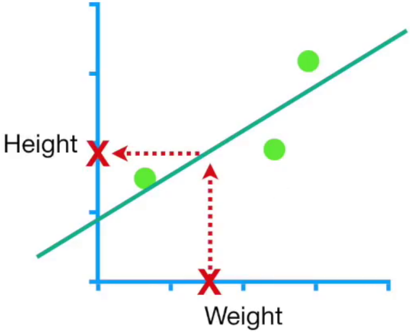
Gradiente Descendiente



Gradiente Descendiente

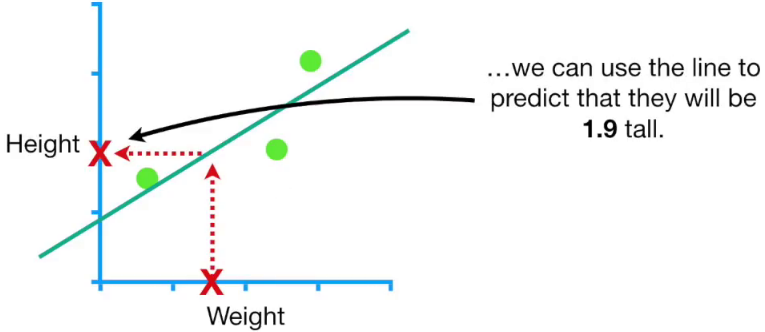


Gradiente Descendiente

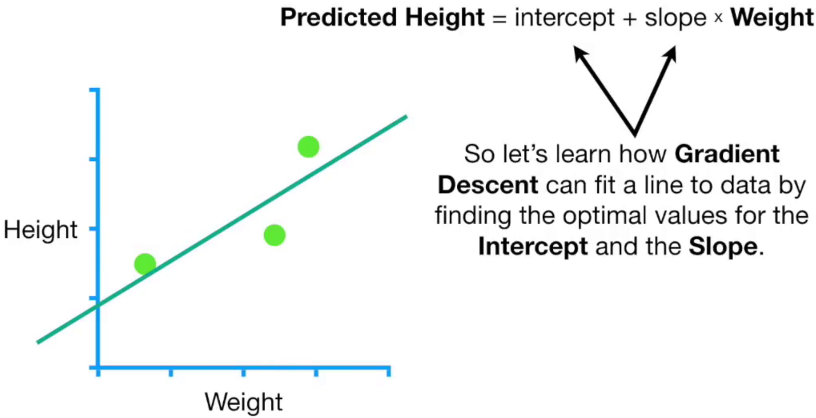


...we can use the line to predict that they will be **1.9** tall.

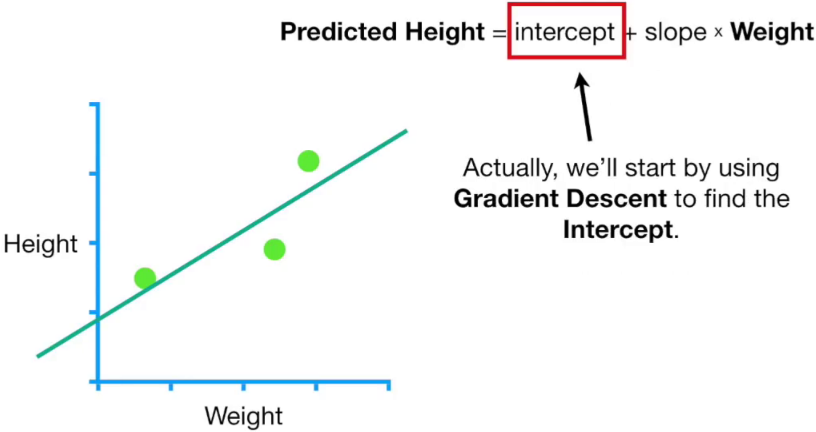
Gradiente Descendiente



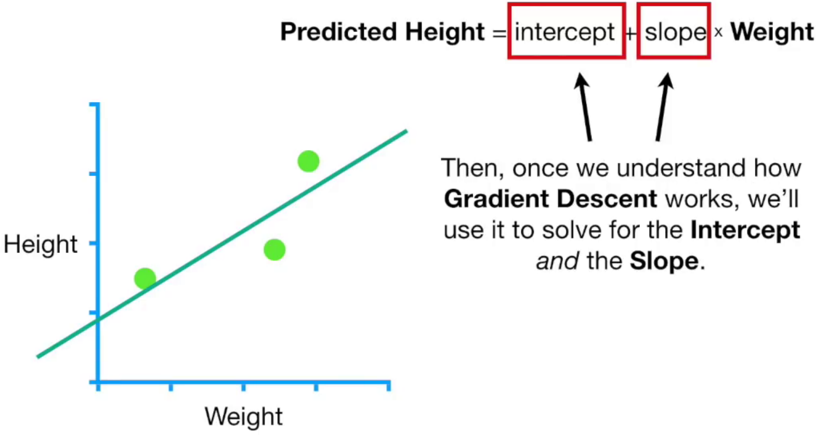
Gradiente Descendiente



Gradiente Descendiente



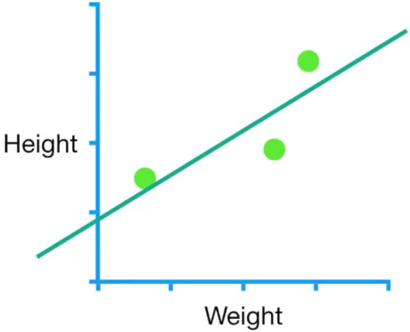
Gradiente Descendiente



Gradiente Descendiente

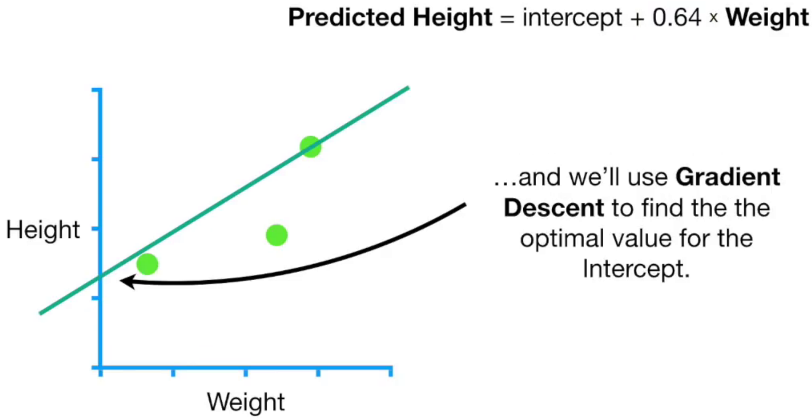
Predicted Height = intercept + slope × **Weight**

So for now, let's just plug in the **Least Squares** estimate for the **Slope**, **0.64**.

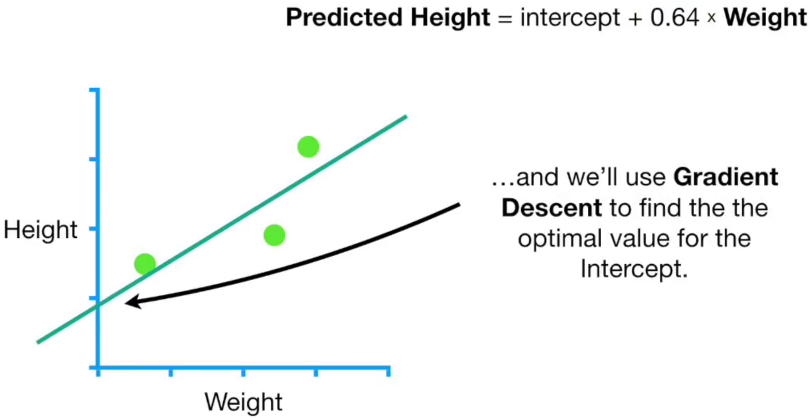


ehinojosa@unsa.edu.pe

Gradiente Descendiente

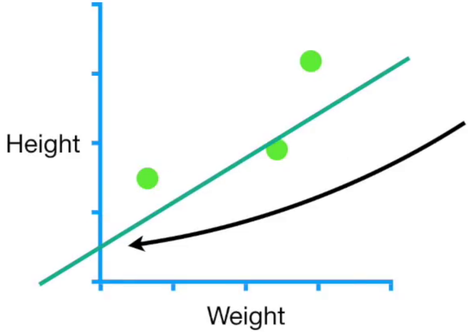


Gradiente Descendiente



Gradiente Descendiente

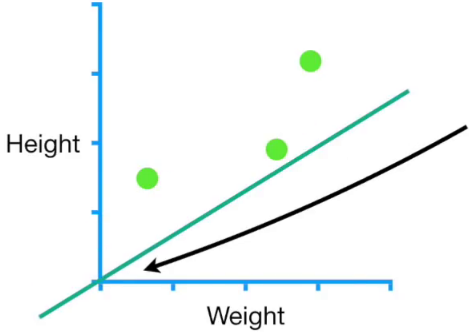
Predicted Height = intercept + 0.64 × **Weight**



...and we'll use **Gradient Descent** to find the the optimal value for the Intercept.

Gradiente Descendiente

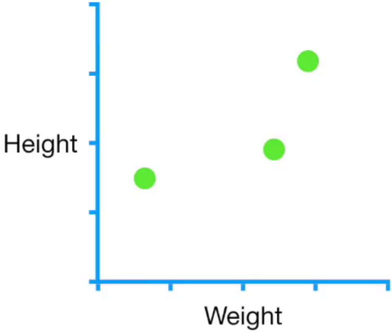
Predicted Height = intercept + 0.64 × **Weight**



...and we'll use **Gradient Descent** to find the the optimal value for the Intercept.

Gradiente Descendiente

Predicted Height = intercept + 0.64 × **Weight**

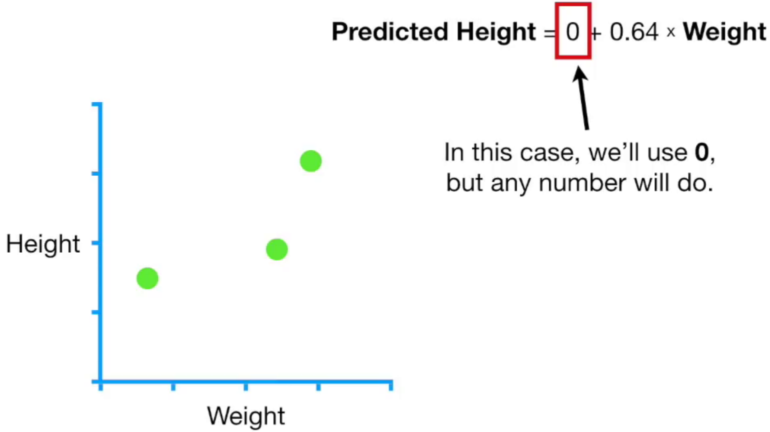


The first thing we do is pick a random value for the **Intercept**.

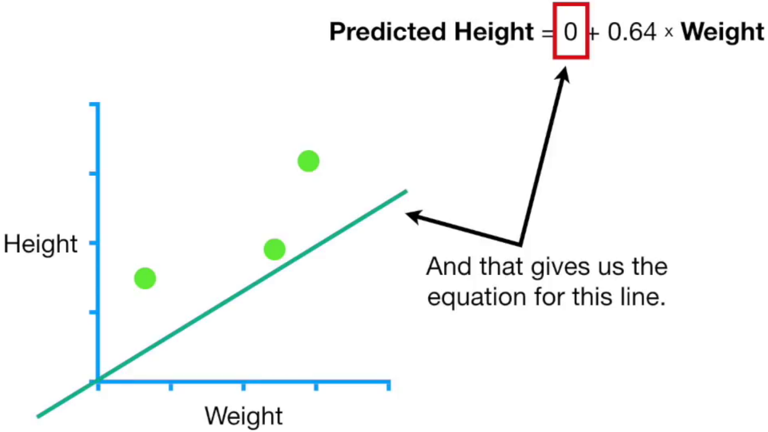
This is just an initial guess that gives **Gradient Descent** something to improve upon.

ehinojosa@unsa.edu.pe

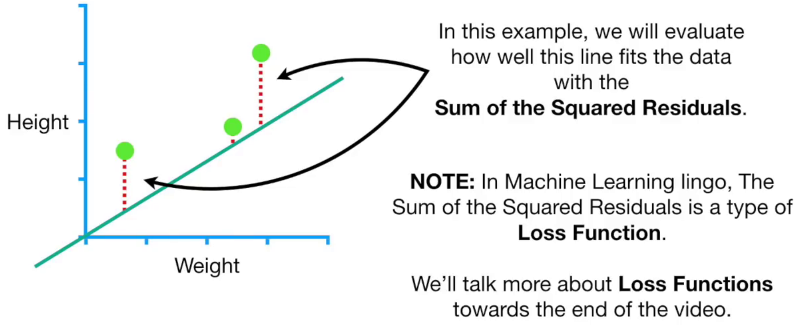
Gradiente Descendiente



Gradiente Descendiente

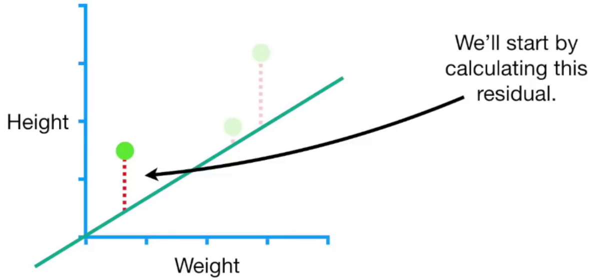


Gradiente Descendiente

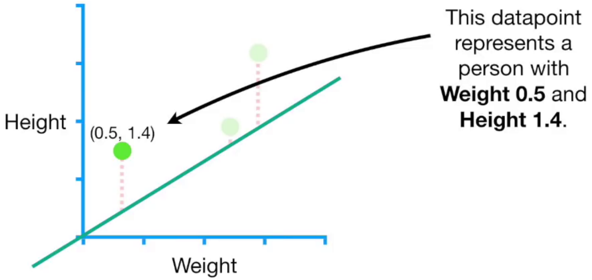


ehinojosa@unsa.edu.pe

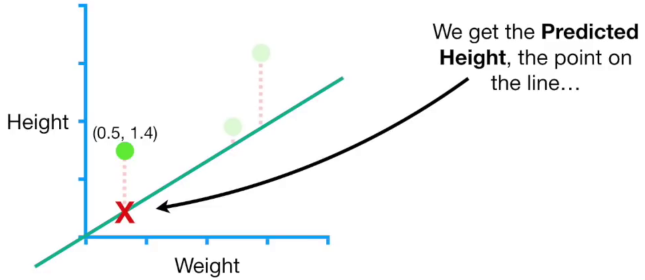
Gradiente Descendiente



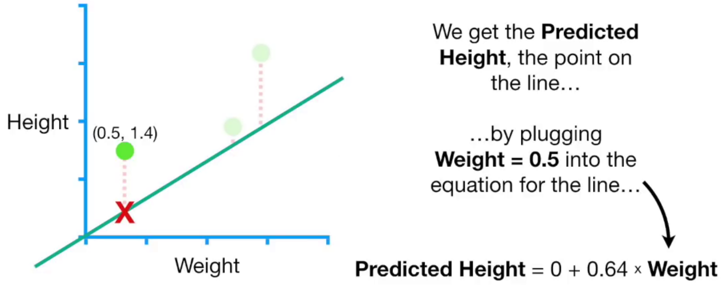
Gradiente Descendiente



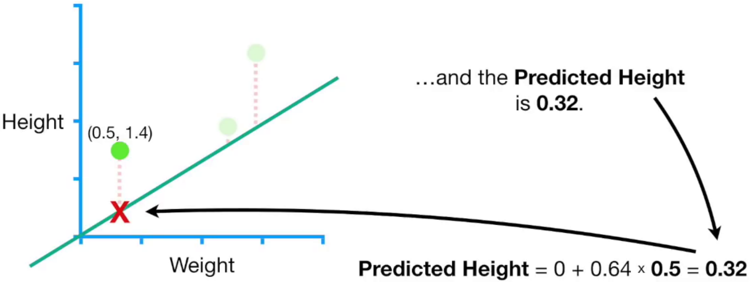
Gradiente Descendiente



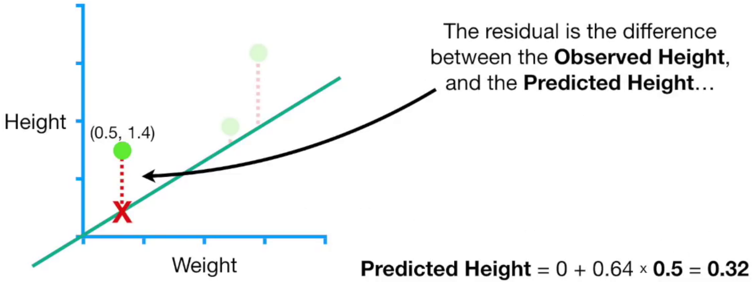
Gradiente Descendiente



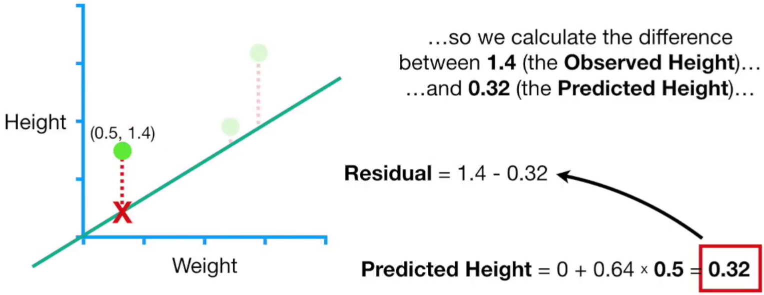
Gradiente Descendiente



Gradiente Descendiente

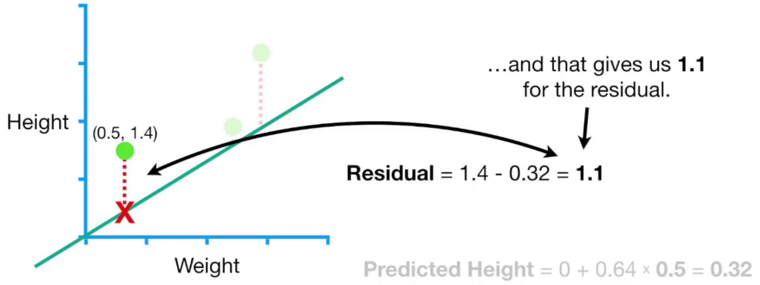


Gradiente Descendiente



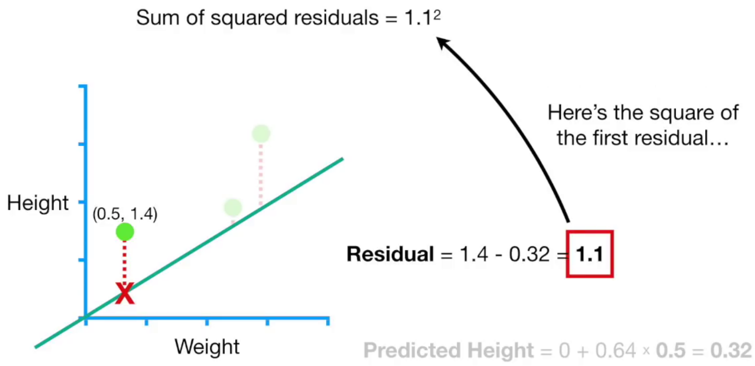
ehinojosa@unsa.edu.pe

Gradiente Descendiente

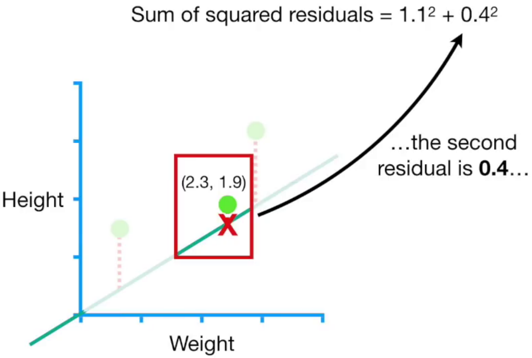


ehinojosa@unsa.edu.pe

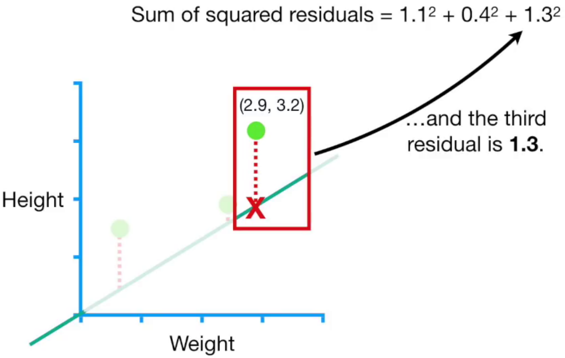
Gradiente Descendiente



Gradiente Descendiente



Gradiente Descendiente

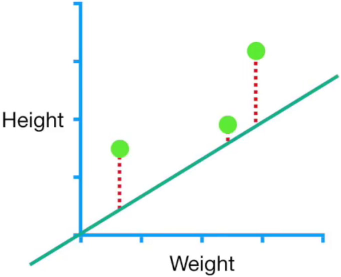


ehinojosa@unsa.edu.pe

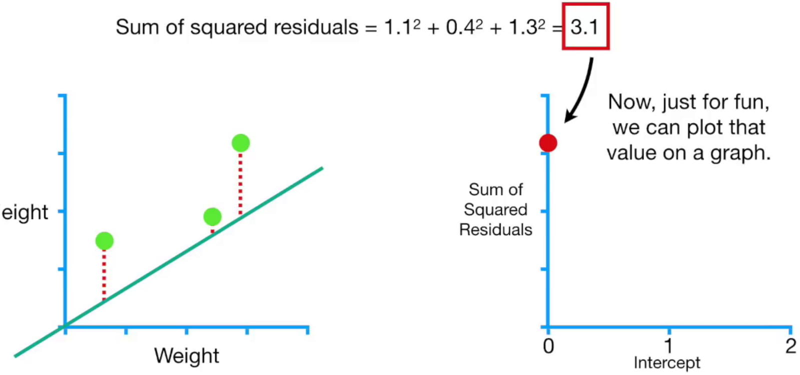
Gradiente Descendiente

Sum of squared residuals = $1.1^2 + 0.4^2 + 1.3^2 = \boxed{3.1}$

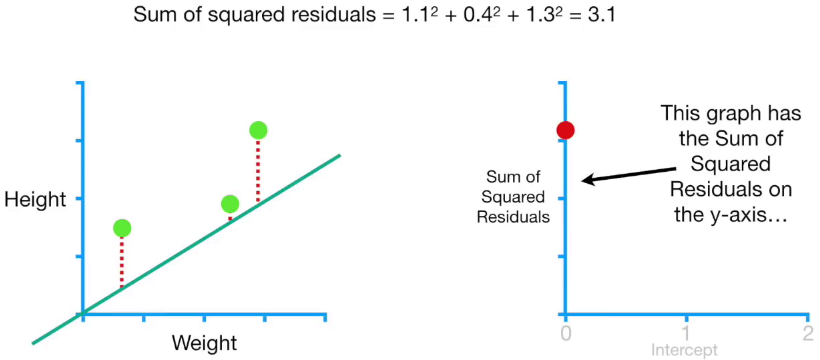
In the end, **3.1** is the Sum of the Squared Residuals.



Gradiente Descendiente

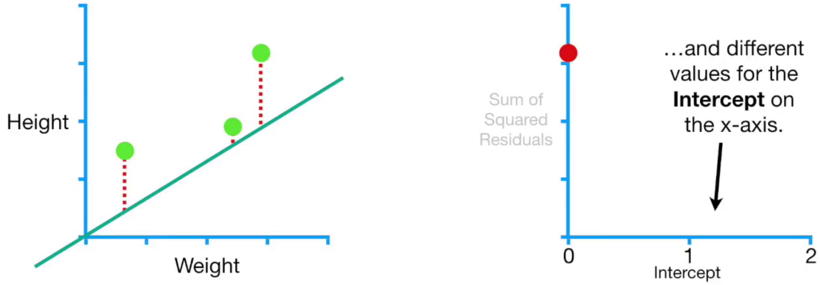


Gradiente Descendiente

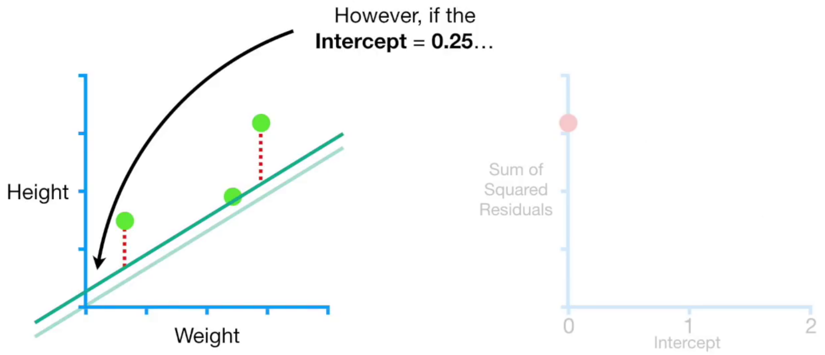


Gradiente Descendiente

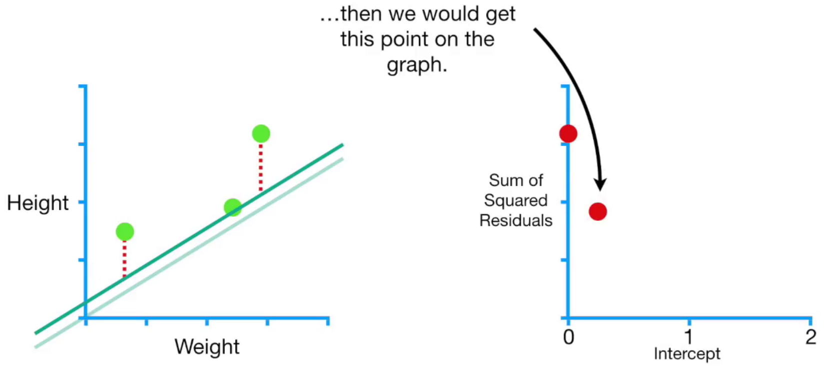
Sum of squared residuals = $1.1^2 + 0.4^2 + 1.3^2 = 3.1$



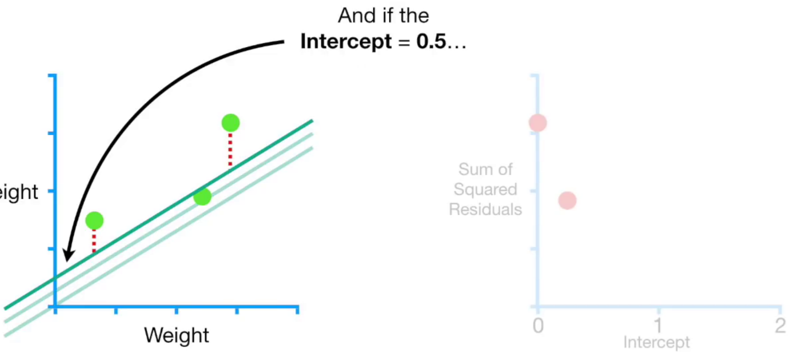
Gradiente Descendiente



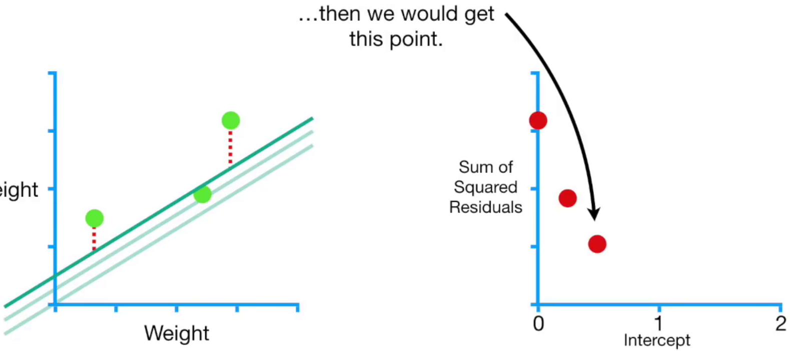
Gradiente Descendiente



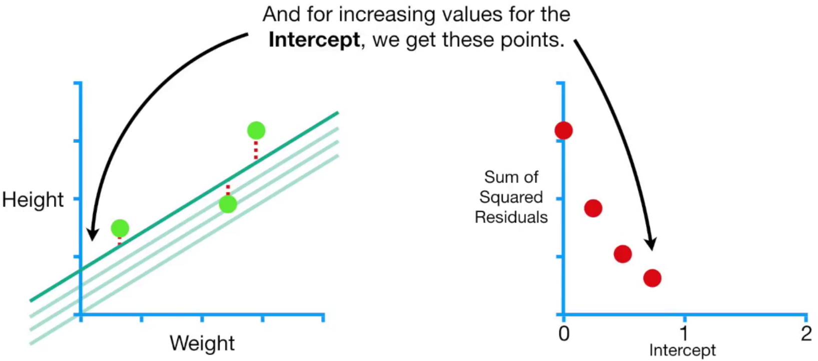
Gradiente Descendiente



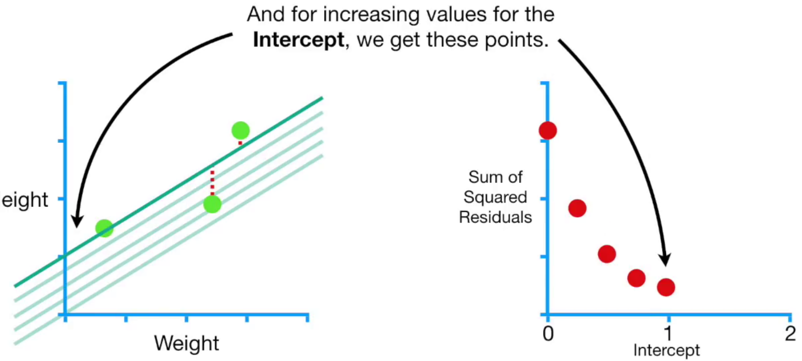
Gradiente Descendiente



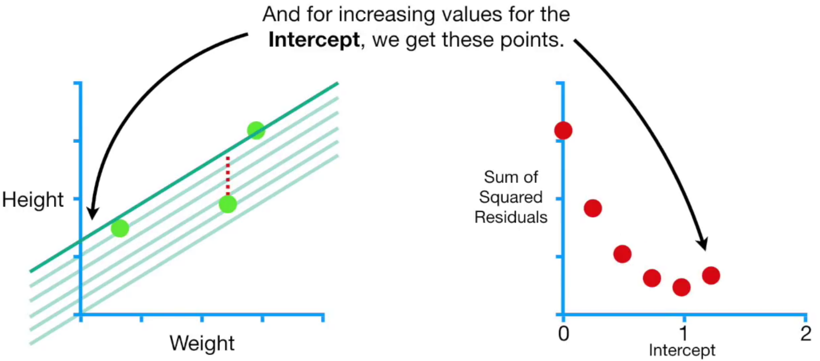
Gradiente Descendiente



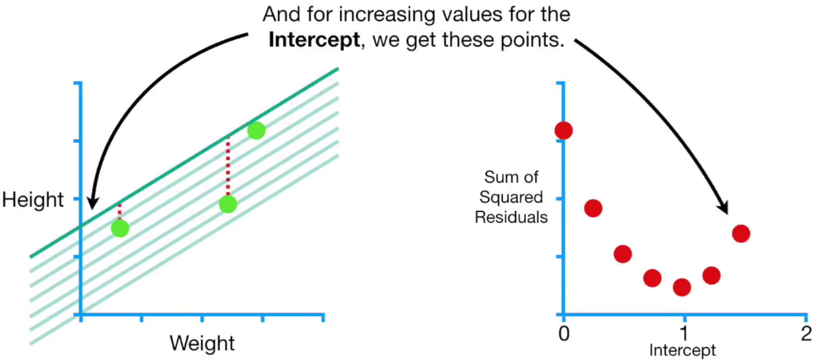
Gradiente Descendiente



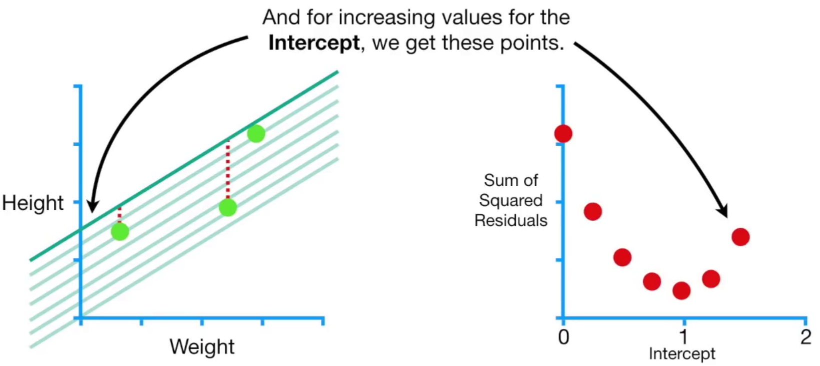
Gradiente Descendiente



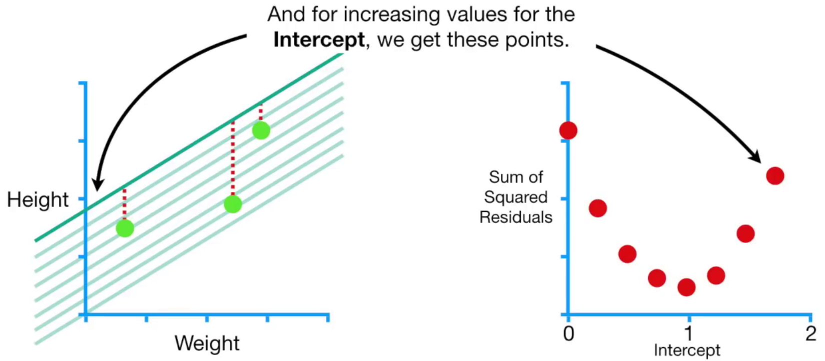
Gradiente Descendiente



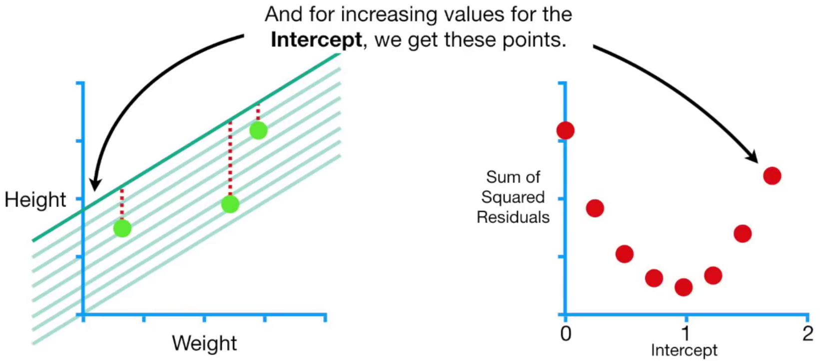
Gradiente Descendiente



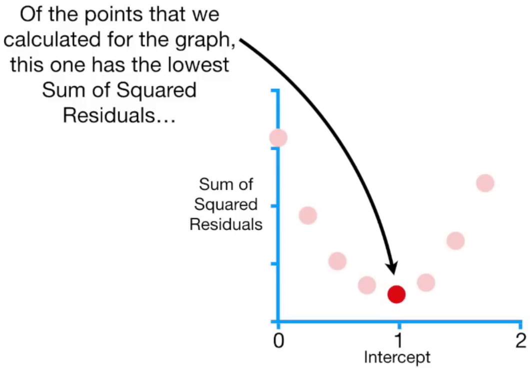
Gradiente Descendiente



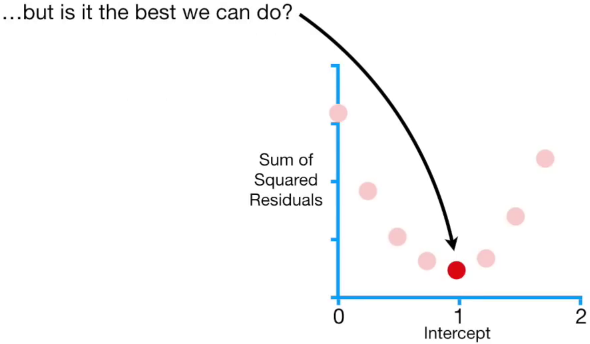
Gradiente Descendiente



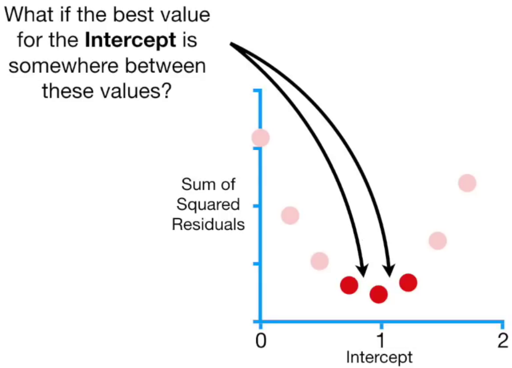
Gradiente Descendiente



Gradiente Descendiente



Gradiente Descendiente

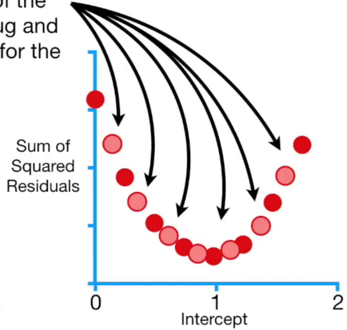


Gradiente Descendiente

A slow and painful method for finding the minimal Sum of the Squared Residuals is to plug and chug a bunch more values for the **Intercept**.

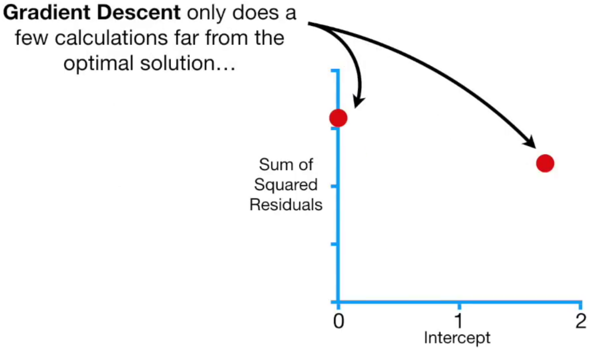
Ugh.

Don't despair! **Gradient Descent** is way more efficient!

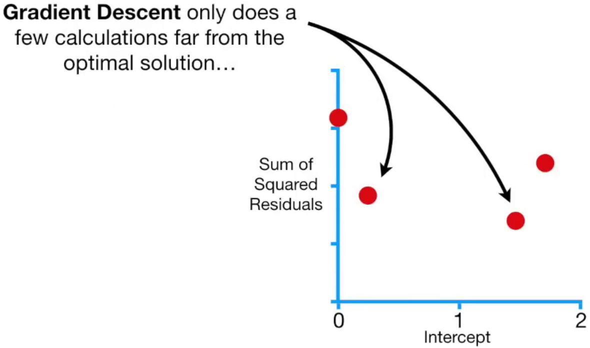


ehinojosa@unsa.edu.pe

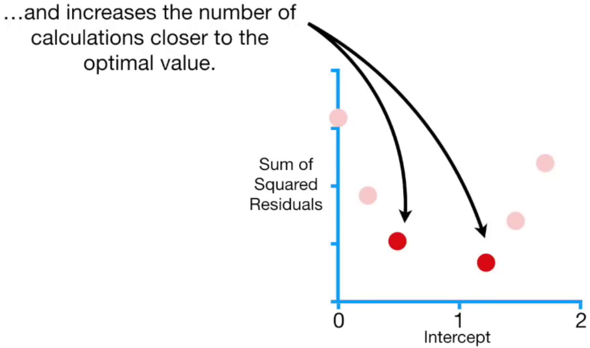
Gradiente Descendiente



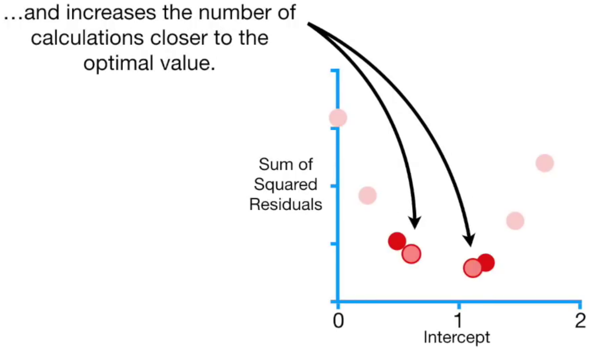
Gradiente Descendiente



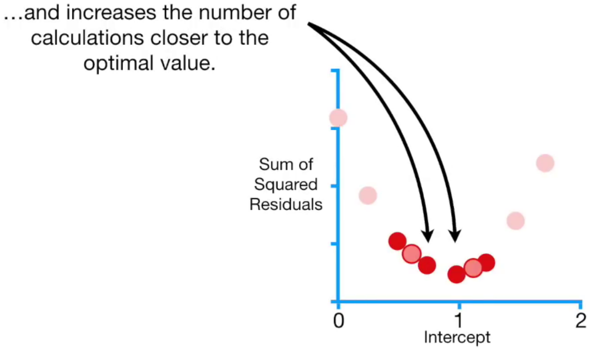
Gradiente Descendiente



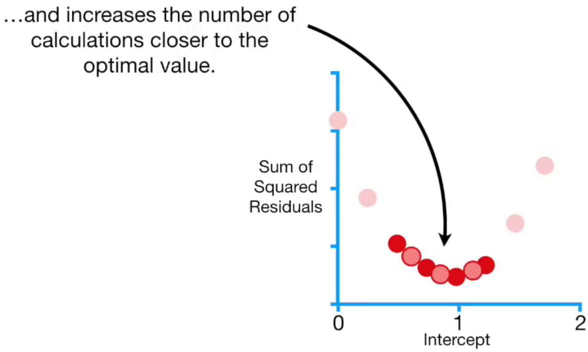
Gradiente Descendiente



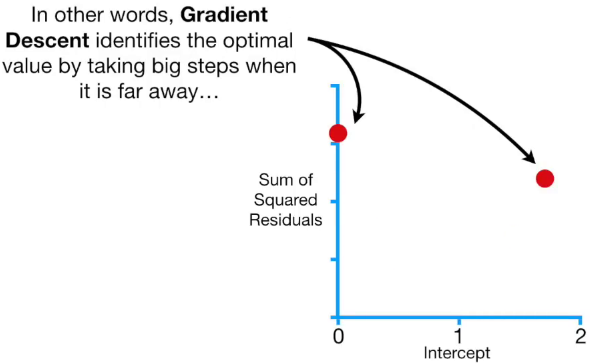
Gradiente Descendiente



Gradiente Descendiente

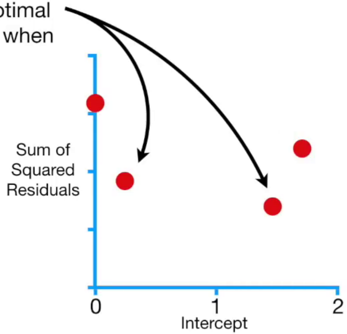


Gradiente Descendiente

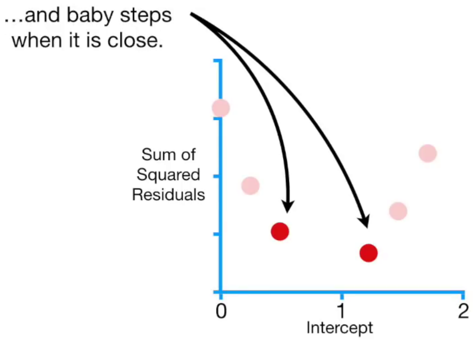


Gradiente Descendiente

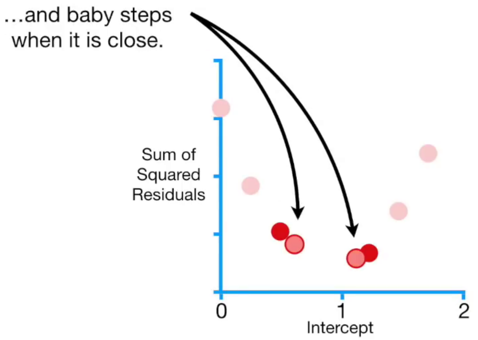
In other words, **Gradient Descent** identifies the optimal value by taking big steps when it is far away...



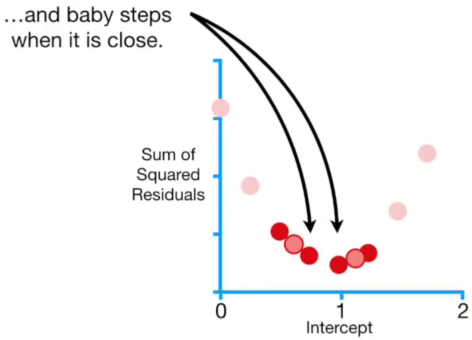
Gradiente Descendiente



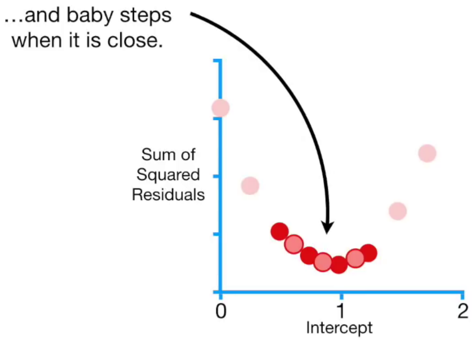
Gradiente Descendiente



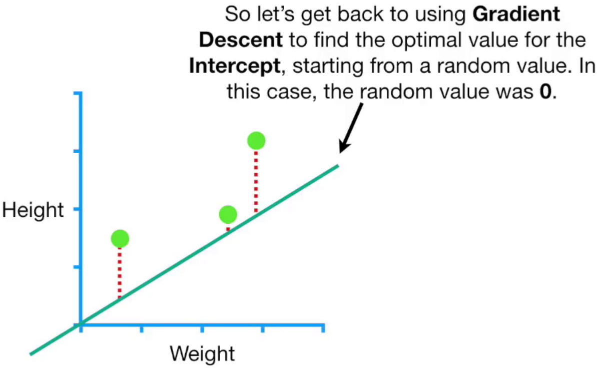
Gradiente Descendiente



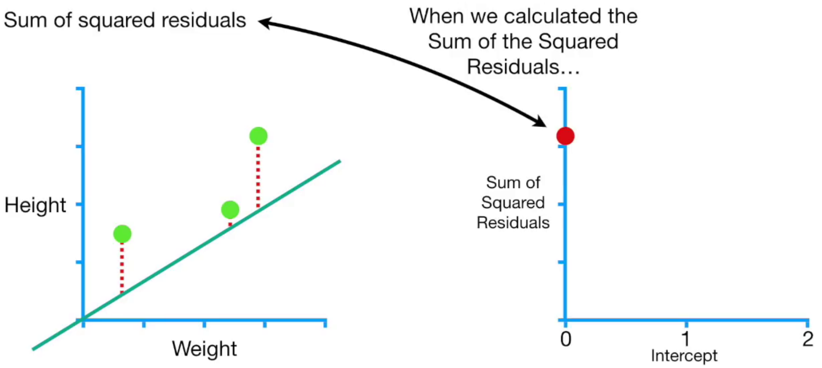
Gradiente Descendiente



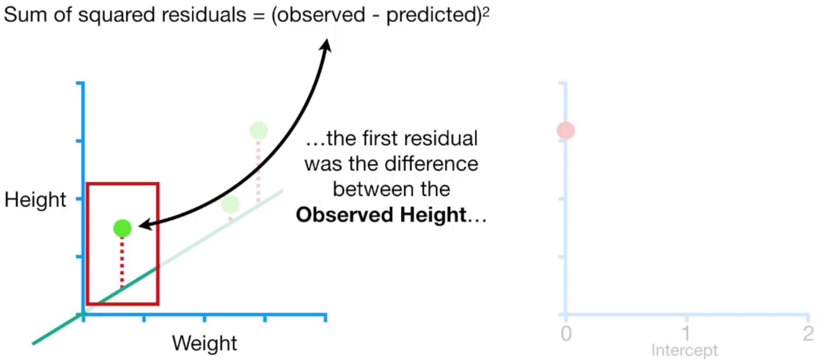
Gradiente Descendiente



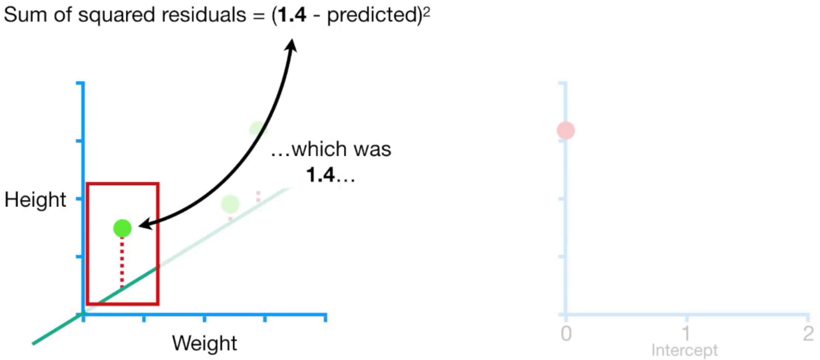
Gradiente Descendiente



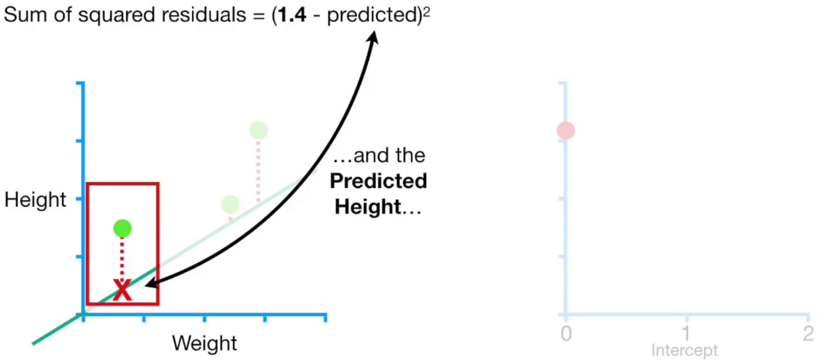
Gradiente Descendiente



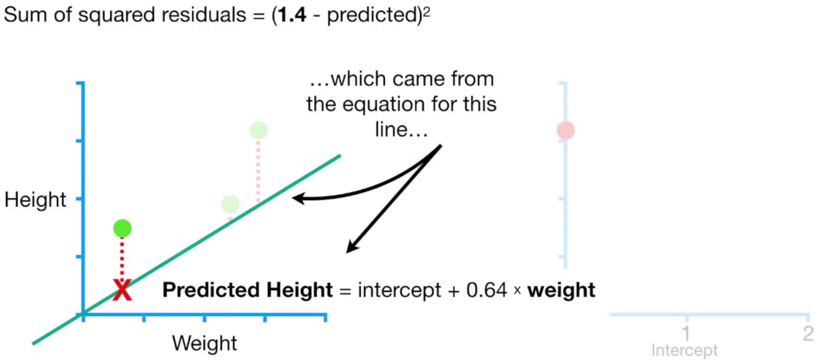
Gradiente Descendiente



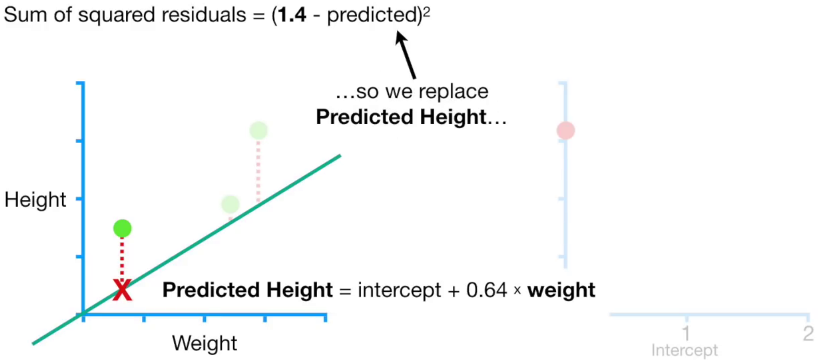
Gradiente Descendiente



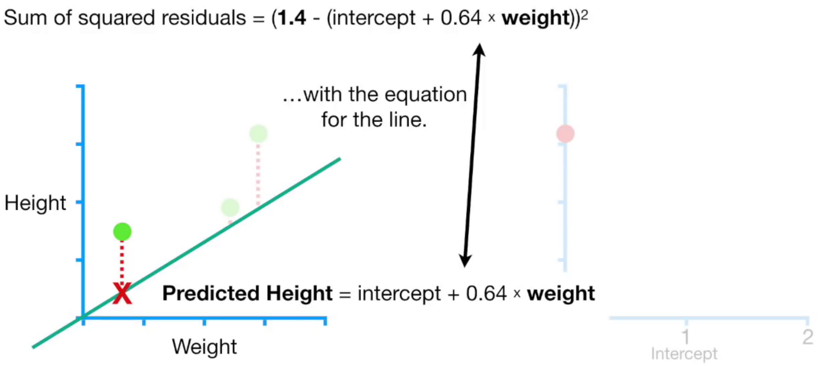
Gradiente Descendiente



Gradiente Descendiente

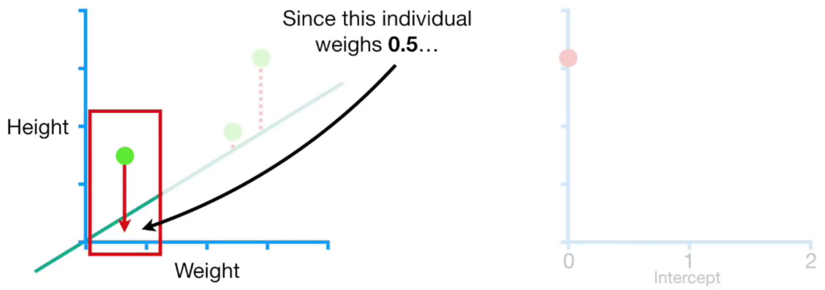


Gradiente Descendiente



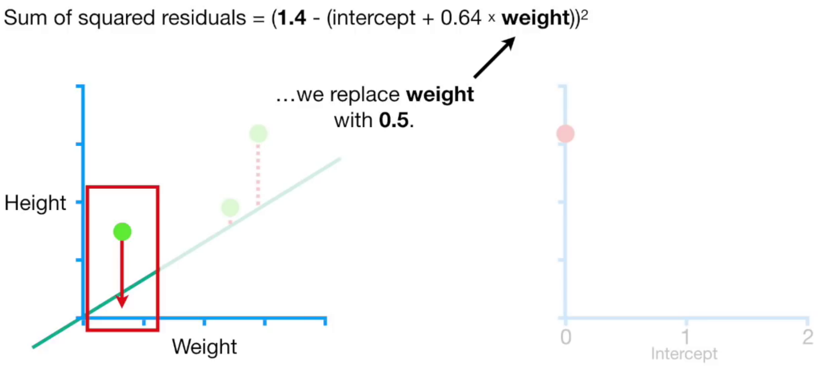
Gradiente Descendiente

Sum of squared residuals = (**1.4** - (intercept + 0.64 × **weight**))^2



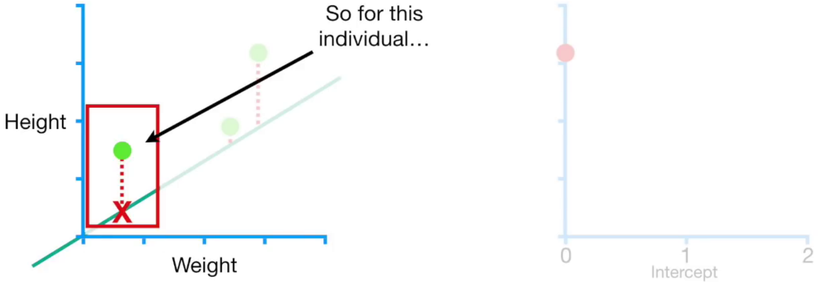
ehinojosa@unsa.edu.pe

Gradiente Descendiente



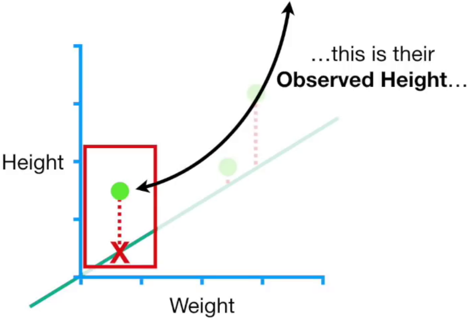
Gradiente Descendiente

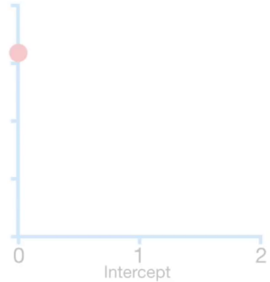
Sum of squared residuals = (**1.4** - (intercept + 0.64 × **0.5**))^2



Gradiente Descendiente

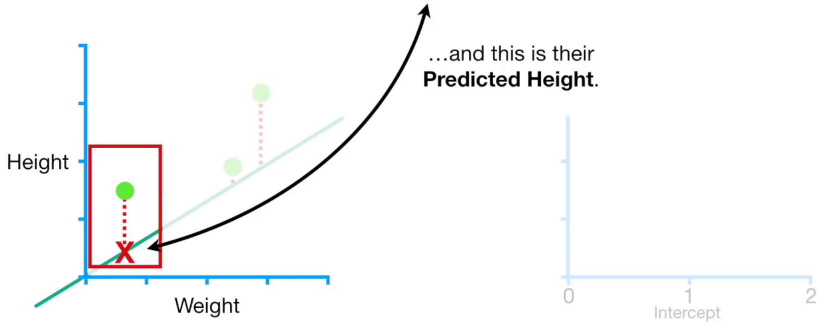
Sum of squared residuals = (**1.4** - (intercept + 0.64 × **0.5**))^2



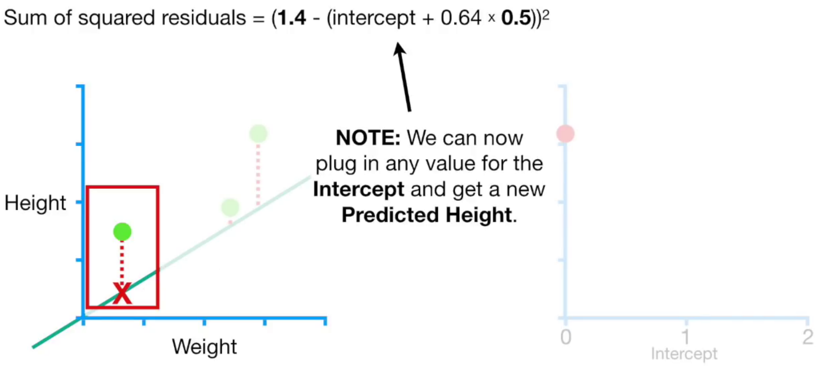


Gradiente Descendiente

Sum of squared residuals = (**1.4** - (intercept + 0.64 × **0.5**))^2

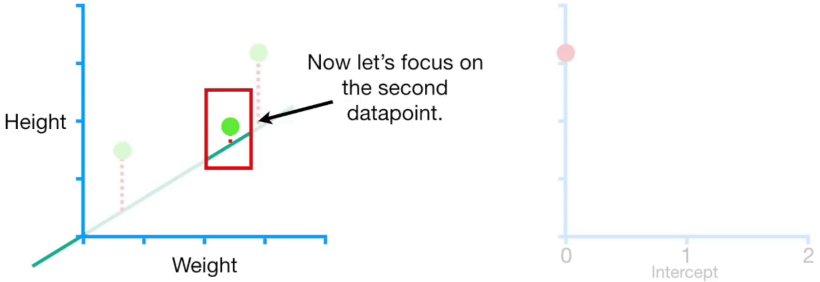


Gradiente Descendiente



Gradiente Descendiente

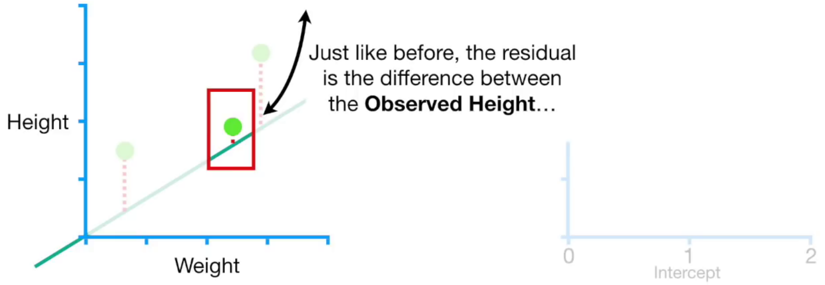
Sum of squared residuals = (**1.4** - (intercept + 0.64 × **0.5**))^2



Gradiente Descendiente

Sum of squared residuals = (**1.4** - (intercept + 0.64 × **0.5**))²

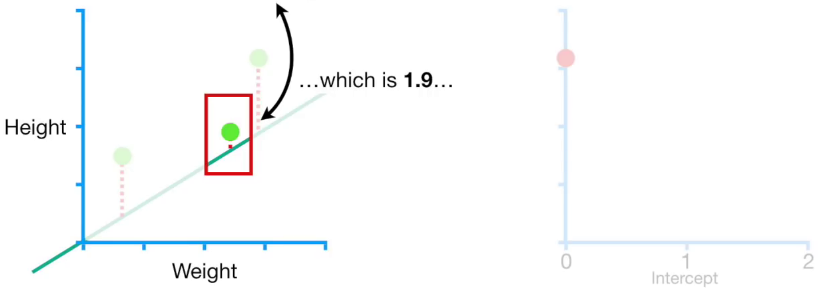
+ (observed - predicted)²



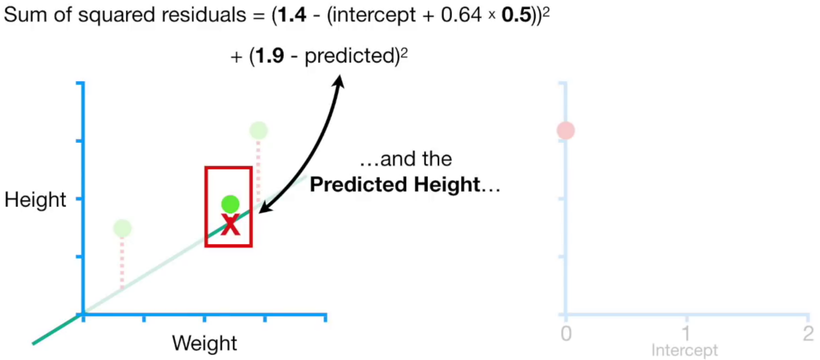
Gradiente Descendiente

Sum of squared residuals = (**1.4** - (intercept + 0.64 × **0.5**))²

+ (**1.9** - predicted)²



Gradiente Descendiente

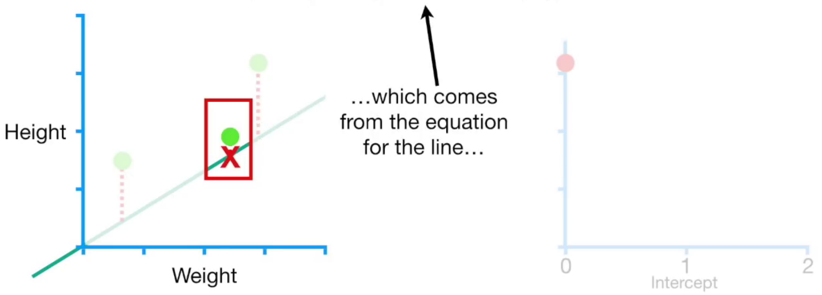


ehinojosa@unsa.edu.pe

Gradiente Descendiente

Sum of squared residuals = (**1.4** - (intercept + 0.64 × **0.5**))²

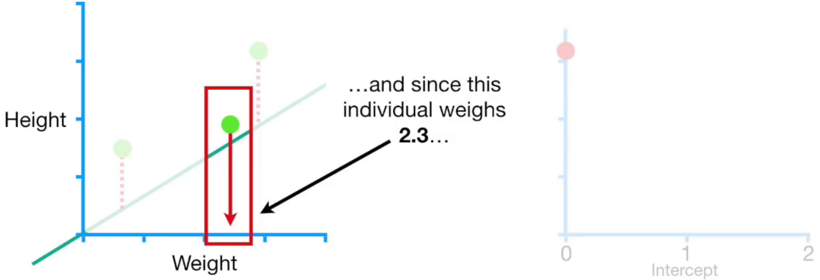
+ (**1.9** - (intercept + 0.64 × **weight**))²



Gradiente Descendiente

Sum of squared residuals = (**1.4** - (intercept + 0.64 × **0.5**))^2

+ (**1.9** - (intercept + 0.64 × **weight**))^2

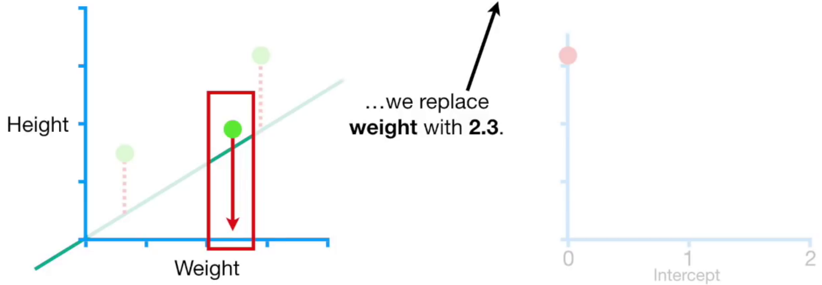


ehinojosa@unsa.edu.pe

Gradiente Descendiente

Sum of squared residuals = (**1.4** - (intercept + 0.64 × **0.5**))^2

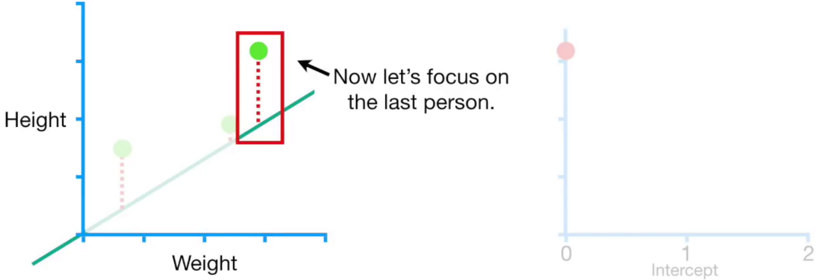
+ (**1.9** - (intercept + 0.64 × **2.3**))^2



Gradiente Descendiente

Sum of squared residuals = (**1.4** - (intercept + 0.64 × **0.5**))^2

+ (**1.9** - (intercept + 0.64 × **2.3**))^2

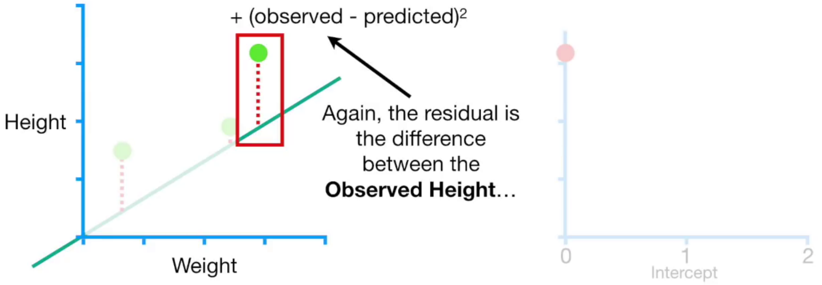


Gradiente Descendiente

Sum of squared residuals = (**1.4** - (intercept + 0.64 × **0.5**))²

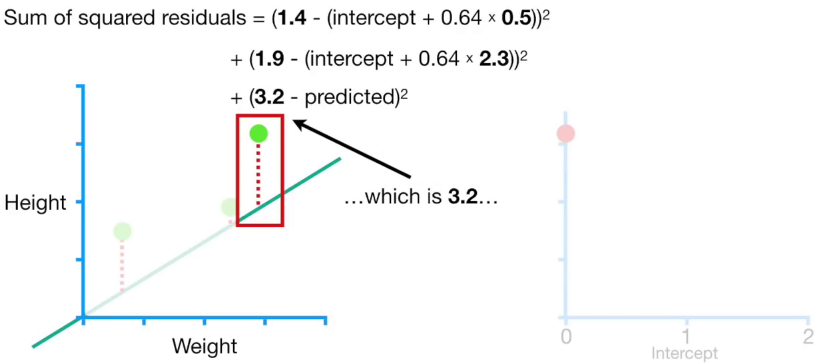
+ (**1.9** - (intercept + 0.64 × **2.3**))²

+ (observed - predicted)²



ehinojosa@unsa.edu.pe

Gradiente Descendiente

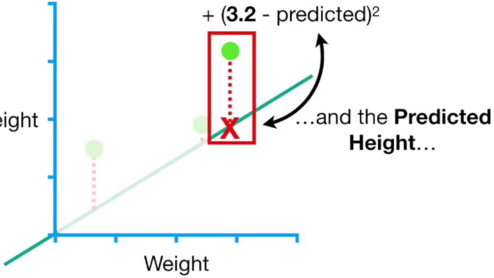


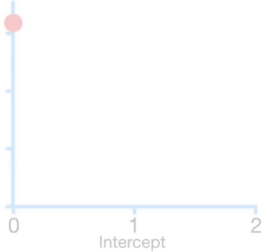
Gradiente Descendiente

Sum of squared residuals = (**1.4** - (intercept + 0.64 × **0.5**))^2

+ (**1.9** - (intercept + 0.64 × **2.3**))^2

+ (**3.2** - predicted)^2





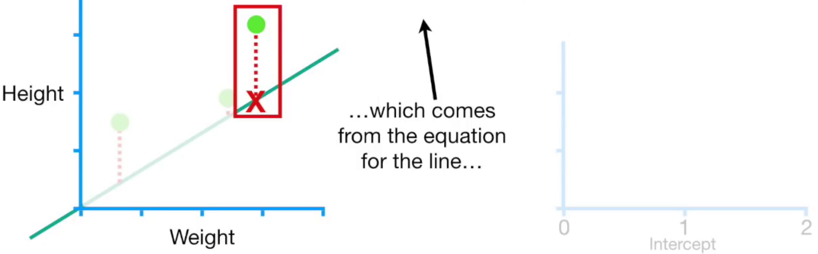
ehinojosa@unsa.edu.pe

Gradiente Descendiente

Sum of squared residuals = (**1.4** - (intercept + 0.64 × **0.5**))^2

+ (**1.9** - (intercept + 0.64 × **2.3**))^2

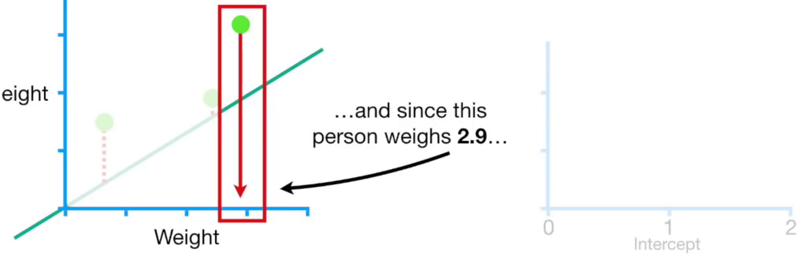
+ (**3.2** - (intercept + 0.64 × **weight**))^2



ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\text{Sum of squared residuals} = (\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))^2$$
$$+ (\mathbf{1.9} - (\text{intercept} + 0.64 \times \mathbf{2.3}))^2$$
$$+ (\mathbf{3.2} - (\text{intercept} + 0.64 \times \mathbf{weight}))^2$$



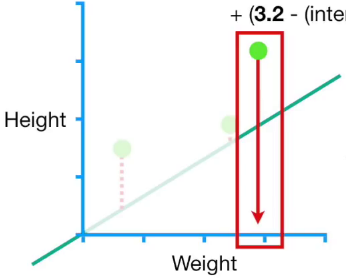
ehinojosa@unsa.edu.pe

Gradiente Descendiente

Sum of squared residuals = (**1.4** - (intercept + 0.64 × **0.5**))^2

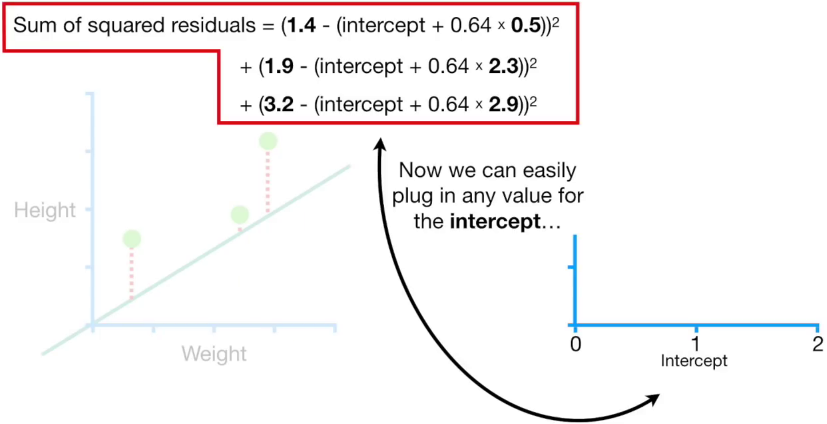
+ (**1.9** - (intercept + 0.64 × **2.3**))^2

+ (**3.2** - (intercept + 0.64 × **weight**))^2

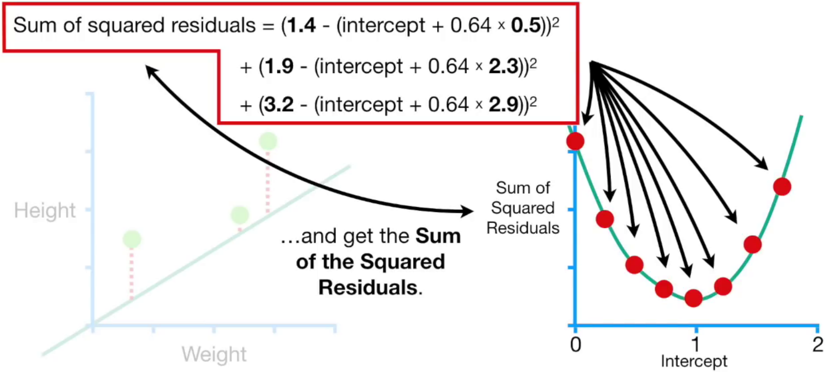


…we'll replace **weight** with **2.9**.

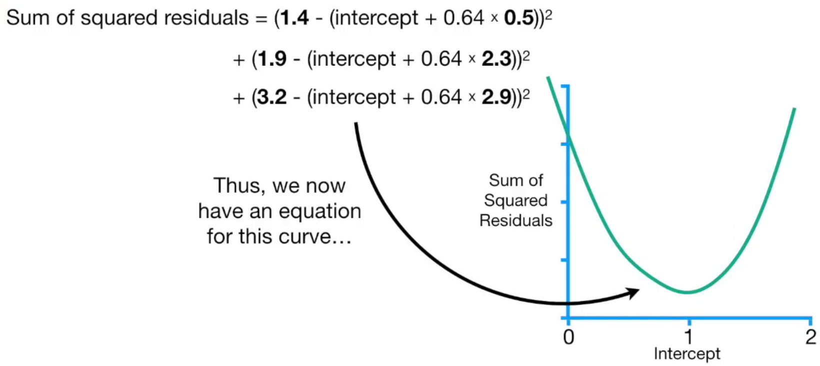
Gradiente Descendiente



Gradiente Descendiente



Gradiente Descendiente



ehinojosa@unsa.edu.pe

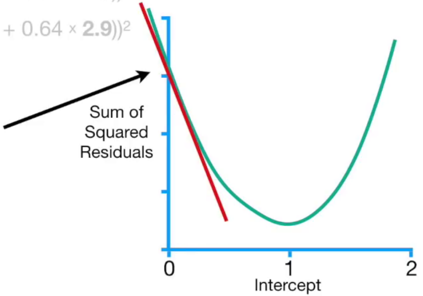
Gradiente Descendiente

Sum of squared residuals = $(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))^2$

$+ (\mathbf{1.9} - (\text{intercept} + 0.64 \times \mathbf{2.3}))^2$

$+ (\mathbf{3.2} - (\text{intercept} + 0.64 \times \mathbf{2.9}))^2$

…and we can take the derivative of this function and determine the slope at any value for the **Intercept**.

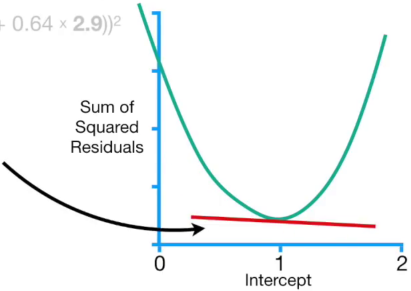


ehinojosa@unsa.edu.pe

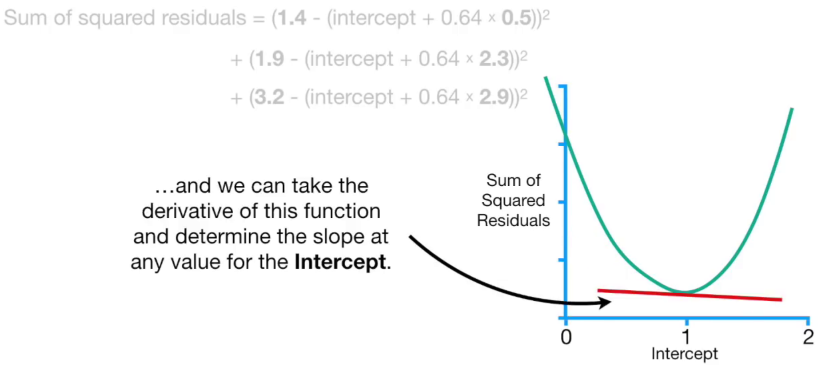
Gradiente Descendiente

$$\text{Sum of squared residuals} = (\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))^2$$
$$+ (\mathbf{1.9} - (\text{intercept} + 0.64 \times \mathbf{2.3}))^2$$
$$+ (\mathbf{3.2} - (\text{intercept} + 0.64 \times \mathbf{2.9}))^2$$

…and we can take the derivative of this function and determine the slope at any value for the **Intercept**.



Gradiente Descendiente



ehinojosa@unsa.edu.pe

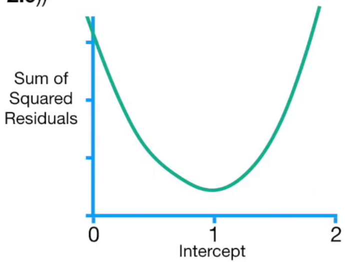
Gradiente Descendiente

Sum of squared residuals = (**1.4** - (intercept + 0.64 × **0.5**))^2

+ (**1.9** - (intercept + 0.64 × **2.3**))^2

+ (**3.2** - (intercept + 0.64 × **2.9**))^2

So let's take the derivative of the Sum of the Squared Residuals with respect to the **Intercept**.



ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\text{Sum of squared residuals} = (\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))^2 + (\mathbf{1.9} - (\text{intercept} + 0.64 \times \mathbf{2.3}))^2 + (\mathbf{3.2} - (\text{intercept} + 0.64 \times \mathbf{2.9}))^2$$

$$\frac{d}{d\ \textit{intercept}}\ \text{Sum of squared residuals} =$$

Gradiente Descendiente

$$\text{Sum of squared residuals} = \boxed{(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))^2}$$
$$+ (\mathbf{1.9} - (\text{intercept} + 0.64 \times \mathbf{2.3}))^2$$
$$+ (\mathbf{3.2} - (\text{intercept} + 0.64 \times \mathbf{2.9}))^2$$

…the derivative of the first part…

$$\frac{d}{d\ \textit{intercept}}\ \text{Sum of squared residuals} = \frac{d}{d\ \textit{intercept}} (\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))^2$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\text{Sum of squared residuals} = (\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))^2$$
$$\boxed{+ (\mathbf{1.9} - (\text{intercept} + 0.64 \times \mathbf{2.3}))^2}$$
$$+ (\mathbf{3.2} - (\text{intercept} + 0.64 \times \mathbf{2.9}))^2$$

...plus the derivative of the second part...

$$\frac{d}{d\ \textit{intercept}} \text{Sum of squared residuals} = \frac{d}{d\ \textit{intercept}} (\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))^2$$
$$+ \frac{d}{d\ \textit{intercept}} (\mathbf{1.9} - (\text{intercept} + 0.64 \times \mathbf{2.3}))^2$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

Sum of squared residuals = $(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))^2$

$+ (\mathbf{1.9} - (\text{intercept} + 0.64 \times \mathbf{2.3}))^2$

$\boxed{+ (\mathbf{3.2} - (\text{intercept} + 0.64 \times \mathbf{2.9}))^2}$

...plus the derivative of the third part.

$$\frac{d}{d\ intercept} \text{Sum of squared residuals} = \frac{d}{d\ intercept}(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))^2$$

$$+ \frac{d}{d\ intercept}(\mathbf{1.9} - (\text{intercept} + 0.64 \times \mathbf{2.3}))^2$$

$$+ \frac{d}{d\ intercept}(\mathbf{3.2} - (\text{intercept} + 0.64 \times \mathbf{2.9}))^2$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

Sum of squared residuals = (**1.4** - (intercept + 0.64 × **0.5**))^2

+ (**1.9** - (intercept + 0.64 × **2.3**))^2

+ (**3.2** - (intercept + 0.64 × **2.9**))^2

Let's start by taking the derivative of the first part.

$$\frac{d}{d\ intercept} \text{ Sum of squared residuals} = \boxed{\frac{d}{d\ intercept} (\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))^2}$$

$$+ \frac{d}{d\ intercept} (\mathbf{1.9} - (\text{intercept} + 0.64 \times \mathbf{2.3}))^2$$

$$+ \frac{d}{d\ intercept} (\mathbf{3.2} - (\text{intercept} + 0.64 \times \mathbf{2.9}))^2$$

Gradiente Descendiente

$$\frac{d}{d\ intercept}(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))^2 =$$

First, we'll move this part of the equation up here so that we have room to work.

$$\frac{d}{d\ intercept}\text{Sum of squared residuals} = \boxed{\frac{d}{d\ intercept}(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))^2}$$

$$+ \frac{d}{d\ intercept}(\mathbf{1.9} - (\text{intercept} + 0.64 \times \mathbf{2.3}))^2$$

$$+ \frac{d}{d\ intercept}(\mathbf{3.2} - (\text{intercept} + 0.64 \times \mathbf{2.9}))^2$$

Gradiente Descendiente

$$\frac{d}{d\ intercept}\boxed{(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))^2} =$$

To take the derivative of this, we need to apply...

$$\frac{d}{d\ intercept}\text{ Sum of squared residuals} = \frac{d}{d\ intercept}(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))^2$$

$$+ \frac{d}{d\ intercept}(\mathbf{1.9} - (\text{intercept} + 0.64 \times \mathbf{2.3}))^2$$

$$+ \frac{d}{d\ intercept}(\mathbf{3.2} - (\text{intercept} + 0.64 \times \mathbf{2.9}))^2$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

The Chain Rule

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ intercept}(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))^{\boxed{2}} = 2(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))$$

So we start by moving the square to the front...

$$\frac{d}{d\ intercept}\text{Sum of squared residuals} = \frac{d}{d\ intercept}(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))^2$$
$$+ \frac{d}{d\ intercept}(\mathbf{1.9} - (\text{intercept} + 0.64 \times \mathbf{2.3}))^2$$
$$+ \frac{d}{d\ intercept}(\mathbf{3.2} - (\text{intercept} + 0.64 \times \mathbf{2.9}))^2$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ intercept}\boxed{(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5})}^2 = 2(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))\boxed{\times -1}$$

…and multiply that by the
derivative of the stuff
inside the parentheses.

Gradiente Descendiente

$$\frac{d}{d\ intercept}\boxed{(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))}^2 = 2(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))\ \boxed{\times -1}$$

$$\frac{d}{d\ intercept}\ \mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5})$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ intercept}\boxed{(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5})}^2 = 2(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))\boxed{\times -1}$$

$$\frac{d}{d\ intercept}\ \mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5})$$

$$\frac{d}{d\ intercept}\ \cancel{\mathbf{1.4}} + (-1)\text{intercept} - \cancel{0.64 \times \mathbf{0.5}} = -1$$

These parts don't contain a term
for the **Intercept**, so they go away.

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ intercept}\boxed{(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5})}^2 = 2(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))\ \boxed{\times -1}$$

$$\frac{d}{d\ intercept}\ \mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5})$$

$$\frac{d}{d\ intercept}\ \mathbf{1.4} + (-1)\text{intercept} - 0.64 \times \mathbf{0.5} = \boxed{-1}$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ intercept}(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))^2 \boxed{= 2(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5})) \times -1}$$

…then we simplify by multiplying 2 by -1…

$$\frac{d}{d\ intercept}\text{ Sum of squared residuals} = \frac{d}{d\ intercept}(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))^2$$

$$+ \frac{d}{d\ intercept}(\mathbf{1.9} - (\text{intercept} + 0.64 \times \mathbf{2.3}))^2$$

$$+ \frac{d}{d\ intercept}(\mathbf{3.2} - (\text{intercept} + 0.64 \times \mathbf{2.9}))^2$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ intercept}(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))^2 = 2(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5})) \times -1$$

$$\boxed{= -2(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))}$$

…then we simplify by multiplying 2 by -1…

$$\frac{d}{d\ intercept}\text{Sum of squared residuals} = \frac{d}{d\ intercept}(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))^2$$

$$+ \frac{d}{d\ intercept}(\mathbf{1.9} - (\text{intercept} + 0.64 \times \mathbf{2.3}))^2$$

$$+ \frac{d}{d\ intercept}(\mathbf{3.2} - (\text{intercept} + 0.64 \times \mathbf{2.9}))^2$$

Gradiente Descendiente

$$\frac{d}{d\ intercept}(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))^2 = 2(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5})) \times -1$$

$$\boxed{= -2(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))}$$

…and this…

$$\frac{d}{d\ intercept}\text{Sum of squared residuals} = \frac{d}{d\ intercept}(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))^2$$

$$+ \frac{d}{d\ intercept}(\mathbf{1.9} - (\text{intercept} + 0.64 \times \mathbf{2.3}))^2$$

$$+ \frac{d}{d\ intercept}(\mathbf{3.2} - (\text{intercept} + 0.64 \times \mathbf{2.9}))^2$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ intercept}(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))^2 = 2(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5})) \times -1$$

$$\boxed{= -2(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))}$$

…and this…

…is the derivative of the first part…

$$\frac{d}{d\ intercept}\text{Sum of squared residuals} = \boxed{\frac{d}{d\ intercept}(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))^2}$$

$$+ \frac{d}{d\ intercept}(\mathbf{1.9} - (\text{intercept} + 0.64 \times \mathbf{2.3}))^2$$

$$+ \frac{d}{d\ intercept}(\mathbf{3.2} - (\text{intercept} + 0.64 \times \mathbf{2.9}))^2$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ intercept}(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))^2 = 2(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5})) \times -1$$

$$\boxed{= -2(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))}$$

...so we plug it in.

$$\frac{d}{d\ intercept}\text{ Sum of squared residuals} = \boxed{-2(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))}$$

$$+ \frac{d}{d\ intercept}(\mathbf{1.9} - (\text{intercept} + 0.64 \times \mathbf{2.3}))^2$$

$$+ \frac{d}{d\ intercept}(\mathbf{3.2} - (\text{intercept} + 0.64 \times \mathbf{2.9}))^2$$

Gradiente Descendiente

Now we need to take the derivative of the next two parts.

$$\frac{d}{d\ intercept} \text{Sum of squared residuals} = -2(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))$$

$$\boxed{\begin{aligned} &+ \frac{d}{d\ intercept}(\mathbf{1.9} - (\text{intercept} + 0.64 \times \mathbf{2.3}))^2 \\ &+ \frac{d}{d\ intercept}(\mathbf{3.2} - (\text{intercept} + 0.64 \times \mathbf{2.9}))^2 \end{aligned}}$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

I'll leave that as an exercise for the viewer.

$$\frac{d}{d\ \textit{intercept}}\ \text{Sum of squared residuals} = -2(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))$$

$$\boxed{\begin{aligned} &+ -2(\mathbf{1.9} - (\text{intercept} + 0.64 \times \mathbf{2.3})) \\ &+ -2(\mathbf{3.2} - (\text{intercept} + 0.64 \times \mathbf{2.9})) \end{aligned}}$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ intercept}\text{Sum of squared residuals} = -2(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))$$

$$+ -2(\mathbf{1.9} - (\text{intercept} + 0.64 \times \mathbf{2.3}))$$

$$+ -2(\mathbf{3.2} - (\text{intercept} + 0.64 \times \mathbf{2.9}))$$

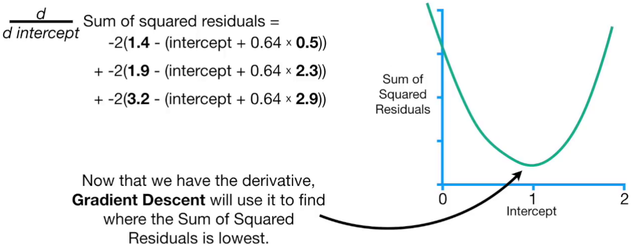
ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ intercept}\ \text{Sum of squared residuals} =$$
$$-2(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))$$
$$+ -2(\mathbf{1.9} - (\text{intercept} + 0.64 \times \mathbf{2.3}))$$
$$+ -2(\mathbf{3.2} - (\text{intercept} + 0.64 \times \mathbf{2.9}))$$

Let's move the derivative up here so that it's not taking up half of the screen.

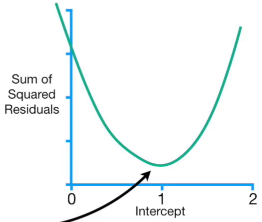
Gradiente Descendiente



ehinojosa@unsa.edu.pe

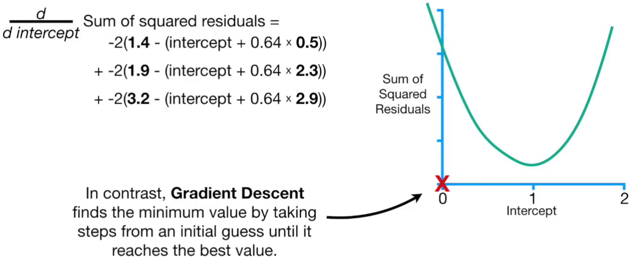
Gradiente Descendiente

$$\frac{d}{d\ intercept}\ \text{Sum of squared residuals} = -2(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5})) + -2(\mathbf{1.9} - (\text{intercept} + 0.64 \times \mathbf{2.3})) + -2(\mathbf{3.2} - (\text{intercept} + 0.64 \times \mathbf{2.9}))$$



NOTE: If we were using **Least Squares** to solve for the optimal value for the **Intercept**, we would simply find where the the slope of the curve = **0**.

Gradiente Descendiente

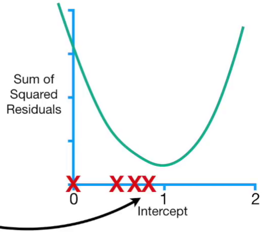


ehinojosa@unsa.edu.pe

Gradiente Descendiente

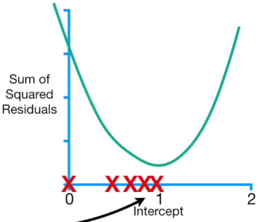
$$\frac{d}{d\ \textit{intercept}}\ \text{Sum of squared residuals} = -2(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5})) + -2(\mathbf{1.9} - (\text{intercept} + 0.64 \times \mathbf{2.3})) + -2(\mathbf{3.2} - (\text{intercept} + 0.64 \times \mathbf{2.9}))$$

In contrast, **Gradient Descent** finds the minimum value by taking steps from an initial guess until it reaches the best value.



Gradiente Descendiente

$$\frac{d}{d\ intercept}\ \text{Sum of squared residuals} = -2(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5})) + -2(\mathbf{1.9} - (\text{intercept} + 0.64 \times \mathbf{2.3})) + -2(\mathbf{3.2} - (\text{intercept} + 0.64 \times \mathbf{2.9}))$$



In contrast, **Gradient Descent** finds the minimum value by taking steps from an initial guess until it reaches the best value.

ehinojosa@unsa.edu.pe

Gradiente Descendiente

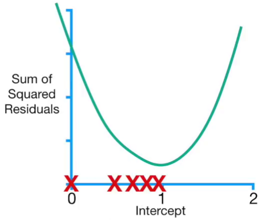
$\frac{d}{d\ intercept}$ Sum of squared residuals =

-2(**1.4** - (intercept + 0.64 × **0.5**))

+ -2(**1.9** - (intercept + 0.64 × **2.3**))

+ -2(**3.2** - (intercept + 0.64 × **2.9**))

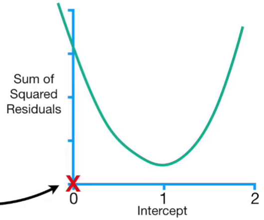
This makes **Gradient Descent** very useful when it is not possible to solve for where the derivative = **0**, and this is why **Gradient Descent** can be used in so many different situations.



Gradiente Descendiente

$$\frac{d}{d\ \textit{intercept}} \text{ Sum of squared residuals} = -2(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5})) + -2(\mathbf{1.9} - (\text{intercept} + 0.64 \times \mathbf{2.3})) + -2(\mathbf{3.2} - (\text{intercept} + 0.64 \times \mathbf{2.9}))$$

Remember, we started by setting the **Intercept** to a random number. In this case, that was **0**.

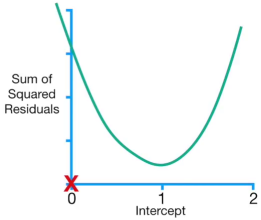


ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ intercept}\ \text{Sum of squared residuals} = -2(\mathbf{1.4} - (0 + 0.64 \times \mathbf{0.5})) + -2(\mathbf{1.9} - (0 + 0.64 \times \mathbf{2.3})) + -2(\mathbf{3.2} - (0 + 0.64 \times \mathbf{2.9}))$$

So we plug **0** into the derivative...



ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ intercept}\ \text{Sum of squared residuals} =$$

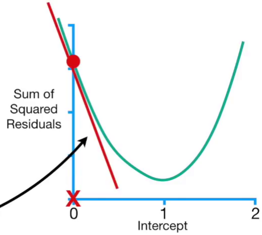
$$-2(\mathbf{1.4} - (0 + 0.64 \times \mathbf{0.5}))$$

$$+ -2(\mathbf{1.9} - (0 + 0.64 \times \mathbf{2.3}))$$

$$+ -2(\mathbf{3.2} - (0 + 0.64 \times \mathbf{2.9}))$$

$$= -5.7$$

So when the **Intercept** = **0**,
the slope of the curve = **-5.7**.

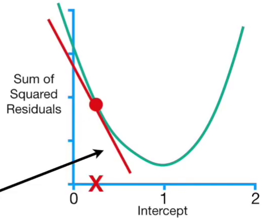


ehinojosa@unsa.edu.pe

Gradiente Descendiente

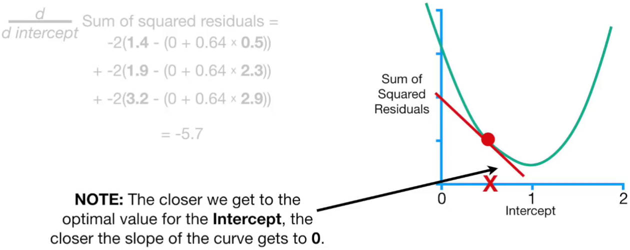
$$\frac{d}{d\ intercept}\ \text{Sum of squared residuals} = -2(\mathbf{1.4} - (0 + 0.64 \times \mathbf{0.5})) + -2(\mathbf{1.9} - (0 + 0.64 \times \mathbf{2.3})) + -2(\mathbf{3.2} - (0 + 0.64 \times \mathbf{2.9})) = -5.7$$

NOTE: The closer we get to the optimal value for the **Intercept**, the closer the slope of the curve gets to **0**.



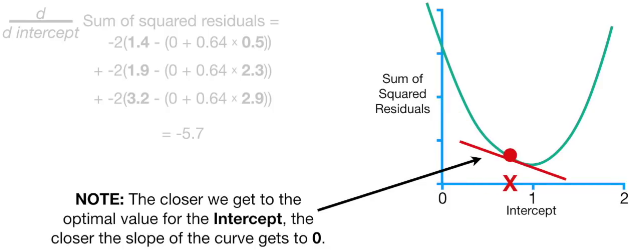
ehinojosa@unsa.edu.pe

Gradiente Descendiente

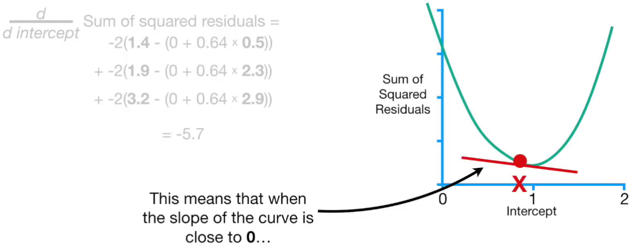


ehinojosa@unsa.edu.pe

Gradiente Descendiente



Gradiente Descendiente

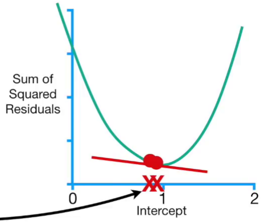


ehinojosa@unsa.edu.pe

Gradiente Descendiente

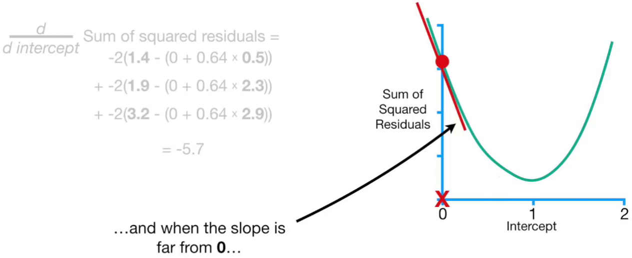
$$\frac{d}{d\ intercept}\ \text{Sum of squared residuals} = -2(\mathbf{1.4} - (0 + 0.64 \times \mathbf{0.5})) + -2(\mathbf{1.9} - (0 + 0.64 \times \mathbf{2.3})) + -2(\mathbf{3.2} - (0 + 0.64 \times \mathbf{2.9})) = -5.7$$

…then we should take baby steps, because we are close to the optimal value…

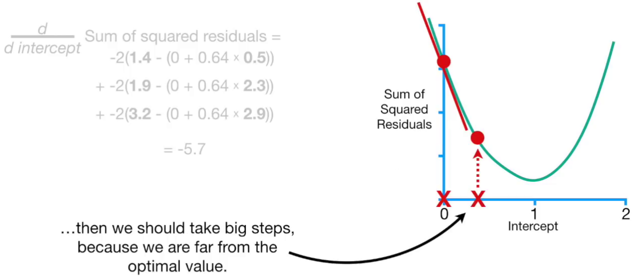


ehinojosa@unsa.edu.pe

Gradiente Descendiente

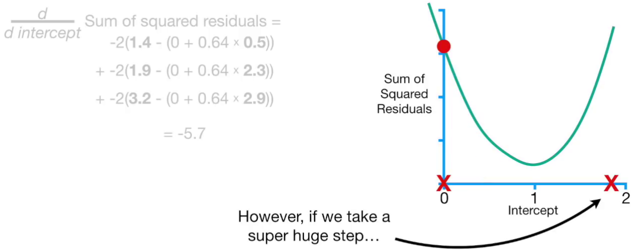


Gradiente Descendiente



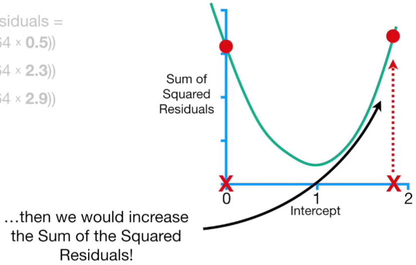
ehinojosa@unsa.edu.pe

Gradiente Descendiente



Gradiente Descendiente

$$\frac{d}{d\ intercept}\ \text{Sum of squared residuals} = -2(\mathbf{1.4} - (0 + 0.64 \times \mathbf{0.5})) + -2(\mathbf{1.9} - (0 + 0.64 \times \mathbf{2.3})) + -2(\mathbf{3.2} - (0 + 0.64 \times \mathbf{2.9})) = -5.7$$

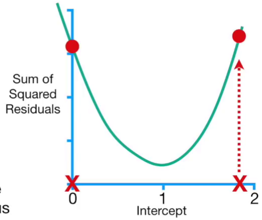


ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ intercept}\ \text{Sum of squared residuals} = -2(\mathbf{1.4} - (0 + 0.64 \times \mathbf{0.5})) + -2(\mathbf{1.9} - (0 + 0.64 \times \mathbf{2.3})) + -2(\mathbf{3.2} - (0 + 0.64 \times \mathbf{2.9}))$$

$$= -5.7$$



So the size of the step should be related to the slope, since it tells us if we should take a baby step or a big step, but we need to make sure the big step is not too big.

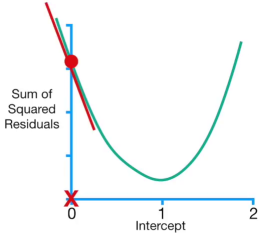
ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ intercept}\ \text{Sum of squared residuals} = -2(\mathbf{1.4} - (0 + 0.64 \times \mathbf{0.5})) + -2(\mathbf{1.9} - (0 + 0.64 \times \mathbf{2.3})) + -2(\mathbf{3.2} - (0 + 0.64 \times \mathbf{2.9})) \boxed{= -5.7}$$

Step Size = -5.7

Gradient Descent determines the **Step Size** by multiplying the **slope**...

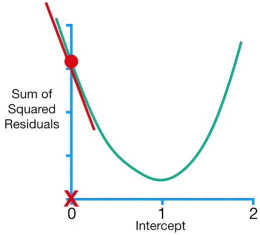


Gradiente Descendiente

$$\frac{d}{d\ intercept}\ \text{Sum of squared residuals} = -2(\mathbf{1.4} - (0 + 0.64 \times \mathbf{0.5})) + -2(\mathbf{1.9} - (0 + 0.64 \times \mathbf{2.3})) + -2(\mathbf{3.2} - (0 + 0.64 \times \mathbf{2.9})) = -5.7$$

Step Size = -5.7 × 0.1

...by a small number called **The Learning Rate**.



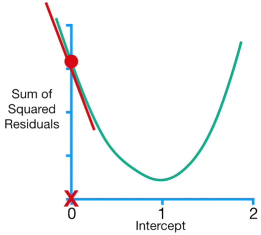
ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ intercept} \text{ Sum of squared residuals} = -2(\mathbf{1.4} - (0 + 0.64 \times \mathbf{0.5})) + -2(\mathbf{1.9} - (0 + 0.64 \times \mathbf{2.3})) + -2(\mathbf{3.2} - (0 + 0.64 \times \mathbf{2.9})) = -5.7$$

Step Size = -5.7 × 0.1 = -0.57

When the **Intercept** = **0**, the **Step Size** = **-0.57**.



Gradiente Descendiente

$\frac{d}{d\ intercept}$ Sum of squared residuals =

-2(**1.4** - (0 + 0.64 × **0.5**))

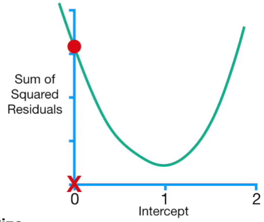
+ -2(**1.9** - (0 + 0.64 × **2.3**))

+ -2(**3.2** - (0 + 0.64 × **2.9**))

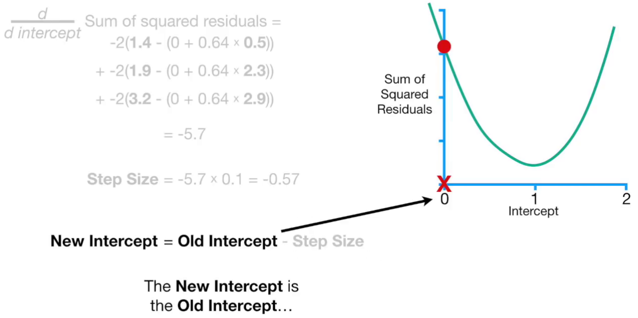
= -5.7

Step Size = -5.7 × 0.1 = -0.57

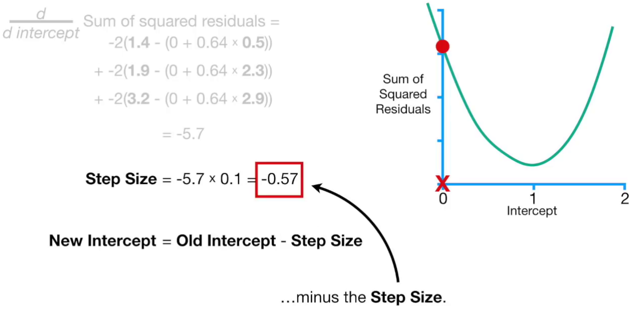
New Intercept = ← With the **Step Size**, we can calculate a **New Intercept**.



Gradiente Descendiente

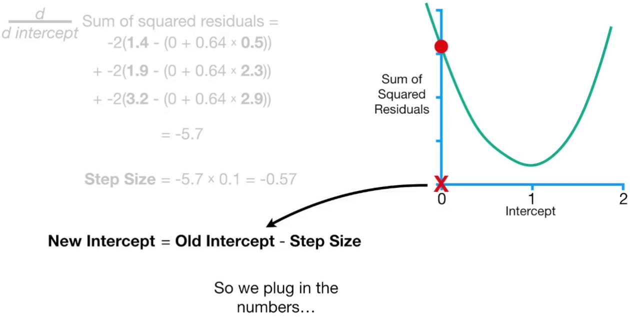


Gradiente Descendiente



ehinojosa@unsa.edu.pe

Gradiente Descendiente



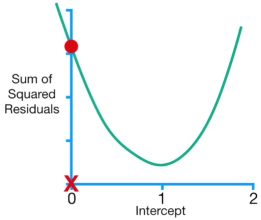
Gradiente Descendiente

$$\frac{d}{d\ intercept} \text{ Sum of squared residuals} = -2(\mathbf{1.4} - (0 + 0.64 \times \mathbf{0.5})) + -2(\mathbf{1.9} - (0 + 0.64 \times \mathbf{2.3})) + -2(\mathbf{3.2} - (0 + 0.64 \times \mathbf{2.9})) = -5.7$$

$$\textbf{Step Size} = -5.7 \times 0.1 = \boxed{-0.57}$$

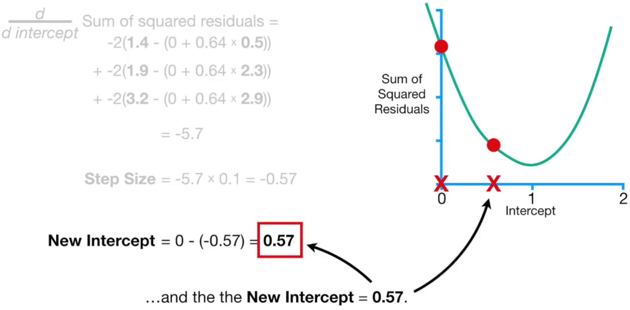
$$\textbf{New Intercept} = 0 - (-0.57)$$

So we plug in the numbers…



ehinojosa@unsa.edu.pe

Gradiente Descendiente



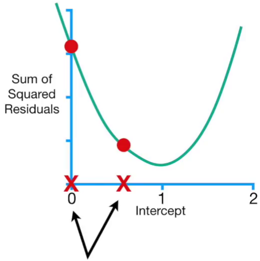
Gradiente Descendiente

$\frac{d}{d\ intercept}$ Sum of squared residuals =
-2(**1.4** - (0 + 0.64 × **0.5**))
+ -2(**1.9** - (0 + 0.64 × **2.3**))
+ -2(**3.2** - (0 + 0.64 × **2.9**))

= -5.7

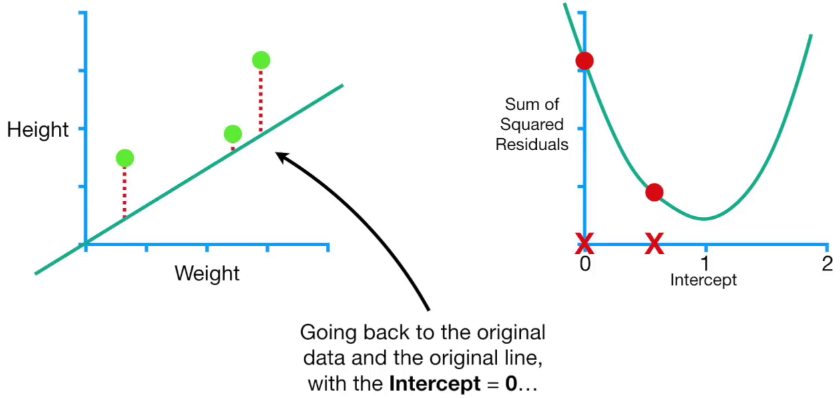
Step Size = -5.7 × 0.1 = -0.57

New Intercept = 0 - (-0.57) = **0.57**

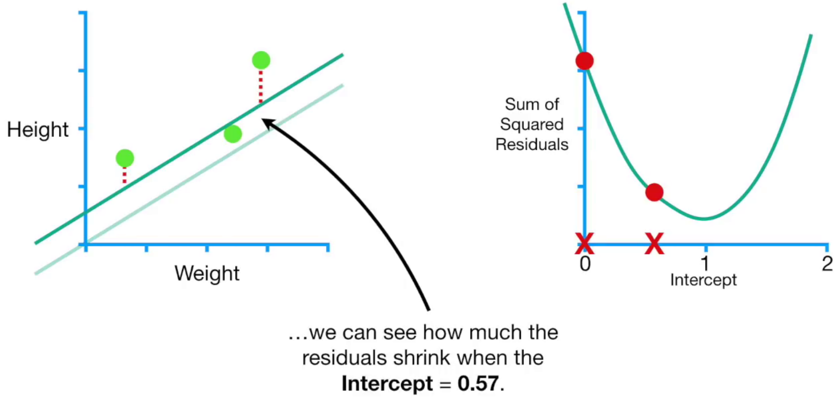


In one big step, we moved much closer to the optimal value for the **Intercept**.

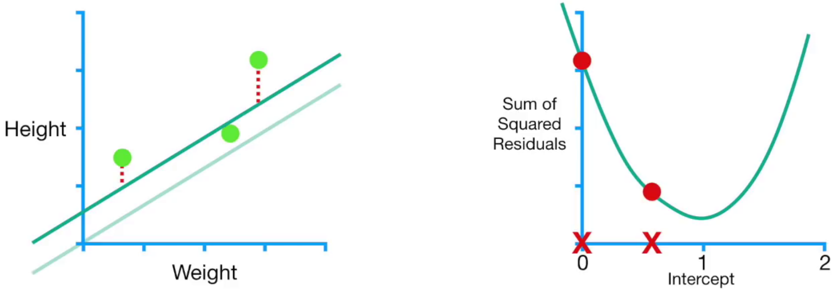
Gradiente Descendiente



Gradiente Descendiente



Gradiente Descendiente

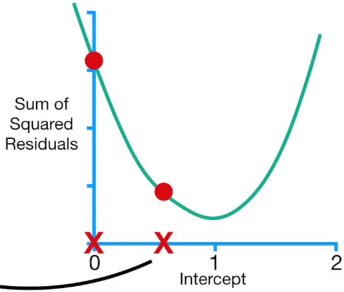


Now let's take another step
closer to the optimal value
for the **Intercept**.

Gradiente Descendiente

$$\frac{d}{d\ intercept}\ \text{Sum of squared residuals} = -2(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5})) + -2(\mathbf{1.9} - (\text{intercept} + 0.64 \times \mathbf{2.3})) + -2(\mathbf{3.2} - (\text{intercept} + 0.64 \times \mathbf{2.9}))$$

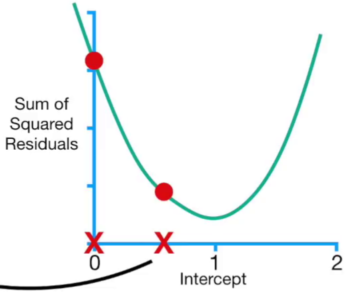
To take another step, we go back to the derivative and plug in the **New Intercept** (**0.57**)...



Gradiente Descendiente

$$\frac{d}{d\ intercept}\ \text{Sum of squared residuals} = -2(\mathbf{1.4} - (0.57 + 0.64 \times \mathbf{0.5})) + -2(\mathbf{1.9} - (0.57 + 0.64 \times \mathbf{2.3})) + -2(\mathbf{3.2} - (0.57 + 0.64 \times \mathbf{2.9}))$$

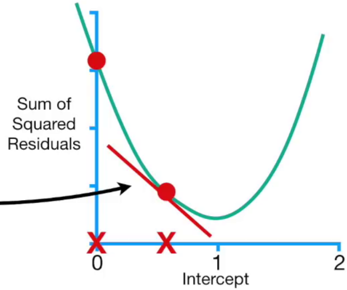
To take another step, we go back to the derivative and plug in the **New Intercept** (**0.57**)…



ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ intercept}\ \text{Sum of squared residuals} = -2(\mathbf{1.4} - (0.57 + 0.64 \times \mathbf{0.5})) + -2(\mathbf{1.9} - (0.57 + 0.64 \times \mathbf{2.3})) + -2(\mathbf{3.2} - (0.57 + 0.64 \times \mathbf{2.9})) = \boxed{-2.3}$$



...and that tells us the slope of the curve = **-2.3**.

ehinojosa@unsa.edu.pe

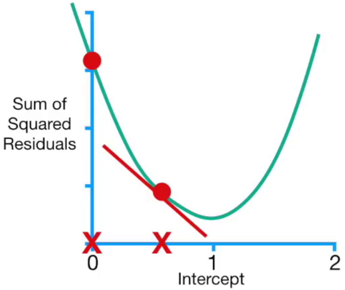
Gradiente Descendiente

$$\frac{d}{d\ intercept}\ \text{Sum of squared residuals} = -2(\mathbf{1.4} - (0.57 + 0.64 \times \mathbf{0.5})) + -2(\mathbf{1.9} - (0.57 + 0.64 \times \mathbf{2.3})) + -2(\mathbf{3.2} - (0.57 + 0.64 \times \mathbf{2.9}))$$

$$= -2.3$$

Step Size = **Slope** × **Learning Rate**

Now let's calculate the **Step Size**...

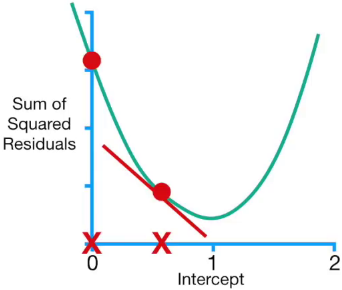


Gradiente Descendiente

$$\frac{d}{d\ intercept} \text{ Sum of squared residuals} = -2(\mathbf{1.4} - (0.57 + 0.64 \times \mathbf{0.5}))$$
$$+ -2(\mathbf{1.9} - (0.57 + 0.64 \times \mathbf{2.3}))$$
$$+ -2(\mathbf{3.2} - (0.57 + 0.64 \times \mathbf{2.9}))$$
$$= \boxed{-2.3}$$

Step Size = -2.3 × **Learning Rate**

...by plugging in **-2.3** for the **Slope**...



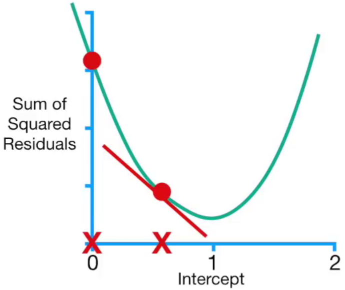
Gradiente Descendiente

$$\frac{d}{d\ intercept} \text{ Sum of squared residuals} = -2(\mathbf{1.4} - (0.57 + 0.64 \times \mathbf{0.5})) + -2(\mathbf{1.9} - (0.57 + 0.64 \times \mathbf{2.3})) + -2(\mathbf{3.2} - (0.57 + 0.64 \times \mathbf{2.9}))$$

$$= -2.3$$

Step Size = -2.3 × 0.1

...and **0.1** for the **Learning Rate**.

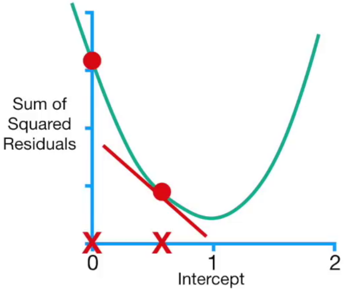


Gradiente Descendiente

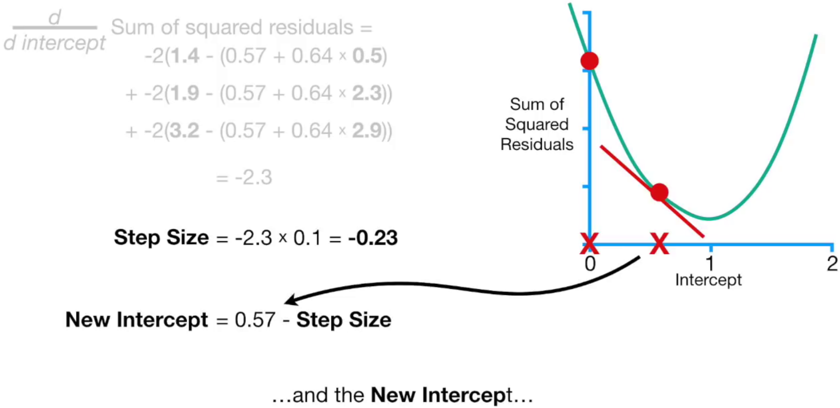
$$\frac{d}{d\ intercept}\ \text{Sum of squared residuals} =$$
$$-2(\mathbf{1.4} - (0.57 + 0.64 \times \mathbf{0.5}))$$
$$+ -2(\mathbf{1.9} - (0.57 + 0.64 \times \mathbf{2.3}))$$
$$+ -2(\mathbf{3.2} - (0.57 + 0.64 \times \mathbf{2.9}))$$
$$= -2.3$$

Step Size = -2.3 × 0.1 = **-0.23**

Ultimately, the **Step Size** is **-0.23**...



Gradiente Descendiente



ehinojosa@unsa.edu.pe

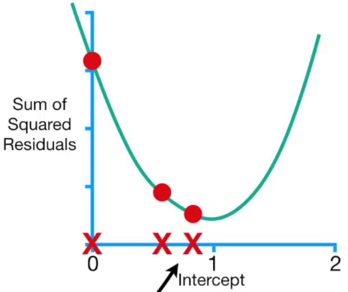
Gradiente Descendiente

$\frac{d}{d\ intercept}$ Sum of squared residuals =
-2(**1.4** - (0.57 + 0.64 × **0.5**)
+ -2(**1.9** - (0.57 + 0.64 × **2.3**))
+ -2(**3.2** - (0.57 + 0.64 × **2.9**))

= -2.3

Step Size = -2.3 × 0.1 = **-0.23**

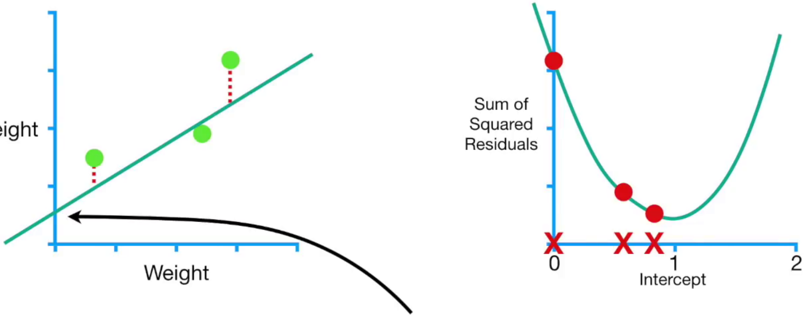
New Intercept = 0.57 - (-0.23) = $\boxed{\textbf{0.8}}$



...and the **New Intercept** = **0.8**

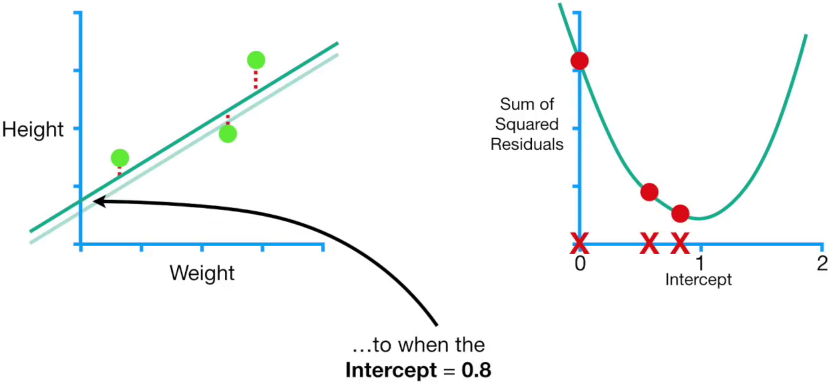
ehinojosa@unsa.edu.pe

Gradiente Descendiente

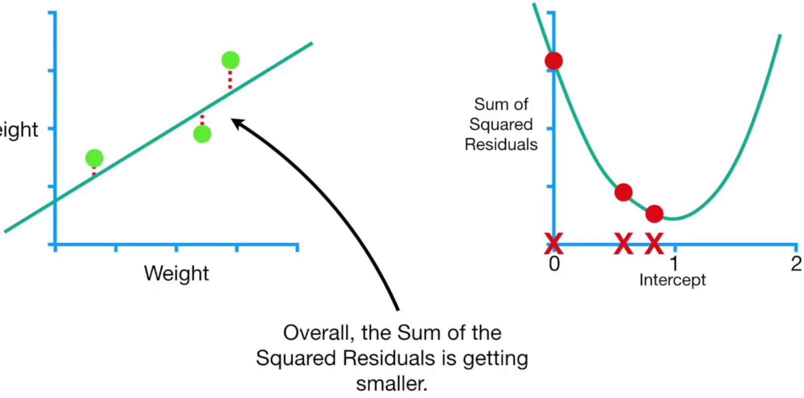


Now we can compare the
residuals when the
Intercept = 0.57...

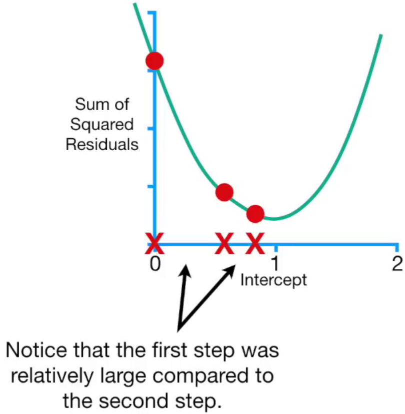
Gradiente Descendiente



Gradiente Descendiente

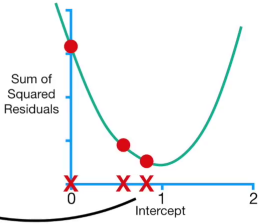


Gradiente Descendiente



Gradiente Descendiente

$$\frac{d}{d\ intercept}\ \text{Sum of squared residuals} =$$
$$-2(\mathbf{1.4} - (\text{intercept} + 0.64 \times \mathbf{0.5}))$$
$$+ -2(\mathbf{1.9} - (\text{intercept} + 0.64 \times \mathbf{2.3}))$$
$$+ -2(\mathbf{3.2} - (\text{intercept} + 0.64 \times \mathbf{2.9}))$$

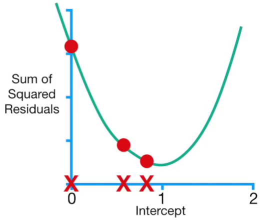


Now let's calculate the
derivative at the
New Intercept (0.8)...

Gradiente Descendiente

$$\frac{d}{d\ intercept} \text{ Sum of squared residuals} =$$
$$-2(\mathbf{1.4} - (0.8 + 0.64 \times \mathbf{0.5}))$$
$$+ -2(\mathbf{1.9} - (0.8 + 0.64 \times \mathbf{2.3}))$$
$$+ -2(\mathbf{3.2} - (0.8 + 0.64 \times \mathbf{2.9}))$$
$$= \boxed{\mathbf{-0.9}}$$

...and we get **-0.9**.

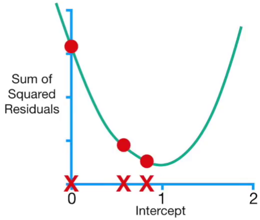


Gradiente Descendiente

$$\frac{d}{d\ intercept}\ \text{Sum of squared residuals} = -2(\mathbf{1.4} - (0.8 + 0.64 \times \mathbf{0.5})) + -2(\mathbf{1.9} - (0.8 + 0.64 \times \mathbf{2.3})) + -2(\mathbf{3.2} - (0.8 + 0.64 \times \mathbf{2.9})) = \mathbf{-0.9}$$

Step Size = **Slope** × **Learning Rate**

The **Step Size**...

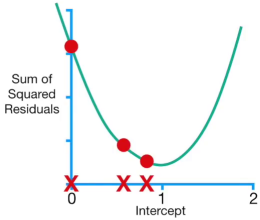


ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ intercept} \text{ Sum of squared residuals} =$$
$$-2(\mathbf{1.4} - (0.8 + 0.64 \times \mathbf{0.5}))$$
$$+ -2(\mathbf{1.9} - (0.8 + 0.64 \times \mathbf{2.3}))$$
$$+ -2(\mathbf{3.2} - (0.8 + 0.64 \times \mathbf{2.9}))$$
$$= \mathbf{-0.9}$$

Step Size = -0.9 × 0.1 = **-0.09**



The **Step Size** = **-0.09**...

Gradiente Descendiente

$\frac{d}{d\ intercept}$ Sum of squared residuals =

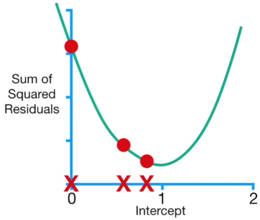
-2(**1.4** - (0.8 + 0.64 × **0.5**))

+ -2(**1.9** - (0.8 + 0.64 × **2.3**))

+ -2(**3.2** - (0.8 + 0.64 × **2.9**))

= **-0.9**

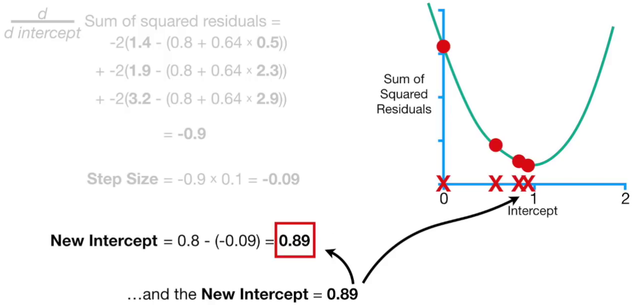
Step Size = -0.9 × 0.1 = **-0.09**



New Intercept = **Old Intercept** - **Step Size**

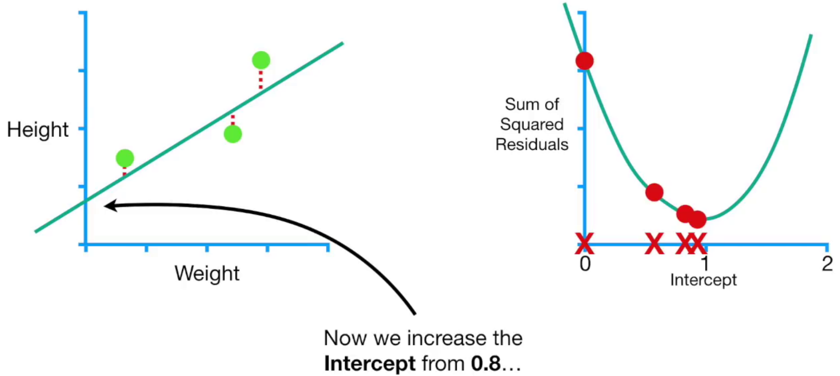
…and the **New Intercept**…

Gradiente Descendiente

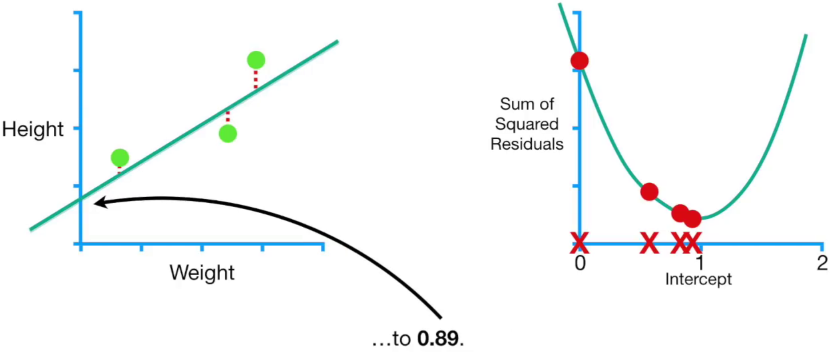


ehinojosa@unsa.edu.pe

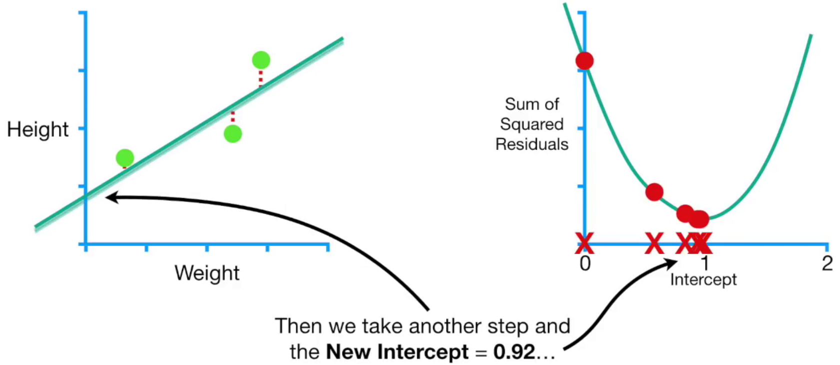
Gradiente Descendiente



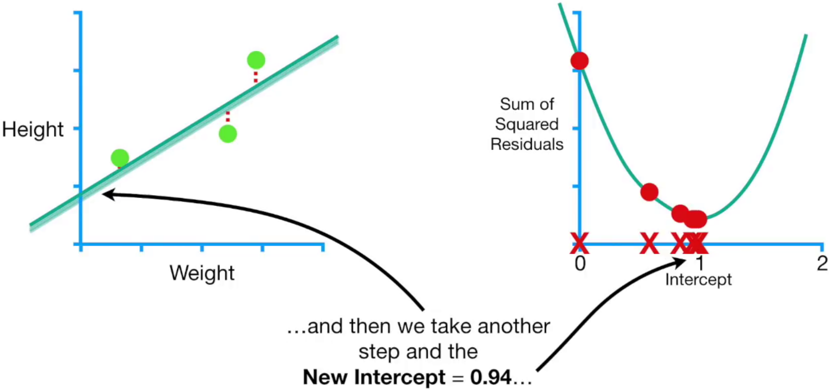
Gradiente Descendiente



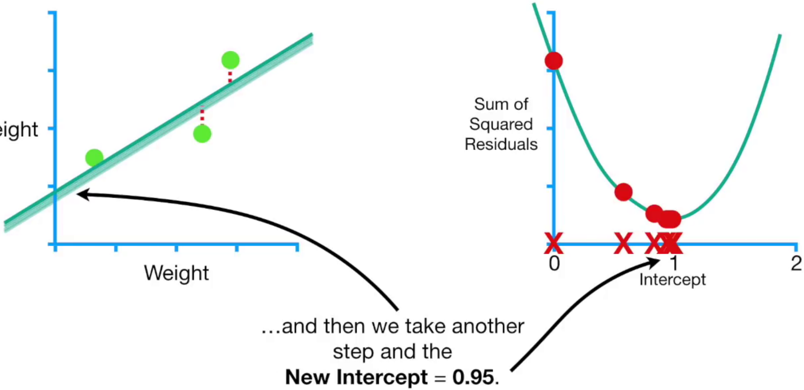
Gradiente Descendiente



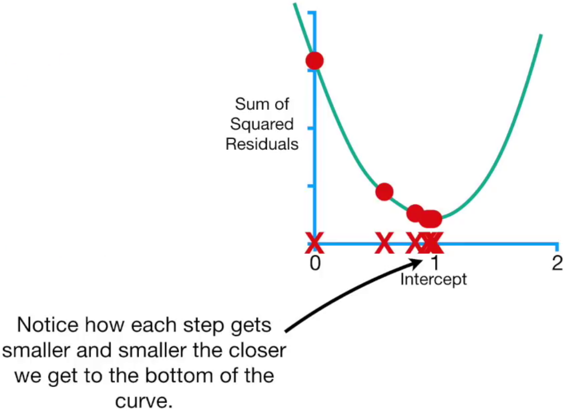
Gradiente Descendiente



Gradiente Descendiente

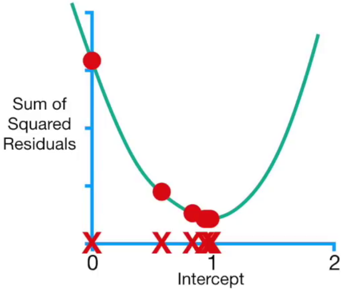


Gradiente Descendiente



Gradiente Descendiente

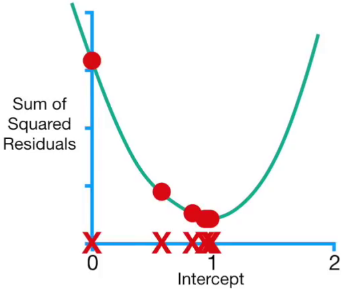
After 6 steps, the **Gradient Descent** estimate for the **Intercept** is **0.95**.



Gradiente Descendiente

After 6 steps, the **Gradient Descent** estimate for the **Intercept** is **0.95**.

NOTE: The **Least Squares** estimate for the intercept is also **0.95**.

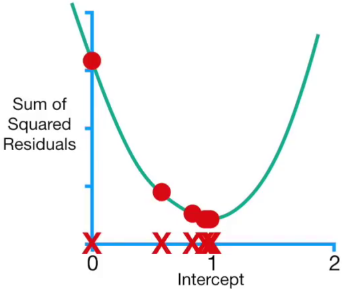


Gradiente Descendiente

After 6 steps, the **Gradient Descent** estimate for the **Intercept** is **0.95**.

NOTE: The **Least Squares** estimate for the intercept is also **0.95**.

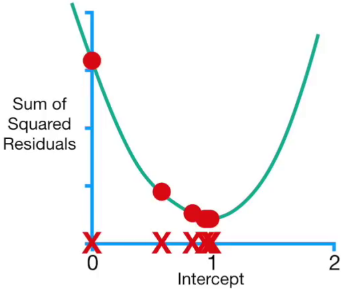
So we know that **Gradient Descent** has done its job, but without comparing its solution to a gold standard, how does **Gradient Descent** know to stop taking steps?



Gradiente Descendiente

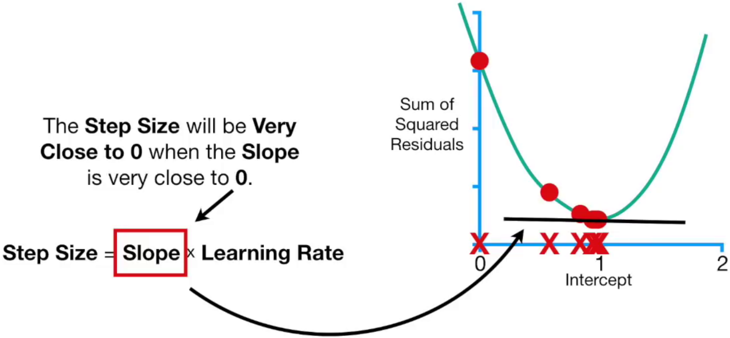
Gradient Descent stops when the **Step Size** is **Very Close To 0**.

Step Size = **Slope** × **Learning Rate**



ehinojosa@unsa.edu.pe

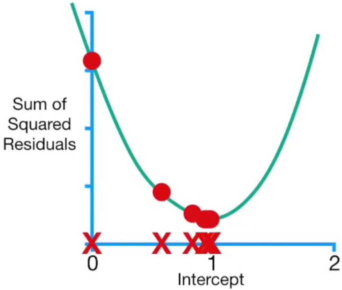
Gradiente Descendiente



Gradiente Descendiente

In practice, the
Minimum Step Size = 0.001
or smaller.

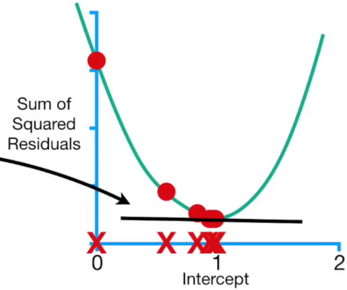
Step Size = Slope × Learning Rate



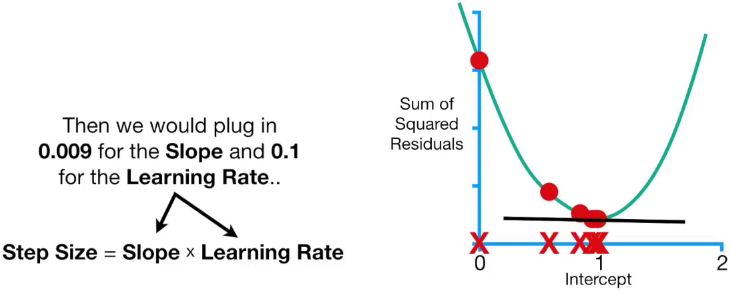
Gradiente Descendiente

So if this **slope** = **0.009**...

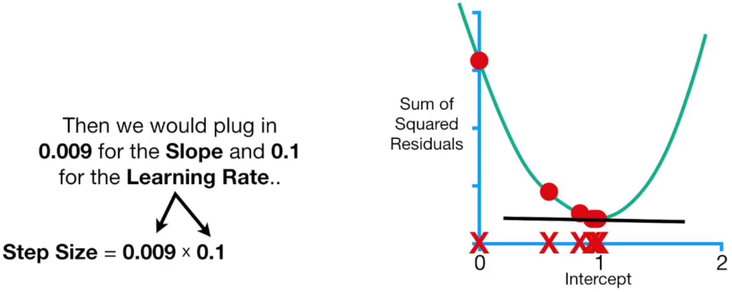
Step Size = **Slope** × **Learning Rate**



Gradiente Descendiente



Gradiente Descendiente

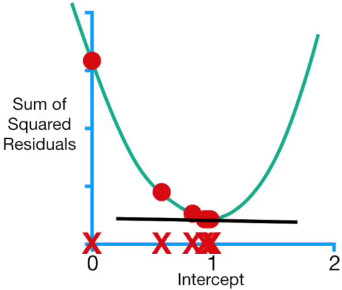


ehinojosa@unsa.edu.pe

Gradiente Descendiente

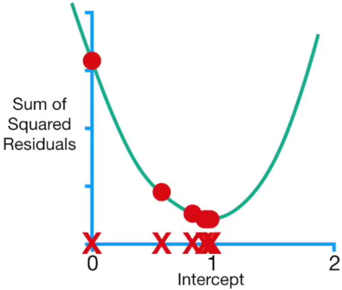
...and get **0.0009**, which is smaller than **0.001**, so **Gradient Descent** would stop.

Step Size = **0.009** × **0.1** = 0.0009



Gradiente Descendiente

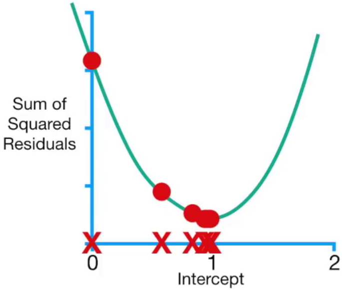
That said, **Gradient Descent** also includes a limit on the number of steps it will take before giving up.



Gradiente Descendiente

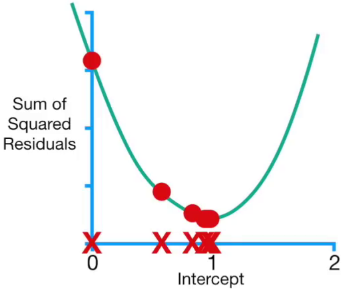
That said, **Gradient Descent** also includes a limit on the number of steps it will take before giving up.

In practice, the **Maximum Number of Steps** = **1,000** or greater.

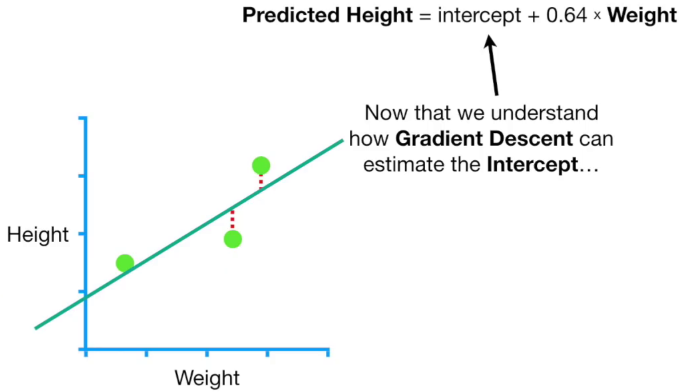


Gradiente Descendiente

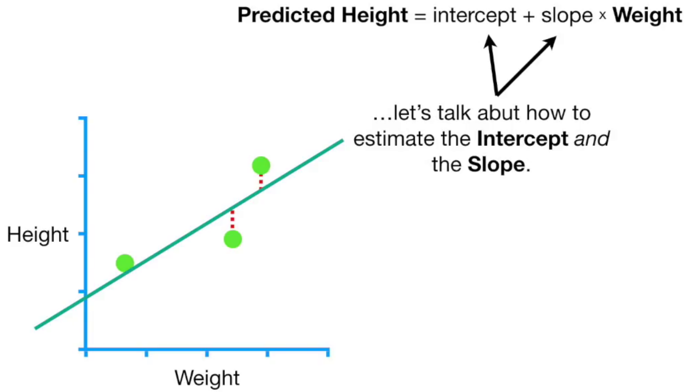
So, even if the **Step Size** is large, if there have been more than the **Maximum Number of Steps**, **Gradient Descent** will stop.



Gradiente Descendiente



Gradiente Descendiente



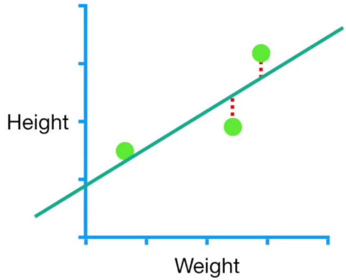
ehinojosa@unsa.edu.pe

Gradiente Descendiente

Sum of squared residuals = (**1.4** - (intercept + slope × **0.5**))^2

+ (**1.9** - (intercept + slope × **2.3**))^2

+ (**3.2** - (intercept + slope × **2.9**))^2



Just like before, we will use the Sum of the Squared Residuals as the **Loss Function**

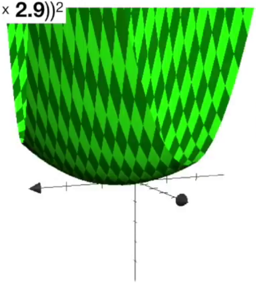
Gradiente Descendiente

Sum of squared residuals = (**1.4** - (intercept + slope × **0.5**))²

+ (**1.9** - (intercept + slope × **2.3**))²

+ (**3.2** - (intercept + slope × **2.9**))²

This is a 3-D graph of the **Loss Function** for different values for the **Intercept** and the **Slope**



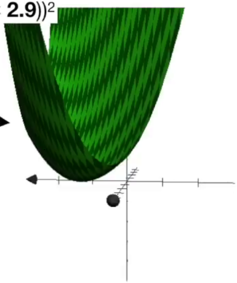
Gradiente Descendiente

Sum of squared residuals = (**1.4** - (intercept + slope × **0.5**))^2

+ (**1.9** - (intercept + slope × **2.3**))^2

+ (**3.2** - (intercept + slope × **2.9**))^2

This is a 3-D graph of the **Loss Function** for different values for the **Intercept** and the **Slope**



ehinojosa@unsa.edu.pe

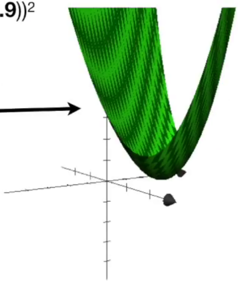
Gradiente Descendiente

Sum of squared residuals = (**1.4** - (intercept + slope × **0.5**))^2

+ (**1.9** - (intercept + slope × **2.3**))^2

+ (**3.2** - (intercept + slope × **2.9**))^2

This is a 3-D graph of the **Loss Function** for different values for the **Intercept** and the **Slope**



ehinojosa@unsa.edu.pe

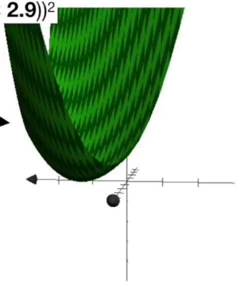
Gradiente Descendiente

Sum of squared residuals = (**1.4** - (intercept + slope × **0.5**))^2

+ (**1.9** - (intercept + slope × **2.3**))^2

+ (**3.2** - (intercept + slope × **2.9**))^2

This is a 3-D graph of the **Loss Function** for different values for the **Intercept** and the **Slope**



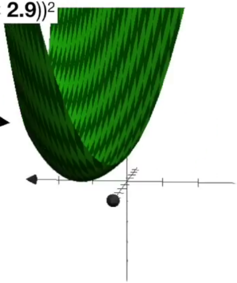
Gradiente Descendiente

Sum of squared residuals = (**1.4** - (intercept + slope × **0.5**))^2

+ (**1.9** - (intercept + slope × **2.3**))^2

+ (**3.2** - (intercept + slope × **2.9**))^2

This is a 3-D graph of the **Loss Function** for different values for the **Intercept** and the **Slope**



ehinojosa@unsa.edu.pe

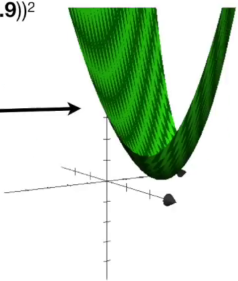
Gradiente Descendiente

Sum of squared residuals = (**1.4** - (intercept + slope × **0.5**))²

+ (**1.9** - (intercept + slope × **2.3**))²

+ (**3.2** - (intercept + slope × **2.9**))²

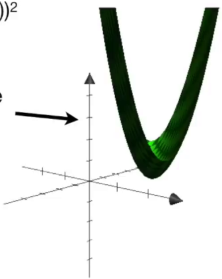
This is a 3-D graph of the **Loss Function** for different values for the **Intercept** and the **Slope**



Gradiente Descendiente

$$\text{Sum of squared residuals} = (\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2$$
$$+ (\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))^2$$
$$+ (\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))^2$$

This axis is the Sum of the Squared Residuals…



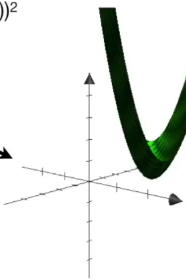
Gradiente Descendiente

Sum of squared residuals = (**1.4** - (intercept + slope × **0.5**))^2

+ (**1.9** - (intercept + slope × **2.3**))^2

+ (**3.2** - (intercept + slope × **2.9**))^2

...this axis represents different values for the **Slope**...



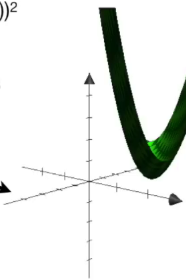
Gradiente Descendiente

Sum of squared residuals = (**1.4** - (intercept + slope × **0.5**))^2

+ (**1.9** - (intercept + slope × **2.3**))^2

+ (**3.2** - (intercept + slope × **2.9**))^2

...and this axis represents different values for the **Intercept**.



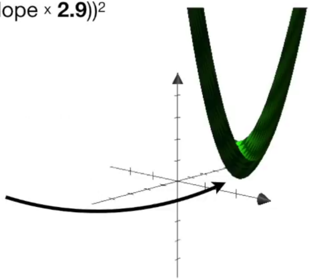
Gradiente Descendiente

Sum of squared residuals = (**1.4** - (intercept + slope × **0.5**))^2

+ (**1.9** - (intercept + slope × **2.3**))^2

+ (**3.2** - (intercept + slope × **2.9**))^2

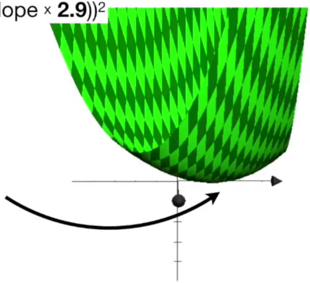
We want to find the values for the **Intercept** and **Slope** that give us the minimum Sum of the Squared Residuals.



Gradiente Descendiente

$$\text{Sum of squared residuals} = (\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2 + (\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))^2 + (\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))^2$$

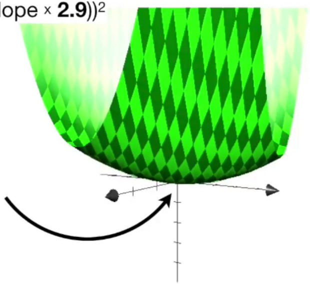
We want to find the values for the **Intercept** and **Slope** that give us the minimum Sum of the Squared Residuals.



Gradiente Descendiente

$$\text{Sum of squared residuals} = (\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2$$
$$+ (\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))^2$$
$$+ (\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))^2$$

We want to find the values for the **Intercept** and **Slope** that give us the minimum Sum of the Squared Residuals.



ehinojosa@unsa.edu.pe

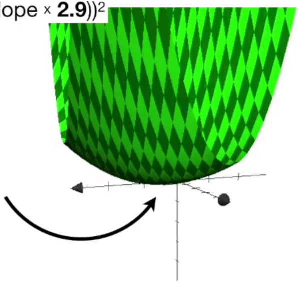
Gradiente Descendiente

Sum of squared residuals = (**1.4** - (intercept + slope × **0.5**))^2

+ (**1.9** - (intercept + slope × **2.3**))^2

+ (**3.2** - (intercept + slope × **2.9**))^2

We want to find the values for the **Intercept** and **Slope** that give us the minimum Sum of the Squared Residuals.



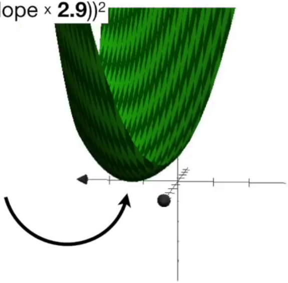
Gradiente Descendiente

Sum of squared residuals = (**1.4** - (intercept + slope × **0.5**))^2

+ (**1.9** - (intercept + slope × **2.3**))^2

+ (**3.2** - (intercept + slope × **2.9**))^2

We want to find the values for the **Intercept** and **Slope** that give us the minimum Sum of the Squared Residuals.



ehinojosa@unsa.edu.pe

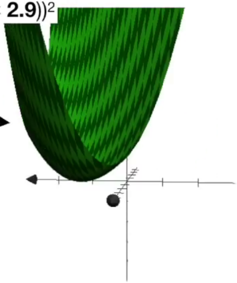
Gradiente Descendiente

Sum of squared residuals = (**1.4** - (intercept + slope × **0.5**))^2

+ (**1.9** - (intercept + slope × **2.3**))^2

+ (**3.2** - (intercept + slope × **2.9**))^2

This is a 3-D graph of the **Loss Function** for different values for the **Intercept** and the **Slope**



ehinojosa@unsa.edu.pe

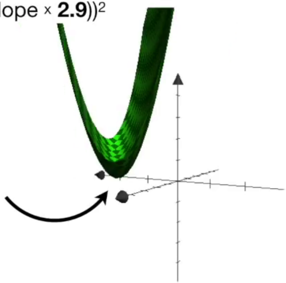
Gradiente Descendiente

Sum of squared residuals = (**1.4** - (intercept + slope × **0.5**))^2

+ (**1.9** - (intercept + slope × **2.3**))^2

+ (**3.2** - (intercept + slope × **2.9**))^2

We want to find the values for the **Intercept** and **Slope** that give us the minimum Sum of the Squared Residuals.

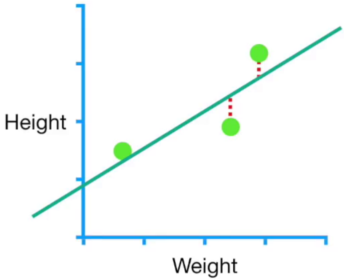


Gradiente Descendiente

Sum of squared residuals = (**1.4** - (intercept + slope × **0.5**))^2

+ (**1.9** - (intercept + slope × **2.3**))^2

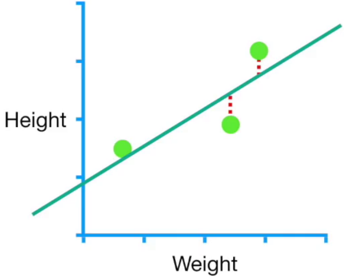
+ (**3.2** - (intercept + slope × **2.9**))^2



So, just like before, we need to take the derivative of this function...

Gradiente Descendiente

$$\text{Sum of squared residuals} = (\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2$$
$$+ (\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))^2$$
$$+ (\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))^2$$

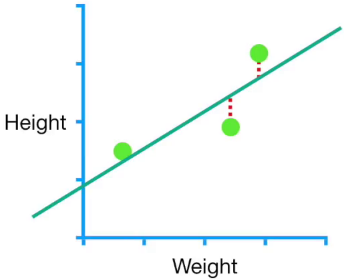


...and just like before, we'll take the derivative with respect to the **Intercept**...

$$\frac{d}{d\ intercept} \text{Sum of squared residuals}$$

Gradiente Descendiente

$$\text{Sum of squared residuals} = (\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2$$
$$+ (\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))^2$$
$$+ (\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))^2$$



…but unlike before, we'll also take the derivative with respect to the **Slope**!

$$\frac{d}{d\ \text{intercept}}\ \text{Sum of squared residuals}$$

$$\frac{d}{d\ \text{slope}}\ \text{Sum of squared residuals}$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

Sum of squared residuals = (**1.4** - (intercept + slope × **0.5**))^2

+ (**1.9** - (intercept + slope × **2.3**))^2

+ (**3.2** - (intercept + slope × **2.9**))^2

We'll start by taking the derivative with respect to the intercept.

$$\frac{d}{d\ intercept} \text{Sum of squared residuals} =$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\text{Sum of squared residuals} = \boxed{(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2}$$
$$+ (\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))^2$$
$$+ (\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))^2$$

Just like before, we take the derivative of each part...

$$\frac{d}{d\ \text{intercept}} \text{Sum of squared residuals} = \boxed{\frac{d}{d\ \text{intercept}} (\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2}$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

Sum of squared residuals = (**1.4** - (intercept + slope × **0.5**))^2

$\boxed{+ (\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))^2}$

+ (**3.2** - (intercept + slope × **2.9**))^2

Just like before, we take the derivative of each part...

$$\frac{d}{d\ intercept} \text{Sum of squared residuals} = \frac{d}{d\ intercept} (\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2$$

$$\boxed{+ \frac{d}{d\ intercept} (\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))^2}$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\text{Sum of squared residuals} = (\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2$$
$$+ (\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))^2$$
$$\boxed{+ (\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))^2}$$

Just like before, we take the derivative of each part...

$$\frac{d}{d\ intercept} \text{Sum of squared residuals} = \frac{d}{d\ intercept} (\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2$$
$$+ \frac{d}{d\ intercept} (\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))^2$$
$$\boxed{+ \frac{d}{d\ intercept} (\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))^2}$$

Gradiente Descendiente

$$\frac{d}{d\ intercept}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2$$

...and just like before,
we'll use...

$$\frac{d}{d\ intercept}\text{ Sum of squared residuals} = \boxed{\frac{d}{d\ intercept}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2}$$

$$+ \frac{d}{d\ intercept}(\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))^2$$

$$+ \frac{d}{d\ intercept}(\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))^2$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ intercept}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^{\boxed{2}} = 2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))$$

...and move the square to the front...

$$\frac{d}{d\ intercept}\text{Sum of squared residuals} = \frac{d}{d\ intercept}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2$$

$$+ \frac{d}{d\ intercept}(\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))^2$$

$$+ \frac{d}{d\ intercept}(\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))^2$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ intercept}\boxed{(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5})}^2 = 2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))\boxed{\times\ -1}$$

...and multiply that by the
derivative of the stuff
inside the parentheses.

Gradiente Descendiente

$$\frac{d}{d\ intercept}\boxed{(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5})}^2 = 2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))\boxed{\times\ -1}$$

$$\frac{d}{d\ intercept}\ \mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5})$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ intercept}\boxed{(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5})}^2 = 2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))\boxed{\times -1}$$

$$\downarrow$$

$$\frac{d}{d\ intercept}\ \mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5})$$

$$\downarrow$$

$$\frac{d}{d\ intercept}\ \mathbf{1.4} + (-1)\text{intercept} - \text{slope} \times \mathbf{0.5}$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ intercept}\boxed{(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2} = 2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))\boxed{\times -1}$$

$$\frac{d}{d\ intercept}\ \mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5})$$

Since we are taking the derivative with respect to the **Intercept**, we treat the **Slope** like a constant, and the derivative of a constant is **0**.

$$\frac{d}{d\ intercept}\ \mathbf{1.4} + (-1)\text{intercept} - \boxed{\text{slope}} \times \mathbf{0.5}$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ intercept}\boxed{(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5})}^2 = 2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))\boxed{\times -1}$$

$$\frac{d}{d\ intercept}\ \mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5})$$

$$\frac{d}{d\ intercept}\ \mathbf{1.4} + (-1)\text{intercept} - \text{slope} \times \mathbf{0.5} = -1$$

So we end up with **-1**, just like before.

Gradiente Descendiente

$$\frac{d}{d\ intercept}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2 = \boxed{2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5})) \times -1}$$

...then we simplify by multiplying 2 by -1...

$$\frac{d}{d\ intercept}\text{Sum of squared residuals} = \frac{d}{d\ intercept}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2$$

$$+ \frac{d}{d\ intercept}(\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))^2$$

$$+ \frac{d}{d\ intercept}(\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))^2$$

Gradiente Descendiente

$$\frac{d}{d\ intercept}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2 = 2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5})) \times -1$$

$$\boxed{= -2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))}$$

…then we simplify by multiplying 2 by -1…

$$\frac{d}{d\ intercept}\text{ Sum of squared residuals} = \frac{d}{d\ intercept}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2$$
$$+ \frac{d}{d\ intercept}(\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))^2$$
$$+ \frac{d}{d\ intercept}(\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))^2$$

Gradiente Descendiente

$$\frac{d}{d\ intercept}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2 = 2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5})) \times -1$$

$$\boxed{= -2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))}$$

…and this…

…is the derivative of the first part…

$$\frac{d}{d\ intercept}\text{Sum of squared residuals} = \boxed{\frac{d}{d\ intercept}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2}$$

$$+ \frac{d}{d\ intercept}(\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))^2$$

$$+ \frac{d}{d\ intercept}(\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))^2$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ intercept}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2 = 2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5})) \times -1$$

$$\boxed{= -2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))}$$

...and this...

...is the derivative of the first part...

$$\frac{d}{d\ intercept}\text{Sum of squared residuals} = \boxed{\frac{d}{d\ intercept}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2}$$

$$+ \frac{d}{d\ intercept}(\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))^2$$

$$+ \frac{d}{d\ intercept}(\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))^2$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ intercept}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2 = 2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5})) \times -1$$

$$\boxed{= -2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))}$$

...so we plug it in.

$$\frac{d}{d\ intercept}\text{Sum of squared residuals} = \boxed{-2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))}$$

$$+ \frac{d}{d\ intercept}(\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))^2$$

$$+ \frac{d}{d\ intercept}(\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))^2$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

Likewise, we replace these terms with their derivatives...

$$\frac{d}{d\ intercept}\text{ Sum of squared residuals} = -2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))$$

$$\boxed{\begin{aligned} &+ \frac{d}{d\ intercept}(\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))^2 \\ &+ \frac{d}{d\ intercept}(\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))^2 \end{aligned}}$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

Likewise, we replace these terms with their derivatives…

$$\frac{d}{d\ intercept}\text{Sum of squared residuals} = -2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))$$

$$\boxed{\begin{array}{l} + -2(\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))^2 \\ \\ + -2(\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))^2 \end{array}}$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

...and this whole thing is the derivative of the Sum of the Squared Residuals with respect to the **Intercept**.

$$\boxed{\frac{d}{d\ \textit{intercept}}\ \text{Sum of squared residuals}} = -2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))$$

$$+ -2(\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))^2$$

$$+ -2(\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))^2$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\text{Sum of squared residuals} = (\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2 + (\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))^2 + (\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))^2$$

Now let's take the derivative of the Sum of the Squared Residuals with respect to the **Slope**.

$$\frac{d}{d\,slope}\ \text{Sum of squared residuals} =$$

Gradiente Descendiente

$$\text{Sum of squared residuals} = \boxed{(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2}$$
$$+ (\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))^2$$
$$+ (\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))^2$$

Just like before, we take the derivative of each part...

$$\frac{d}{d\,slope} \text{Sum of squared residuals} = \boxed{\frac{d}{d\,slope} (\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2}$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

Sum of squared residuals = (**1.4** - (intercept + slope × **0.5**))^2

$\boxed{+ (\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))^2}$

+ (**3.2** - (intercept + slope × **2.9**))^2

Just like before, we take the derivative of each part...

$$\frac{d}{d\,slope}\text{ Sum of squared residuals} = \frac{d}{d\,slope}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2$$

$$\boxed{+ \frac{d}{d\,slope}(\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))^2}$$

Gradiente Descendiente

Sum of squared residuals = (**1.4** - (intercept + slope × **0.5**))^2

+ (**1.9** - (intercept + slope × **2.3**))^2

$\boxed{+ (\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))^2}$

Just like before, we take the derivative of each part…

$$\frac{d}{d\ slope}\ \text{Sum of squared residuals} = \frac{d}{d\ slope}\ (\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2$$

$$+ \frac{d}{d\ slope}\ (\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))^2$$

$$\boxed{+ \frac{d}{d\ slope}\ (\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))^2}$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\,slope}\,(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2$$

…and just like before, we'll use…

$$\frac{d}{d\,slope}\ \text{Sum of squared residuals} = \boxed{\frac{d}{d\,slope}\,(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2}$$

$$+ \frac{d}{d\,slope}\,(\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))^2$$

$$+ \frac{d}{d\,slope}\,(\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))^2$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ slope}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^{\boxed{2}} = 2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))$$

...to move the square to the front...

$$\frac{d}{d\ slope}\text{ Sum of squared residuals} = \frac{d}{d\ slope}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2$$

$$+ \frac{d}{d\ slope}(\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))^2$$

$$+ \frac{d}{d\ slope}(\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))^2$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ slope}\ \boxed{(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5})}^2 = 2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))\ \boxed{\times\ \mathbf{-0.5}}$$

…and multiply that by the
derivative of the stuff
inside the parentheses.

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\,slope}\boxed{(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5})}^2 = 2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))\boxed{\times\ \mathbf{-0.5}}$$

$$\frac{d}{d\,slope}\ \mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5})$$

Gradiente Descendiente

$$\frac{d}{d\ slope}\boxed{(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5})}^2 = 2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))\boxed{\times\ \mathbf{-0.5}}$$

$$\frac{d}{d\ slope}\ \mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5})$$

$$\frac{d}{d\ slope}\ \mathbf{1.4} + (-1)\text{intercept} - \text{slope} \times \mathbf{0.5}$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ \mathit{slope}}\ \boxed{(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))}^2 = 2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))\ \boxed{\times\ \mathbf{-0.5}}$$

$$\frac{d}{d\ \mathit{slope}}\ \mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5})$$

Since we are taking the derivative with respect to the **Slope**, we treat the **Intercept** like a constant, and the derivative of a constant is **0**.

$$\frac{d}{d\ \mathit{slope}}\ \mathbf{1.4} + \boxed{(-1)\text{intercept}} - \text{slope} \times \mathbf{0.5}$$

Gradiente Descendiente

$$\frac{d}{d\ slope}\boxed{(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5})}^2 = 2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))\boxed{\times\ \mathbf{-0.5}}$$

$$\frac{d}{d\ slope}\ \mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5})$$

So we end up with **-0.5**.

$$\frac{d}{d\ slope}\ \mathbf{1.4} + (-1)\text{intercept} - \text{slope} \times \mathbf{0.5} = \mathbf{-0.5}$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\,slope}\boxed{(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5})}^2 = 2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))\boxed{\times\ \mathbf{-0.5}}$$

$$\frac{d}{d\,slope}\ \mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5})$$

$$\frac{d}{d\,slope}\ \mathbf{1.4} + (-1)\text{intercept} - \text{slope} \times \mathbf{0.5} = \boxed{\mathbf{-0.5}}$$

Gradiente Descendiente

$$\frac{d}{d\ slope}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2 = \boxed{2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5})) \times \mathbf{-0.5}}$$

…then we simplify by moving the -0.5 to the front

$$\frac{d}{d\ slope}\text{ Sum of squared residuals} = \frac{d}{d\ slope}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2$$

$$+ \frac{d}{d\ slope}(\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))^2$$

$$+ \frac{d}{d\ slope}(\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))^2$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ slope}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2 = 2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5})) \times \mathbf{-0.5}$$

$$\boxed{= -2 \times \mathbf{0.5}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))}$$

…then we simplify by moving the -0.5 to the front

$$\frac{d}{d\ slope}\text{Sum of squared residuals} = \frac{d}{d\ slope}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2$$

$$+ \frac{d}{d\ slope}(\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))^2$$

$$+ \frac{d}{d\ slope}(\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))^2$$

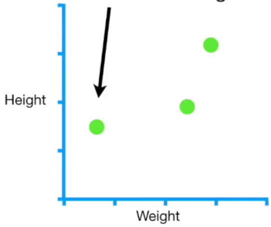
ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ slope}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2 = 2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5})) \times -\mathbf{0.5}$$

$$\boxed{= -2 \times \mathbf{0.5}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))}$$

NOTE: I left the **0.5** in bold instead of multiplying it by 2 to remind us that **0.5** is the weight for the first sample.



Gradiente Descendiente

$$\frac{d}{d\ slope}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2 = 2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5})) \times \mathbf{-0.5}$$

$$\boxed{= -2 \times \mathbf{0.5}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))}$$

...and this...

$$\frac{d}{d\ slope}\text{Sum of squared residuals} = \frac{d}{d\ slope}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2$$

$$+ \frac{d}{d\ slope}(\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))^2$$

$$+ \frac{d}{d\ slope}(\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))^2$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\,slope}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2 = 2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5})) \times -\mathbf{0.5}$$

$$\boxed{= -2 \times \mathbf{0.5}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))}$$

…and this…

…is the derivative of the first part…

$$\frac{d}{d\,slope}\text{ Sum of squared residuals} = \boxed{\frac{d}{d\,slope}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2}$$

$$+ \frac{d}{d\,slope}(\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))^2$$

$$+ \frac{d}{d\,slope}(\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))^2$$

Gradiente Descendiente

$$\frac{d}{d\ slope}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2 = 2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5})) \times \mathbf{-0.5}$$

$$\boxed{= -2 \times \mathbf{0.5}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))}$$

…so we plug it in.

$$\frac{d}{d\ slope}\text{Sum of squared residuals} = \boxed{\frac{d}{d\ slope}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))^2}$$

$$+ \frac{d}{d\ slope}(\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))^2$$

$$+ \frac{d}{d\ slope}(\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))^2$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

Likewise, we replace these terms with their derivatives.

$$\frac{d}{d\ slope}\text{ Sum of squared residuals} = -2 \times \mathbf{0.5}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))$$

$$\boxed{\begin{aligned} &+ \frac{d}{d\ slope}(\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))^2 \\ &+ \frac{d}{d\ slope}(\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))^2 \end{aligned}}$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

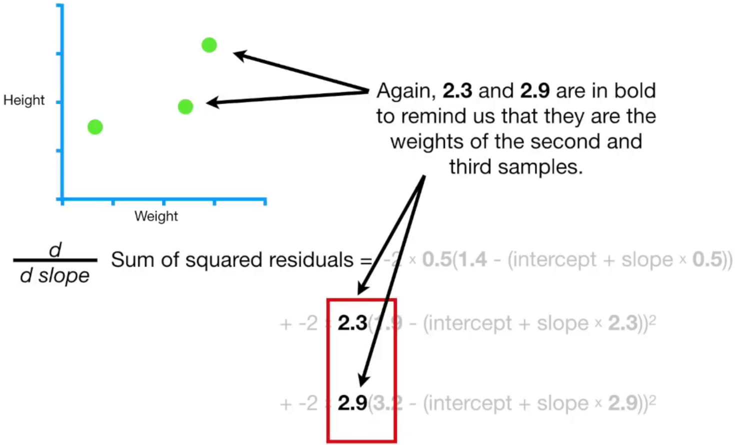
Likewise, we replace these terms with their derivatives.

$$\frac{d}{d\,slope}\ \text{Sum of squared residuals} = -2 \times \mathbf{0.5}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))$$

$$\boxed{\begin{array}{l} + -2 \times \mathbf{2.3}(\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))^2 \\ \\ + -2 \times \mathbf{2.9}(\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))^2 \end{array}}$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente



Gradiente Descendiente

$$\frac{d}{d\ intercept}\text{Sum of squared residuals} =$$
$$-2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))$$
$$+ -2(\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))$$
$$+ -2(\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))$$

Here's the derivative of the Sum of the Squared Residuals with respect to the **Intercept**...

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ \textit{intercept}} \text{ Sum of squared residuals} = -2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5})) + -2(\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3})) + -2(\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))$$

Here's the derivative of the Sum of the Squared Residuals with respect to the **Intercept**...

...and here's the derivative with respect to the **Slope**.

$$\frac{d}{d\ \textit{slope}} \text{ Sum of squared residuals} = -2 \times \mathbf{0.5}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5})) + -2 \times \mathbf{2.9}(\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9})) + -2 \times \mathbf{2.3}(\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ intercept} \text{Sum of squared residuals} = -2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5})) + -2(\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3})) + -2(\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))$$

NOTE: When you have two or more derivatives of the same function, they are called a **Gradient**.

$$\frac{d}{d\ slope} \text{Sum of squared residuals} = -2 \times \mathbf{0.5}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5})) + -2 \times \mathbf{2.9}(\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9})) + -2 \times \mathbf{2.3}(\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ intercept} \text{ Sum of squared residuals} =$$
$$-2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))$$
$$+ -2(\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))$$
$$+ -2(\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))$$

We will use this **Gradient** to **descend** to lowest point in the **Loss Function**, which, in this case, is the Sum of the Squared Residuals...

$$\frac{d}{d\ slope} \text{ Sum of squared residuals} =$$
$$-2 \times \mathbf{0.5}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))$$
$$+ -2 \times \mathbf{2.9}(\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))$$
$$+ -2 \times \mathbf{2.3}(\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))$$

Gradiente Descendiente

$$\frac{d}{d\ intercept} \text{ Sum of squared residuals} = -2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5})) + -2(\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3})) + -2(\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))$$

We will use this **Gradient** to **descend** to lowest point in the **Loss Function**, which, in this case, is the Sum of the Squared Residuals…

…thus, this is why this algorithm is called **Gradient Descent**!

$$\frac{d}{d\ slope} \text{ Sum of squared residuals} = -2 \times \mathbf{0.5}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5})) + -2 \times \mathbf{2.9}(\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9})) + -2 \times \mathbf{2.3}(\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ intercept}\ \text{Sum of squared residuals} =$$
$$-2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))$$
$$+ -2(\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))$$
$$+ -2(\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))$$

Just like before, we will start by picking a random number for the **Intercept**. In this case we'll set the **Intercept** = **0**…

$$\frac{d}{d\ slope}\ \text{Sum of squared residuals} =$$
$$-2 \times \mathbf{0.5}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))$$
$$+ -2 \times \mathbf{2.9}(\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))$$
$$+ -2 \times \mathbf{2.3}(\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ intercept} \text{ Sum of squared residuals} = -2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5})) + -2(\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3})) + -2(\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))$$

Just like before, we will start by picking a random number for the **Intercept**. In this case we'll set the **Intercept** = **0**…

…and we'll pick a random number for the **Slope**. In this case we'll set the **Slope** = **1**.

$$\frac{d}{d\ slope} \text{ Sum of squared residuals} = -2 \times \mathbf{0.5}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5})) + -2 \times \mathbf{2.9}(\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9})) + -2 \times \mathbf{2.3}(\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$\frac{d}{d\ intercept}$ Sum of squared residuals =

-2(**1.4** - (intercept + slope × **0.5**))

+ -2(**1.9** - (intercept + slope × **2.3**))

+ -2(**3.2** - (intercept + slope × **2.9**))

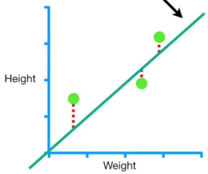
$\frac{d}{d\ slope}$ Sum of squared residuals =

-2 × **0.5**(**1.4** - (intercept + slope × **0.5**))

+ -2 × **2.9**(**3.2** - (intercept + slope × **2.9**))

+ -2 × **2.3**(**1.9** - (intercept + slope × **2.3**))

Thus, this line, with **Intercept** = **0** and **Slope** = **1**, is where we will start.



Gradiente Descendiente

$$\frac{d}{d\ intercept}\ \text{Sum of squared residuals} =$$
$$-2(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))$$
$$+ -2(\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))$$
$$+ -2(\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))$$

Now let's plug in **0** for the **Intercept** and **1** for the **Slope**...

$$\frac{d}{d\ slope}\ \text{Sum of squared residuals} =$$
$$-2 \times \mathbf{0.5}(\mathbf{1.4} - (\text{intercept} + \text{slope} \times \mathbf{0.5}))$$
$$+ -2 \times \mathbf{2.9}(\mathbf{3.2} - (\text{intercept} + \text{slope} \times \mathbf{2.9}))$$
$$+ -2 \times \mathbf{2.3}(\mathbf{1.9} - (\text{intercept} + \text{slope} \times \mathbf{2.3}))$$

Gradiente Descendiente

$$
\begin{aligned}
\frac{d}{d\ \mathit{intercept}}\ \text{Sum of squared residuals} = \; & -2(\mathbf{1.4} - (0 + 1 \times \mathbf{0.5})) \\
& + -2(\mathbf{1.9} - (0 + 1 \times \mathbf{2.3})) \\
& + -2(\mathbf{3.2} - (0 + 1 \times \mathbf{2.9})) \ \boxed{= \mathbf{-1.6}}
\end{aligned}
$$

...and that gives us two **Slopes**...

$$
\begin{aligned}
\frac{d}{d\ \mathit{slope}}\ \text{Sum of squared residuals} = \; & -2 \times \mathbf{0.5}(\mathbf{1.4} - (0 + 1 \times \mathbf{0.5})) \\
& + -2 \times \mathbf{2.9}(\mathbf{3.2} - (0 + 1 \times \mathbf{2.9})) \\
& + -2 \times \mathbf{2.3}(\mathbf{1.9} - (0 + 1 \times \mathbf{2.3})) \ \boxed{= \mathbf{-0.8}}
\end{aligned}
$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ intercept} \text{ Sum of squared residuals} = -2(\mathbf{1.4} - (0 + 1 \times \mathbf{0.5})) + -2(\mathbf{1.9} - (0 + 1 \times \mathbf{2.3})) + -2(\mathbf{3.2} - (0 + 1 \times \mathbf{2.9})) \boxed{= \mathbf{-1.6}}$$

Step Size$_{\text{Intercept}}$ = **Slope** × **Learning Rate**

...now we plug the **Slopes** into the **Step Size** formulas...

$$\frac{d}{d\ slope} \text{ Sum of squared residuals} = -2 \times \mathbf{0.5}(\mathbf{1.4} - (0 + 1 \times \mathbf{0.5})) + -2 \times \mathbf{2.9}(\mathbf{3.2} - (0 + 1 \times \mathbf{2.9})) + -2 \times \mathbf{2.3}(\mathbf{1.9} - (0 + 1 \times \mathbf{2.3})) \boxed{= \mathbf{-0.8}}$$

Step Size$_{\text{Slope}}$ = **Slope** × **Learning Rate**

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ intercept}\text{ Sum of squared residuals} = -2(\mathbf{1.4} - (0 + 1 \times \mathbf{0.5})) + -2(\mathbf{1.9} - (0 + 1 \times \mathbf{2.3})) + -2(\mathbf{3.2} - (0 + 1 \times \mathbf{2.9})) \boxed{= \mathbf{-1.6}}$$

Step Size$_{\text{Intercept}}$ = -1.6 × **Learning Rate**

...now we plug the **Slopes** into the **Step Size** formulas...

$$\frac{d}{d\ slope}\text{ Sum of squared residuals} = -2 \times \mathbf{0.5}(\mathbf{1.4} - (0 + 1 \times \mathbf{0.5})) + -2 \times \mathbf{2.9}(\mathbf{3.2} - (0 + 1 \times \mathbf{2.9})) + -2 \times \mathbf{2.3}(\mathbf{1.9} - (0 + 1 \times \mathbf{2.3})) \boxed{= \mathbf{-0.8}}$$

Step Size$_{\text{Slope}}$ = -0.8 × **Learning Rate**

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ intercept}\text{ Sum of squared residuals} = -2(\mathbf{1.4} - (0 + 1 \times \mathbf{0.5})) + -2(\mathbf{1.9} - (0 + 1 \times \mathbf{2.3})) + -2(\mathbf{3.2} - (0 + 1 \times \mathbf{2.9})) = \mathbf{-1.6}$$

Step Size$_{\text{Intercept}}$ = -1.6 × **Learning Rate**

…and multiply by the **Learning Rate**, which this time we set to **0.01**…

$$\frac{d}{d\ slope}\text{ Sum of squared residuals} = -2 \times \mathbf{0.5}(\mathbf{1.4} - (0 + 1 \times \mathbf{0.5})) + -2 \times \mathbf{2.9}(\mathbf{3.2} - (0 + 1 \times \mathbf{2.9})) + -2 \times \mathbf{2.3}(\mathbf{1.9} - (0 + 1 \times \mathbf{2.3})) = \mathbf{-0.8}$$

Step Size$_{\text{Slope}}$ = -0.8 × **Learning Rate**

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$\frac{d}{d\ intercept}$ Sum of squared residuals =

$-2(\mathbf{1.4} - (0 + 1 \times \mathbf{0.5}))$

$+ -2(\mathbf{1.9} - (0 + 1 \times \mathbf{2.3}))$

$+ -2(\mathbf{3.2} - (0 + 1 \times \mathbf{2.9})) = \mathbf{-1.6}$

Step Size$_{\text{Intercept}} = -1.6 \times 0.01$

…and multiply by the **Learning Rate**, which this time we set to **0.01**…

Step Size$_{\text{Slope}} = -0.8 \times 0.01$

$\frac{d}{d\ slope}$ Sum of squared residuals =

$-2 \times \mathbf{0.5}(\mathbf{1.4} - (0 + 1 \times \mathbf{0.5}))$

$+ -2 \times \mathbf{2.9}(\mathbf{3.2} - (0 + 1 \times \mathbf{2.9}))$

$+ -2 \times \mathbf{2.3}(\mathbf{1.9} - (0 + 1 \times \mathbf{2.3})) = \mathbf{-0.8}$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$\frac{d}{d\ intercept}$ Sum of squared residuals =

-2(**1.4** - (0 + 1 × **0.5**))

+ -2(**1.9** - (0 + 1 × **2.3**))

+ -2(**3.2** - (0 + 1 × **2.9**)) = **-1.6**

Step Size$_{\text{Intercept}}$ = -1.6 × 0.01

NOTE: The larger **Learning Rate** that we used in the first example doesn't work this time. Even after a bunch of steps, **Gradient Descent** doesn't arrive at the correct answer.

$\frac{d}{d\ slope}$ Sum of squared residuals =

-2 × **0.5**(**1.4** - (0 + 1 × **0.5**))

+ -2 × **2.9**(**3.2** - (0 + 1 × **2.9**))

+ -2 × **2.3**(**1.9** - (0 + 1 × **2.3**)) = **-0.8**

Step Size$_{\text{Slope}}$ = -0.8 × 0.01

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ intercept} \text{ Sum of squared residuals} = -2(\mathbf{1.4} - (0 + 1 \times \mathbf{0.5})) + -2(\mathbf{1.9} - (0 + 1 \times \mathbf{2.3})) + -2(\mathbf{3.2} - (0 + 1 \times \mathbf{2.9})) = \mathbf{-1.6}$$

Step Size$_{\text{Intercept}}$ = -1.6 × 0.01

This means that **Gradient Descent** can be very sensitive to the **Learning Rate**.

$$\frac{d}{d\ slope} \text{ Sum of squared residuals} = -2 \times \mathbf{0.5}(\mathbf{1.4} - (0 + 1 \times \mathbf{0.5})) + -2 \times \mathbf{2.9}(\mathbf{3.2} - (0 + 1 \times \mathbf{2.9})) + -2 \times \mathbf{2.3}(\mathbf{1.9} - (0 + 1 \times \mathbf{2.3})) = \mathbf{-0.8}$$

Step Size$_{\text{Slope}}$ = -0.8 × 0.01

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ intercept}\ \text{Sum of squared residuals} = -2(\mathbf{1.4} - (0 + 1 \times \mathbf{0.5})) + -2(\mathbf{1.9} - (0 + 1 \times \mathbf{2.3})) + -2(\mathbf{3.2} - (0 + 1 \times \mathbf{2.9})) = \mathbf{-1.6}$$

Step Size$_{\text{Intercept}}$ = -1.6 × 0.01

The good news is that in practice, a reasonable **Learning Rate** can be determined automatically by starting large and getting smaller with each step.

$$\frac{d}{d\ slope}\ \text{Sum of squared residuals} = -2 \times \mathbf{0.5}(\mathbf{1.4} - (0 + 1 \times \mathbf{0.5})) + -2 \times \mathbf{2.9}(\mathbf{3.2} - (0 + 1 \times \mathbf{2.9})) + -2 \times \mathbf{2.3}(\mathbf{1.9} - (0 + 1 \times \mathbf{2.3})) = \mathbf{-0.8}$$

Step Size$_{\text{Slope}}$ = -0.8 × 0.01

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ \mathit{intercept}} \text{ Sum of squared residuals} = -2(\mathbf{1.4} - (0 + 1 \times \mathbf{0.5})) + -2(\mathbf{1.9} - (0 + 1 \times \mathbf{2.3})) + -2(\mathbf{3.2} - (0 + 1 \times \mathbf{2.9})) = \mathbf{-1.6}$$

Step Size$_{\text{Intercept}}$ = -1.6 × 0.01

So, in theory, you shouldn't have to worry too much about the **Learning Rate**.

$$\frac{d}{d\ \mathit{slope}} \text{ Sum of squared residuals} = -2 \times \mathbf{0.5}(\mathbf{1.4} - (0 + 1 \times \mathbf{0.5})) + -2 \times \mathbf{2.9}(\mathbf{3.2} - (0 + 1 \times \mathbf{2.9})) + -2 \times \mathbf{2.3}(\mathbf{1.9} - (0 + 1 \times \mathbf{2.3})) = \mathbf{-0.8}$$

Step Size$_{\text{Slope}}$ = -0.8 × 0.01

Gradiente Descendiente

$$\frac{d}{d\ intercept}\ \text{Sum of squared residuals} = -2(\mathbf{1.4} - (0 + 1 \times \mathbf{0.5})) + -2(\mathbf{1.9} - (0 + 1 \times \mathbf{2.3})) + -2(\mathbf{3.2} - (0 + 1 \times \mathbf{2.9})) = \mathbf{-1.6}$$

Step Size$_{\text{Intercept}}$ = -1.6 × 0.01 = **-0.016**

Anyway, we do the math and get two **Step Sizes**.

$$\frac{d}{d\ slope}\ \text{Sum of squared residuals} = -2 \times \mathbf{0.5}(\mathbf{1.4} - (0 + 1 \times \mathbf{0.5})) + -2 \times \mathbf{2.9}(\mathbf{3.2} - (0 + 1 \times \mathbf{2.9})) + -2 \times \mathbf{2.3}(\mathbf{1.9} - (0 + 1 \times \mathbf{2.3})) = \mathbf{-0.8}$$

Step Size$_{\text{Slope}}$ = -0.8 × 0.01 = **-0.008**

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ intercept}\ \text{Sum of squared residuals} = -2(\mathbf{1.4} - (0 + 1 \times \mathbf{0.5})) + -2(\mathbf{1.9} - (0 + 1 \times \mathbf{2.3})) + -2(\mathbf{3.2} - (0 + 1 \times \mathbf{2.9})) = \mathbf{-1.6}$$

Step Size$_{\text{Intercept}}$ = -1.6 × 0.01 = **-0.016**

New Intercept = **Old Intercept** - **Step Size**

Now we calculate the **New Intercept** and **New Slope** by plugging in the **Old Intercept** and the **Old Slope**…

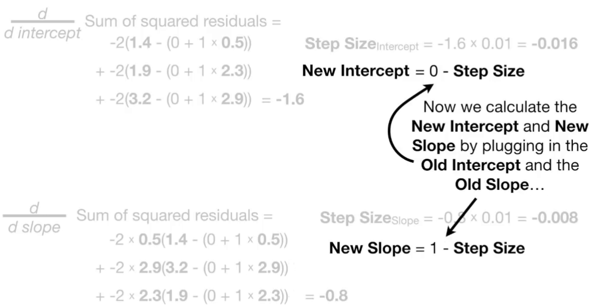
$$\frac{d}{d\ slope}\ \text{Sum of squared residuals} = -2 \times \mathbf{0.5}(\mathbf{1.4} - (0 + 1 \times \mathbf{0.5})) + -2 \times \mathbf{2.9}(\mathbf{3.2} - (0 + 1 \times \mathbf{2.9})) + -2 \times \mathbf{2.3}(\mathbf{1.9} - (0 + 1 \times \mathbf{2.3})) = \mathbf{-0.8}$$

Step Size$_{\text{Slope}}$ = -0.8 × 0.01 = **-0.008**

New Slope = **Old Slope** - **Step Size**

ehinojosa@unsa.edu.pe

Gradiente Descendiente



ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ intercept}\text{ Sum of squared residuals} = -2(\mathbf{1.4} - (0 + 1 \times \mathbf{0.5})) + -2(\mathbf{1.9} - (0 + 1 \times \mathbf{2.3})) + -2(\mathbf{3.2} - (0 + 1 \times \mathbf{2.9})) = \mathbf{-1.6}$$

$$\textbf{Step Size}_{\text{Intercept}} = -1.6 \times 0.01 = \boxed{\mathbf{-0.016}}$$

$$\textbf{New Intercept} = 0 - \textbf{Step Size}$$

…and the
Step Sizes…

$$\frac{d}{d\ slope}\text{ Sum of squared residuals} = -2 \times \mathbf{0.5}(\mathbf{1.4} - (0 + 1 \times \mathbf{0.5})) + -2 \times \mathbf{2.9}(\mathbf{3.2} - (0 + 1 \times \mathbf{2.9})) + -2 \times \mathbf{2.3}(\mathbf{1.9} - (0 + 1 \times \mathbf{2.3})) = \mathbf{-0.8}$$

$$\textbf{Step Size}_{\text{Slope}} = -0.8 \times 0.01 = \boxed{\mathbf{-0.008}}$$

$$\textbf{New Slope} = 1 - \textbf{Step Size}$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$$\frac{d}{d\ intercept}\ \text{Sum of squared residuals} = -2(\mathbf{1.4} - (0 + 1 \times \mathbf{0.5})) + -2(\mathbf{1.9} - (0 + 1 \times \mathbf{2.3})) + -2(\mathbf{3.2} - (0 + 1 \times \mathbf{2.9})) = \mathbf{-1.6}$$

Step Size$_{\text{Intercept}}$ = -1.6 × 0.01 = $\boxed{\mathbf{-0.016}}$

New Intercept = 0 - (-0.016)

…and the **Step Sizes**…

$$\frac{d}{d\ slope}\ \text{Sum of squared residuals} = -2 \times \mathbf{0.5}(\mathbf{1.4} - (0 + 1 \times \mathbf{0.5})) + -2 \times \mathbf{2.9}(\mathbf{3.2} - (0 + 1 \times \mathbf{2.9})) + -2 \times \mathbf{2.3}(\mathbf{1.9} - (0 + 1 \times \mathbf{2.3})) = \mathbf{-0.8}$$

Step Size$_{\text{Slope}}$ = -0.8 × 0.01 = $\boxed{\mathbf{-0.008}}$

New Slope = 1 - (-0.008)

ehinojosa@unsa.edu.pe

Gradiente Descendiente

$\frac{d}{d\ intercept}$ Sum of squared residuals =

-2(**1.4** - (0 + 1 × **0.5**))

+ -2(**1.9** - (0 + 1 × **2.3**))

+ -2(**3.2** - (0 + 1 × **2.9**)) = **-1.6**

Step Size$_{\text{Intercept}}$ = -1.6 × 0.01 = **-0.016**

New Intercept = 0 - (-0.016) = 0.016

...and we end up with a **New Intercept** and a **New Slope**.

$\frac{d}{d\ slope}$ Sum of squared residuals =

-2 × **0.5**(**1.4** - (0 + 1 × **0.5**))

+ -2 × **2.9**(**3.2** - (0 + 1 × **2.9**))

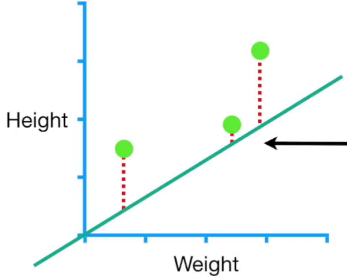
+ -2 × **2.3**(**1.9** - (0 + 1 × **2.3**)) = **-0.8**

Step Size$_{\text{Slope}}$ = -0.8 × 0.01 = **-0.008**

New Slope = 1 - (-0.008) = 1.008

ehinojosa@unsa.edu.pe

Gradiente Descendiente

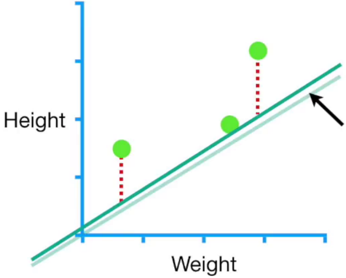


New Intercept = 0 - (-0.016) = 0.016

This is the line we started with... (**Slope = 1** and **Intercept = 0**)

New Slope = 1 - (-0.008) = 1.008

Gradiente Descendiente

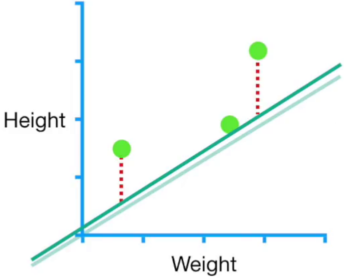


New Intercept = 0 - (-0.016) = 0.016

...and this is the new line (with **Slope = 1.008** and **Intercept = 0.016**) after the first step.

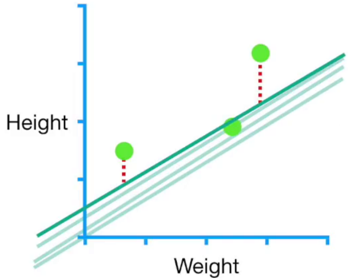
New Slope = 1 - (-0.008) = 1.008

Gradiente Descendiente



Now we just repeat what we did until all of the **Steps Sizes** are very small or we reach the **Maximum Number of Steps**.

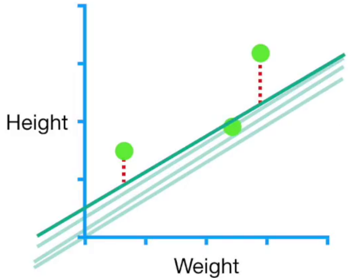
Gradiente Descendiente



Now we just repeat what we did until all of the **Steps Sizes** are very small or we reach the **Maximum Number of Steps**.

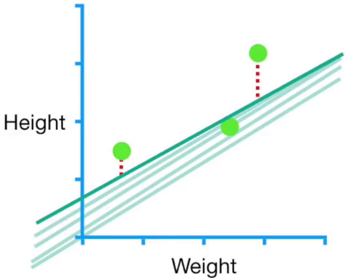
ehinojosa@unsa.edu.pe

Gradiente Descendiente



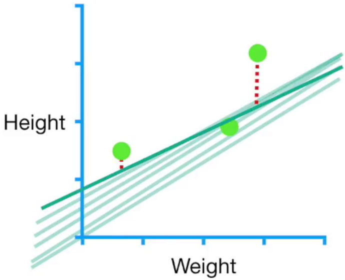
Now we just repeat what we did until all of the **Steps Sizes** are very small or we reach the **Maximum Number of Steps**.

Gradiente Descendiente



Now we just repeat what we did until all of the **Steps Sizes** are very small or we reach the **Maximum Number of Steps**.

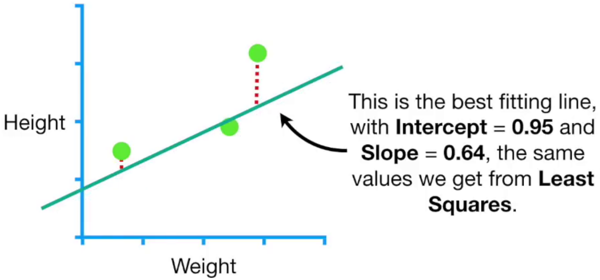
Gradiente Descendiente



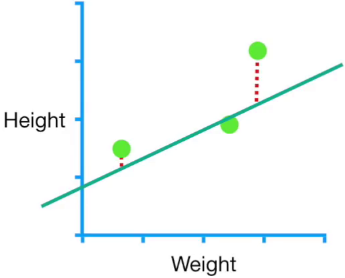
Now we just repeat what we did until all of the **Steps Sizes** are very small or we reach the **Maximum Number of Steps**.

ehinojosa@unsa.edu.pe

Gradiente Descendiente

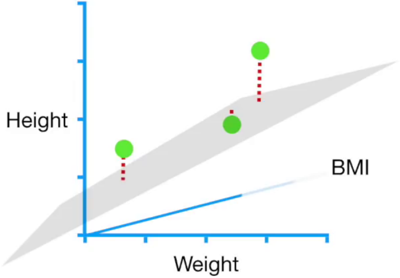


Gradiente Descendiente



We now know how **Gradient Descent** optimizes two parameters, the **Slope** and **Intercept**.

Gradiente Descendiente



If we had more parameters, then we'd just take more derivatives and everything else stays the same.

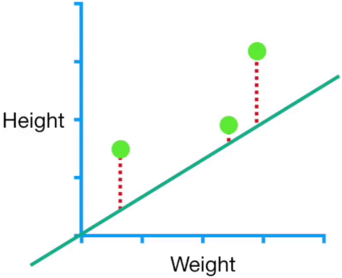
Gradiente Descendiente

Sum of squared residuals = (**1.4** - (intercept + 0.64 × **0.5**))²

+ (**1.9** - (intercept + 0.64 × **2.3**))²

+ (**3.2** - (intercept + 0.64 × **2.9**))²

NOTE: The Sum of the Squared Residuals is just one type of **Loss Function.**



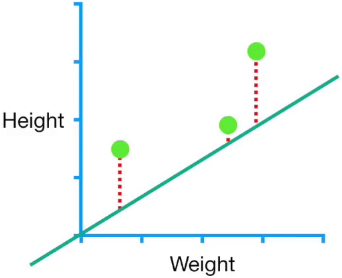
Gradiente Descendiente

Sum of squared residuals = (**1.4** - (intercept + 0.64 × **0.5**))^2

+ (**1.9** - (intercept + 0.64 × **2.3**))^2

+ (**3.2** - (intercept + 0.64 × **2.9**))^2

However, there are tons of other **Loss Functions** that work with other types of data.



Gradiente Descendiente

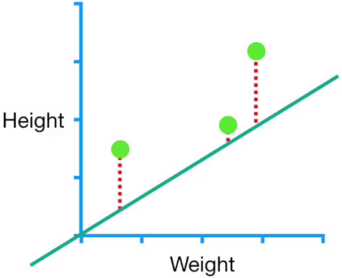
Sum of squared residuals = (**1.4** - (intercept + 0.64 × **0.5**))^2

+ (**1.9** - (intercept + 0.64 × **2.3**))^2

+ (**3.2** - (intercept + 0.64 × **2.9**))^2

However, there are tons of other **Loss Functions** that work with other types of data.

Regardless of which **Loss Function** you use, **Gradient Descent** works the same way.



Gradiente Descendiente

$\mathbf{w} \leftarrow$ any point in the parameter space

while not converged **do**

  for each w_i **in w do**

$$w_i \leftarrow w_i - \alpha \frac{\partial}{\partial w_i} Loss\left(\mathbf{w}\right)$$

$$\frac{\partial}{\partial w_0} Loss(\mathbf{w}) = -2\left(y - h_{\mathbf{w}}\left(x\right)\right); \qquad \frac{\partial}{\partial w_1} Loss\left(\mathbf{w}\right) = -2\left(y - h_{\mathbf{w}}\left(x\right)\right) \times x.$$

ehinojosa@unsa.edu.pe

Gradiente Descendiente

```
Iteración 1
Intercepto anterior = 0.0
Pendiente anterior = 1.0
Tasa de aprendizaje = 0.01
Derivada intercepto = -1.6
Derivada pendiente = -0.8
Intercepto nuevo = 0.016
Pendiente nueva = 1.008

Iteración 2
Intercepto anterior = 0.016
Pendiente anterior = 1.008
Tasa de aprendizaje = 0.01
Derivada intercepto = -1.4128
Derivada pendiente = -0.3944
Intercepto nuevo = 0.0301
Pendiente nueva = 1.0119

Iteración 3
Intercepto anterior = 0.030128
Pendiente anterior = 1.011944
Tasa de aprendizaje = 0.01
Derivada intercepto = -1.28307
Derivada pendiente = -0.123303
Intercepto nuevo = 0.043
Pendiente nueva = 1.0132

Iteración 4
Intercepto anterior = 0.042959
Pendiente anterior = 1.013177
Tasa de aprendizaje = 0.01
Derivada intercepto = -1.19203
Derivada pendiente = 0.057368
Intercepto nuevo = 0.0549
Pendiente nueva = 1.0126

Iteración 5
Intercepto anterior = 0.054879
Pendiente anterior = 1.012603
Tasa de aprendizaje = 0.01
Derivada intercepto = -1.127048
Derivada pendiente = 0.177254
Intercepto nuevo = 0.0661
Pendiente nueva = 1.0108
```

ehinojosa@unsa.edu.pe

Gradiente Descendiente

```
Pendiente anterior = 0.64103
Tasa de aprendizaje = 0.01
Derivada intercepto = -1.1E-5
Derivada pendiente = 5.0E-6
Intercepto nuevo = 0.9487
Pendiente nueva = 0.641

Iteración 997
Intercepto anterior = 0.948708
Pendiente anterior = 0.64103
Tasa de aprendizaje = 0.01
Derivada intercepto = -1.1E-5
Derivada pendiente = 5.0E-6
Intercepto nuevo = 0.9487
Pendiente nueva = 0.641

Iteración 998
Intercepto anterior = 0.948708
Pendiente anterior = 0.64103
Tasa de aprendizaje = 0.01
Derivada intercepto = -1.1E-5
Derivada pendiente = 5.0E-6
Intercepto nuevo = 0.9487
Pendiente nueva = 0.641

Iteración 999
Intercepto anterior = 0.948708
Pendiente anterior = 0.64103
Tasa de aprendizaje = 0.01
Derivada intercepto = -1.1E-5
Derivada pendiente = 5.0E-6
Intercepto nuevo = 0.9487
Pendiente nueva = 0.641

Iteración 1000
Intercepto anterior = 0.948708
Pendiente anterior = 0.64103
Tasa de aprendizaje = 0.01
Derivada intercepto = -1.1E-5
Derivada pendiente = 5.0E-6
Intercepto nuevo = 0.9487
Pendiente nueva = 0.641
```

ehinojosa@unsa.edu.pe


¡GRACIAS!