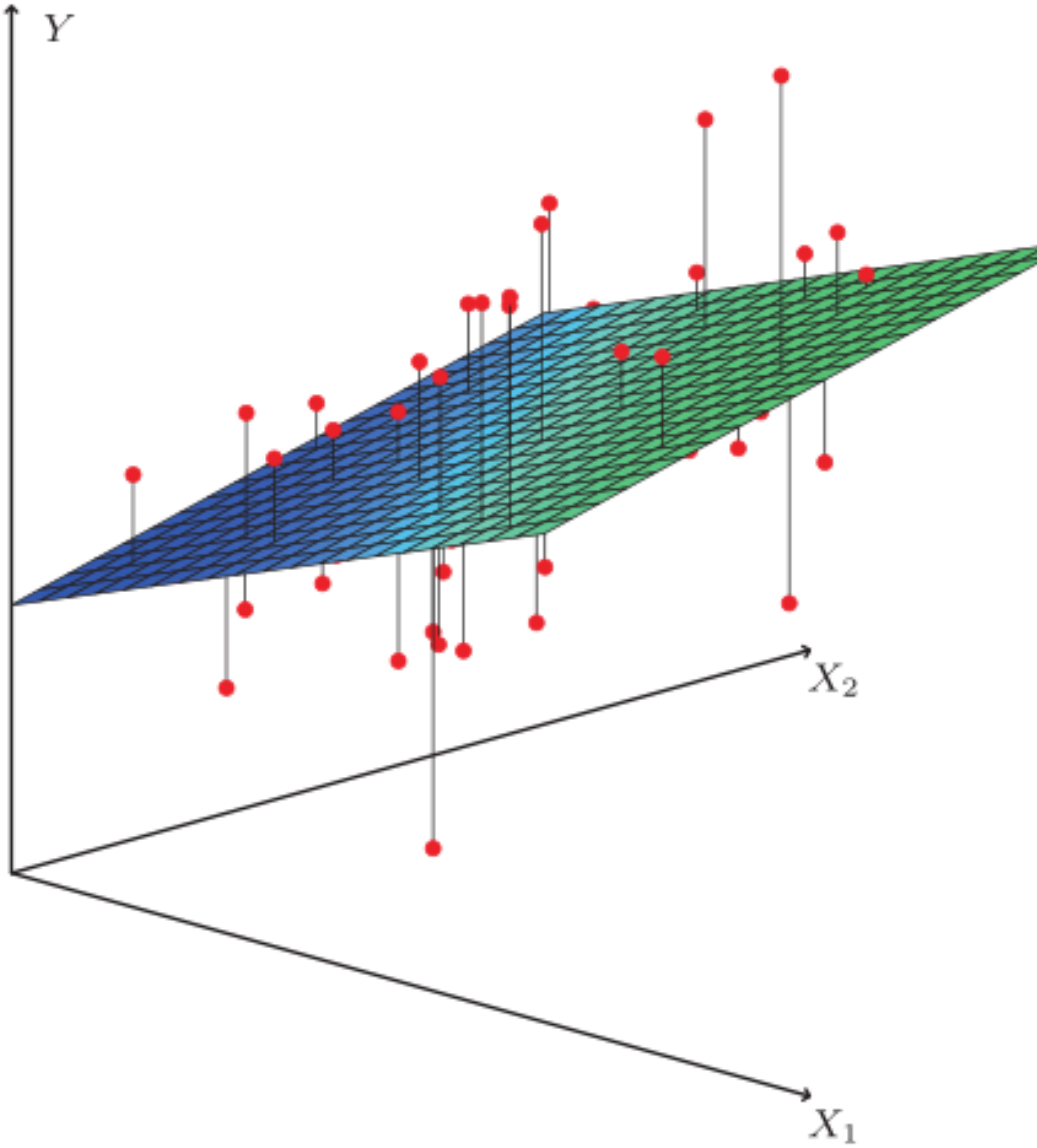
Curso Data Science

Listo para convertirte en el mejor Data Science de la Historia?

Emanuel Lemos | Let's coder

Regression Lineal

Simple y Multiple



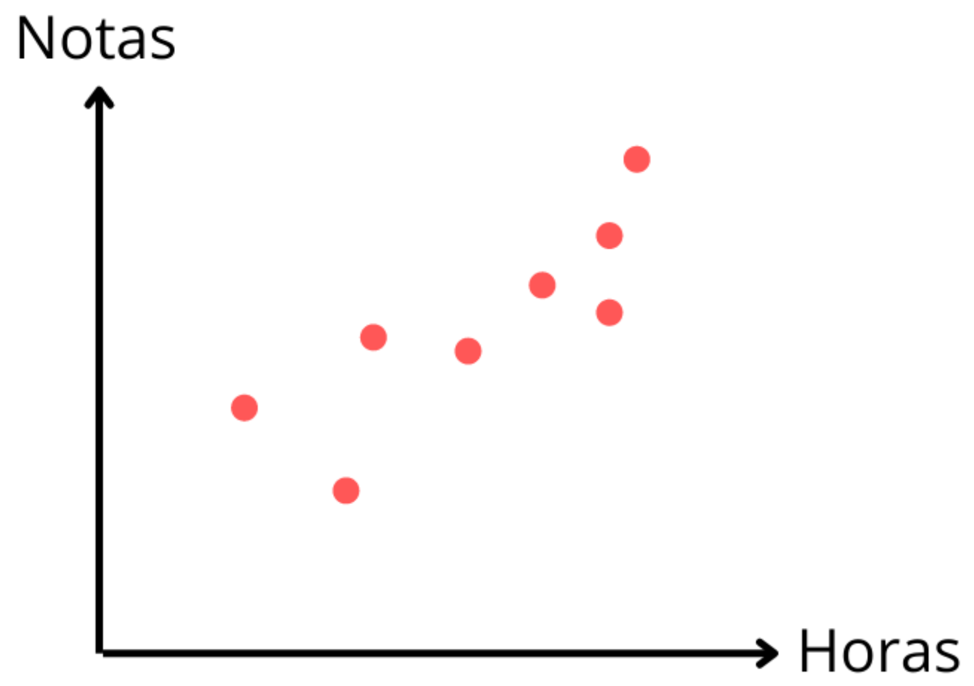
Regresión Lineal
Regresión Lineal Simple

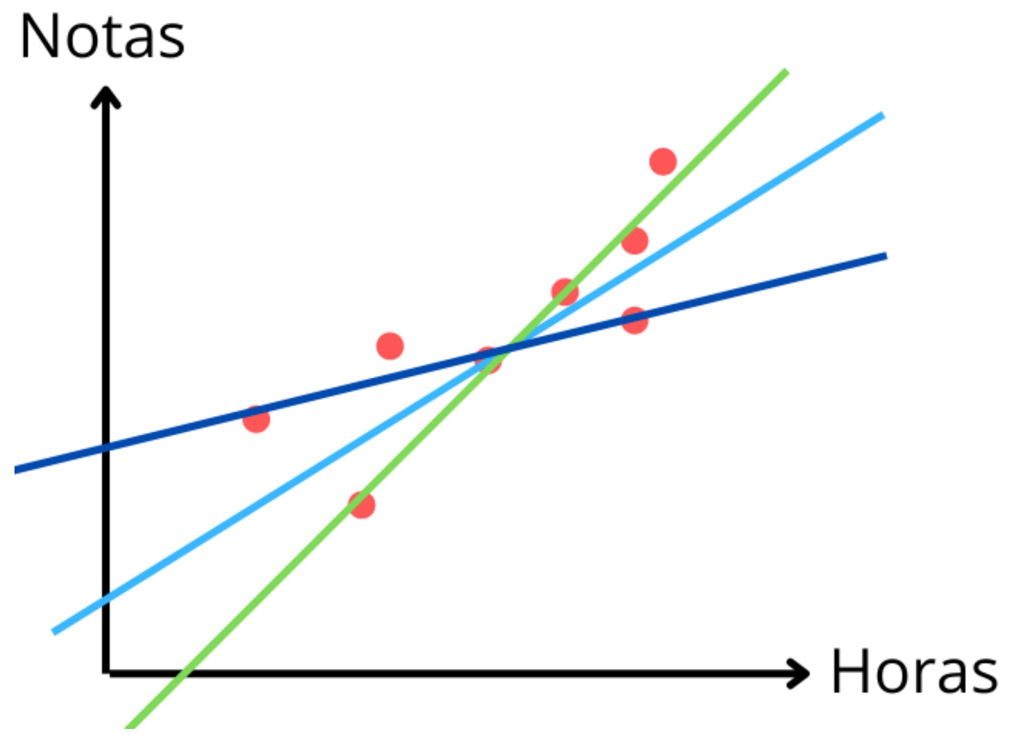
- En particular, la regresión lineal es una herramienta útil para predecir una respuesta cuantitativa.
  Por ejemplo predecir una cantidad:
  -Tiempo de demora de un vuelo
  -El precio de una propiedad
- Dentro de la Regresión Lineal tenemos la Simple y la Múltiple.

Regresión Lineal

Regresión Lineal Simple

- Sabemos que la ecuación de una recta es y= mx+b . Donde "m" es la pendiente y "b" la intercepción con el eje y
- Supongamos que queremos graficar las notas de los alumnos en base a la cantidad de horas que han estudiado. Podríamos graficarlo mediante ejes cartesianos (x1,y1) (x2,y2) … (xn,yn) lo cual obtendríamos un par de puntos como el siguiente:





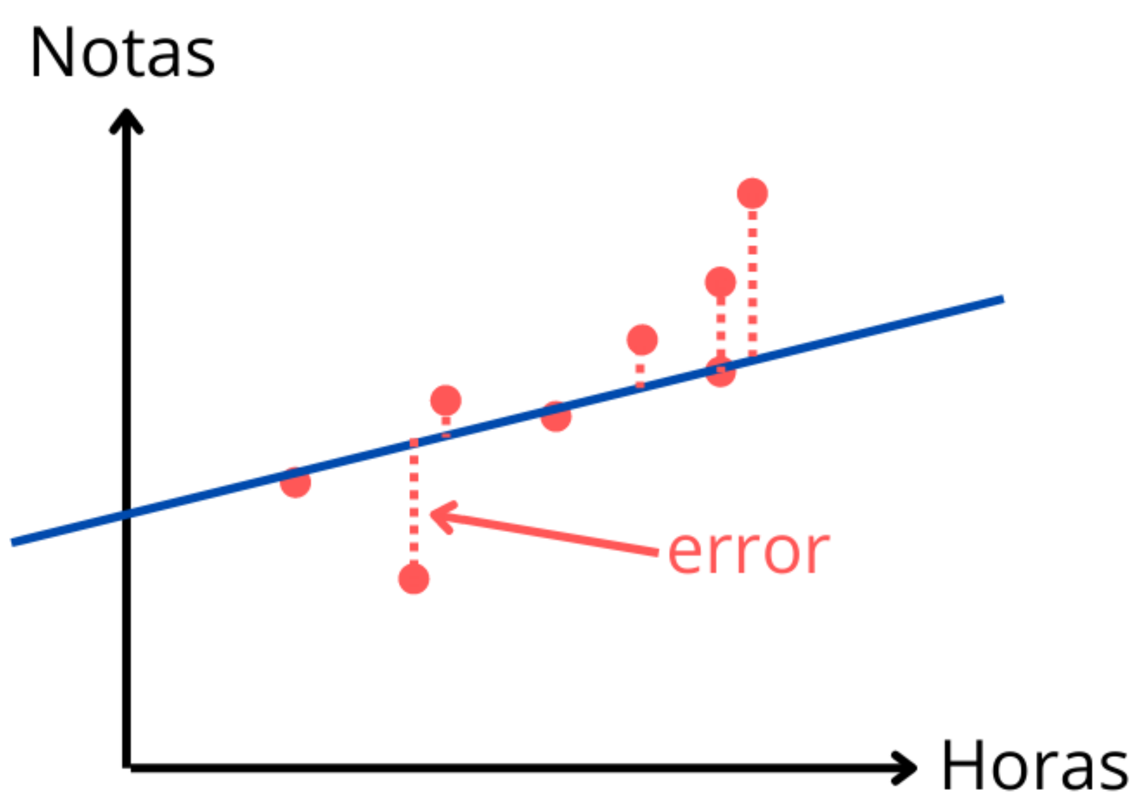
se puede observar que hay una tendencia a sacarse una buena nota si se invierte más horas de estudio
Pero si queremos representar esto en una recta como y=mx+b es imposible, ya que no hay recta que se adapte a TODOS los puntos (alguna llega a interceptar 3 puntos otras 2 pero ninguna recta TODOS)

- Pero sin embargo hay una recta que se ajusta mejor a los datos, a esto le vamos a llamar ecuación estimada $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x,$
el sombrero de la y indica una predicción / estimación

Regresión Lineal
Regresión Lineal Simple

- Pero ¿Cómo hacemos para encontrar la mejor recta que se adapte a nuestros puntos?
- Primero veamos que es el "Error" . Llamamos error o residuo a la diferencia entre "y" real e "y" predicho.

$$e_i = y_i - \hat{y}_i$$



en palabras fáciles, es la distancia desde nuestros puntos hasta la recta

- Si sumamos todos los errores (de todos los puntos) tendríamos la suma. A esto lo vamos a llamar RSS (residual sum of squares)

$$\mathrm{RSS} = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 \cdot x_i)^2$$

lo elevamos ^2 ya que también tenemos resultados negativos y para que esto no ocurra se eleva al cuadrado (ejemplo -2^2=4)

- Es lógico que si reducimos esta suma de errores estaremos en frente de la mejor recta que se adapta a nuestros puntos

Regresión Lineal

Regresión Lineal Simple

mínimos cuadrados

- Entonces, ¿Cómo hacemos para reducir lo máximo posible la suma de estos errores?. Para ello podemos usar el método de **mínimos cuadrados**.
  Veamos , nuestro error es igual a $(y_n-\hat{y})^2$
  Sabemos que $\hat{y}=mx+b$
  podemos remplazarlo. Llamermos EC al error cuadrático

  $$EC =(y_1-(m_1+b))^2 + (y_2-(m_2+b))^2 + \ldots\ldots + (y_n-(m_n+b))^2$$

  $$\begin{aligned} EC = \; & y_1^2 - 2y_1(mx_1+b)+(mx_1+b)^2 \\ & + y_2^2 - 2y_2(mx_2+b)+(mx_2+b)^2 \\ & + \ldots\ldots\ldots\ldots\ldots\ldots \\ & + y_n^2 - 2y_n(mx_n+b)+(mx_n+b)^2 \end{aligned}$$

  $$\begin{aligned} EC = \; & y_1^2 - 2y_1mx_1 - 2y_1b +m^2 x_1^2 + 2mx_1b + b^2 \\ & + y_2^2 - 2y_2mx_2 - 2y_2b +m^2 x_2^2 + 2mx_2b + b^2 \\ & + \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ & + y_n^2 - 2y_nmx_n - 2y_nb +m^2 x_n^2 + 2mx_nb + b^2 \end{aligned}$$

Regresión Lineal

Regresión Lineal Simple

mínimos cuadrados

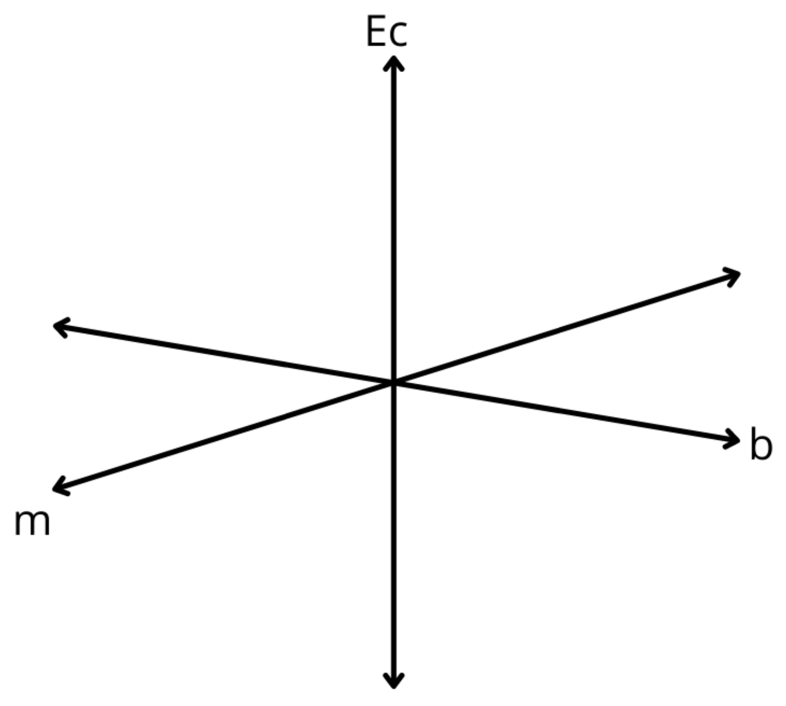
- Lo que voy a hacer ahora es juntar un poco los términos . Marcare cada uno con colores así será más fácil de seguir
  EC = y_1^2 - 2y_1mx_1 - 2y_1b +m^2 x_1^2 + 2mx_1b + b^2
  + y_2^2 - 2y_2mx_2 - 2y_2b +m^2 x_2^2 + 2mx_2b + b^2
  + ...
  + y_n^2 - 2y_nmx_n - 2y_nb +m^2 x_n^2 + 2mx_nb + b^2
- Si juntamos cada termino nos quedaría de la siguiente forma
  EC= (y_1^2 + y_2^2 +.... + y_n^2) - 2m(y_1x_1+y_2x_2+....+y_nx_n) - 2b(y_1+y_2+..+y_n) + m^2 (x_1^2+x_2^2+...+x_n^2) + 2mb(x_1+x_2+...+x_n) + n b^2
- Ahora te pido que hagamos unas pequeñas cuentas aparte, te prometo que ya volvemos a la ecuación anterior.
  sabemos que la media de y ($\bar{y}$) se calcula de la siguiente forma
  $\bar{y}$ ^2 = (y_1^2 + y_2^2 +.... + y_n^2) / n
  por lo tanto
  ($\bar{y}$ ^2) n = (y_1^2 + y_2^2 +.... + y_n^2)
  Y si nos fijamos es lo que tenemos en el primer termino de la ecuación anterior . Lo mismo para el segundo termino
  $\bar{y}$ $\bar{x}$ = (y_1x_1+y_2x_2+....+y_nx_n) / n
  ($\bar{y}$ $\bar{x}$) n = (y_1x_1+y_2x_2+....+y_nx_n)
  Dejare que ustedes hagan los demás para practicar

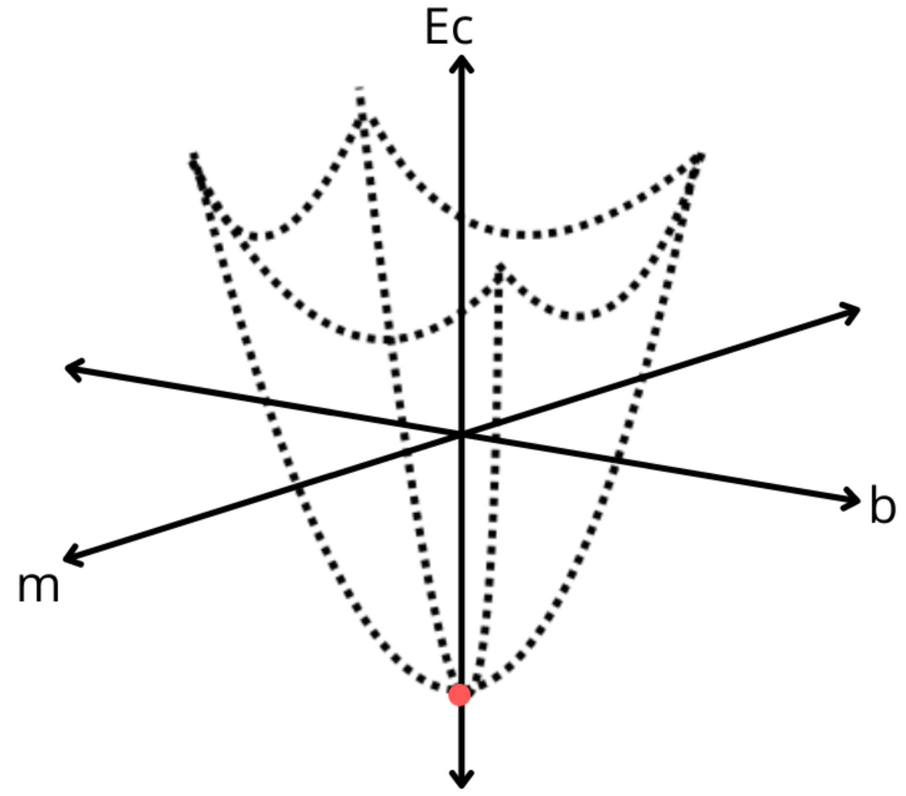
Regresión Lineal

Regresión Lineal Simple

mínimos cuadrados

- Ahora remplazamos todo en la ecuación que teníamos anteriormente, nos quedaría de la siguiente manera
  $EC = n(\bar{y}^2) - 2m(n(\bar{y}\bar{x})) - 2b(n\bar{y}) + m^2(n(\bar{x}^2)) + 2mb(n\bar{x}) + n b^2$
- Ahora si la ecuación esta mucho más ordenada que antes. Lo que debemos hacer ahora es intentar optimizar esta función o en otras palabras encontrar los mejores valores de “m” y “b” que permitan reducir el error cuadrático al máximo.
- Antes de eso grafiquemos esta función del error cuadrático, como vemos es una función de solo 2 variables (“m” y “b”) todo los demás son constantes.





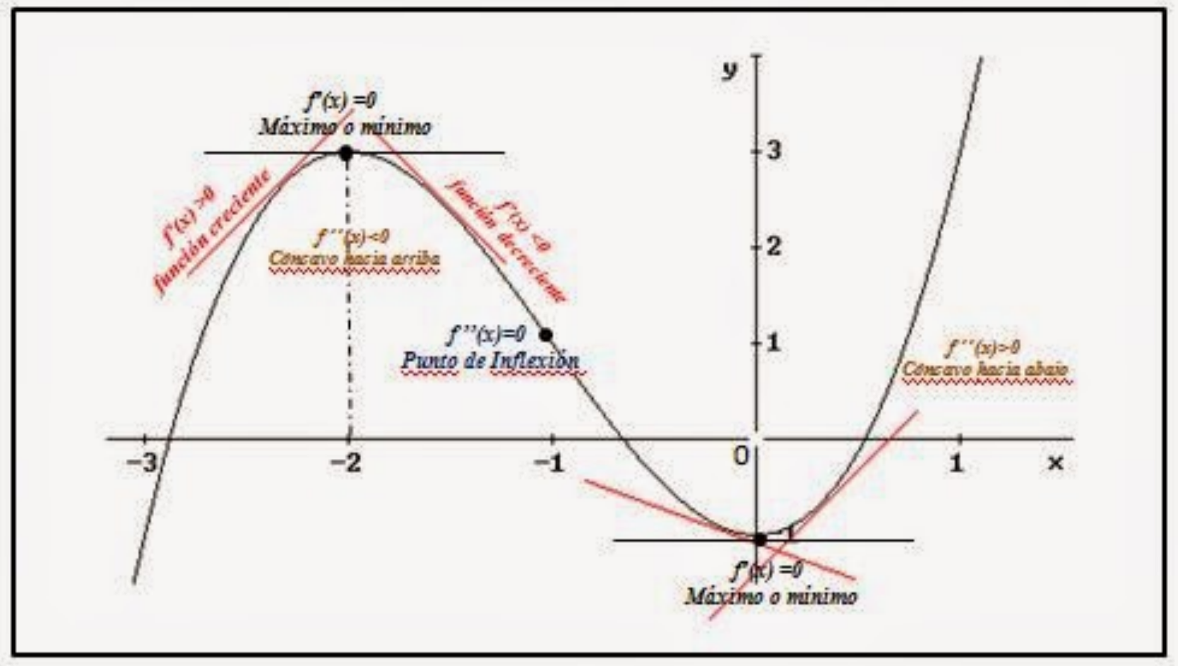
- Si hiciéramos todas las combinaciones posibles de “m” y “b” y las graficariamos sobre unos ejes cartesianos nos quedaría de la siguiente manera . (Una especie de cono / punta de calcetín /pañuelo) como vemos en el segundo gráfico.
- Como dijimos anteriormente , la idea es encontrar el mínimo de esta función (punto rojo que se ve en el gráfico). Cómo ? Utilizando derivadas parciales e igualando a 0

Regresión Lineal

Regresión Lineal Simple

mínimos cuadrados

- Recordemos que cuando calculamos derivadas parciales e igualamos a 0 estamos buscando un mínimo . Ya que la derivada nos indica la pendiente y si igualamos ésta a 0 estaríamos buscando una pendiente plana (ósea un mínimo o mínimo local).



  haremos lo mismo con nuestra función del error cuadrático . Primero la derivada parcial del error cuadrático con respecto a "m" (δSE / δm) = 0 y luego la derivada pacial del error cuadrático con respecto "b" (δSE / δb) = 0

- La derivada parcial con respecto a m
  $(\delta SE / \delta m) = -2n \bar{x} \bar{y} + 2n \bar{x}^2 m + 2 bn \bar{x} = 0$
- La derivada parcial con respecto a b
  $(\delta SE / \delta b) = -2n \bar{y} + 2mn \bar{x} + 2 bn = 0$
- Tenemos un sistema lineal de ecuaciones (2 ecuaciones y 2 variables)
  entonces podemos dividir ambas ecuaciones entre 2n
- Quedaria
  $(\delta SE / \delta m) = \bar{x} \bar{y} + m \bar{x}^2 + b \bar{x} = 0$
  $(\delta SE / \delta b) = - \bar{y} + mn \bar{x} + b = 0$

Regresión

Regresión Lineal Simple

mínimos cuadrados

- Teníamos :
  $(\delta SE / \delta m) = \bar{x}\ \bar{y} + m\ \bar{x}\text{^}2 + b\ \bar{x} = 0$
  $(\delta SE / \delta b) = -\ \bar{y} + mn\ \bar{x} + b = 0$
- Sumo $\bar{x}\ \bar{y}$ en la primera ecuación:
  $(\delta SE / \delta m) = \bar{x}\ \bar{y} - \bar{x}\ \bar{y} + m\ \bar{x}\text{^}2 + b\ \bar{x} = 0 + \bar{x}\ \bar{y}$
  $(\delta SE / \delta m) = m\ \bar{x}\text{^}2 + b\ \bar{x} = \bar{x}\ \bar{y}$
  divido ambos lados por $\bar{x}$
  $(\delta SE / \delta m) = m\ (\bar{x}\text{^}2 / \bar{x}) + b\ \bar{x} = (\bar{x}\ \bar{y}) / \bar{x}$
- Sumo $\bar{y}$ en la segunda ecuación:
  $(\delta SE / \delta b) = \bar{y} - \bar{y} + mn\ \bar{x} + b = 0 + \bar{y}$
  $(\delta SE / \delta b) = mn\ \bar{x} + b = \bar{y}$

Regresión

Regresión Lineal Simple

mínimos cuadrados

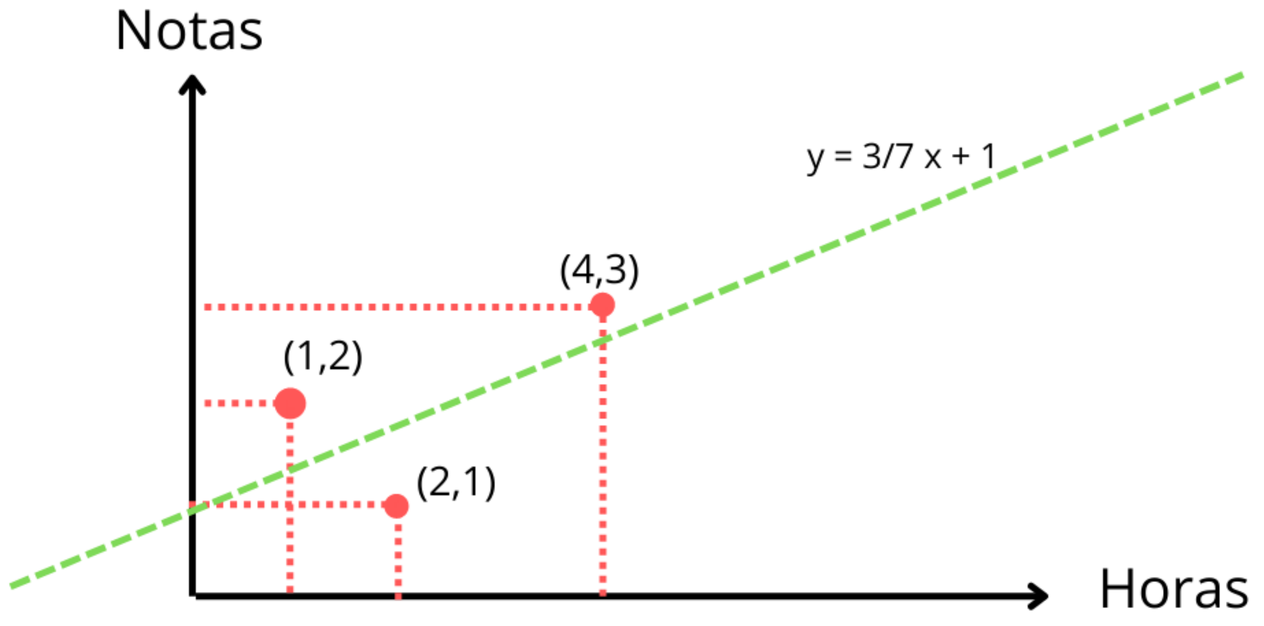
- Ya casi terminamos , por ultimo lo que vamos a hacer es restar estas dos ecuaciones que hemos dejado.
  $(\delta SE / \delta m) = m (\bar{x}^2 / \bar{x}) + b\ \bar{x} = (\bar{x}\ \bar{y}) / \bar{x}$
  $- \ (\delta SE / \delta b) = mn\ \bar{x} + b = \bar{y}$
- Nos quedaria lo siguiente
  $m(\bar{x} - (\bar{x}^2/\bar{x})) = \bar{y} - (\bar{x}\ \bar{y}/ \bar{x})$
  $m= (\bar{y} - (\bar{x}\ \bar{y}/ \bar{x})) / (\bar{x} - (\bar{x}^2/\bar{x}))$
  simplifico multiplicando y dividiendo por $\bar{x}$
  $m = ((\bar{y} - (\bar{x}\ \bar{y}/ \bar{x})) / (\bar{x} - (\bar{x}^2/\bar{x})))\ \ \bar{x}/\bar{x}$
  $m = (\bar{x}\ \bar{y} - \bar{x}\ \bar{y}) / ((\bar{x})^2) -\bar{x}^2$
  Ya m nos quedo como una constante
- Ahora vamos a hallar b , para eso remplazamos
  $b = \bar{y}-m\bar{x}$

Regresión

Regresión Lineal Simple

mínimos cuadrados

- Ya tenemos las 2 formulas , para “m” y para “b”. Vamos con un ejemplo:
- Supongamos que tenemos los puntos (1,2) , (2,1) , (4,3). Deseamos buscar la recta que mejor se adapte a nuestros puntos.
  - primero calculamos la media de las x => $\bar{x} = 7/3$
  - ahora la media de las y => $\bar{y} = 2$
  - la media de las xy => $\bar{x}\,\bar{y} = 16/3$
  - la media de x^2 => $\bar{x}^2 = 7$
- Entonces ,busquemos la pendiente con la ecuación anterior
  $m = (\bar{x}\,\bar{y} - \bar{x}\,\bar{y}) / ((\bar{x})^2) -\bar{x}^2$
  $m = (((7/3)\ 2)- (16/3)) / ((7/3)^2)- 7$
  $m = 3/7$
- Ahora busquemos la ordenada al origen
  $b = \bar{y}-m\bar{x}$
  $b = 2 - ((3/7)\ (7/3))$
  $b = 1$
- Entonces para finalizar la ecuación de la recta nos quedaría de la siguiente manera
  $y = mx+b$
  $y = 3/7\ x + 1$

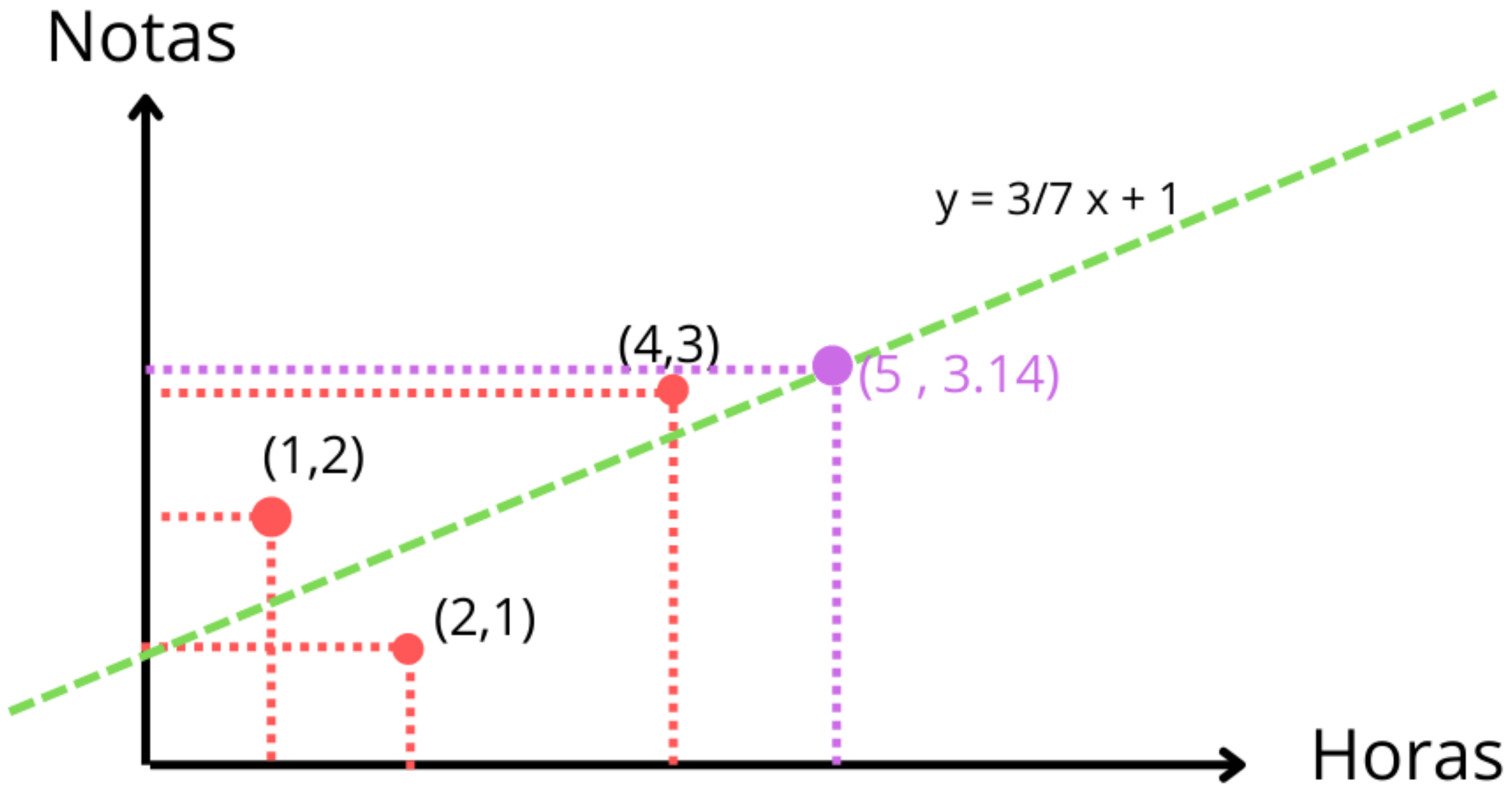


Regresión

Regresión Lineal Simple

mínimos cuadrados

- Una vez que tenemos la regresión lineal ya podemos hacer una predicción ! Nuestra primera predicción . Solo hace falta pasarle un dato de “x” y la función nos dará su predicción.
- Por ejemplo paseémosle un valor de 5 horas a x



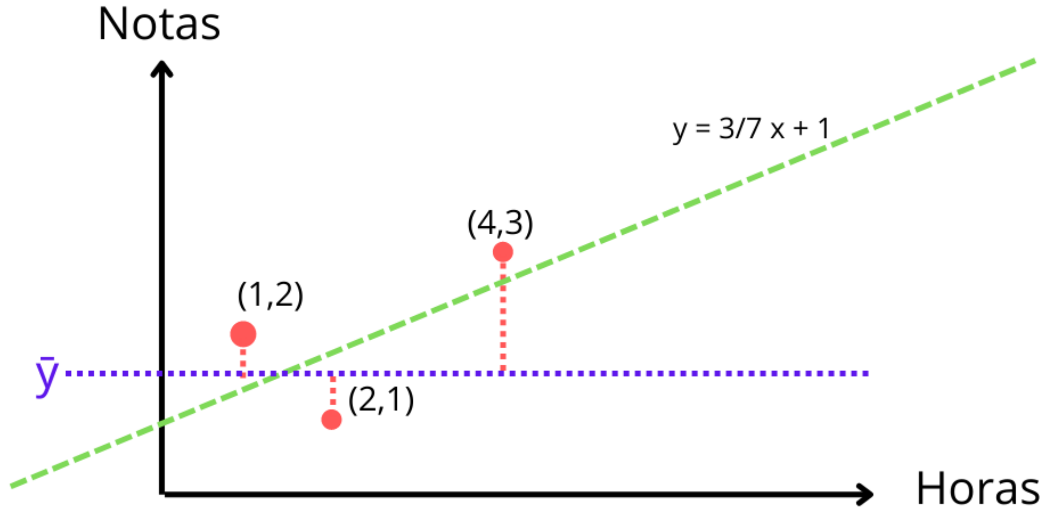
como vemos , la predicción nos devuelve que su nota seria 3.14

Regresión

Regresión Lineal Simple

Evaluación del Modelo

- Ahora queremos cuantificar hasta qué punto el modelo se ajusta a los datos. El modelo es confiable ? Veamos.
- Primero hagámonos la pregunta ¿Cuánto (qué %) de la variación total en "y" está descrita por la variación en "x"?
  Para eso calculemos la variación total en y (la distancia de sus puntos con respecto a su media al cuadrado) a esto lo llamaremos TSS (Total Sum of Squares)



  la ecuación viene dada

$$\mathrm{TSS} = \sum_{i=1}^{n} (y_i - \overline{y})^2$$

  Recuerden que tenemos la variabilidad no explicada por el modelo (RSS) el error cuadrático que habíamos llamado EC.

$$\mathrm{RSS} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

- Ahora la pregunta sería ¿Cuánto de la variación total NO esta descrita por la recta? A eso lo llamaremos **R^2 (R cuadrado) o coeficiente de determinación** y su ecuación viene dada por

$$R^2 = \frac{\mathrm{TSS} - \mathrm{RSS}}{\mathrm{TSS}}$$

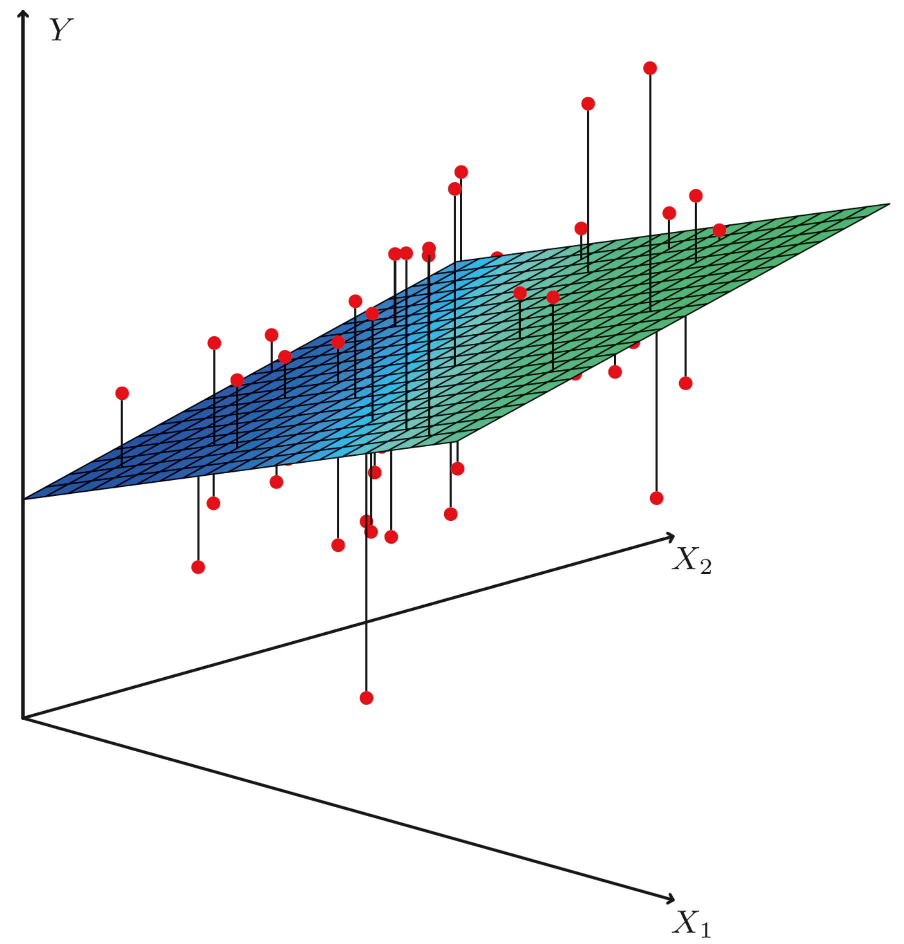
$R^2 \rightarrow 0$ cuando el modelo explica poco de la variabilidad de los datos.

$R^2 \rightarrow 1$ cuando el modelo explica mucho de la variabilidad de los datos.

Regresión

Regresión Lineal Múltiple

- Bien ya vimos que es la regresión lineal Simple, y aprender las bases nos ayuda un montón a luego entender con facilidad el resto. La realidad es que en la práctica tenemos que lidiar con más de un predictor. En el ejemplo anterior la nota de los alumnos tal vez dependa no solo de la cantidad de horas estudiadas, sino también de la edad, facilidad para la comprensión de matemáticas, etc.
- Por suerte ya aprendimos de arriba abajo las matemáticas para la regresión por lo que será bastante fácil explicar la Múltiple. En la lineal graficabamos una recta, en la Múltiple como estamos hablando de varias variables vamos a graficar un Plano.



$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_p x_p.$$

Regresión

Regresión Lineal Múltiple

Descenso del gradiente

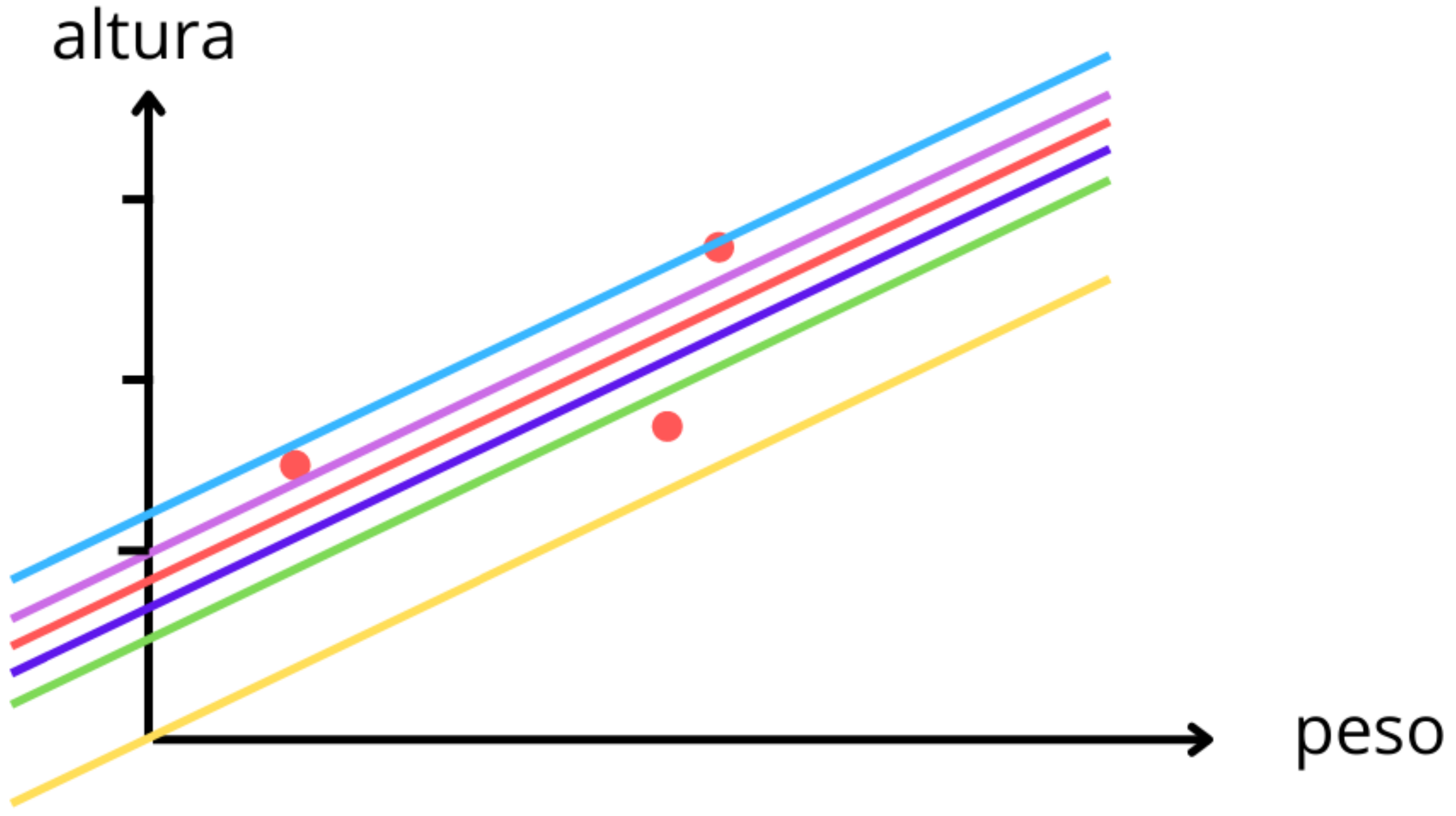
- Al igual que la regresión lineal simple , aquí buscaremos reducir el error cuadrático.
- En vez de mínimos cuadrados acá utilizaremos el Descenso del Gradiente (este nos servirá para la optimización problemas de regresión lineal, logística , cluster, ect)
- Bien haremos esto con un ejemplo así se entiende fácil
- Supongamos que queremos predecir la altura con respecto al peso de una persona, tenemos 3 puntos ((0.5 , 1.4) (2.3 , 1.9) (2.9 , 3.2)) .
  Lo primero que tenemos que hacer es calcular la función de Costo / la suma de los residuos.
  Para eso sabemos que nuestra recta seria algo como :
  Altura Predicha = intercepto + pendiente * peso
- Por ahora hagamos de cuenta que la pendiente es 0.64
  Altura Predicha = intercepto + 0.64 * peso

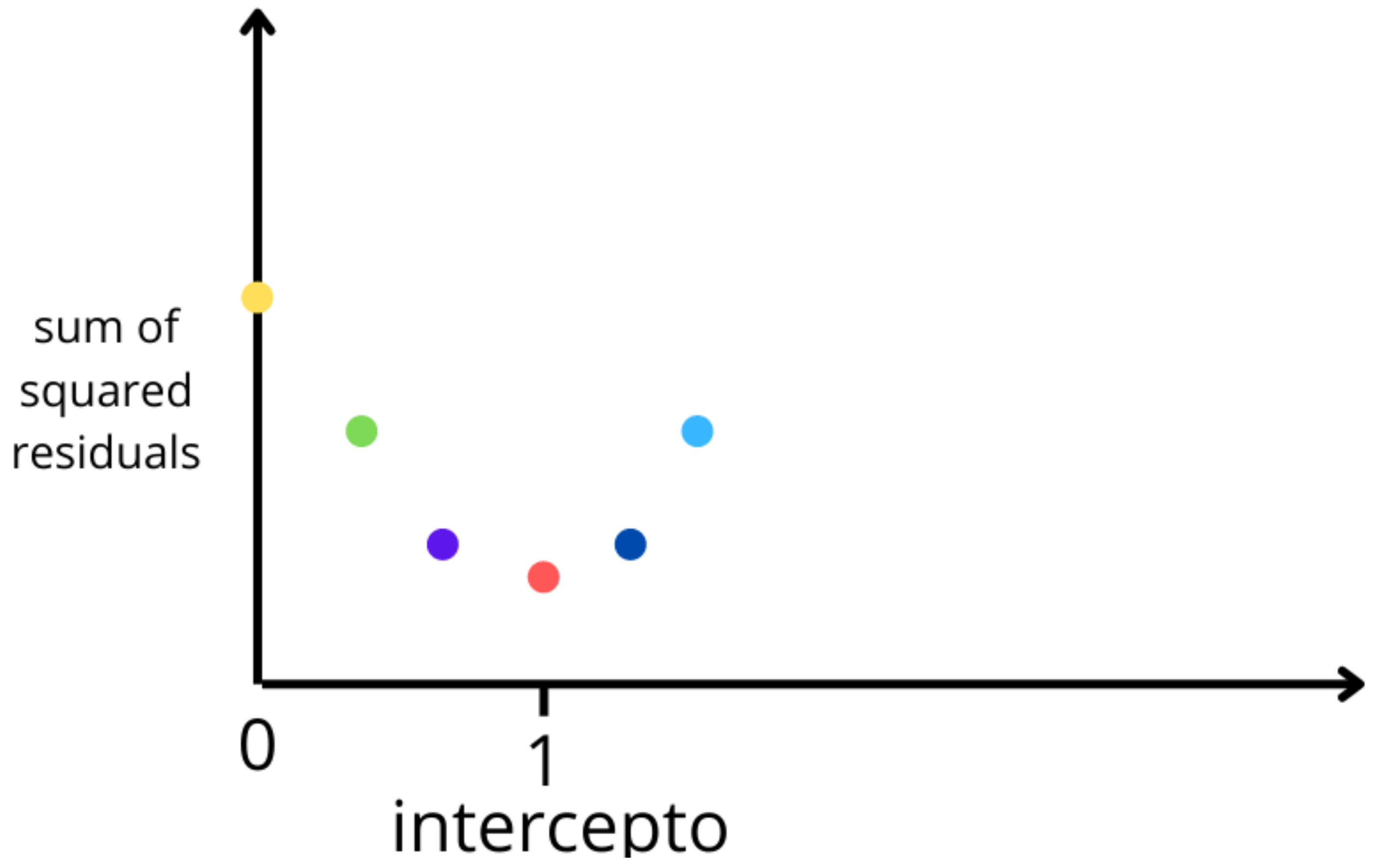
Regresión

Regresión Lineal Múltiple

Descenso del gradiente

- Ahora si , escogemos un valor random para el intercepto . Esto es una suposición inicial que le da al descenso de gradiente algo para mejorar. Será 0 pero podríamos utilizar cualquier valor.





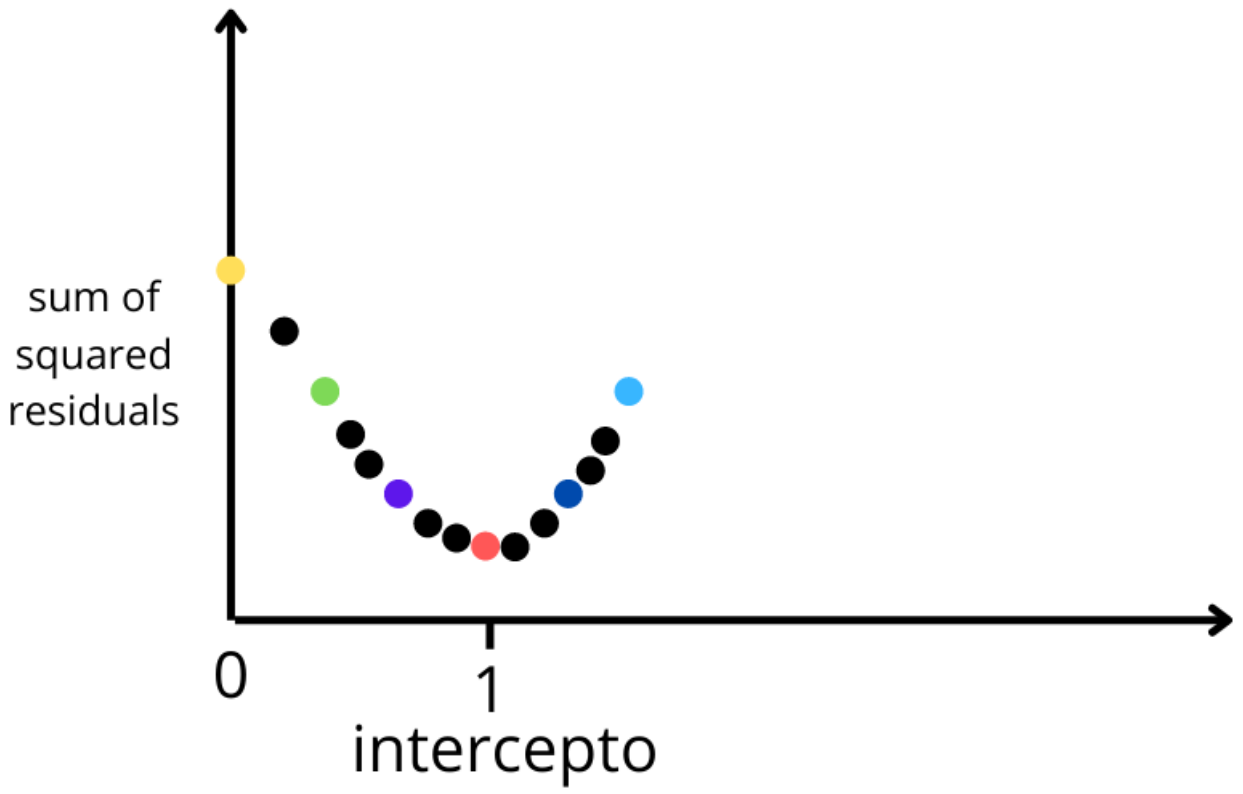
- Se puede observar en el primer gráfico que como la pendiente la dejamos constante, el gradiente solo va probando ajustar la linea para arriba y para abajo (cambiando el valor del intercepto) . Cada linea tiene un color representado por un punto en el gráfico de la función de costo. Podemos ver que para la recta Amarilla (intercepto =0) su punto en la función de costo esta muy elevado, y que la recta roja tiene el menor valor en sum of squared residuals

Regresión

Regresión Lineal Múltiple

Descenso del gradiente

- Pero estamos seguros que el punto rojo es el mejor ? Trafiquemos mas puntos



lo que hace el gradiente descendente es identificar el valor optimo haciendo pequeños "pasos" hasta llegar al mínimo.

- Sigamos , conocíamos la ecuación de la suma de errores:

$$\mathrm{RSS} = \sum_{i=1}^{n} (y_i - \hat{y_i})^2$$

RRS = (1.4 - (intercepto + 0.64 * peso))^2 + (1.9 - (intercepto + 0.64 * peso))^2 + (3.2 - (intercepto + 0.64 * peso))^2
RRS = (1.4 - (intercepto + 0.64 * 0.5))^2 + (1.9 - (intercepto + 0.64 * 2.3))^2 + (3.2 - (intercepto + 0.64 * 2.9))^2

Regresión

Regresión Lineal Múltiple

Descenso del gradiente

- Conociendo lo anterior , tomamos la derivada de la función de costo
  RRS = (1.4 - (intercepto + 0.64 * 0.5))^2 + (1.9 - (intercepto + 0.64 * 2.3))^2 + (3.2 - (intercepto + 0.64 * 2.9))^2
- Derivadas parciales
  (δ/δ intercepto) RRS = (δ/δ intercepto) (1.4 - (intercepto + 0.64 * 0.5))^2
  + (δ/δ intercepto) (1.9 - (intercepto + 0.64 * 2.3))^2
  + (δ/δ intercepto) (3.2 - (intercepto + 0.64 * 2.9))^2
  (δ/δ intercepto) RRS = -2 (1.4 - (intercepto + 0.64 * 0.5))
  -2 (1.9 - (intercepto + 0.64 * 2.3))
  -2 (3.2 - (intercepto + 0.64 * 2.9))

Regresión

Regresión Lineal Múltiple

Descenso del gradiente

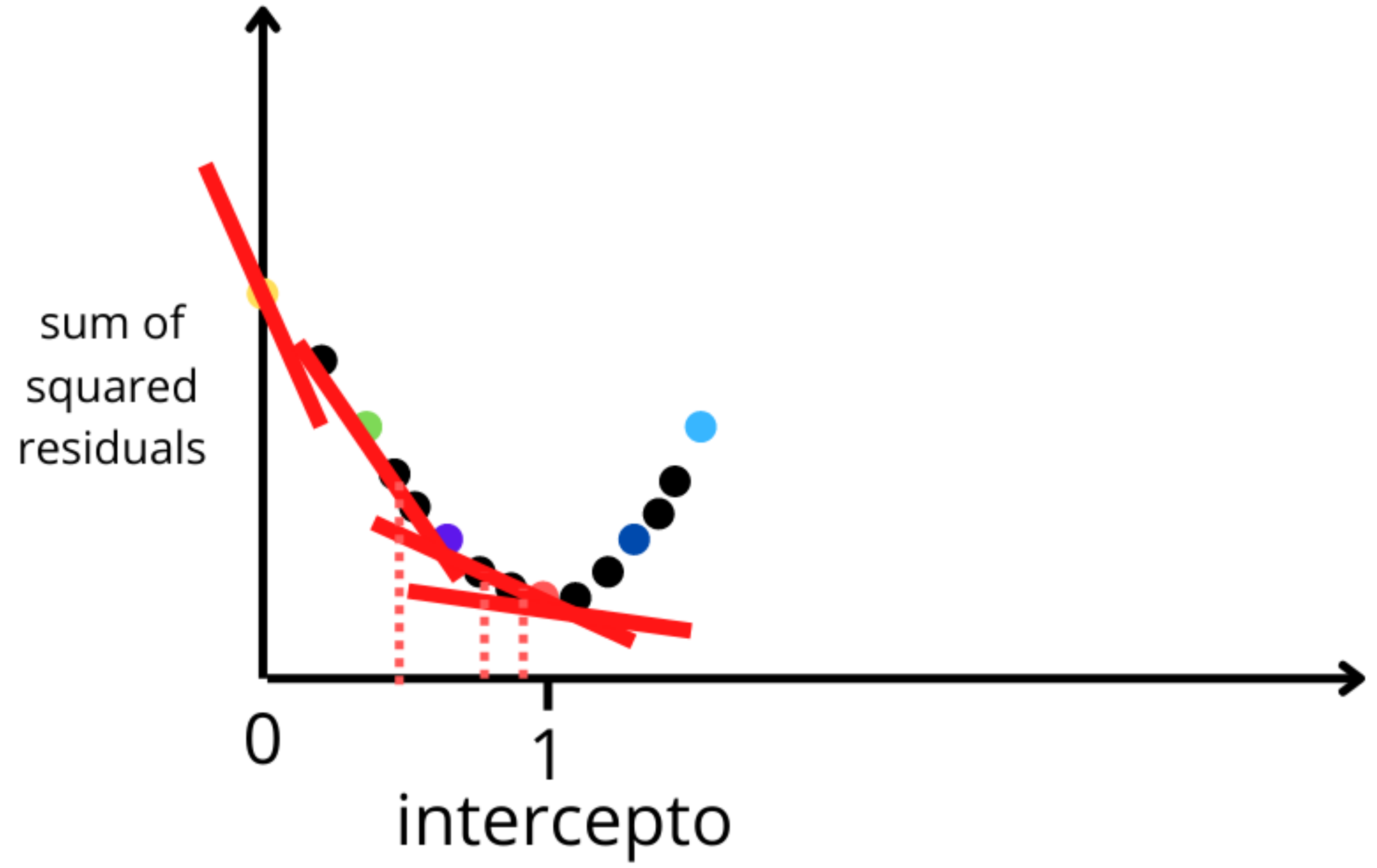
- Una vez que tenemos las derivadas parciales hechas , usaremos el Gradiente descendiente para encontrar el mínimo de la función. La diferencia con mínimos cuadrados, es que en mínimos cuadrados encontrábamos el mínimo igualando a 0 , y el gradiente descendiente lo encuentra realizando pequeños pasos. Esto hace que el gradiente descendiente sea verdaderamente útil cuando no podemos derivar e igualar a cero.
- Volvamos a la ecuación anterior
  (δ/δ intercepto) RRS = -2 (1.4 - (intercepto + 0.64 * 0.5))
  -2 (1.9 - (intercepto + 0.64 * 2.3))
  -2 (3.2 - (intercepto + 0.64 * 2.9))
- Comencemos con un intercepto de 0
  (δ/δ intercepto) RRS = -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

Regresión

Regresión Lineal Múltiple

Descenso del gradiente

- (δ/δ intercepto) RRS = -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
- Cuando el intercepto = 0 , el la pendiente de la curva es -5.7



vemos como va dando los pasos y a medida que nos acercamos a una pendiente = 0 da pequeños pasos.

- Entonces para calcular el tamaño de los pasos:
  step Size = -5.7 x Learning rate
  por ahora diremos que el Learning rate es = 0.1
  step Size = -5.7 x 0.1

Regresión

Regresión Lineal Múltiple

Descenso del gradiente

- step Size = -5.7 x 0.1
- Para tomar un nuevo valor de intercepto daremos 1 paso
  nuevo intercepto = viejo intercepto - step size
  nuevo intercepto = 0 - (-0.57) = 0.57
- Volvemos a hacer la cuenta
  (δ/δ intercepto) RRS = -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
  pendiente de la curva -2.3 .... Todavía seguimos alejados del cero
- Repito
  nuevo intercepto = 0.57 - (-0.23) = 0.8
  (δ/δ intercepto) RRS = -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
  Pendiente de la curva -0.9 .... Ya nos vamos acercando y cada vez los pasos se están dando más cortos.... Y asi repetiremos
- Y obviamente para que no se vuelva un loop infinito definiremos como interparámetro : Maximo número de pasos (recordar que también estaba el Learning step)
- Ahora si ya comprendes como funciona el descenso del gradiente . Al comienzo definimos como constante la pendiente (para poder explicar mejor su funcionamiento) pero la realidad es que solo basta con tomar derivadas parciales del intercepto y de la pendiente. De esta manera seria repetir el proceso con estas dos variables

Regresión

Regresión Lineal Múltiple

Descenso del gradiente

- Repasamos los pasos:
- Paso1 : Tomamos la derivada de la función de coste de CADA parámetro
- Paso2: Tomamos un valor random para cada parámetro
- Paso3: Pondremos esos valores dentro de las derivadas (comienza Gradiente Descendente)
- Paso4: Calculamos el Step Size (Step Size= pendiente + learning Rate)
- Paso5: Calculamos los nuevos Parámetros (Nuevos Parametros = parámetro anterior - Step Size)
- Repetimos paso 3…. Hasta el 5
- Volvemos a repetir hasta encontrar un valor óptimo