Universidad Sergio Arboleda
Segundo Parcial HPC-1
Diego Alejandro Bermúdez González 1000470149 27/04/2022

• Demostración Función de costo de MCO

• Partiendo del supuesto de $\Rightarrow \underbrace{y = b + mx}$

↳ Función de la recta
↳ Regresión lineal de los datos

• Teniendo como el error asociado a cada predicción con su valor real que:

$$\text{error} = y_a - (mx_a + b)$$

• Dado que se empleara el modelo de MCO tenemos que elevar al cuadrado el valor de cada error (esto con el fin de evitar que errores positivos de cancelen con los negativos)

$$\Rightarrow \text{error}^2 = (y_a - (mx_a + b))^2$$

• Ahora bien, la Función de costo será el resultado de la sumatoria de todos los errores asociados a la estimación con su valor real.

$$\Rightarrow EC = (y_1 - (mx_1 + b))^2 + (y_2 - (mx_2 + b))^2 + \ldots + (y_n - (mx_n + b))^2$$

• Para minimizar el valor de esta función tenemos que: los valores m y b deben ser calculados convenientemente.

$$\Rightarrow EC = y_1^2 - 2y_1(mx_1 + b) + (mx_1 + b)^2 + y_2^2 - 2y_2(mx_2 + b) + (mx_2 + b)^2 + \ldots$$
$$\ldots + y_n^2 - 2y_n(mx + b) + (mx_n + b)^2$$

$$= \begin{array}{cccccccccccc} \boxed{y_1^2} & - & \boxed{2y_1 m x_1} & - & \boxed{2y_1 b} & + & \boxed{m^2 x_1^2} & + & \boxed{2mx_1 b} & + & \boxed{b^2} & + \\ y_2^2 & - & 2y_2 m x_2 & - & 2y_2 b & + & m^2 x_2^2 & + & 2mx_2 b & + & b^2 & + \\ \vdots & & & & \vdots & & & & & & & \\ y_n^2 & - & 2y_n m x_n & - & 2y_n b & + & m^2 x_n^2 & + & 2mx_n b & + & b^2 & \end{array}$$

(se suman Terminos semejantes)

$$EC = \overbrace{(y_1^2 + y_2^2 + \ldots + y_n^2)}^{n\overline{y^2}} - 2(\overbrace{x_1 y_1 + x_2 y_2 + \ldots + x_n y_n}^{n\overline{xy}}) - 2b(\overbrace{y_1 + y_2 + \ldots + y_n}^{n\overline{y}}) +$$

$$m^2(\overbrace{x_1^2 + x_2^2 + \ldots + x_n^2}^{n\overline{x^2}}) + 2mb(\overbrace{x_1 + x_2 + \ldots + x_n}^{n\overline{x}}) + nb^2$$

$$\overline{y^2} = \frac{y_1^2 + y_2^2 + \ldots + y_n^2}{n} \Rightarrow n\overline{y^2} = y_1^2 + y_2^2 + \ldots + y_n^2 \quad (\text{tambien para } x)$$

$$\frac{x_1 y_1 + x_2 y_2 + \ldots + x_n y_n}{n} = \overline{xy} \Rightarrow \overline{xy}\, n = x_1 y_1 + x_2 y_2 + \ldots x_n y_n$$

→ Al hacer el cambio de variable correspondiente tenemos que

$$EC = n\overline{y^2} - 2mn\,\overline{xy} - 2bn\overline{y} + m^2 n\overline{x^2} + 2mbn\overline{x} + nb^2$$

• Para hallar m y b tal que EC sea mínima debemos hallar:

$$\frac{\partial EC}{\partial m}=0 \qquad \frac{\partial EC}{\partial b}=0$$

→ Por lo tanto tenemos que:

$$\begin{aligned} -\cancel{2n}\,\overline{xy} + \cancel{2n}\,\overline{x^2}\,m + \cancel{2}b\cancel{n}\,\bar{x} &= 0 \quad (1) \\ -\cancel{2n}\,\bar{y} + \cancel{2}m\cancel{n}\,\bar{x} + \cancel{2}b\cancel{n} &= 0 \quad (2) \end{aligned} \Bigg/ 2n \Rightarrow \begin{aligned} -\overline{xy} + \overline{x^2}\,m + b\bar{x} &= 0 \\ -\bar{y} + m\bar{x} + b &= 0 \end{aligned}$$

De 1 tenemos que

$$\overline{xy} - \overline{xy} + m\overline{x^2} + b\bar{x} = \overline{xy} \Rightarrow \overline{xy} = m\overline{x^2} + b\bar{x} \quad (3)$$

De 2 tenemos que

$$\bar{y} - \bar{y} + m\bar{x} + b = \bar{y} \Rightarrow \bar{y} = m\bar{x} + b \Rightarrow b = \bar{y} - m\bar{x} \quad (4)$$

De 4 en 3 tenemos que

$$\overline{xy} = m\overline{x^2} + (\bar{y} - m\bar{x})\,\bar{x} \Rightarrow m\overline{x^2} + \bar{x}\,\bar{y} - m(\bar{x})^2 = \overline{xy}$$

$$m = \frac{\overline{xy} - \bar{x}\,\bar{y}}{\overline{x^2} - (\bar{x})^2} \quad (5)$$

De 5 en 4 tenemos que

$$b = \bar{y} - \left(\frac{\overline{xy} - \bar{x}\,\bar{y}}{\overline{x^2} - (\bar{x})^2}\right)\bar{x}$$

Demostración gradiente descendiente

El objetivo de este metodo consiste en partir de una función generalizada e ir haciendo cambios hasta encontrar la mejor versión de la regresión "costo" (minima), es decir optimizar el supuesto inicial.

Por lo cual y para el uso practico de una regresión lineal "recta" partimos de los siguientes supuestos:

$$y = mx + b$$

Donde el caso más simple es el cual en el que $b=0$

Por lo cual tenemos que:

• $y = mx$

Donde m será un valor arbitrario elegido a "ojo" sobre los datos: por ejemplo $m = w$

Por lo cual tenemos que:

$$y = wx$$

↳ Ahora bien, gracias al uso del error cuadratico medio, se:

→ Calcula su valor, es decir: Suponemos un w como el mejor valor de m; a partir de ello optimizamos este valor hasta que encontremos un valor w tal que el MSE resultante sea el mejor posible.

→ Este proceso puede ser de incremeto o decremento.

Por ejemplo:

Para un DataSet A tenemos que:

Para un $w = 3, 5, 7, 16, 20, \ldots, 100, 110$

Tenemos un valor del $MSE = 20, 15, 13, 12, 10 \ldots, 0.7, 2$ Respectivamente

↳ Por lo cual encontramos un valor w para obtener el minimo valor de MSE correspondiente. (se podría optimizar más, de ser necesario)

↳ Eso quiere decir que tendríamos una función de regresión lineal tentativa cuando $b = 0$

↳ Es decir la función $y = 100x$

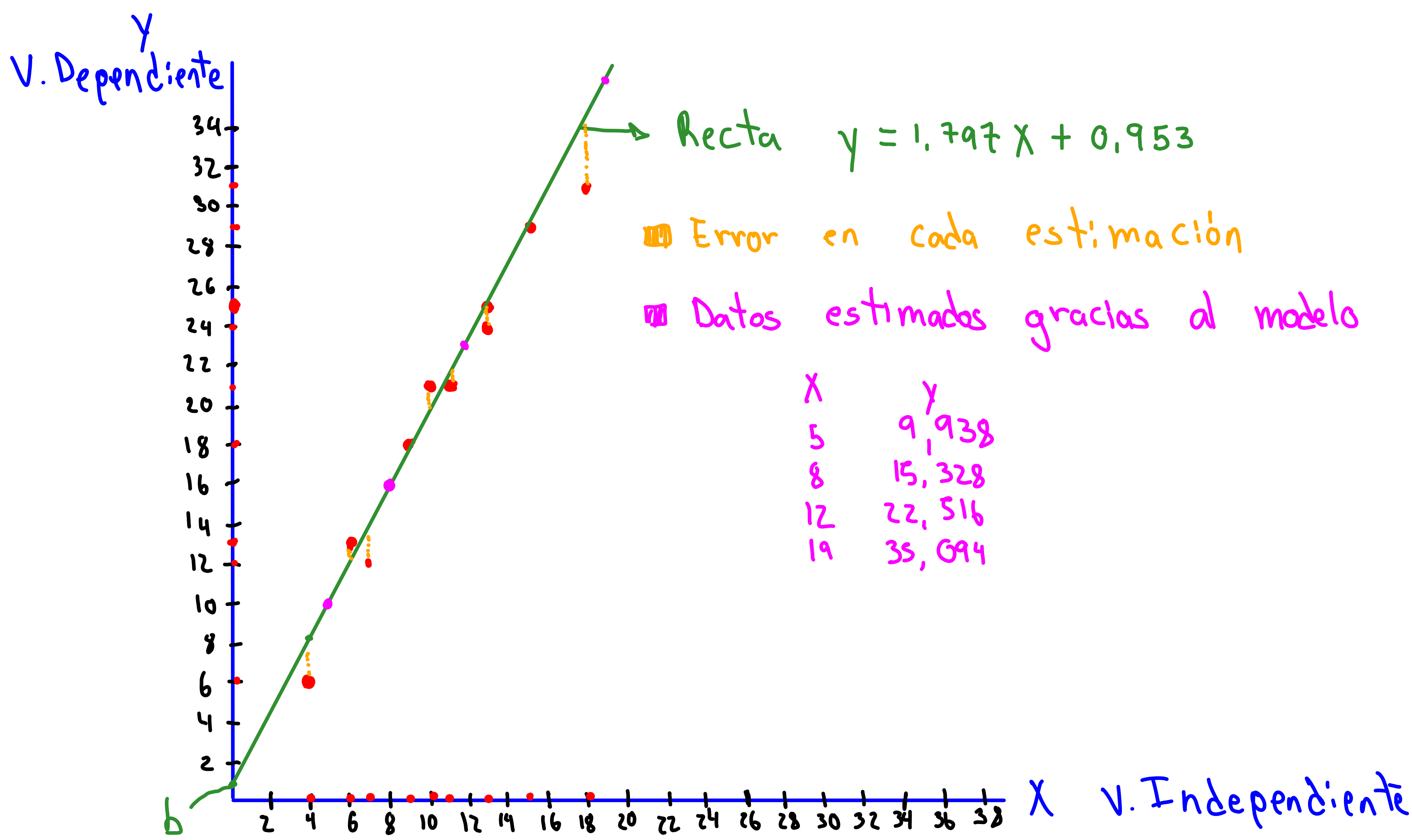
Nota: es importante que el ratio de aprendizaje "variación en w" no sea redundante; se recomienda que los incrementos o decrementos sean cortos.

Nota$_2$: Es posible que el gradiente descendiente se quede atrapado en un minimo local, pero es tan baja esta probabilidad que en muchas ocasiones (la mayoria) se descarta dicha posibilidad.

Creación Dataset de 10 valores

Variable Independiente X	Variable Dependiente y
7	12
11	21
9	18
4	6
13	25
13	24
18	31
6	13
10	21
15	29

Graficación del Dataset



Calculo de m y b Tal que la Función de costo tenga su valor mínimo!

$$b = \bar{y} - \left(\frac{\overline{xy} - \bar{x}\,\bar{y}}{\overline{x^2} - (\bar{x})^2} \right) \bar{x} \qquad m = \frac{\overline{xy} - \bar{x}\,\bar{y}}{\overline{x^2} - (\bar{x})^2}$$

Para lo cual tenemos que:

$\bar{X} = 10{,}6$ $\qquad \overline{Xy} = 241{,}9$ $\qquad \sum X = 106$ $\qquad \overline{X^2} = 129$

$\bar{y} = 20$ $\qquad \bar{X}\,\bar{y} = 212$ $\qquad \sum y = 200$ $\qquad \bar{X}^2 = 112{,}36$

Por lo tanto:

$$b = 20 - \left(\frac{241{,}9 - 212}{129 - 112{,}36} \right)(10{,}6) = 0{,}953 \qquad m = \frac{241{,}9 - 212}{129 - 112{,}36} = 1{,}797$$

Todos los datos presentados anteriormente se sustentan en la hoja de Calculo anexa.

■ Metrica de rendimiento para evaluar el modelo

- Error cuadrático medio (MSE)

En esta metrica de rendimiento tenemos que evaluar la siguiente formula

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

Donde un modelo perfecto tendria como resultado un valor de 0

↳ en este modelo solo podremos obtener valores positivos, por lo cual, cuanto más grande sea el valor obtenido, peor será el modelo

↳ Para el modelo en cuestion tenemos que:

$$MSE = 2{,}073$$

↳ Por lo cual se induce que el modelo, gracias a la poca cantidad de datos y por la naturaleza de la metrica utilizada (castiga demasiado los datos alejados) tiene una Fiabilidad baja.

Esta metrica Fue castigada en gran medida por la minima cantidad de datos. Este resultado Fue sesgado por la naturaleza del mismo.