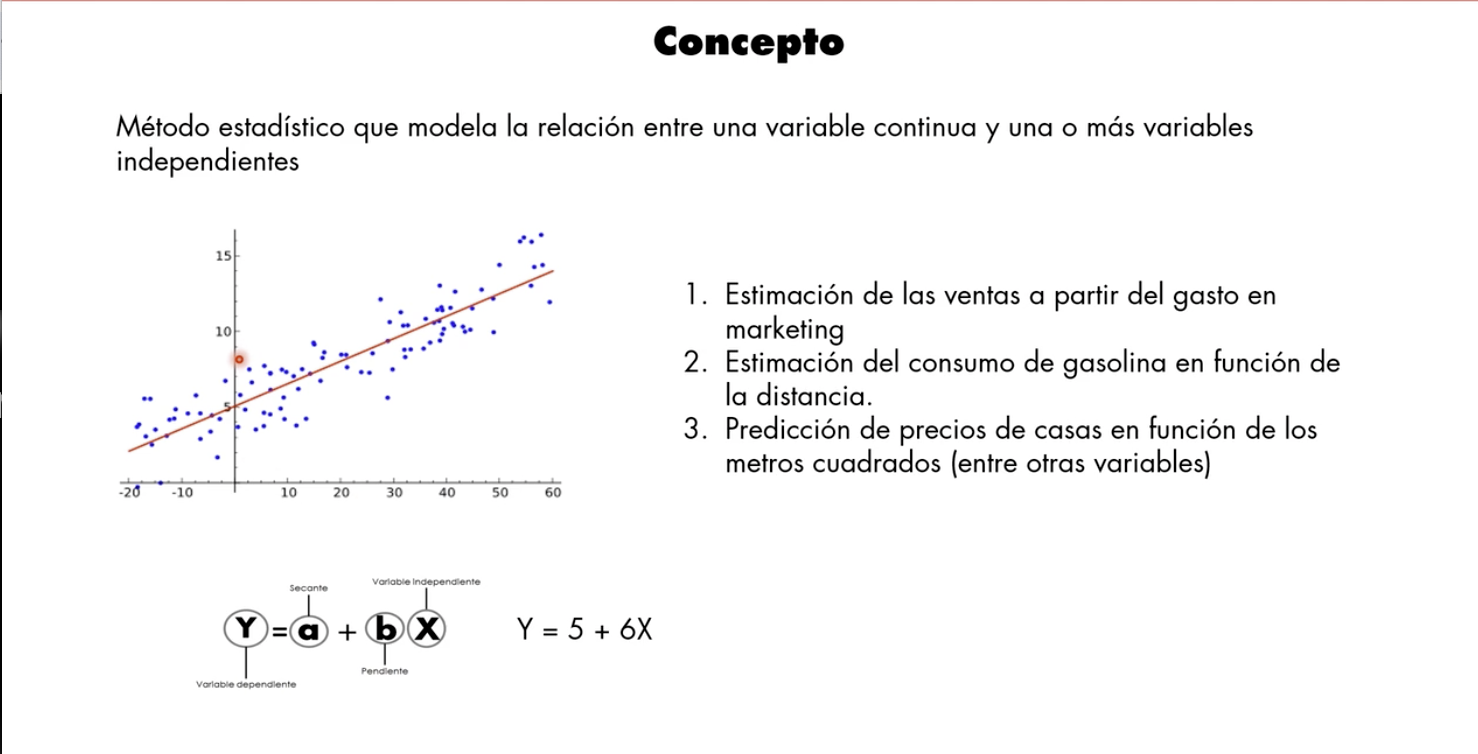
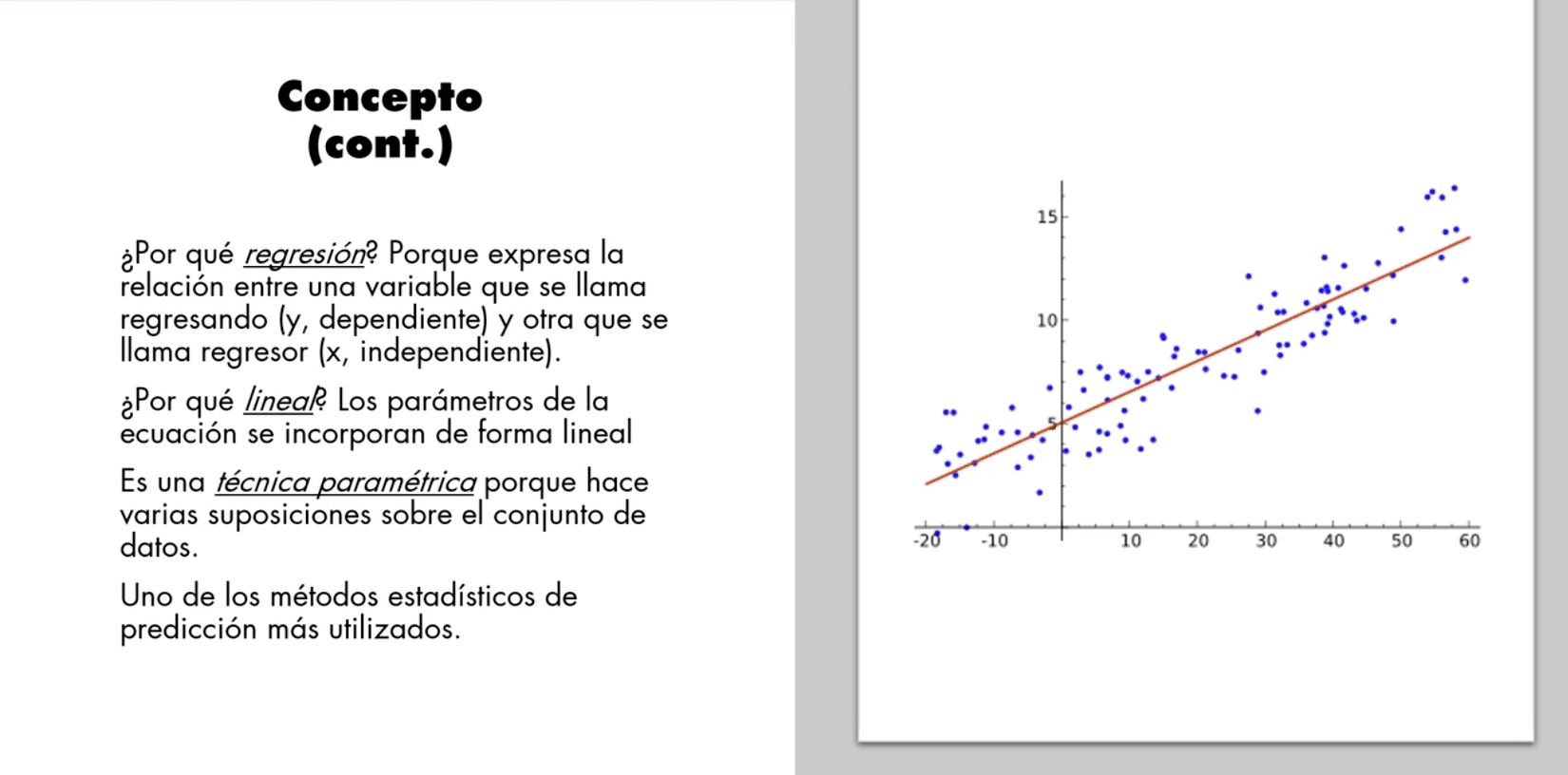
El concepto de regresion lineal es el metodo estadistico que modela la relación entre una variable continua y una o mas variables independientes. Por ejemplo, puede ser la estimación de las ventas a partir de los gastos en marketing, o el consumo de combustible según la distancia o por ejemplo también puede ser la predicción de precios de casas según sus metros cuadrados.

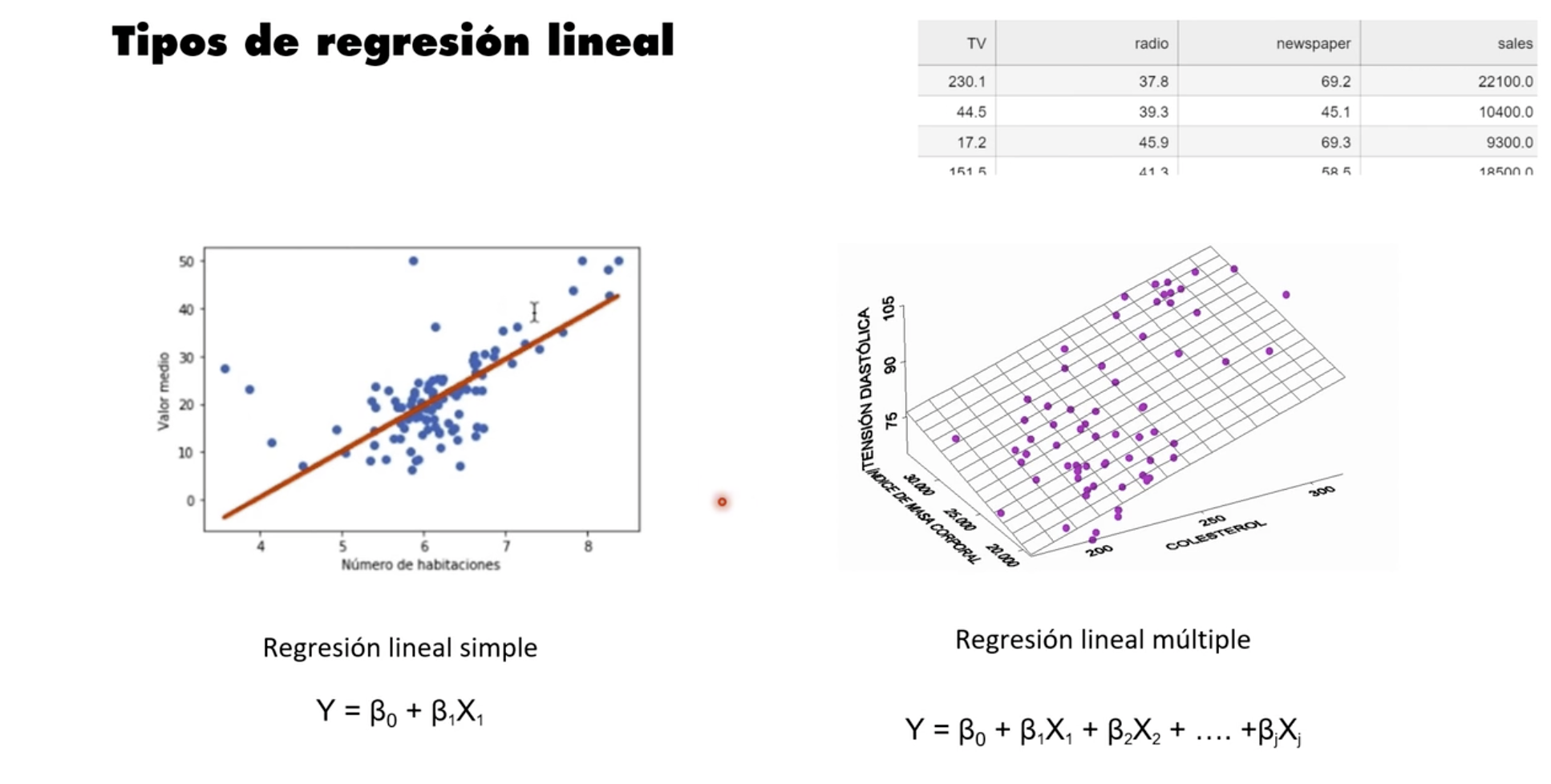


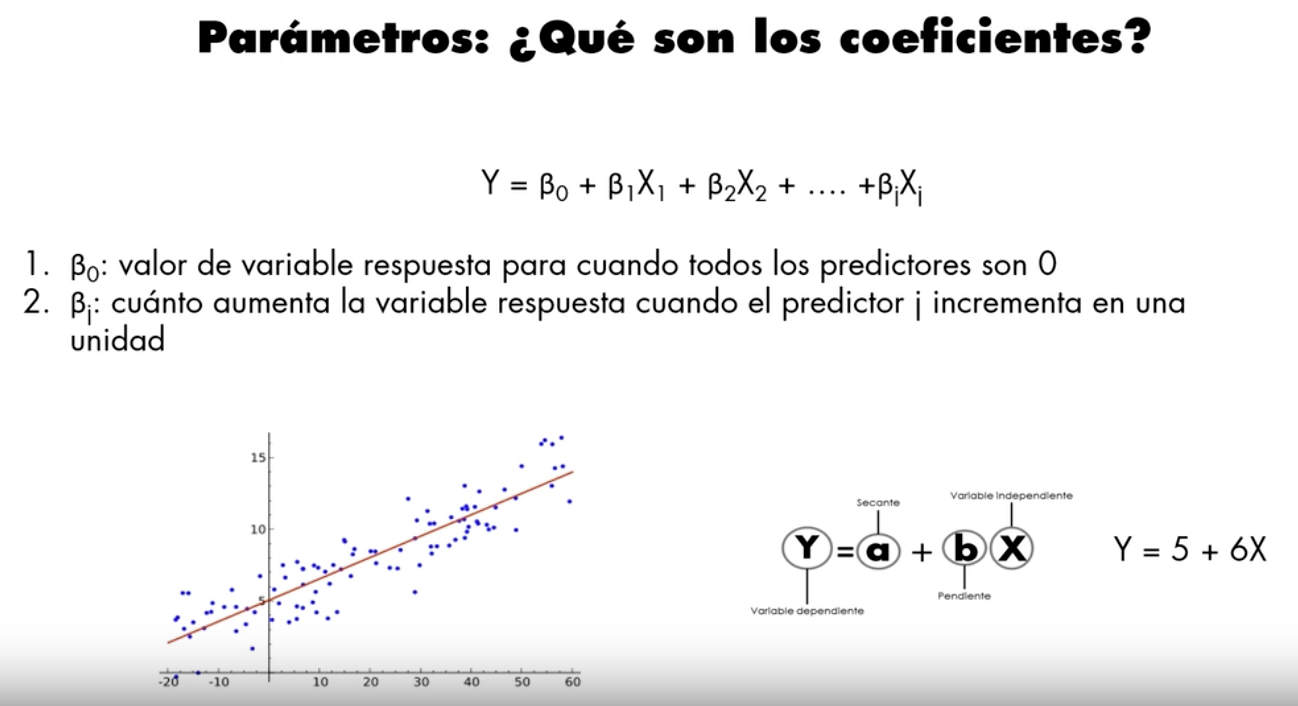
En el grafico, el eje x representa el gasto en marketing, y el y las ventas.



Hay dos tipos de regresión lineal, la simple y la multiple.

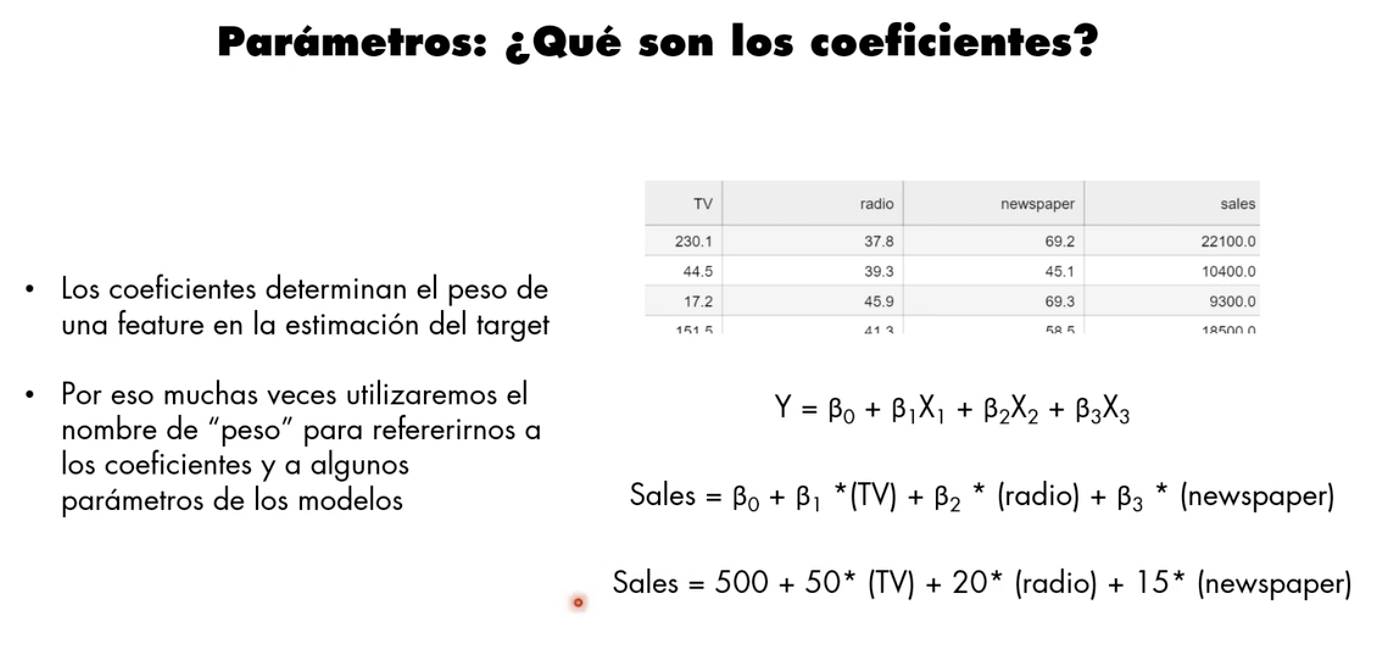
Simple es cuando solo tiene un feature, y multiple cuando tiene mas de 1.





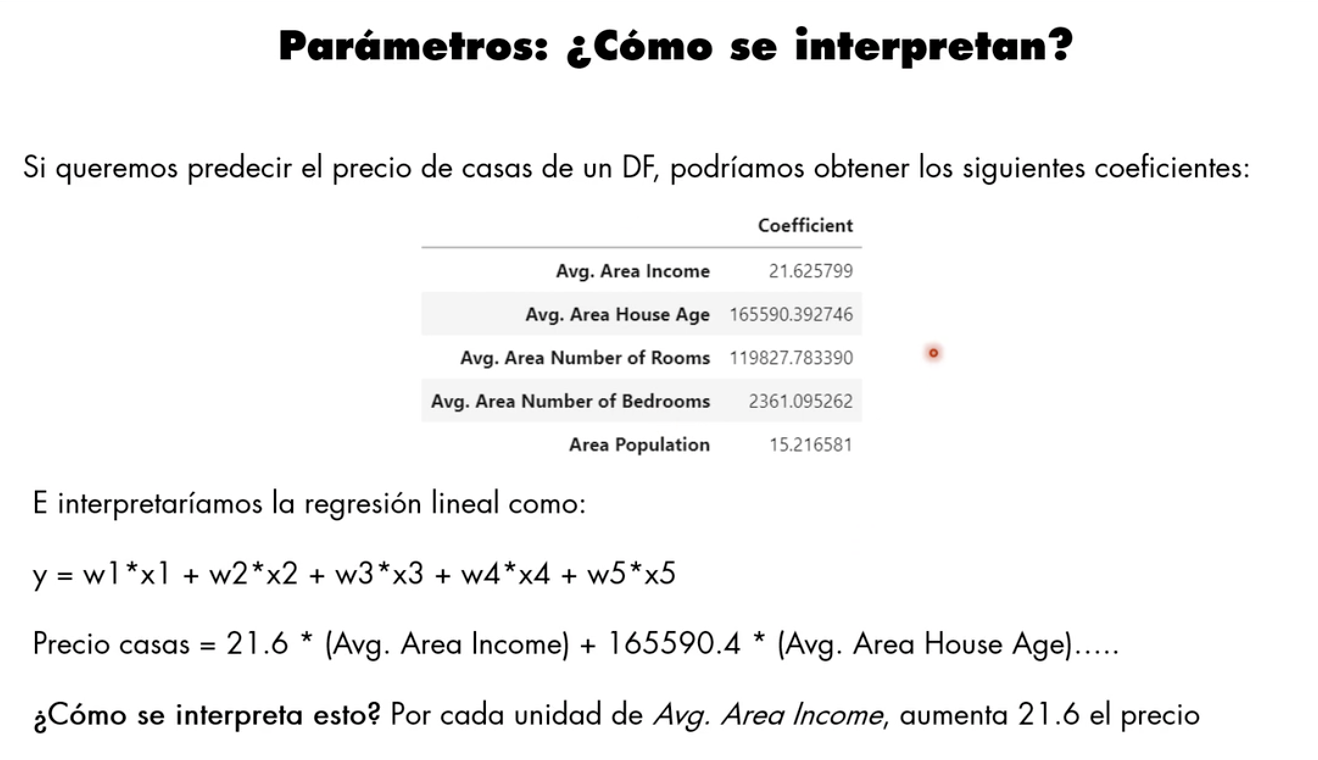
Los parámetros son aquellos coeficientes que multiplican a aquellas features para determinar la variable independiente/target. A las variables features también se les denomina predictores.

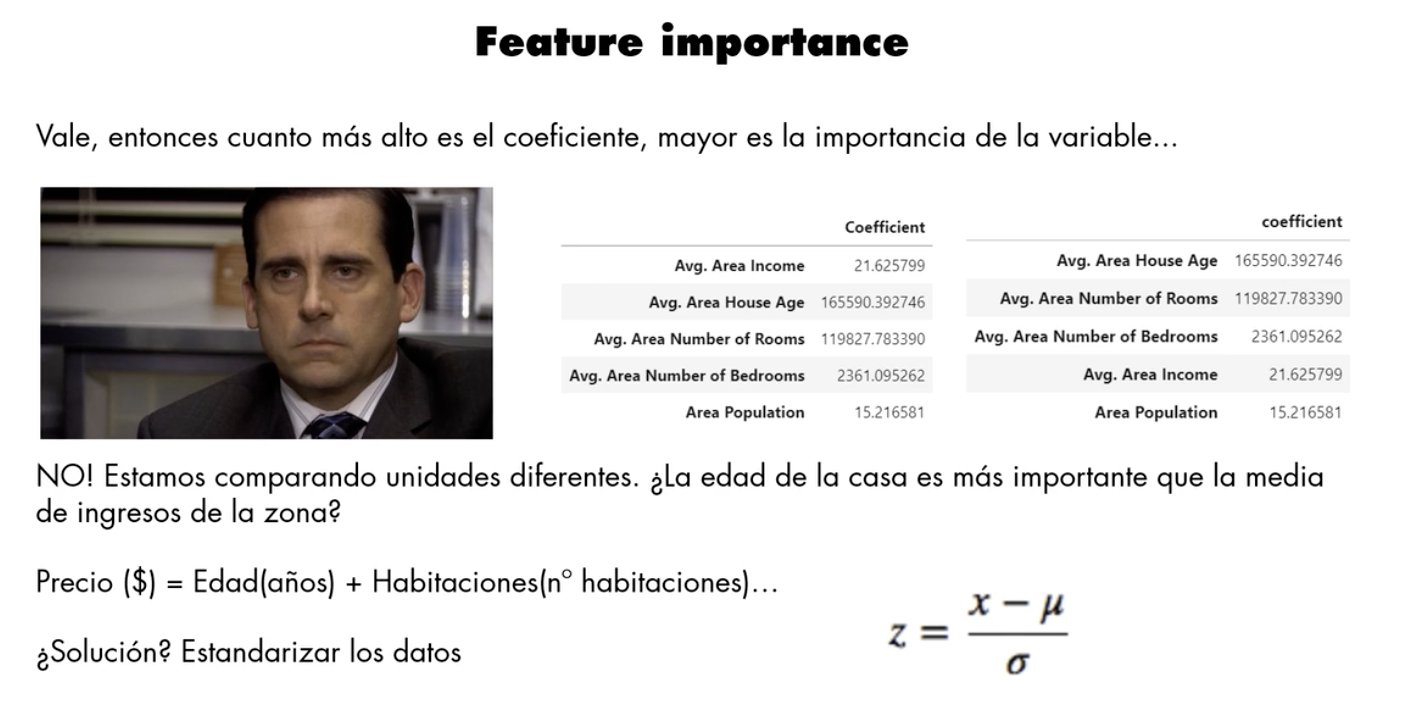
El beta0 determina la intersección en el eje Y mientras que el resto nos dice el grado de variación de nuestro target según las features que tengamos.



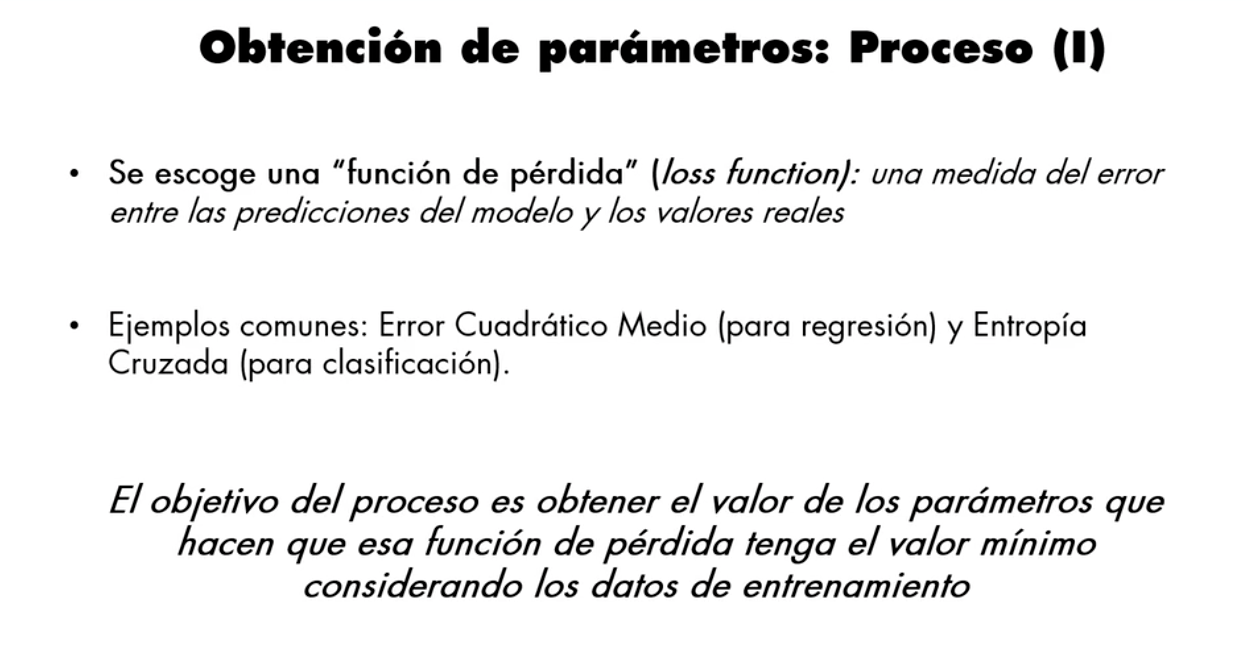
Los coeficientes determinan el peso de una feature en la estimación del target.

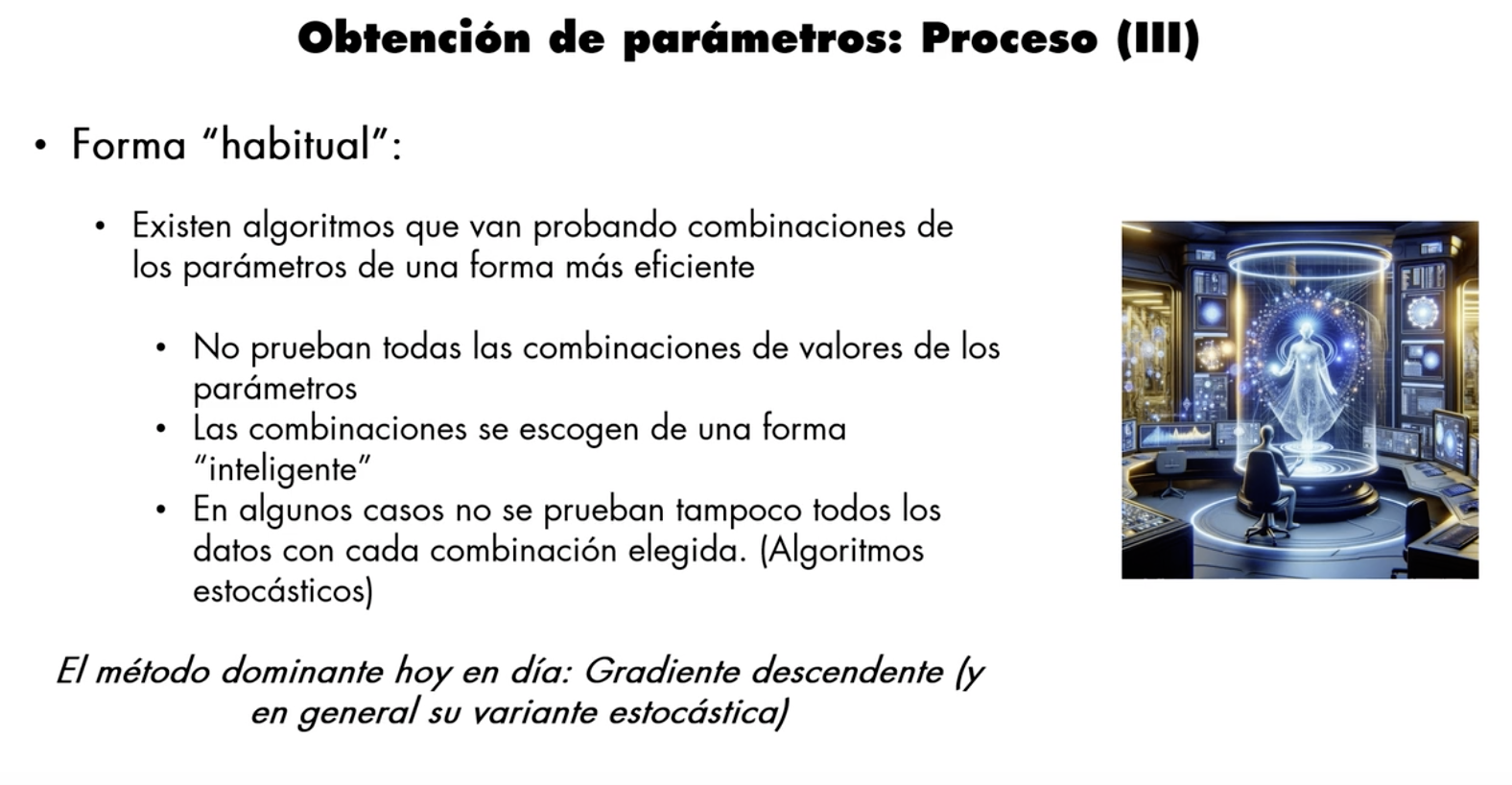
En la regresión multiple que se presenta abajo.. por cada unidad invertida de tv vamos a ganar 50 euros.

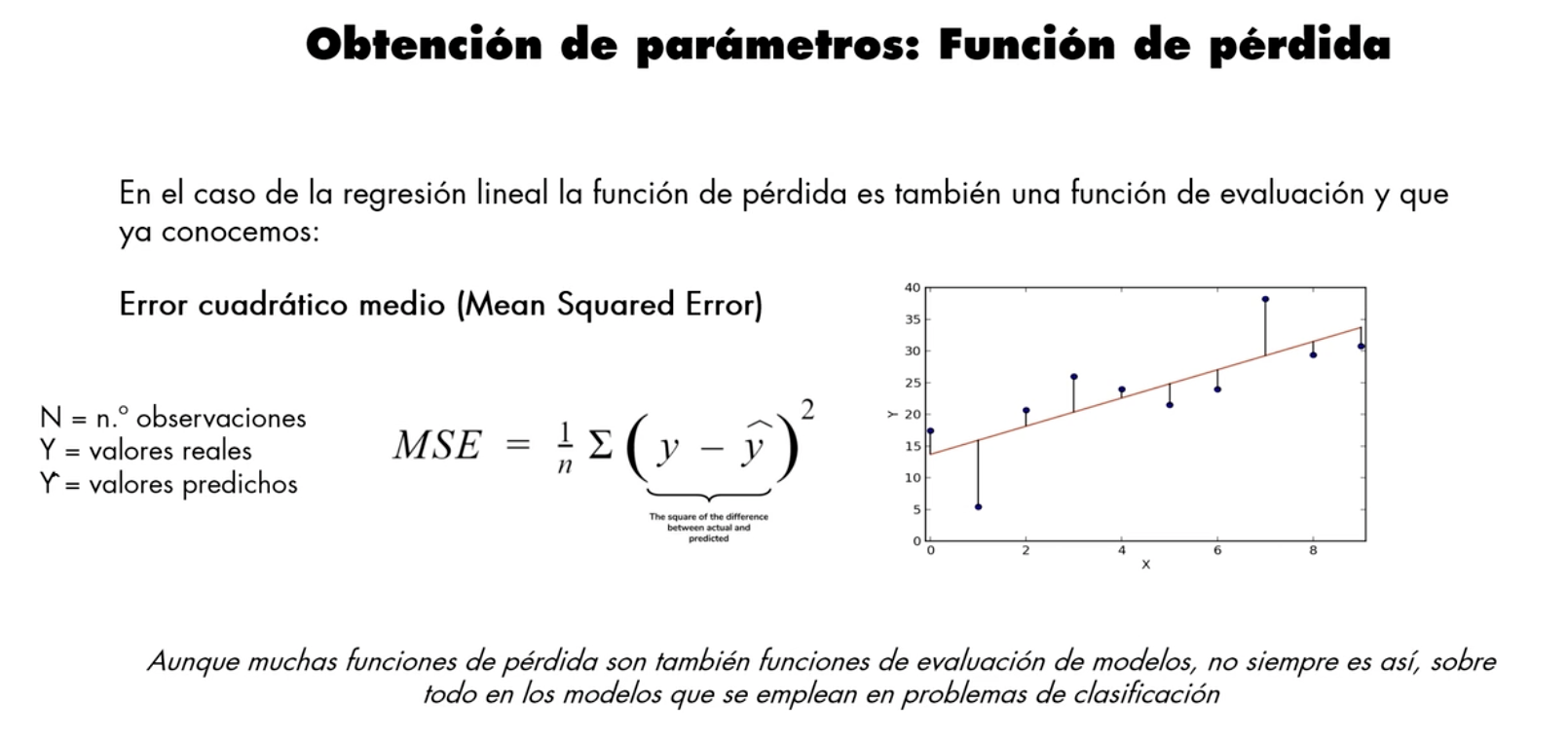


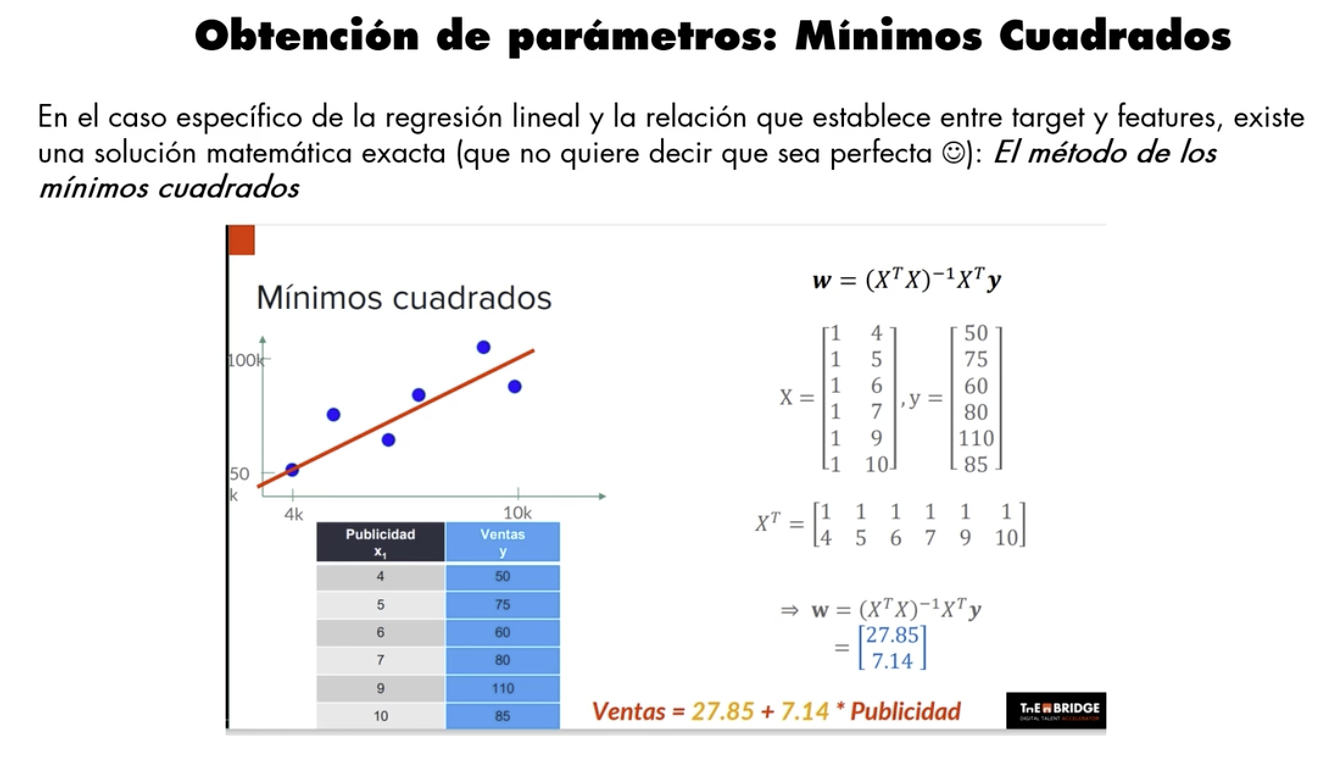


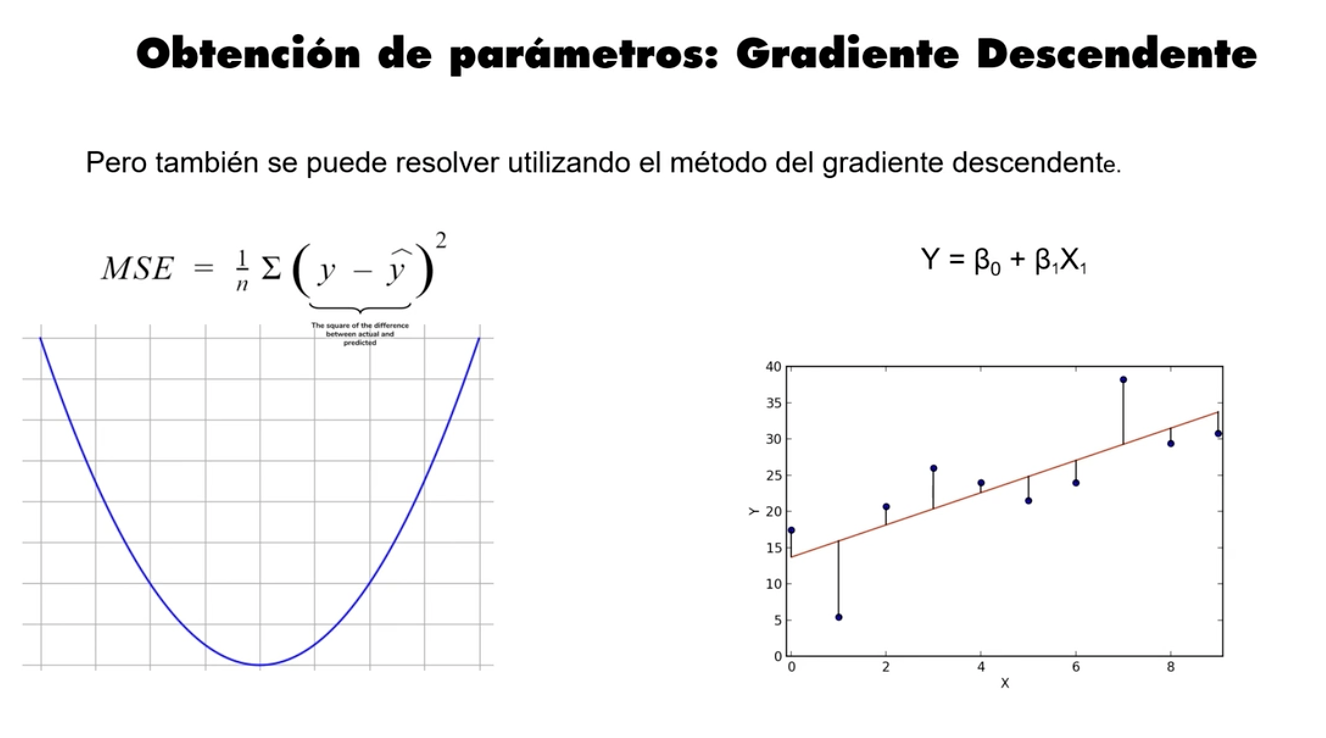


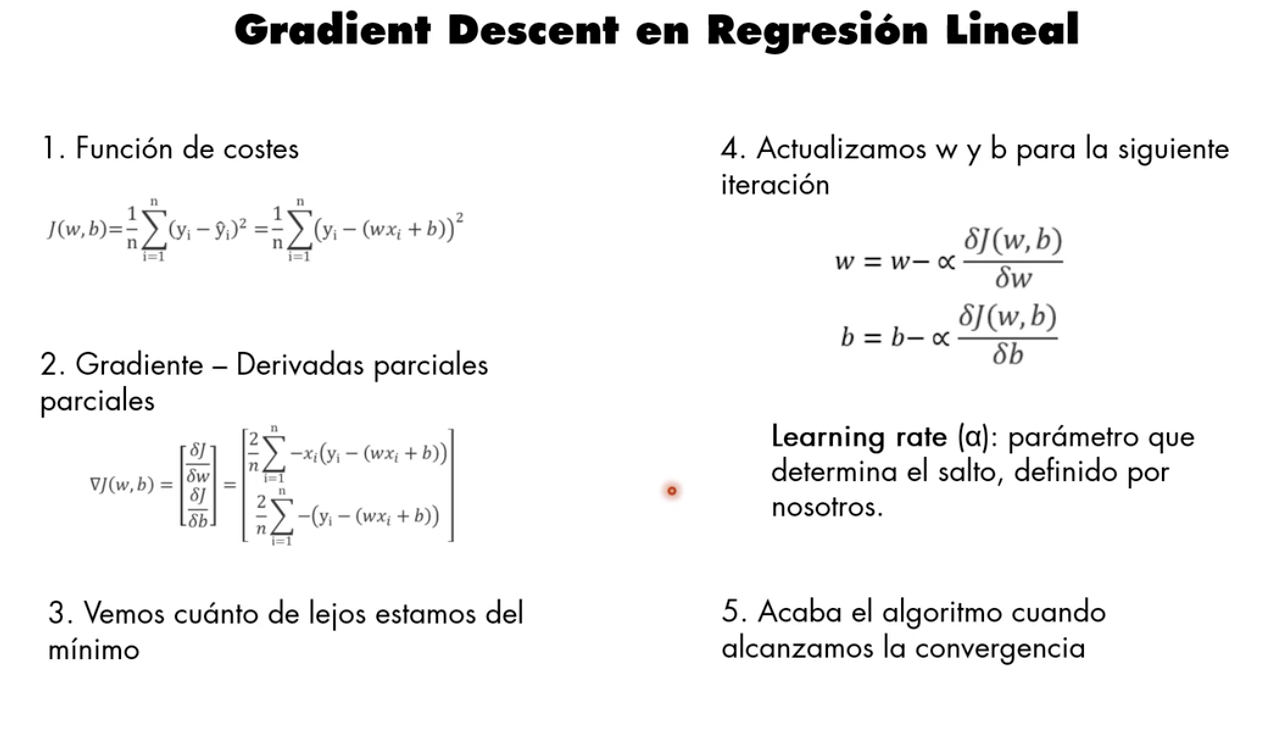












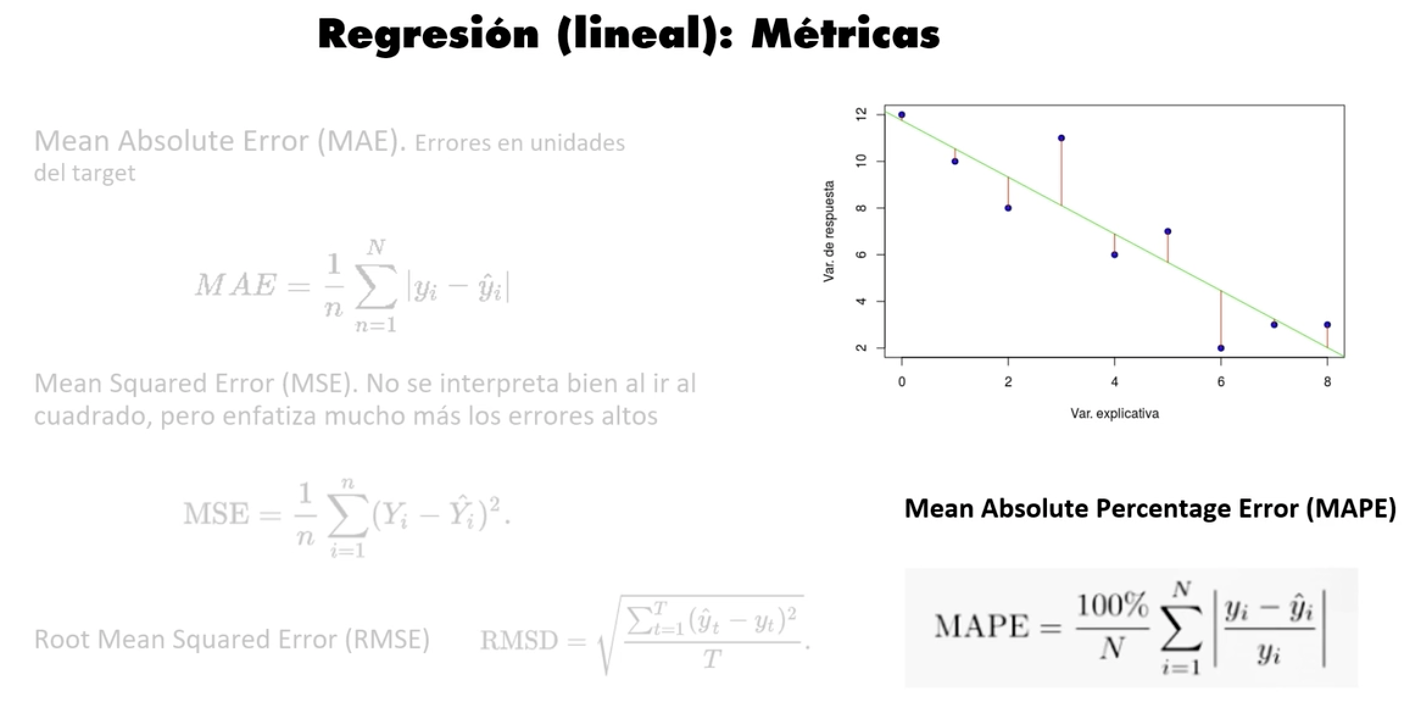
Se elije una función de perdida, no necesariamente la función de coste del error cuadratico medio..

Se calcula el gradiente a fin de sustituir lo que el modelo estima (en este caso es la función de regresión lineal)

Se hace una estimación de las derivadas parciales de esa función de costes.

Y luego aplico a w y b valores para que vayan disminuyendo en las funciones del punto 4, que esta definida como el learning rate.

Lo importante es escoger un buen learning rate para que el algoritmo no falle y elegir bien los valores de los parámetros.



Metricas de evaluación para regresión.

MAE error absoluto medio: Util cuando queremos medir errores en unidades de la variable target. Te dice que error puedes esperar en la predicción de la variable. Muy bueno cuando hay outliers también. Ya que lo que vamos a ver, lo vamos a ver en la medida que corresponda.

MSE Error cuadratico medio: Toma los errores los eleva al cuadrado y luego hace el promedio de los mismos. Esto significa que penaliza mas los errores grandes. Es muy útil para optimización. Mas que una metrica, se usa mucho como una función de perdida.

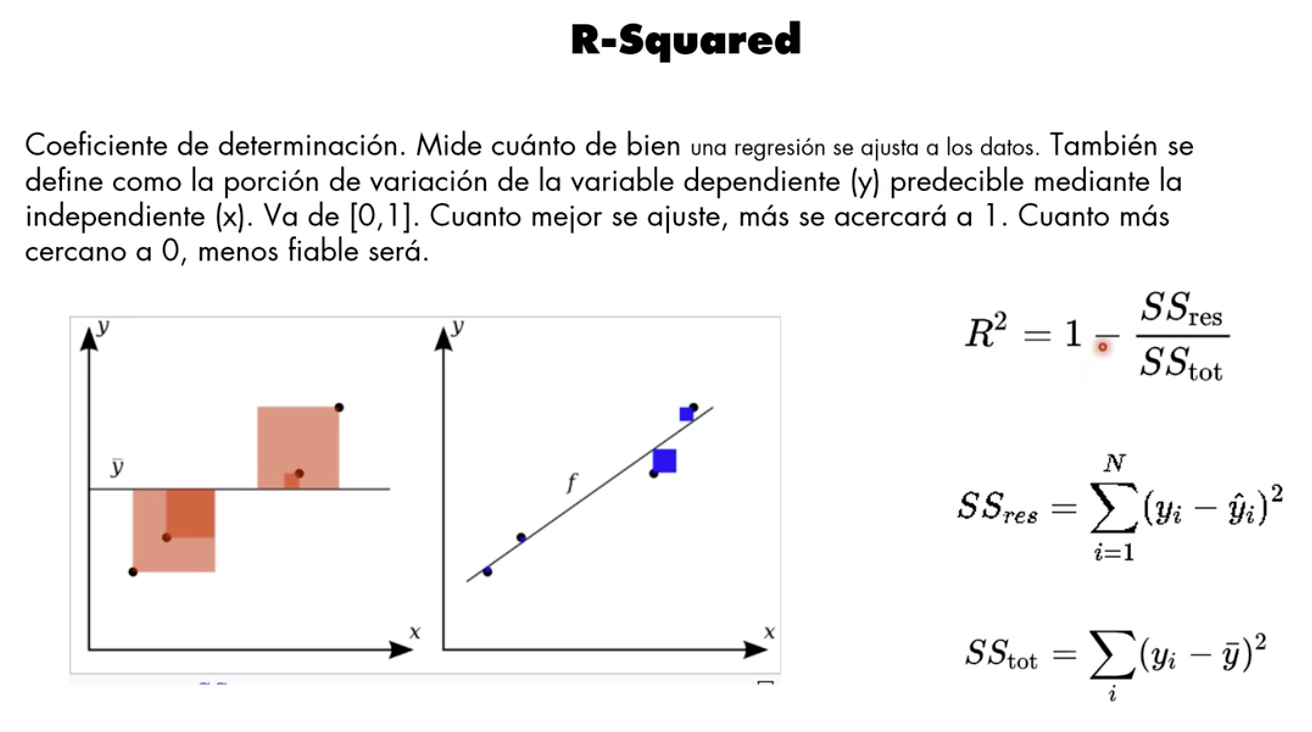
RMSE La raíz del error cuadratico medio: El que realmente se utiliza como medida de evaluación ya que además de panealizar los errores grandes, tiene la ventaja de estar en la misma unidad que nuestro objetivo. Es lo que mas utilizaremos en los problemas de regresión.

MAPE Error medio absoluto en porcentaje: Metrica acotada entre 0 y 1. Proporciona en media el x porciento de erro para cada punto respecto su valor real (bastante interesante).

Cuando queremos comparar modelos con diferentes targets es interesante usar el MAPE como por ejemplo en series temporales. Otras métricas no nos sirven porque van en función de cada target. Con el MAPE se puede comparar con la misma unidad de medida.

R Cuadrado o coeficiente de determinación:

Cuanto mas se acerque a 1 mejor es nuestro modelo, y lo contrario cuando se acerque a 0.



Condiciones para aplicar la regresión lineal.

