Regresión

DRA. CONSUELO VARINIA GARCÍA MENDOZA

Limitaciones del algoritmo de gradiente descendiente

El algoritmo de gradiente descendiente permite ajustar los pesos de forma iterativa

En cada iteración del cálculo del gradiente es necesario recorrer todos los puntos del conjunto de datos

Para conjuntos de datos grandes esto puede ser muy costoso y el algoritmo de gradiente descendiente se vuelve muy lento

La versión del algoritmo que considera todos los puntos en su cálculo se llama gradiente descendiente por lotes (BGD)

$$w := w - \alpha \frac{\partial}{\partial f(w)} f(w)$$

$$\frac{\partial}{\partial f(w)} f(w) = \sum_{i=1}^{m} 2(w \cdot x - y) \cdot x$$

Gradiente descendente estocástico (SGD)

Existe una variante del algoritmo que considera sólo un punto para el ajuste de pesos en cada iteración

Con este único punto se ajustan los pesos y se intenta una aproximación al óptimo

Este algoritmo reduce el número de operaciones que se realizan en cada iteración por lo que puede manejar grandes cantidades de datos

Sin embargo, se debe considerar que no siempre se encontrarán los pesos óptimos, pero si una buena aproximación

Gradiente descendente

$$w_i = w_i - \alpha \frac{\partial}{\partial w_i} f(w)$$

BGD

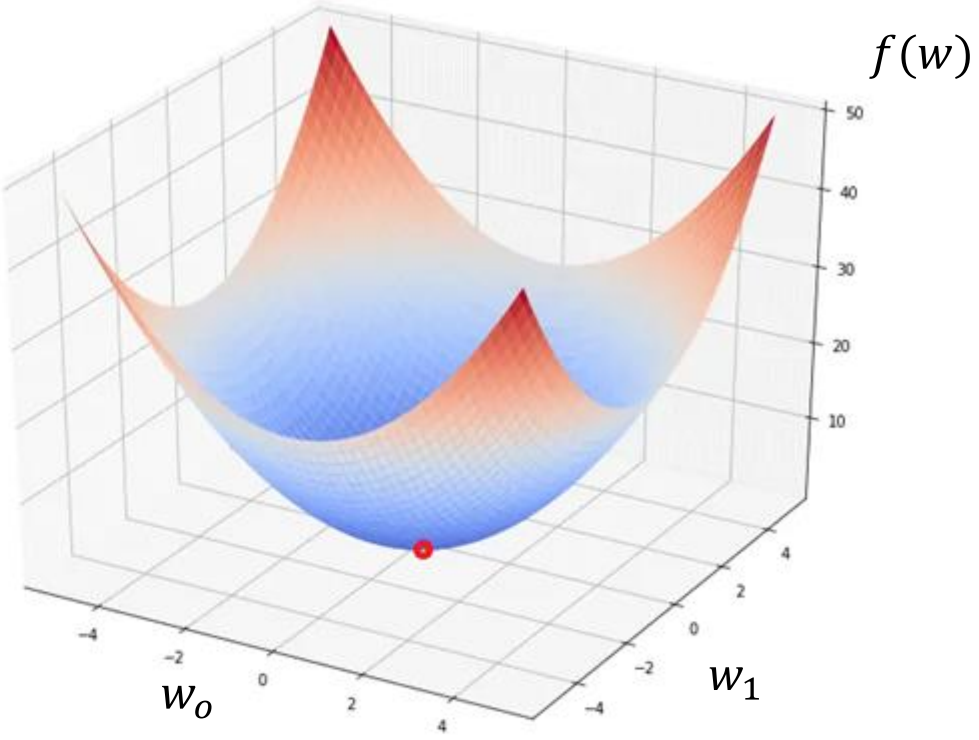
$$f(w) = \sum_{i=1}^{m} (w \cdot x - y)^2$$

$$\frac{\partial}{\partial f(w)} f(w) = \sum_{i=1}^{m} 2(w \cdot x - y) \cdot x$$

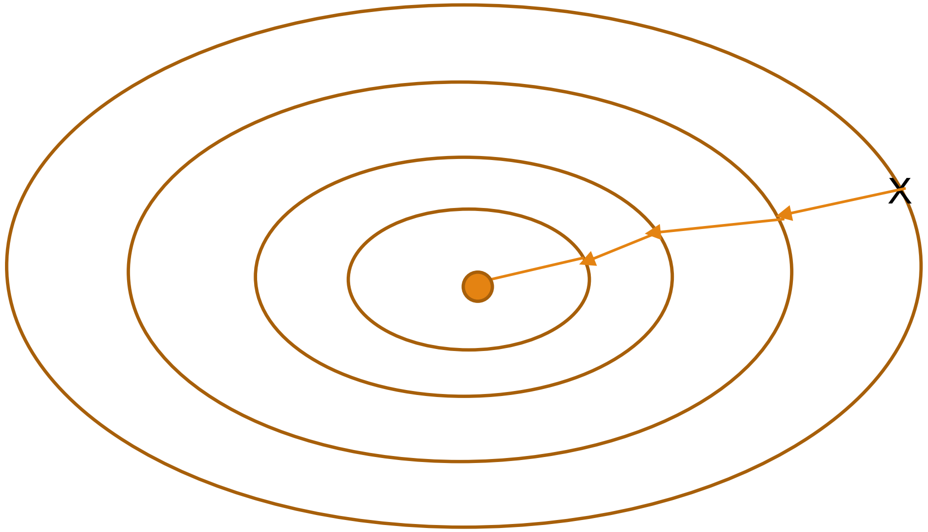
SGD

$$f(w) = (w \cdot x - y)^2$$

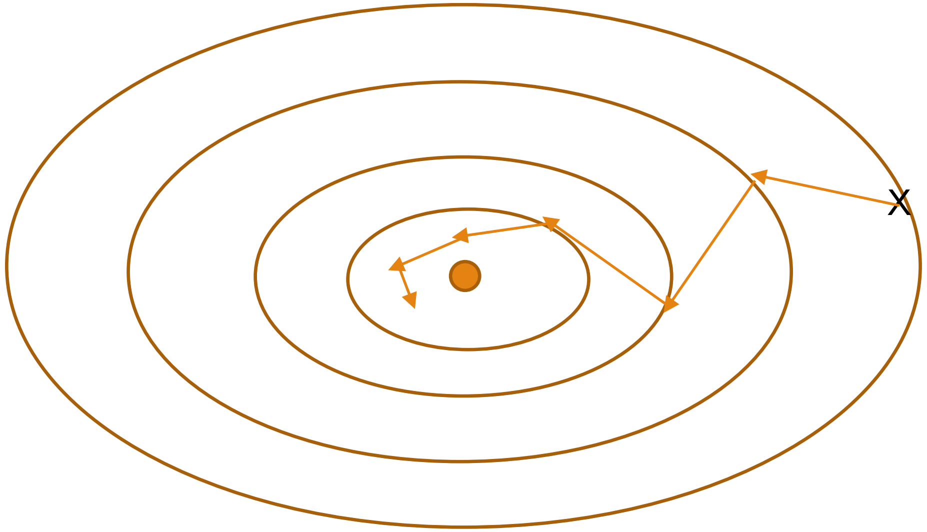
$$\frac{\partial}{\partial f(w)} f(w) = 2(w \cdot x - y) \cdot x$$



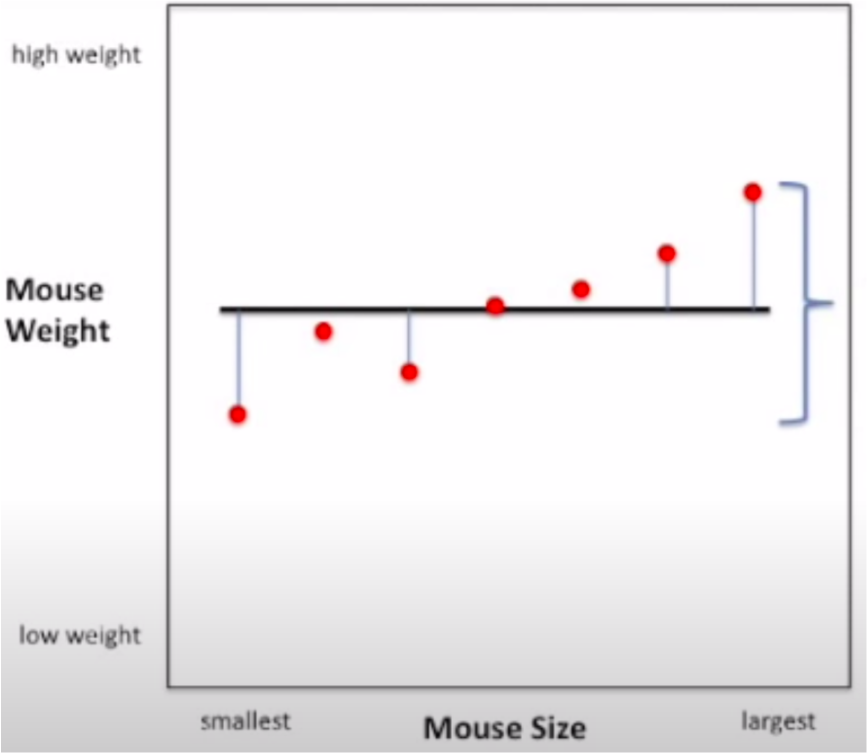
BGD



SGD

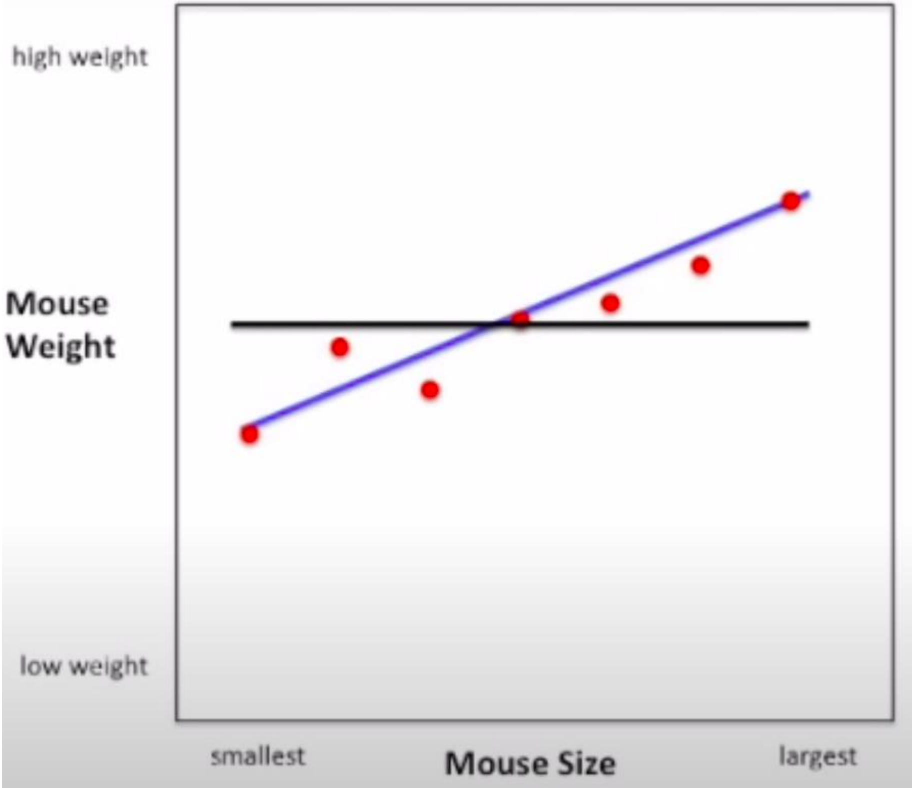


Diferencias usando el promedio de los valores



$$\frac{\sum_{I=1}^{n} |y_i - \hat{y}_i|}{n}$$

Promedio vs línea ajustada



Coeficiente de determinación R^2

El coeficiente de determinación, también llamado R cuadrado, refleja que tan bueno es el ajuste de un modelo con respecto a la variable que pretender explicar

Es la proporción de la varianza total de la variable explicada por la regresión

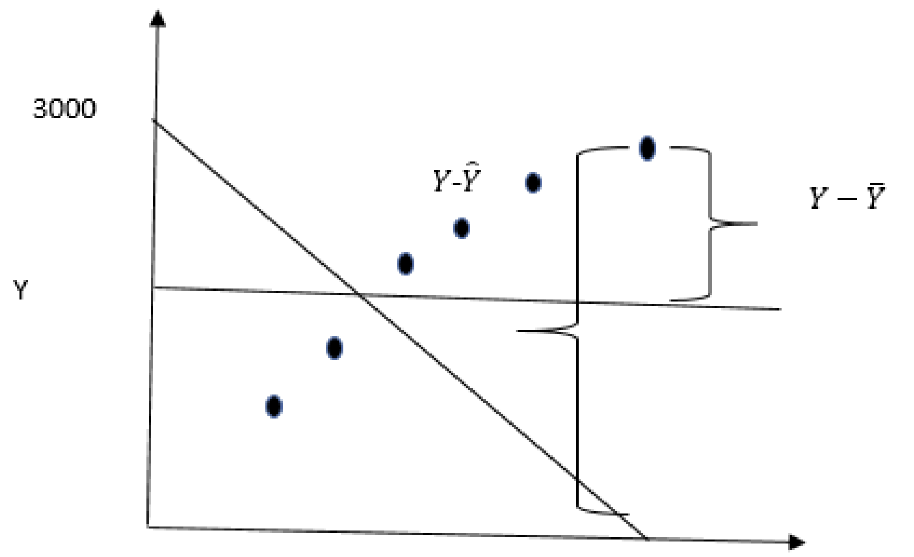
Este coeficiente toma valores entre 0 y 1, cuanto más cerca de 1 está, mejor será el ajuste del modelo

De forma inversa, cuanto más cerca de cero, menos ajustado estará el modelo y, por tanto, menos fiable será

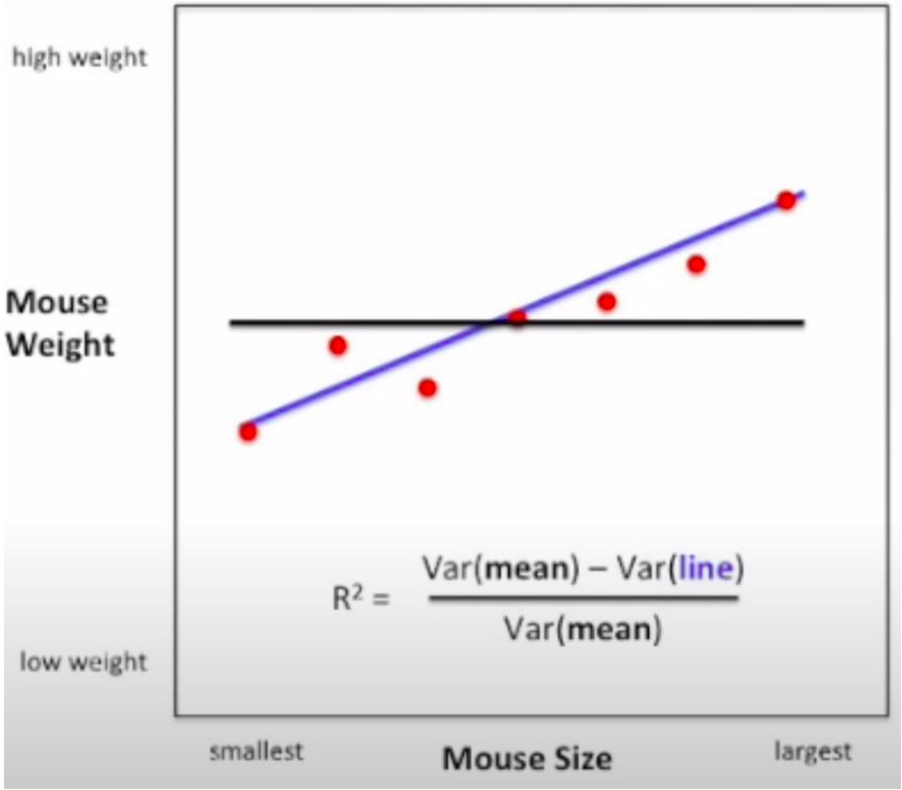
Coeficiente de determinación R^2

El valor del coeficiente también puede tomar valores negativos cuando el modelo creado es arbitrariamente incorrecto

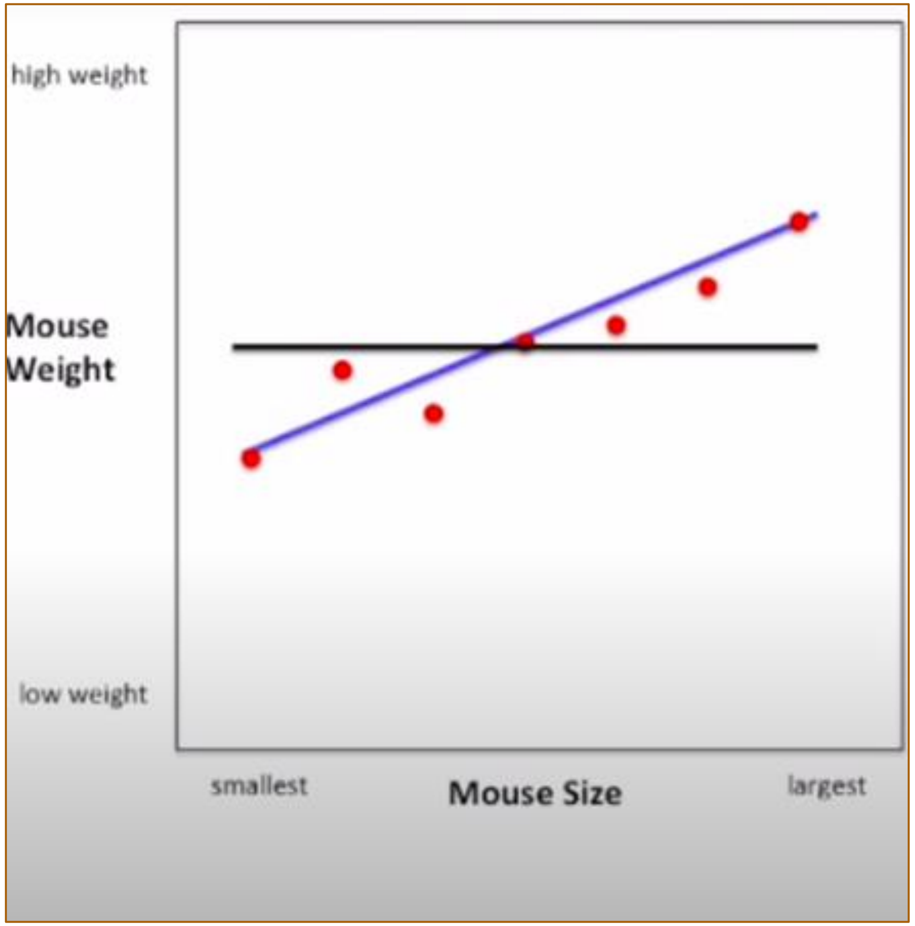
Esto se manifiesta con líneas que no siguen la tendencia de los datos



Coeficiente de determinación R^2



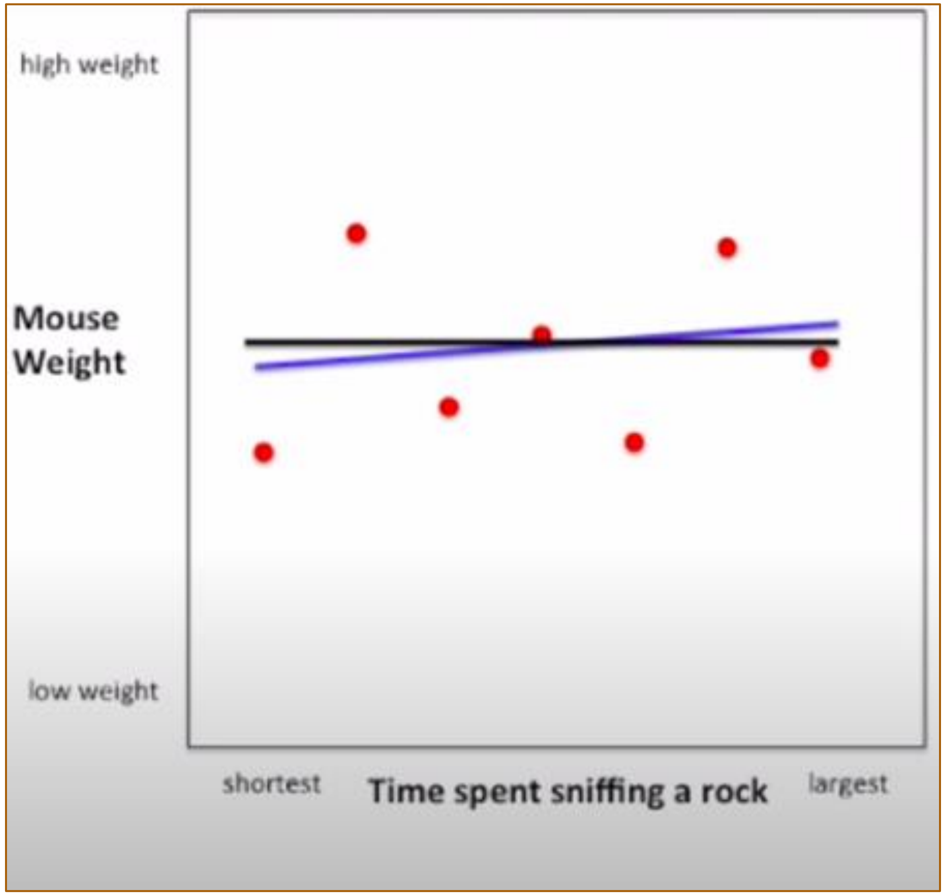
Ejemplo 1. Coeficiente de determinación R^2



$$\text{Var}(\textbf{mean}) = 32$$
$$\text{Var}(\textbf{line}) = 6$$
$$R^2 = \frac{\text{Var}(\textbf{mean}) - \text{Var}(\textbf{line})}{\text{Var}(\textbf{mean})}$$
$$R^2 = \frac{32 - 6}{32}$$
$$R^2 = \frac{26}{32} = 0.81 = 81\%$$

- Existe un 81% menos de variación entre la línea ajustada y el promedio
- La variable independiente y la salida tienen una alta correlación

Ejemplo 2. Coeficiente de determinación R^2



$$\text{Var}(\textbf{mean}) = 32$$

$$\text{Var}(\textbf{line}) = 30$$

$$R^2 = \frac{\text{Var}(\textbf{mean}) - \text{Var}(\textbf{line})}{\text{Var}(\textbf{mean})}$$

$$R^2 = \frac{32 - 30}{32}$$

$$R^2 = \frac{2}{32} = 0.06 = 6\%$$

- Sólo hay un 6% menos de variación entre la línea ajustada y el promedio
- La variable independiente y la salida no estás correlacionadas

Regresión lineal multivariable

La regresión lineal univariable puede ser extendida para manejar múltiples variables

La función de hipótesis *h* para una instancia en particular (*j*) se expresa mediante la siguiente ecuación

$$h_w(x_j) = w_0 + w_1 x_{j,1} + \dots + w_n x_{j,n} \qquad h_w(x_j) = w_0 + \sum_{i}^{n} w_i x_{j,i}$$

- Si se asigna un valor ficticio (dummy) de 1 a $x_{j,0}$ que multiplica a w_0. De esta forma se puede generalizar la ecuación de la siguiente manera

$$h_w(x_j) = \sum_{i}^{n} w_i x_{j,i}$$

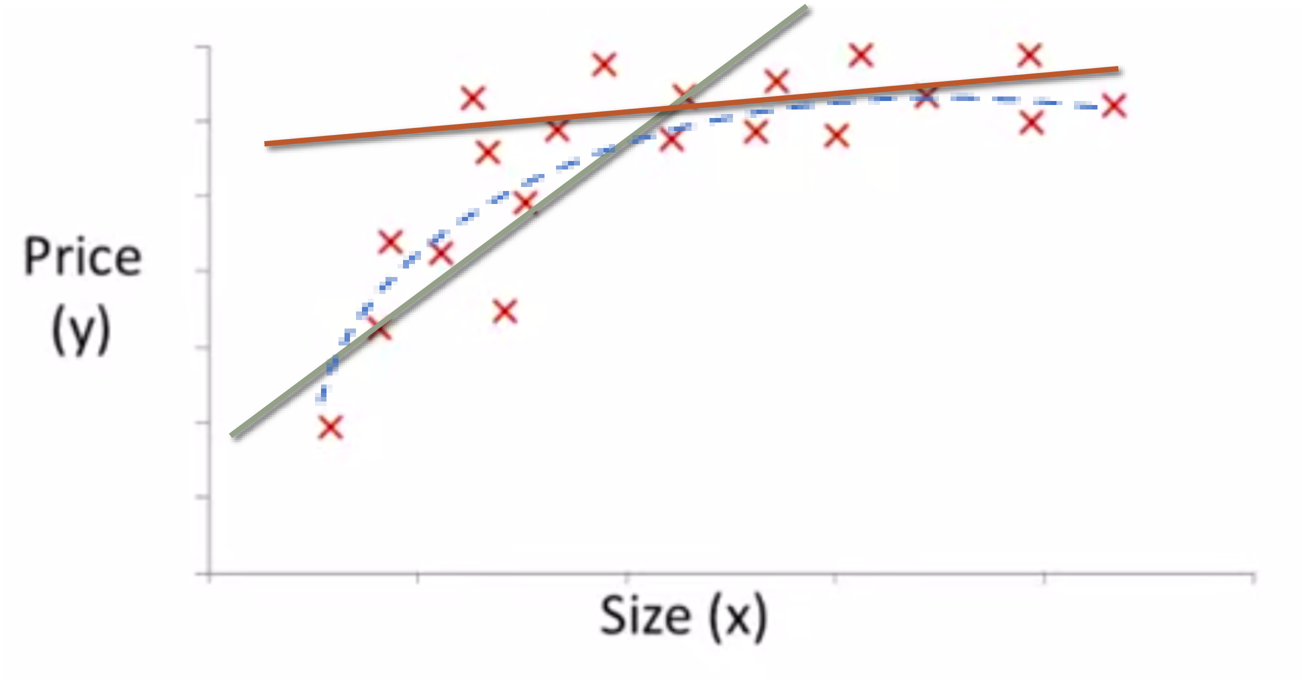
Limitaciones de la regresión lineal

Los modelos creados mediante regresión vistos hasta ahora presentan un buen desempeño cuando los datos siguen una tendencia lineal

Sin embargo, pueden existir datos que sigan una tendencia no lineal

Con este tipo de datos los modelos de regresión lineal obtendrán un bajo desempeño

Datos con tendencia no lineal



$$w_0 + w_1 x + w_2 x^2$$

Regresión polinomial

La regresión polinomial permite generar modelos que se ajusten a datos con tendencia no lineal

En particular esta regresión ayuda cuando se aprecia una tendencia curvilínea entre las variables y el dato a predecir

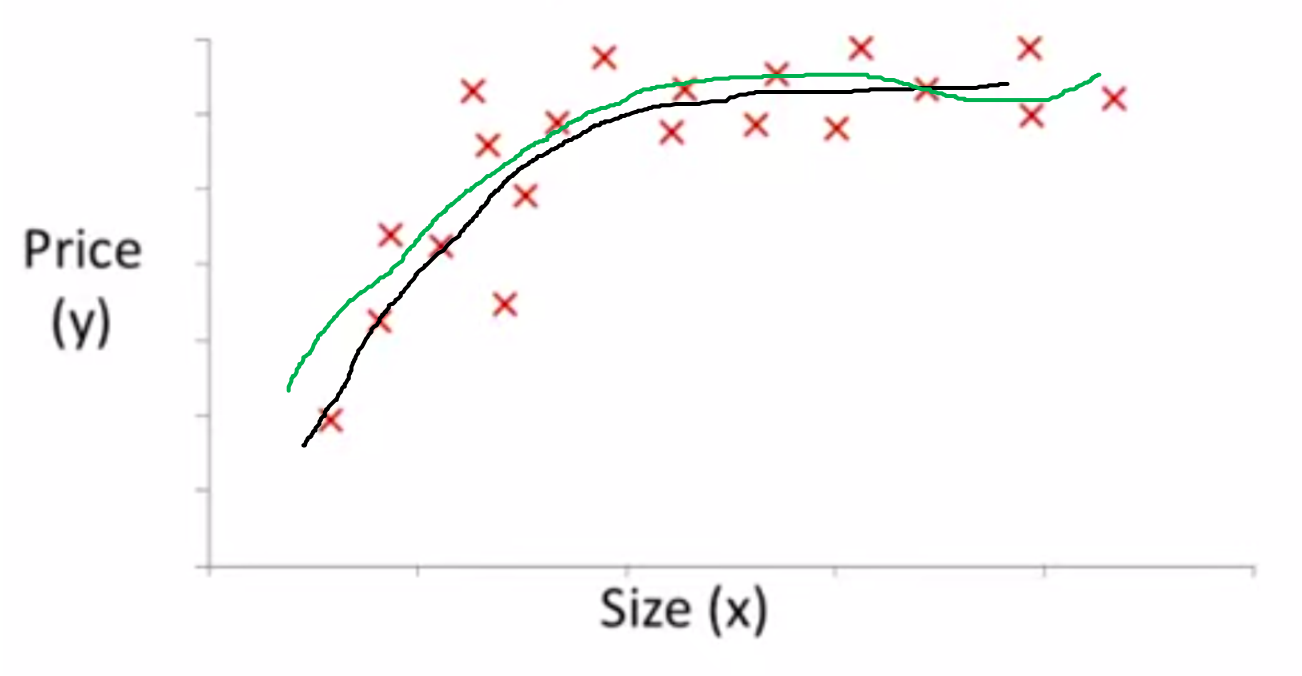
Es importante aclarar que la regresión polinomial, en un sentido estricto, se sigue considerando un problema de estimación estadística lineal

Lo anterior se debe a que la función de regresión sigue siendo lineal a los parámetros (pesos) que se desean estimar a partir de los datos

Es por eso que la regresión polinomial se considera un caso especial de la regresión lineal multivariable

Regresión polinomial

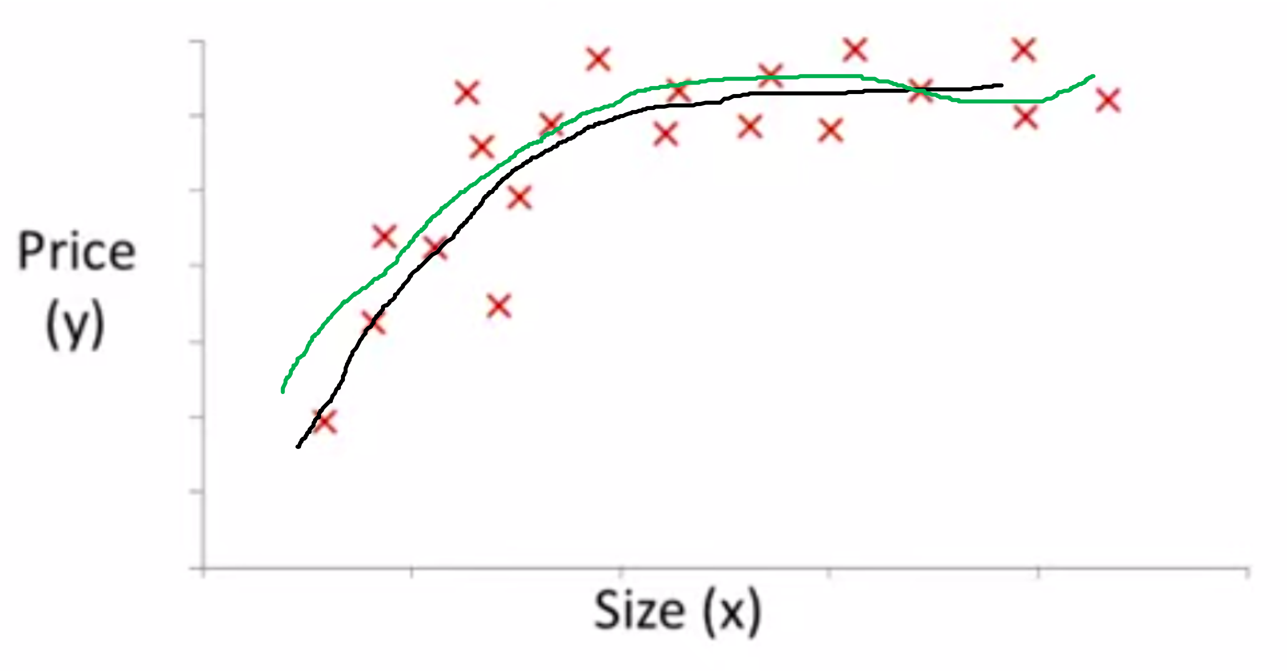
En la regresión polinomial se puede modificar el grado del polinomio para ajustarse a distintas tendencias de los datos



$$w_0 + w_1 x + w_2 x^2$$

$$w_0 + w_1 x + w_2 x^2 + w_3 x^3$$

Regresión polinomial



$$H_w(x) = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_3$$

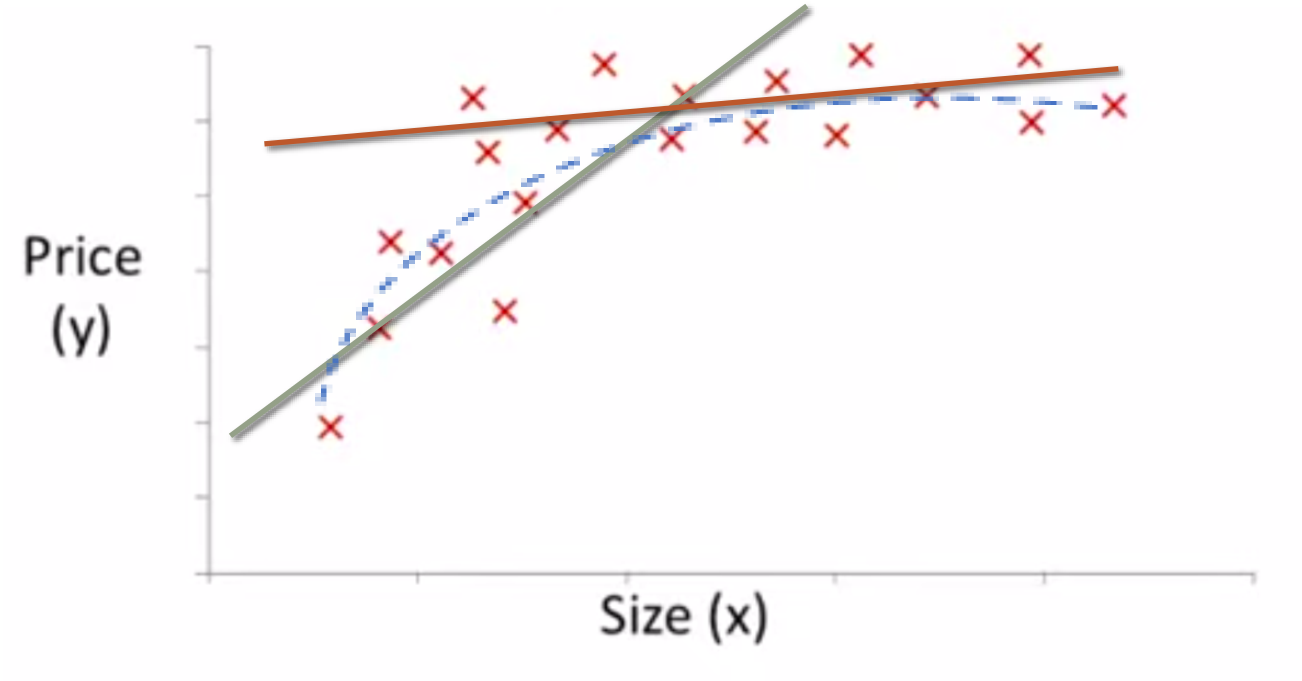
$$= w_0 + w_1(size) + w_2(size)^2 + w_3(size)^3$$

$$x_1 = (size)$$

$$x_2 = (size)^2$$

$$x_3 = (size)^3$$

Datos con tendencia no lineal



$$w_0 + w_1 x + w_2 x^2$$

Escalamiento de los datos

Algo a considerar en la regresión polinomial es que al elevar a diferentes potencias las variables, sus valores crecerán demasiado

Esto puede traer como consecuencia que rebasen la capacidad de precisión que se puede manejar en la computadora

Por ejemplo:

- x = 1-1,000
- X^2 = 1 - 1,000,000
- X^3 = 1 - 1,000,000,000

Además, algunos algoritmos de aprendizaje automático como regresión y redes neuronales muestran sesgo hacia las variables con valores muy grandes

Escalamiento de datos

El escalamiento de los datos nos permite establecer un mismo rango de valores para las diferentes las variables

De esta forma se evita que existan variables con valores muy grandes y otras con valores muy pequeños

Existen varias técnicas de escalamiento de datos, algunos utilizan la media como medida base para definir el rango, otros utilizan los valores mínimos y máximos para establecer un umbral