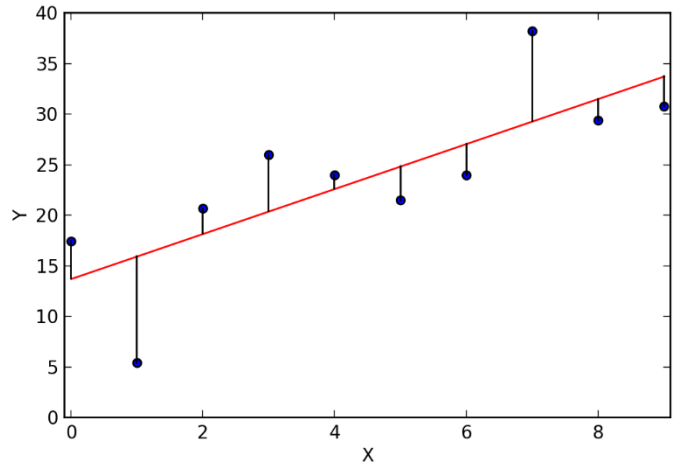
MACHINE LEARNING APLICADO

Problemas de Regresión

UNIVERSIDAD SIMÓN BOLÍVAR
BARRANQUILLA Y CÚCUTA - COLOMBIA | VIGILADA MINEDUCACIÓN
FACULTAD DE INGENIERÍAS

Regresión

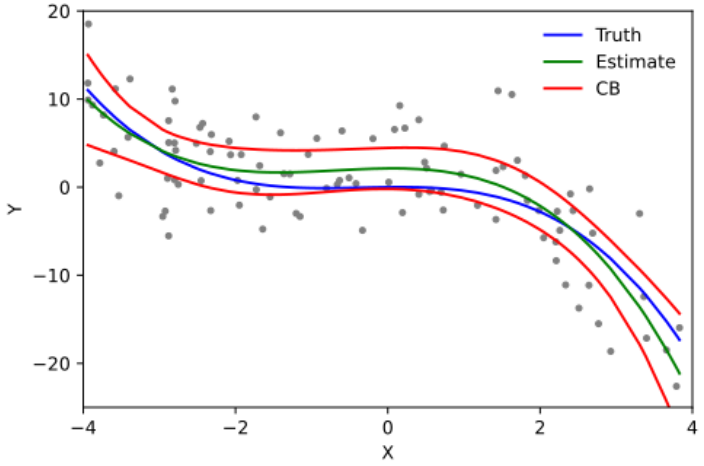
- Es un enfoque de aprendizaje supervisado
- Consiste en predecir el valor de una variable continua en base a los valores de otras variables.
- La variable dependiente se suele representar como Y y las variables independientes como X_i.
- Se crea un modelo o función matemática a partir de los datos históricos



Tipos de Regresión

- Regresión Simple: existe una sola variable independiente X.
  - Regresión Lineal Simple
  - Regresión No-Lineal Simple

- Regresión Múltiple: existen varias variables independientes $X_{i.}$
  - Regresión Lineal Múltiple
  - Regresión No-Lineal Múltiple



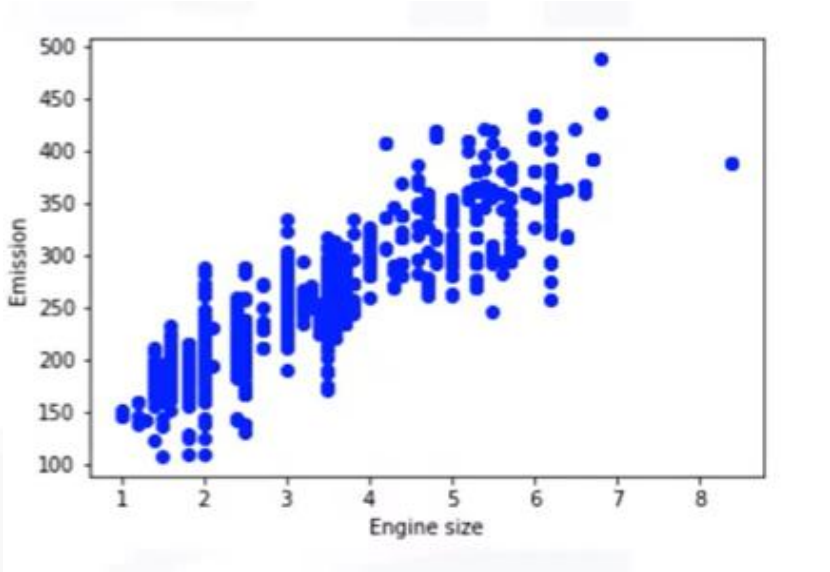
Aplicaciones de Regresión

- Predicción de Ventas
- Predicción del Precio de un Producto(Casas, u otros)
- Predecir salarios
- Predecir cantidad de lluvia al dia siguiente
- Predecir el valor de la presión arterial en base a otras variables fisiológicas.

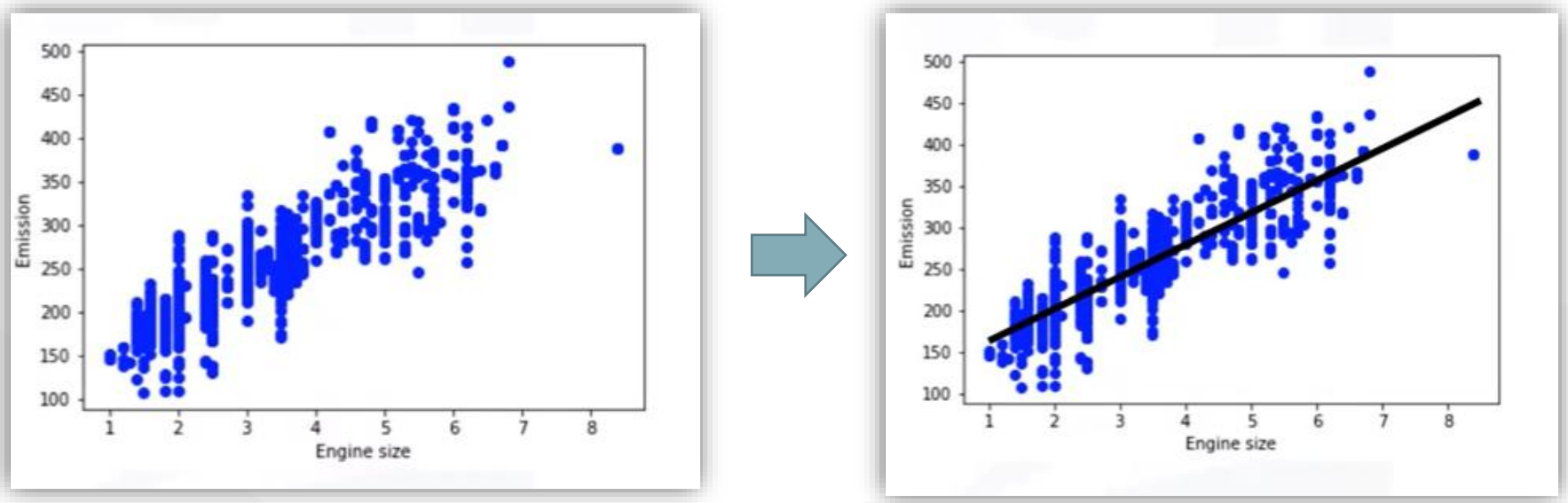
Regresión Lineal Simple

- Involucra solo dos variables: la variable dependiente Y y una variable independiente X. La variable Y que se quiere predecir es continua.

- Para funcionar correctamente las variables deben tener una relación lineal.



Regresión Lineal Simple



Regresión Lineal Simple

- Una regresión lineal simple tiene la forma:

$$\hat{Y} = \theta_0 + \theta_1 x$$

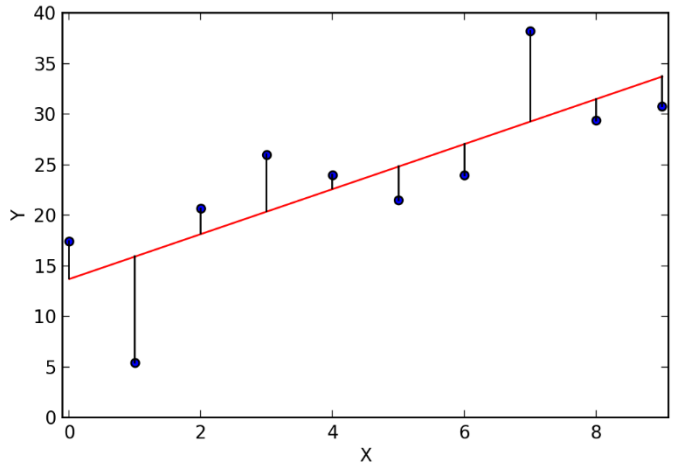
- $\hat{Y}$: Variable dependiente u objetivo
- θ_0: Intercepto
- θ_1: Pendiente o Coeficiente
- x: Variable Independiente

La regresión lineal busca calcular los valores de Θ_0 y Θ_1 que describan la línea que mejor se ajusta al conjunto de datos.

Regresión Lineal Simple

¿Cómo se haya la mejor línea?

Se elige la línea que minimice los errores, es decir el MSE:



Error = Y – Ŷ

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

Regresión Lineal Simple

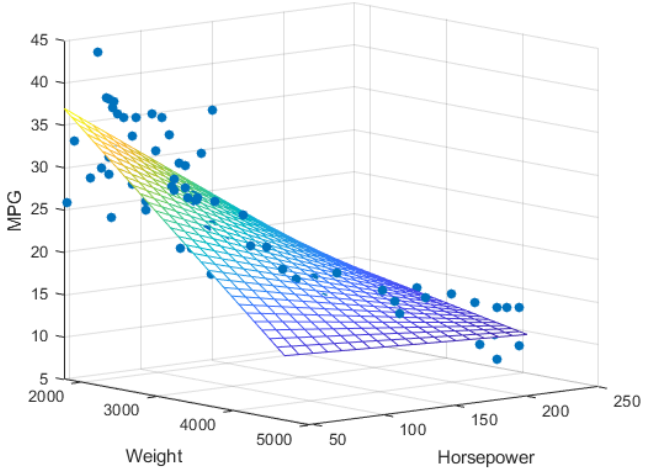
- Para la regresión lineal simple se pueden utilizar las siguientes ecuaciones para hallar Θ_0 y Θ_1.

$$\theta_1 = \frac{\sum_{i=1}^{s}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{s}(x_i - \bar{x})^2}$$

$$\theta_0 = \bar{y} - \theta_1 \bar{x}$$

Regresión Lineal Múltiple

- Es una regresión lineal que involucra varias variables independientes.
- La regresión lineal múltiple permite evaluar el impacto de cada una de las variables independientes sobre el valor de la variable objetivo y predecir como impactaría un cambio en una de estas.
- Un modelo de regresión lineal múltiple tiene la forma:



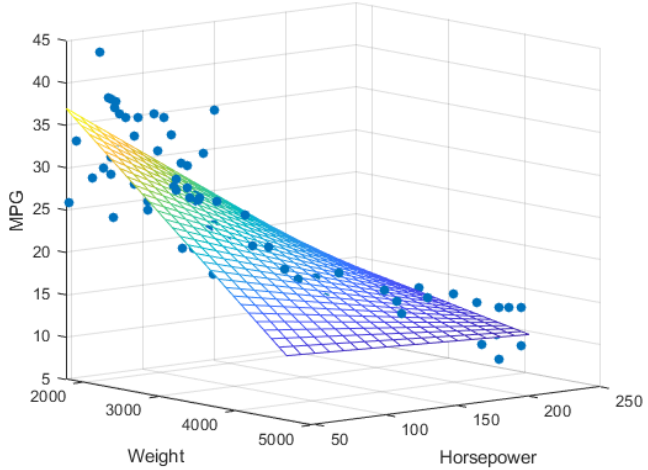
$$\hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_n x_n$$

Regresión Lineal Múltiple

¿Cómo calcular los coeficientes θ ?

- Mínimos cuadrados ordinarios
  - Operaciones de Algebra Lineal
  - Toma bastante tiempo para datasets grandes

- Algoritmo de Optimización
  - Descenso del Gradiente
  - Más apropiado para datasets grandes



Métricas de Evaluación

Mean Absolute Error

$$MAE = \frac{1}{n}\sum_{j=1}^{n}|y_j - \hat{y}_j|$$

Mean Squared Error

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

Root Mean Squared Error

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}$$

Métricas de Evaluación

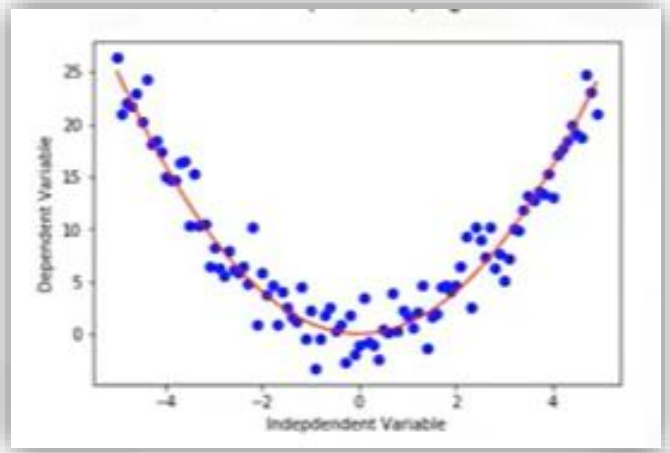
Relative Squared Error

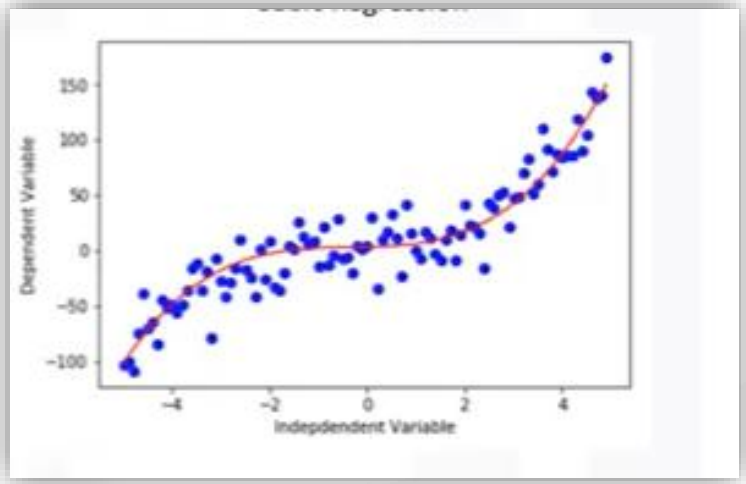
$$RSE = \frac{\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}{\sum_{j=1}^{n}(y_j - \bar{y})^2}$$

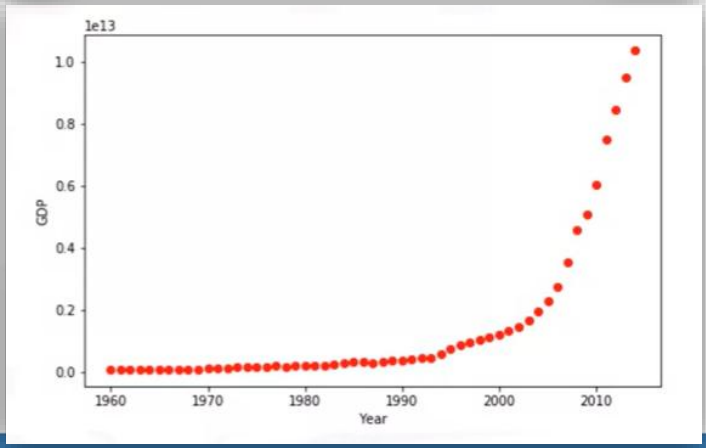
$$R^2 = 1 - RSE$$

- **Entre más cercano sea R^2 a 1, mejor será el ajuste del modelo**

Regresión No-Lineal





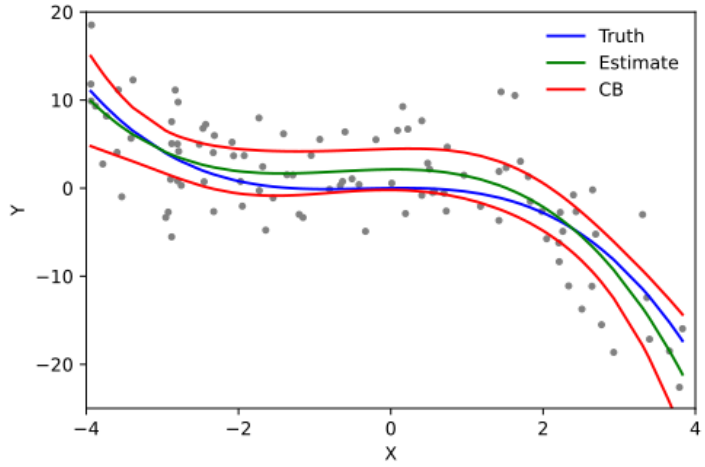


Regresión Polinómica

- Una regresión polinómica sigue una curva y tiene la forma de:

$$\hat{Y} = \boldsymbol{\theta_0} + \boldsymbol{\theta_1 x} + \boldsymbol{\theta_2 x^2} + \boldsymbol{\theta_3 x^3} \ldots . + \boldsymbol{\theta_n x^n}$$

- Un modelo de regresión polinómica puede ser expresado como un modelo de regresión lineal.



- $X_1 = X$
- $X_2 = X^2$
- $X_3 = X^3$

UNIVERSIDAD SIMÓN BOLÍVAR
BARRANQUILLA Y CÚCUTA - COLOMBIA | VIGILADA MINEDUCACIÓN

FACULTAD DE INGENIERÍAS

¿Preguntas?

Puedes publicar tu pregunta en el foro de Inquietudes y Novedades

O escribir al correo:
silvia.moreno@unisimon.edu.co

Referencias

- Géron, Aurélien. Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media. 2017.

- Hastie, Trevor, Tibshirani, Robert y Friedman, Jerome. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer. 2016

- Russell, Stuart J. Inteligencia artificial, un enfoque moderno. Segunda Edición. México, Prentice-Hall Hispanoamericana. 2013