



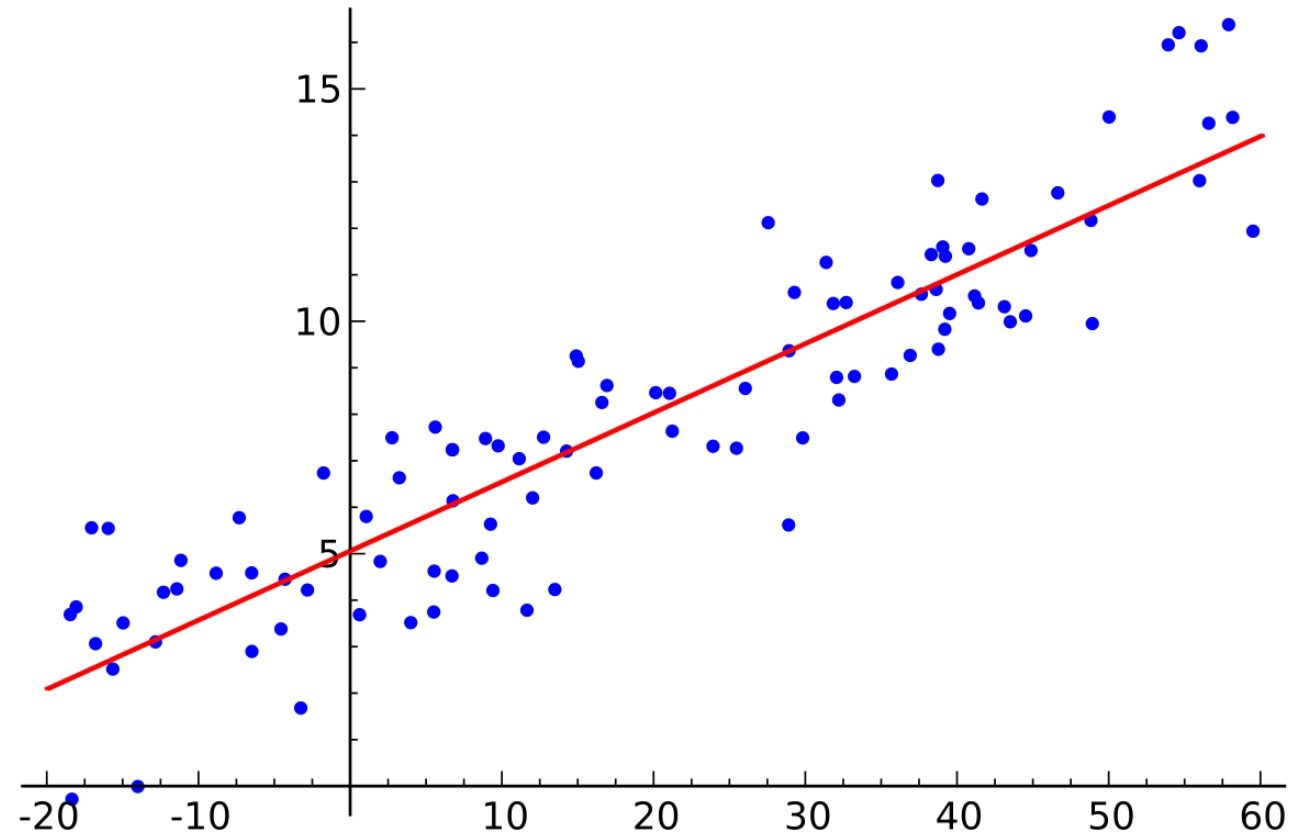
ESCUELA DE INGENIERÍA  
FACULTAD DE INGENIERÍA

# IIC2433 Minería de Datos

## Regresión lineal

Profesor: Mauricio Arriagada

# REGRESIÓN LINEAL



# OBJETIVO

- ▶ Entender cómo predecir variables numéricas
- ▶ Estudiar regresiones lineales simples y múltiples

# Algunas aplicaciones

- ▶ Electricidad
  - ▶ obtener el valor de una resistencia en un circuito y su error mediante un ajuste de regresión lineal
- ▶ Sensores
  - ▶ Calibración de un sensor de temperatura en función de la caída de tensión y la temperatura
- ▶ Construcción
  - ▶ Para caracterizar diversas cualidades del hormigón. A partir del módulo de elasticidad es posible predecir la resistencia a la compresión de una determinada composición de un hormigón
- ▶ Negocios
  - ▶ Para estimar compras en base a la cantida de productos comprados anteriormente

# Métodos de aprendizaje

- ▶ No supervisados
  - ▶ Donde no hay información previa
  - ▶ Reglas de asociación: busca patrones en los datos
- ▶ Supervisados
  - ▶ Existe información previa
  - ▶ Por ejemplo: etiquetas, tags, etc,
  - ▶ La evaluación se hace sobre un conjunto llamado testeo

# Modelos de Regresión

# Lineales

- ▶ Simple
- ▶ Múltiple

# Regresión lineal simple (RLS)

- ▶ Técnica de análisis estadístico utilizada para estimar el efectos de una variable (independientes /predictores) en otra variable cuantitativa (dependiente/ predicha/ respuesta).
- ▶ Técnica de análisis estadístico utilizada para predecir una variable dependiente, a partir de otra independiente.



# Motivación

- ▶ Determinar cómo afecta la variable independiente en la variable dependiente
- ▶ Predecir la variable dependiente a partir de la variable independiente

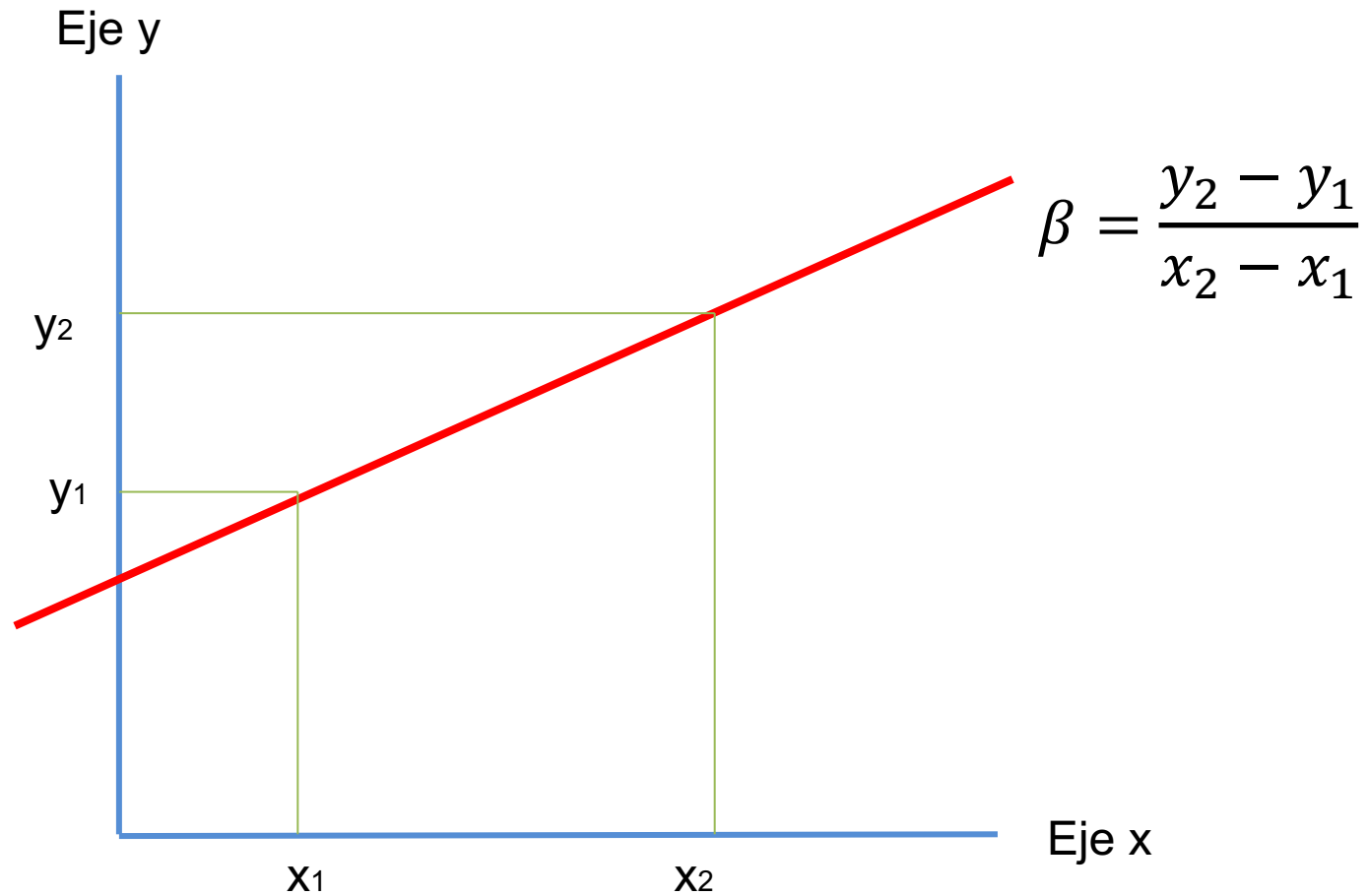
# El modelo

$$Y = \alpha + \beta X$$

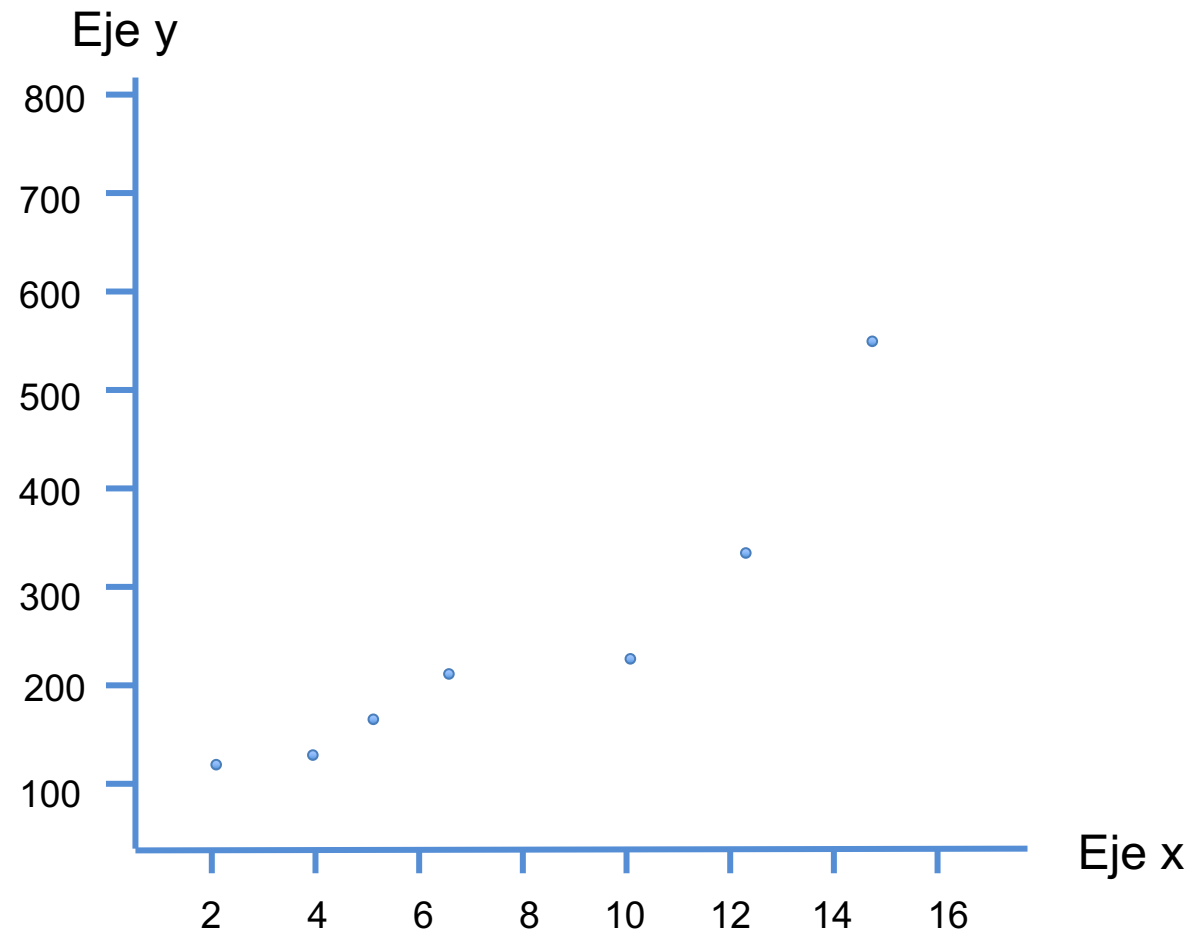
donde:

$\alpha$  = donde corta en Y

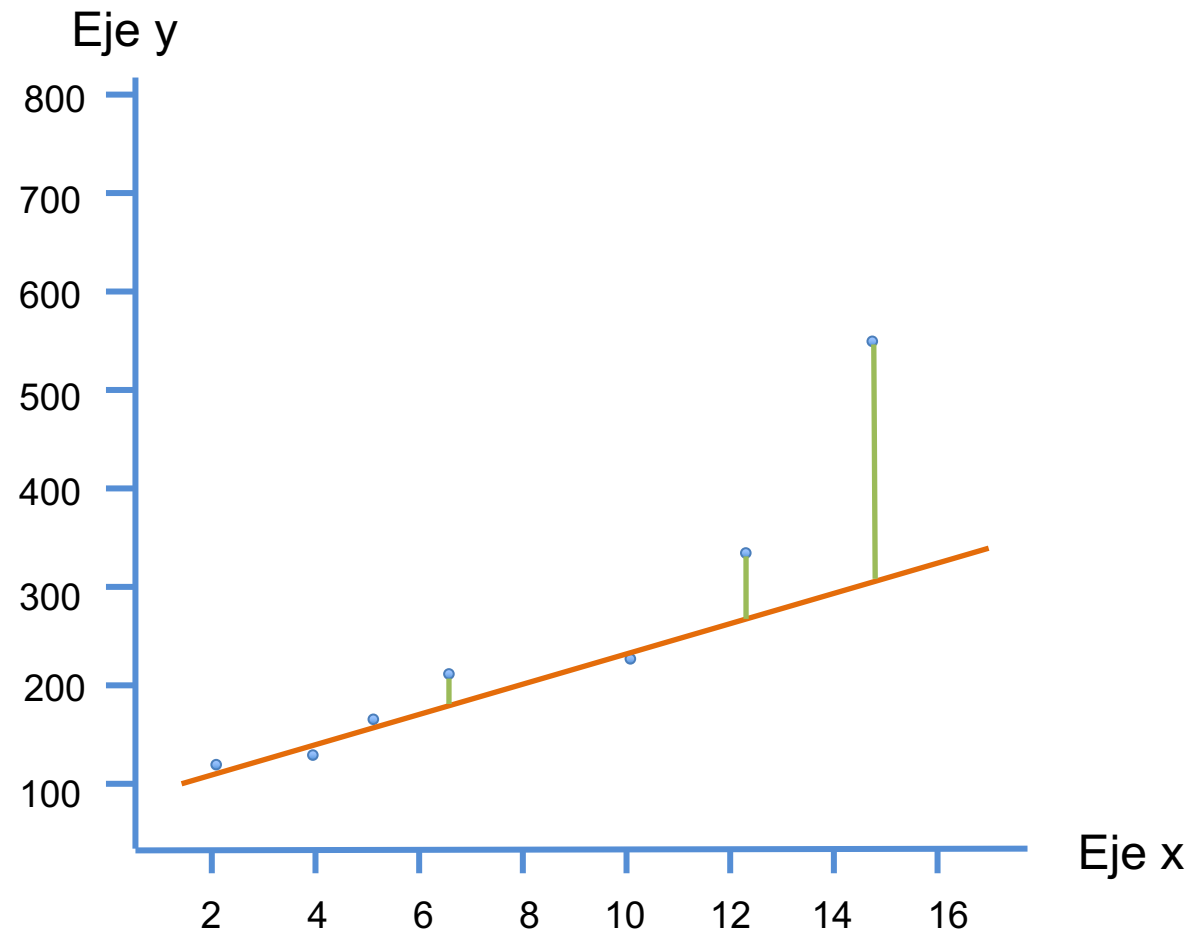
$\beta$  = pendiente de la recta



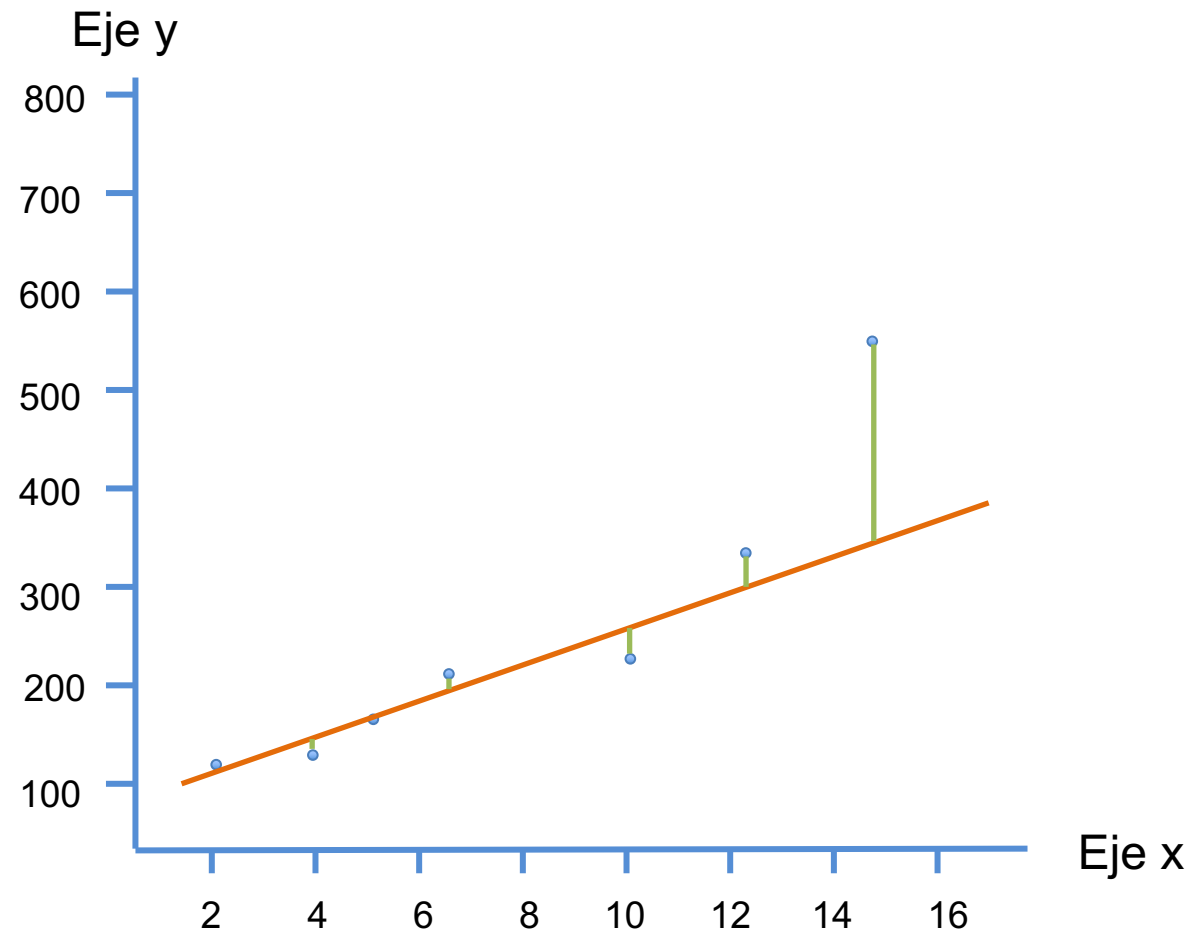
# El modelo



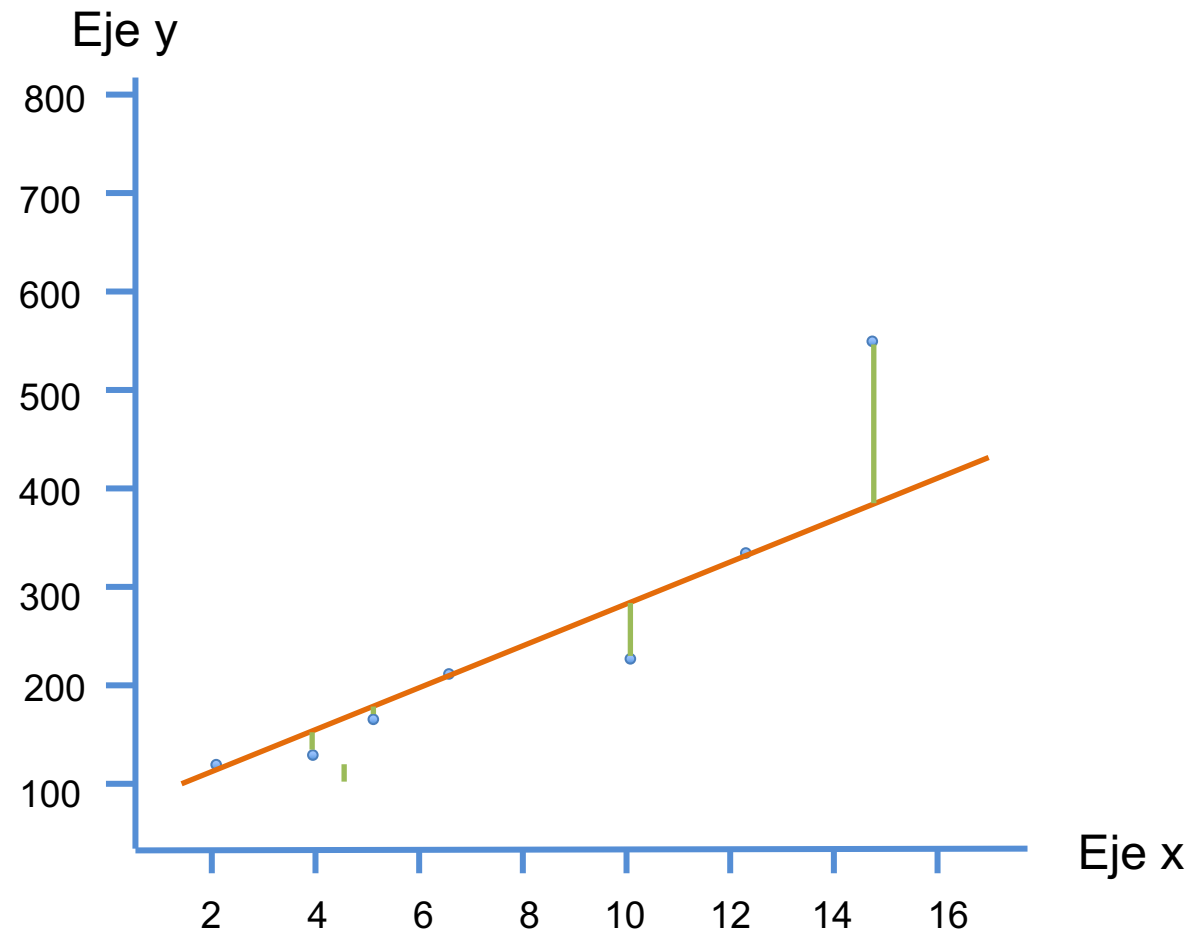
# El modelo



# El modelo



# El modelo



# Regresión lineal multiple (RLM)

- ▶ Técnica de análisis estadístico utilizada para estimar el efectos de varias variables (independientes /predictores) en otra variable cuantitativa (dependiente/ predicha/ respuesta).
- ▶ Técnica de análisis estadístico utilizada para predecir una variable dependiente, a partir de otras independientes.

# El modelo

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon$$

donde:

$X_1 \dots X_k$  = variables independientes

$Y$  = variable dependiente

$\alpha$  = intersección en eje Y

$\beta_1 \dots \beta_k$  = efecto de X en Y

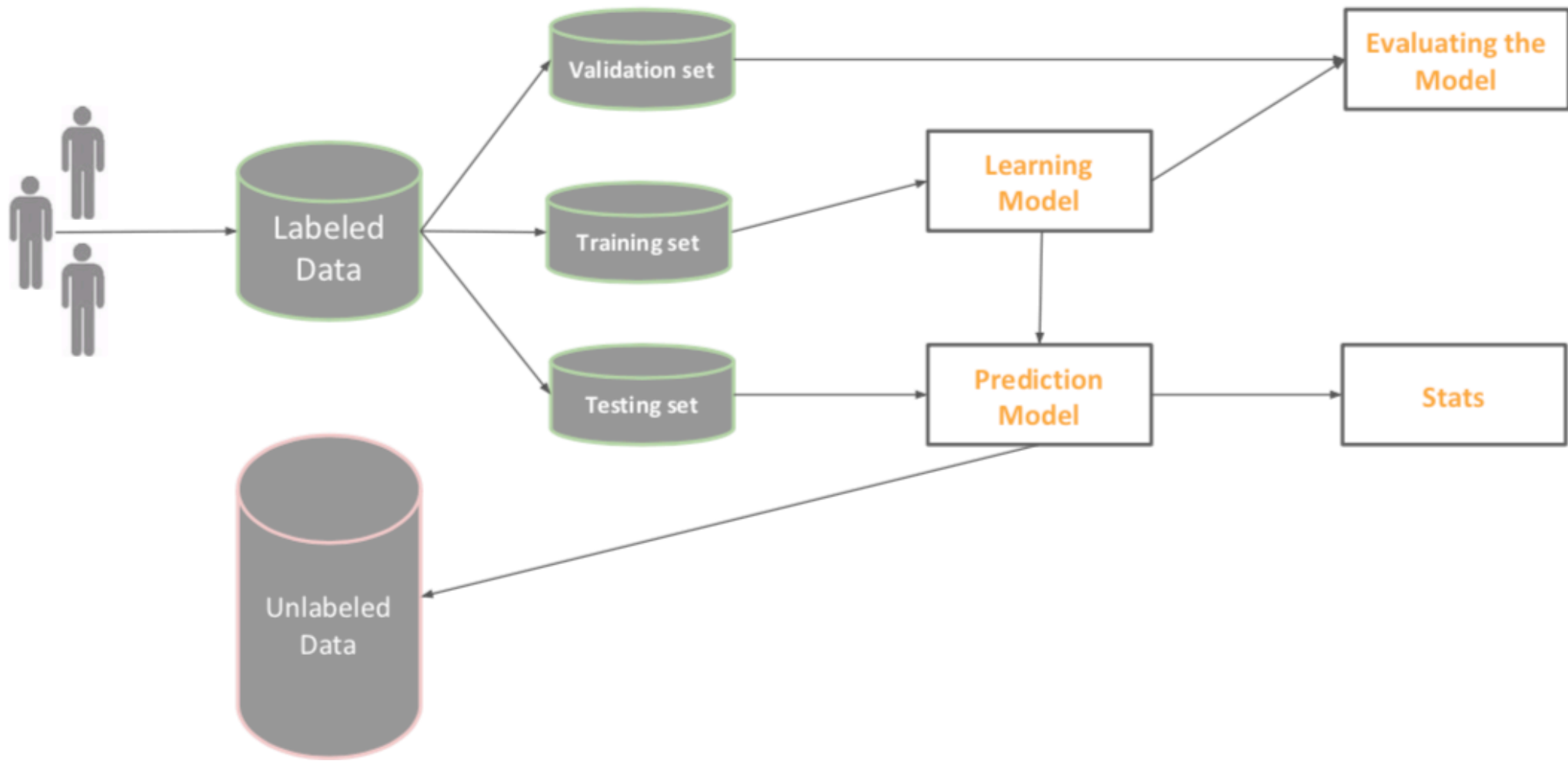


# No lineales

- ▶ Combinaciones lineales
- ▶ Cambio de base
- ▶ No implica ajustar una recta

# Consideraciones de un modelo supervisado

- ▶ Training set
- ▶ Validation set
- ▶ Testing set
- ▶ Unlabeled data



# REFERENCIAS

- ▶ Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.
- ▶ Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.
- ▶ Hand, D. J. (2006). Data Mining. *Encyclopedia of Environmetrics*, 2.