Clase 6:
Modelado Data-driven

Modelado Data-driven

- Motivación en el marco de la materia
- Vamos a hacer en los Colabs
- Reconstrucción de ODEs
  - Regresión
  - Regresión LASSO
  - SINDy
- Bibliografía

Motivación en el marco de la materia

- Sistemas dinámicos = ODEs

$$\frac{d\vec{x}}{dt} = \vec{f}\,(t,\vec{x}) \quad \xRightarrow{\quad\int\quad} \quad \vec{x}(t) \longrightarrow \text{variables}$$

dato

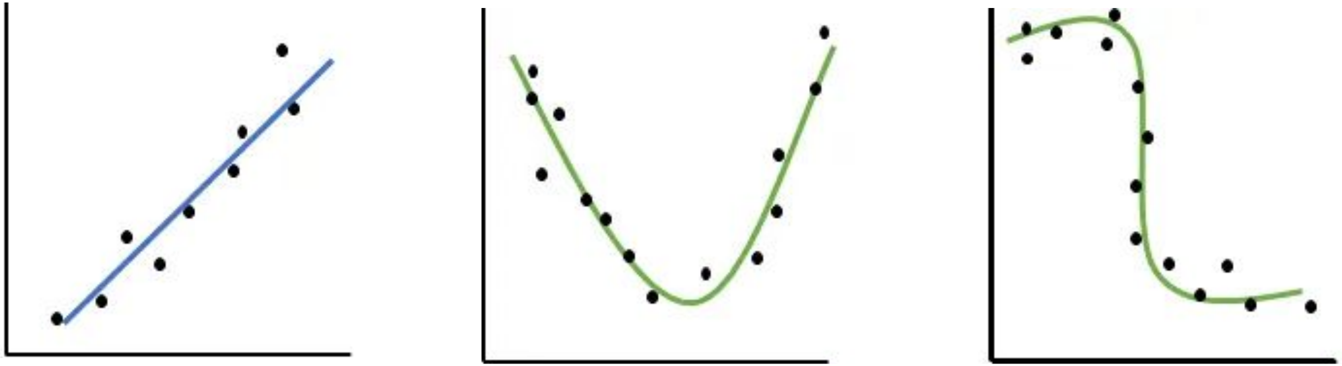
Cómo hago el proceso inverso?

Reconstrucción de ODEs

Vamos a hacer en los Colabs

- Reconstrucción de ODEs
  - Atractor de Lorenz
    - LASSO
    - SINDy
  - Oscilador de relajación de Van der Pol
    - Dificultades del método?

Regresión



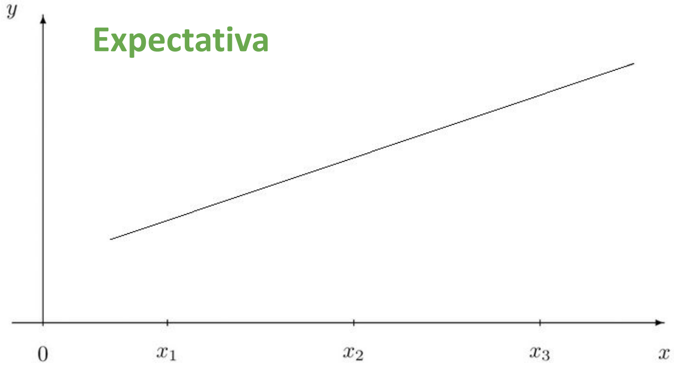
Regresión lineal Regresión no lineal (polinomio, etc)

Propongo un **modelo** a partir del cual hay una **relación entre variables (variables independientes y variables dependientes)**

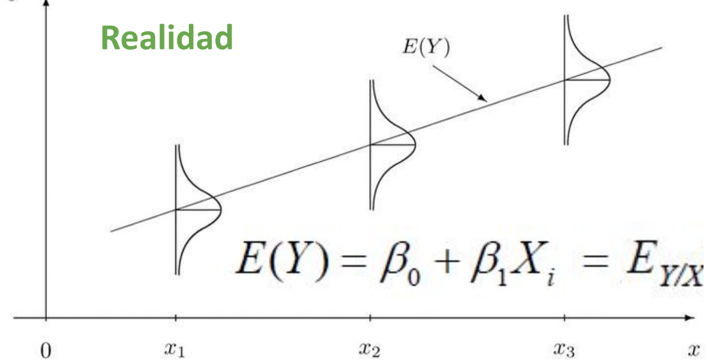
Estimo los **parámetros** que hacen que el **modelo** se **ajuste** a mis **datos**

Regresión lineal

$$Y_i = \beta_0 + \beta_1 X_i$$



$$Y_i = \beta_0 + \beta_1 X_i + \boxed{\varepsilon_i}$$

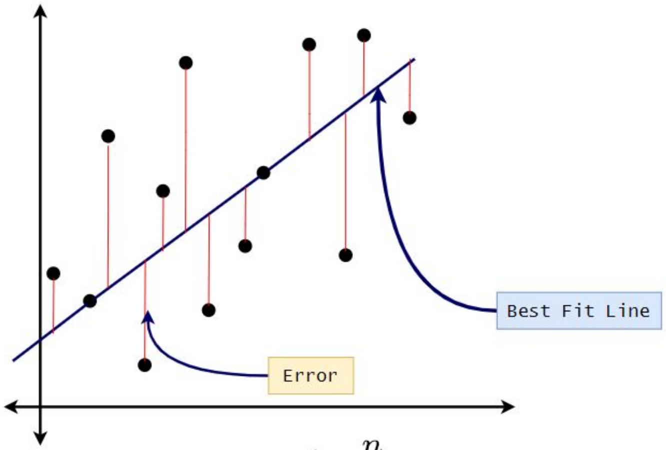


En datos reales tengo una componente aleatoria de ruido (puede ser natural)

Necesito resolver el sistema sobredeterminado

Voy a ajustar con una recta que no pasa exactamente por todos los puntos

Regresión lineal



Variable dependiente

Parámetros

Variable independiente

$$Y_i = \underbrace{\beta_0 + \beta_1}_{\text{Parámetros}} X_i$$

Ajuste del modelo (estimación)

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \cdot x$$

$$J(\hat{\beta}) = \frac{1}{n} \sum_{i=1}^{n} \Big(\underbrace{y^{(i)} - \hat{y}^{(i)}}_{\text{error}} \Big)^2$$

Error cuadrático medio

Función costo

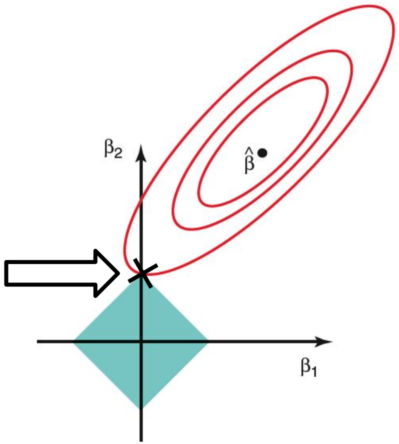
Ajustar un modelo es básicamente resolver un problema de **optimización** que consiste en encontrar los **parámetros que minimizan una función costo**

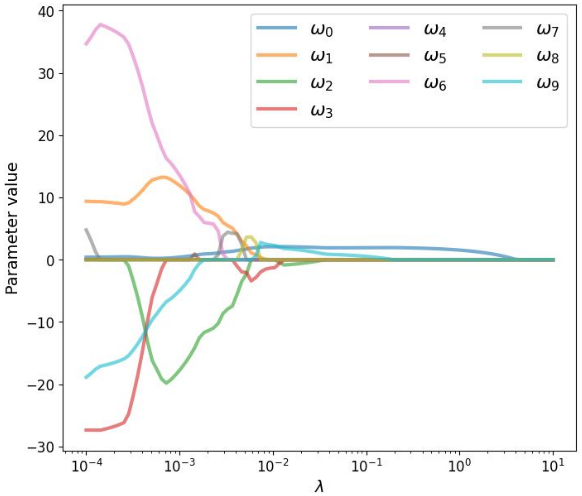
Regresión LASSO

Regularización LASSO

$$J(\hat{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\underbrace{\left(y^{(i)} - \hat{y}^{(i)}\right)}_{\text{error}}^{2} + \lambda\sum_{j=1}^{M}\left\|\hat{\beta}_j\right\|_p$$

Agregamos un término de **penalización** en la función costo





Cuando penalizo, es posible que parámetros se hagan cero (se cancelan términos)

SINDy

Sparse Identification of Nonlinear Dynamical systems

Sistema dinámico $\longrightarrow$ $$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}, t; \mu)$$

Puedo plantear el campo vector como un desarrollo de variables en una base de transformaciones

$$\dot{\mathbf{X}} = \Theta(\mathbf{X})\Xi$$

Parámetros (coeficientes)

Derivadas

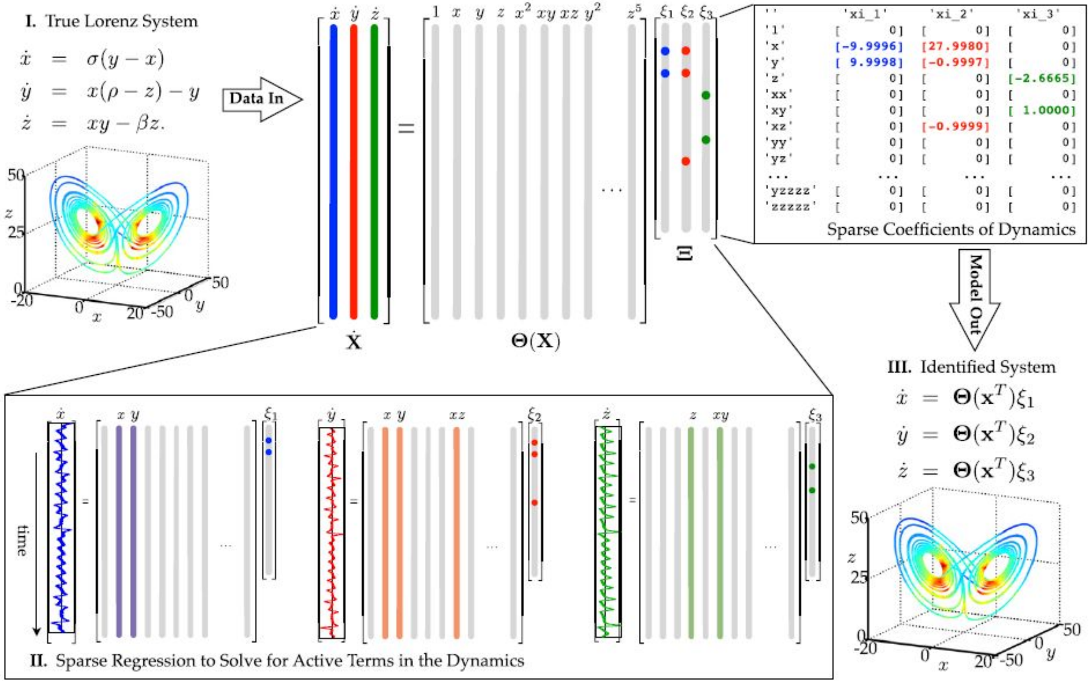
Variables transformadas (por ejemplo x^2.z)

Si tengo las variables y sus derivadas, puedo hacer una regresión lineal

En particular, regresión de tipo LASSO, agrego penalización que selecciona parámetros y los hace 0

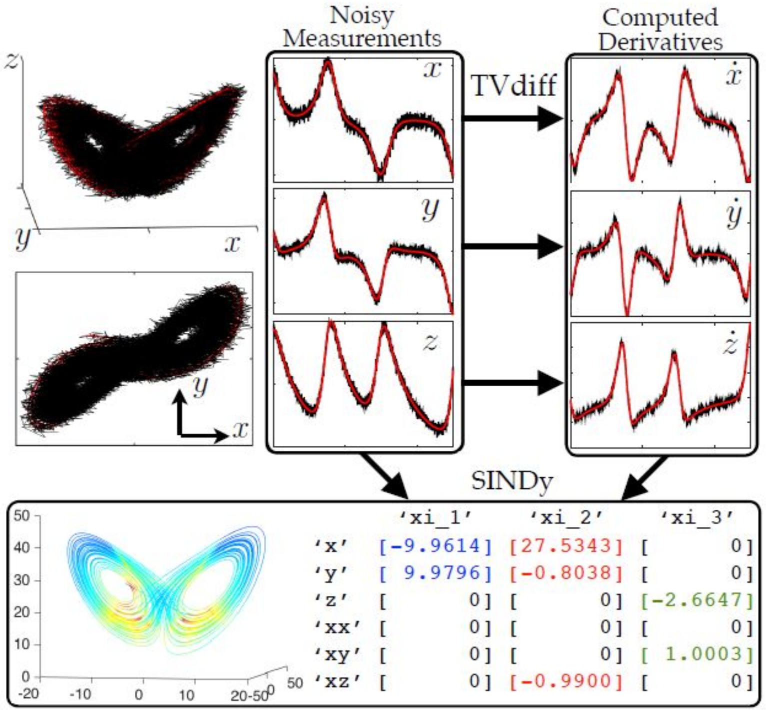
Puedo obtener una representación esparsa de las relaciones y reconstruir las ecuaciones del sistema

SINDy



Brunton, Proctor, & Kutz (2016). *Discovering governing equations from data by sparse identification of nonlinear dynamical systems*. PNAS, *113*(15), 3932-3937.

SINDy



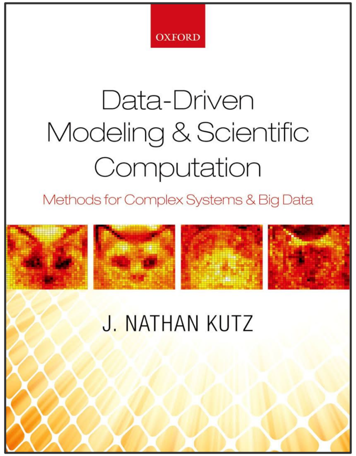
Funciona también

- para datos ruidosos
- cuando no tengo la derivada pero la puedo calcular numéricamente con cierta confianza

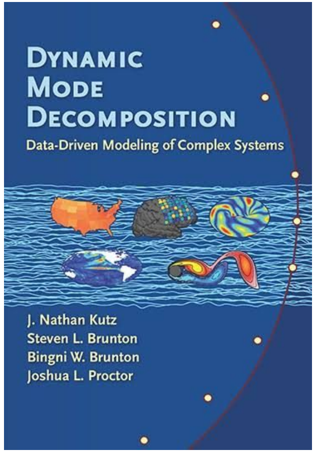
Qué limitaciones tiene esto?

Brunton, Proctor, & Kutz (2016). *Discovering governing equations from data by sparse identification of nonlinear dynamical systems*. PNAS, *113*(15), 3932-3937.

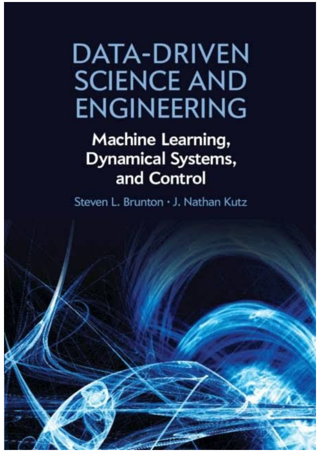
Bibliografía recomendada



Kutz 2013



Kutz et al 2016



Brunton & Kutz 2019

Google
stackoverflow
towards data science
YouTube