
Series de Tiempo ₂

JIVB - 2022

“... all models are approximations. Essentially, all models are wrong, but some are useful. However, the approximate nature of the model must always be borne in mind....”

— George Box

1. Repaso
 2. Random Walk
 3. Q-Tests
 4. Box & Jenkins
 5. Métricas de Ajuste
 6. Métricas de Performance
-

Repasso

Random Walk

Q-Tests

Box & Jenkins

Métricas de Ajuste

Métricas de Performance

Repaso

1. Comentarios de las Presentaciones

- a. Cumplimiento de Consignas / Entregables / Plazos
- b. Estabilidad
 - i. Raíces de Polinomio Característico de Matriz del Proceso Dinámico
 - ii. Raíces de Polinomio Característico de Matriz del Proceso Dinámico expresado con operador L
 - iii. Coeficientes del proceso autorregresivo AR(p) // Serie Geométrica / Convergente
- c. Estacionariedad
 - i. Procesos No Estacionarios, Estacionarios Débiles, Estacionarios Estrictos
 - ii. Débil / de Covarianzas != Fuerte / Estricta != Simple (?)
- d. Invertibilidad
 - i. Wold no es una demostración
 - ii. Aplica sólo a un
 - iii. Es un análogo al requisito de estabilidad para los procesos MA(q)
- e. Ergodicidad
 - i. Memoria Larga y AR(p)

Repaso

Random Walk

Q-Tests

Box & Jenkins

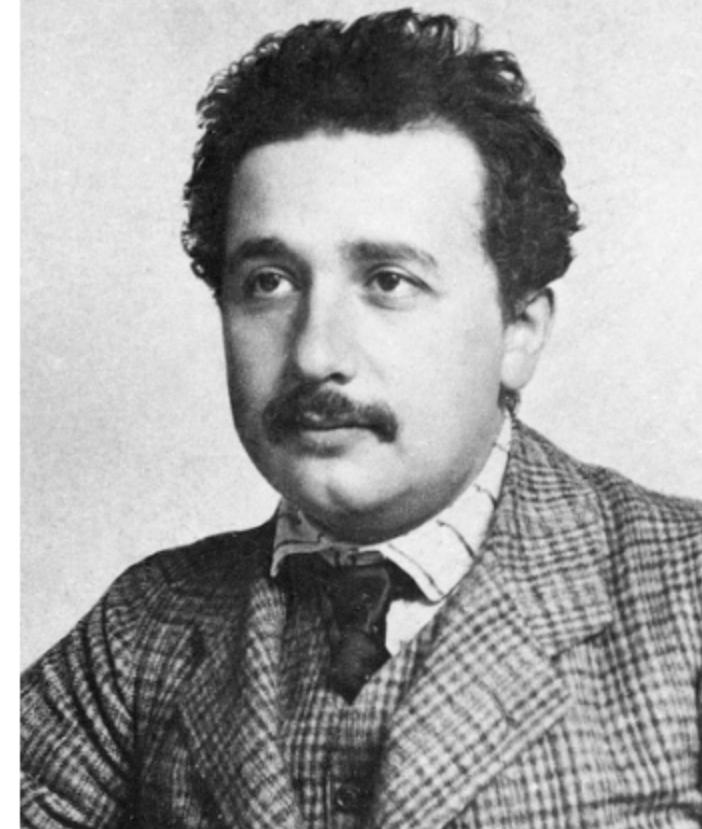
Métricas de Ajuste

Métricas de Performance

1905

Annus mirabilis

Annus mirabilis - 4 papers de Einstein



1. The first paper explained the [photoelectric effect](#), which established the energy of the light quanta, and was the only specific discovery mentioned in the citation awarding Einstein the [Nobel Prize in Physics](#).^[1]
2. **The second paper explained Brownian motion, which established the Einstein relation and led reluctant physicists to accept the existence of atoms.**
3. The third paper introduced Einstein's [theory of special relativity](#), which established the [universal constant speed of light](#) for all reference frames and a theory of [spacetime](#).
4. The fourth, a consequence of the theory of special relativity, developed the principle of [mass-energy equivalence](#), expressed in the famous equation and which led to the discovery and use of [atomic energy](#).

Karl Pearson - 1905 - Revista Nature

The Problem of the Random Walk.

CAN any of your readers refer me to a work wherein I should find a solution of the following problem, or failing the knowledge of any existing solution provide me with an original one? I should be extremely grateful for aid in the matter.

A man starts from a point O and walks l yards in a straight line; he then turns through any angle whatever and walks another l yards in a second straight line. He repeats this process n times. I require the probability that after these n stretches he is at a distance between r and $r + \delta r$ from his starting point, O.

The problem is one of considerable interest, but I have only succeeded in obtaining an integrated solution for two stretches. I think, however, that a solution ought to be found, if only in the form of a series in powers of $1/n$, when n is large.

KARL PEARSON.

The Gables, East Ilsley, Berks.

Random Walk

$$X_t = X_{t-1} + e_t,$$

Si lo descomponemos

$$\begin{aligned} X_t &= X_{t-1} + e_t \\ &= (X_{t-2} + e_{t-1}) + e_t. \end{aligned}$$

obtenemos que

$$X_t = X_0 + e_1 + e_2 + \dots + e_{t-1} + e_t.$$

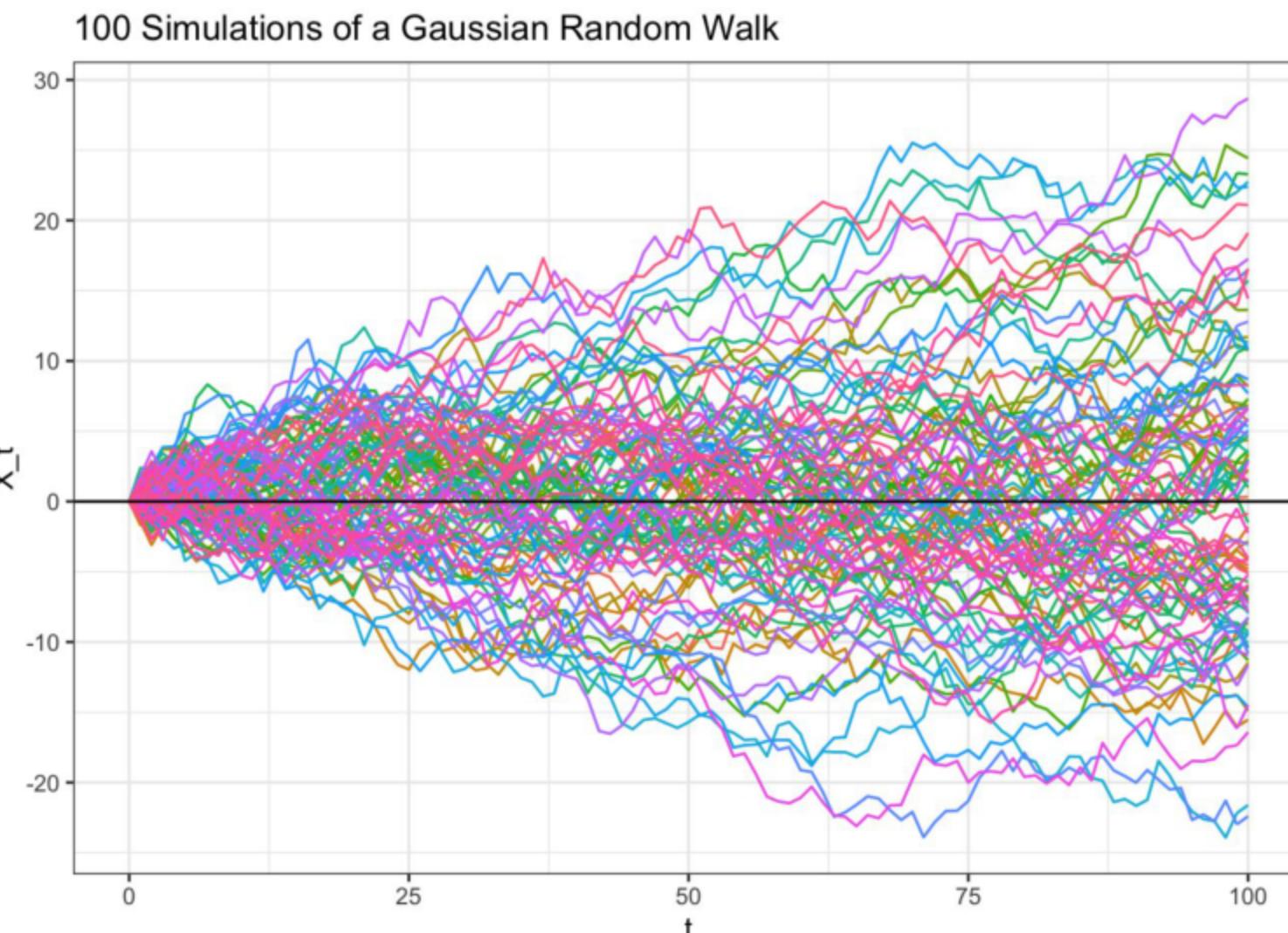
Con media

$$\begin{aligned} E(X_t | X_0) &= E(X_0 + e_1 + e_2 + \dots + e_t | X_0) \\ &= X_0 + E(e_1 | X_0) + E(e_2 | X_0) + \dots + E(e_t | X_0) \\ &= X_0. \end{aligned}$$

Y varianza

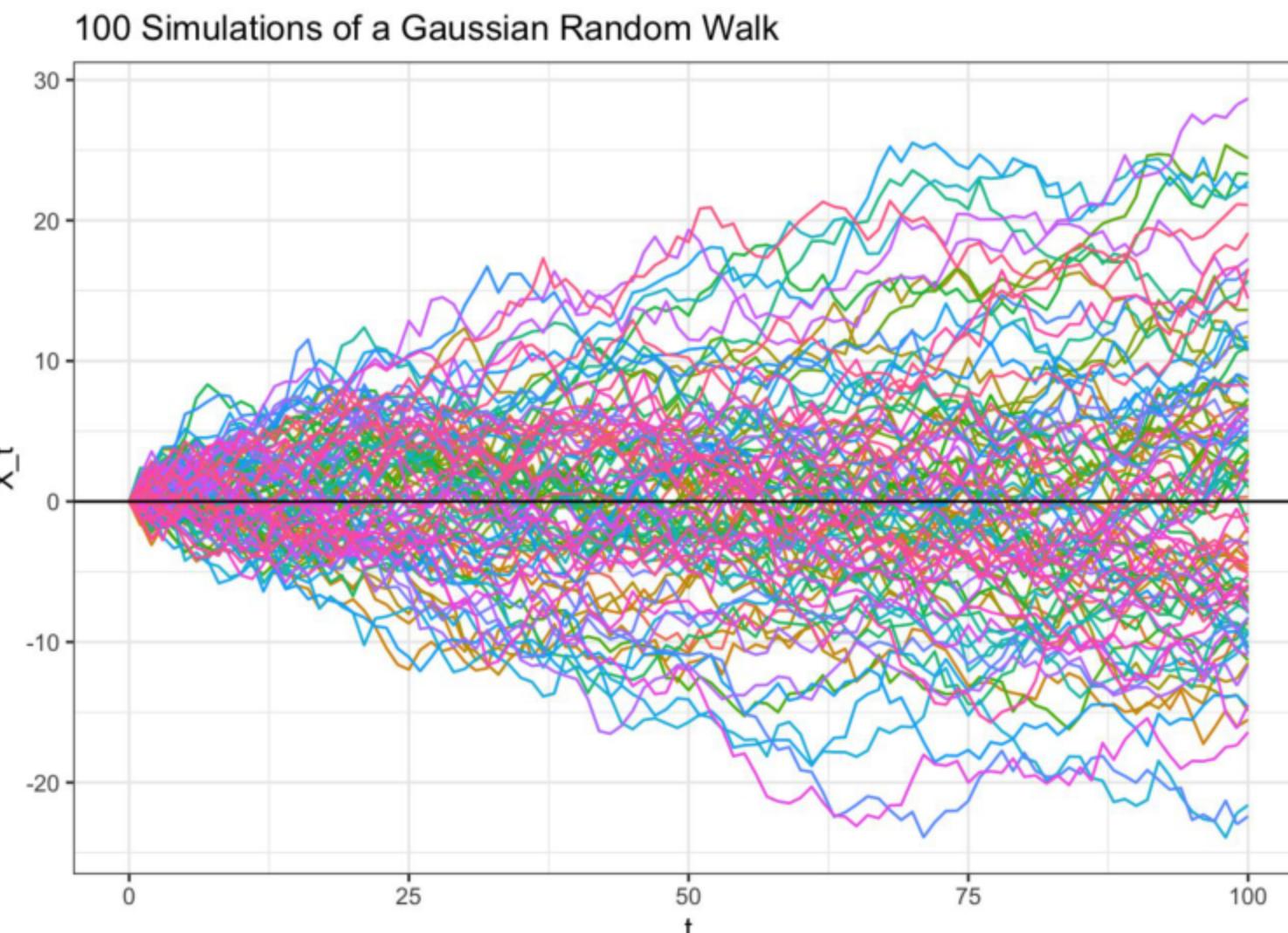
$$\begin{aligned} Var(X_t) &= Var(X_0) + Var(e_1) + Var(e_2) + \dots + Var(e_{t-1}) + Var(e_t) \\ &= 0 + \sigma^2 + \sigma^2 + \dots + \sigma^2 + \sigma^2 \\ &= t\sigma^2. \end{aligned}$$

NO ESTACIONARIO



Random Walk

Si tomamos la primera diferencia de un proceso **RW**



$$X_t - X_{t-1} = X_{t-1} - X_{t-1} + e_t$$

$$Z_t = e_t,$$

obtenemos un proceso **WN** (Estacionario)

Repaso

Random Walk

Q-Tests

Box & Jenkins

Métricas de Ajuste

Métricas de Performance

Q - Tests : Box and Pierce, 1970

1 . Box-Pierce Q-test

$$Q = T \sum_{i=1}^k \hat{\rho}_i^2,$$

donde $\hat{\rho}_i$ son los elementos de la FAC muestral.

Bajo H_0 todos los $\hat{\rho}_i$ hasta k son iguales a cero.

El estadístico Q se distribuye asintóticamente como una chi cuadrado con k grados de libertad, siempre que la serie de tiempo sea generada por un proceso ARMA estacionario.

Q - Tests: Ljung and Box, 1978

2 . Ljung-Box Q-test

$$Q = T(T+2) \sum_{i=1}^k \frac{\hat{\rho}_i^2}{T-i},$$

donde $\hat{\rho}_i$ son los elementos de la FAC muestral.

Bajo H_0 todos los $\hat{\rho}_i$ hasta k son iguales a cero.

El estadístico Q se distribuye asintóticamente como una chi cuadrado con k grados de libertad, siempre que la serie de tiempo sea generada por un proceso ARMA estacionario.

¿Cómo implementarlos?

El test de Ljung Box se ajusta mejor que el de Box-Pierce a muestras de menor tamaño.

Al momento de buscar el lag k óptimo partimos de 1 hasta $T/4$ donde T es el tamaño de la muestra.

Si el estadístico supera el umbral χ_k^2 podemos decir que al menos una autocorrelación $\{\rho_i\}_{i=1}^k$ es (estadísticamente hablando) significativamente diferente de 0.

En la práctica, estos estadísticos se usan como umbrales de referencia para detectar los órdenes p y q, pero sólo se los toma considerando estrictamente en resultados al momento de evaluar los residuos.

Repaso
Random Walk
Metodología B&J
Métricas de Ajuste
Métricas de Performance

El sentido de la metodología B&J

1. No busca ser
 - a. un modelo causal
 - b. identificar exactamente el proceso dinámico subyacente
2. Si busca
 - a. **Parsimonia**
 - i. en la selección de los parámetros
 - b. **Rigor Estadístico**
 - i. criterio en la selección del modelo (información estadística)
 - ii. criterio en la evaluación de los residuos (tests de Q)

Pasos B&J

1. Transformaciones

- a. ¿El proceso es estacionario? ¿podemos hacerlo estacionario en diferencias?
 - i. Muchas veces si podemos y necesitamos diferenciar la serie (d)

2. Identificación del proceso generador subyacente

- a. Exploración de correlogramas: FAC y FACP
- b. Definimos si es un AR(p) / un MA(q) o un ARMA(p,q)

3. Modelamos

- a. Estimación de Parámetros (i.e.: Yule - Walker)

4. Pruebas de Diagnóstico

- a. Criterio de Información (AIC, BIC)

5. Validación

- a. revisamos resultados, residuos, etc. ok...
- b. si los resultados no sirven: repetimos

Repasso

Random Walk

M. Box & Jenkins

Métricas de Ajuste

Métricas de Performance

Criterio de Información

Criterio de información de Akaike

$$AIC = -2\mathcal{L} + 2K$$

"desajuste" penalización

Criterio de información de Bayes

$$BIC = -2\mathcal{L} + \ln(T)K$$

"desajuste" penalización

Ejemplo

AIC

	q=0	q=1	q=2
p=0	128.02	89.62	83.77
p=1	64.31	30.74	31.92
p=2	46.27	32.21	34.44
p=3	24.64	26.51	28.32
p=4	26.50	28.43	25.10

BIC

	q=0	q=1	q=2
p=0	134.22	98.92	96.17
p=1	73.61	43.14	47.42
p=2	58.67	47.71	53.04
p=3	40.14	45.11	50.02
p=4	45.10	50.13	49.90

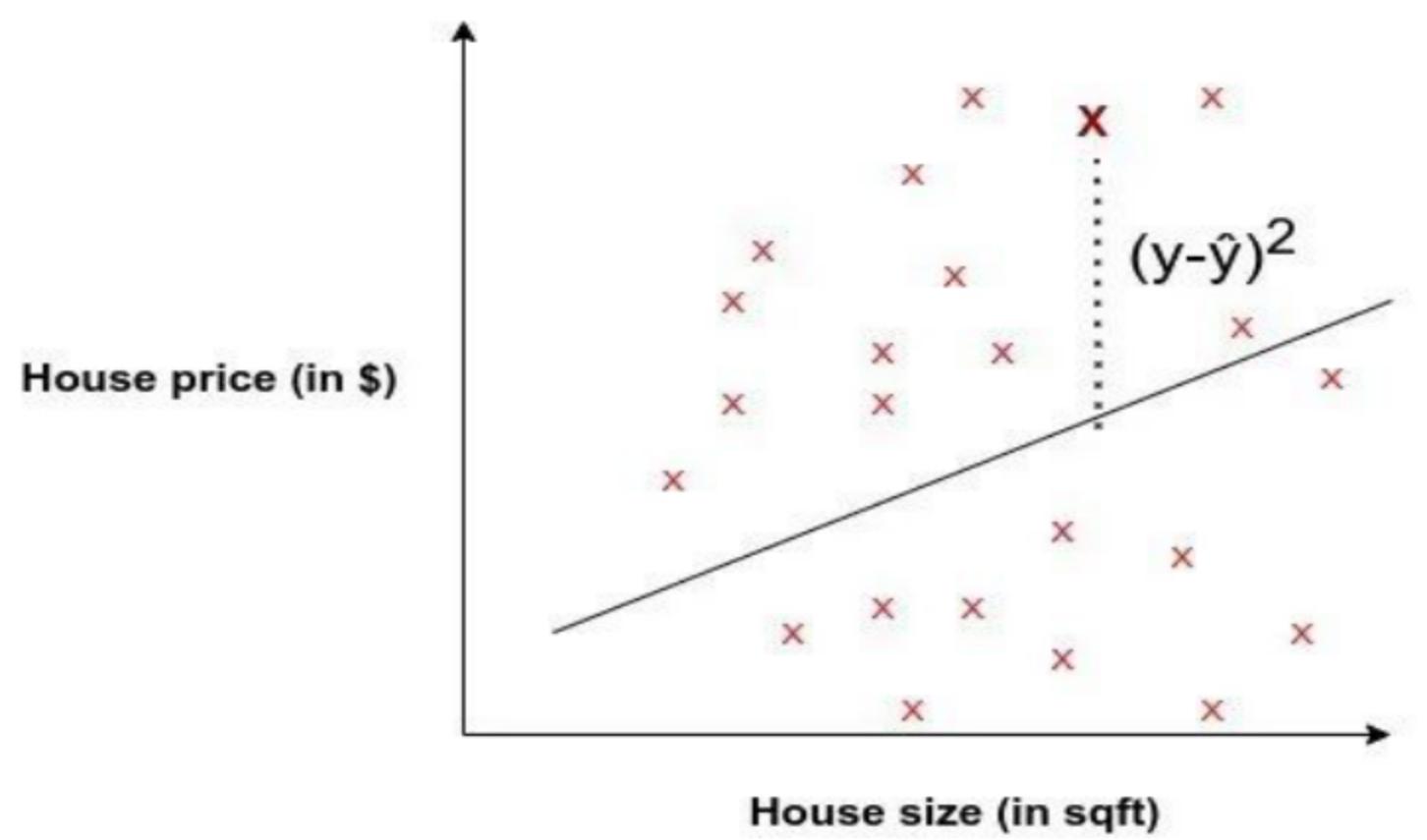
Repasso
Random Walk
M. Box & Jenkins
Métricas de Ajuste
Métricas de Performance

Métricas de Performance Puntual

Scale Dependent Errors

a. Mean Squared Error

$$MSE = \frac{1}{N} \sum_{j=1}^N (y_j - \hat{y}_j)^2$$

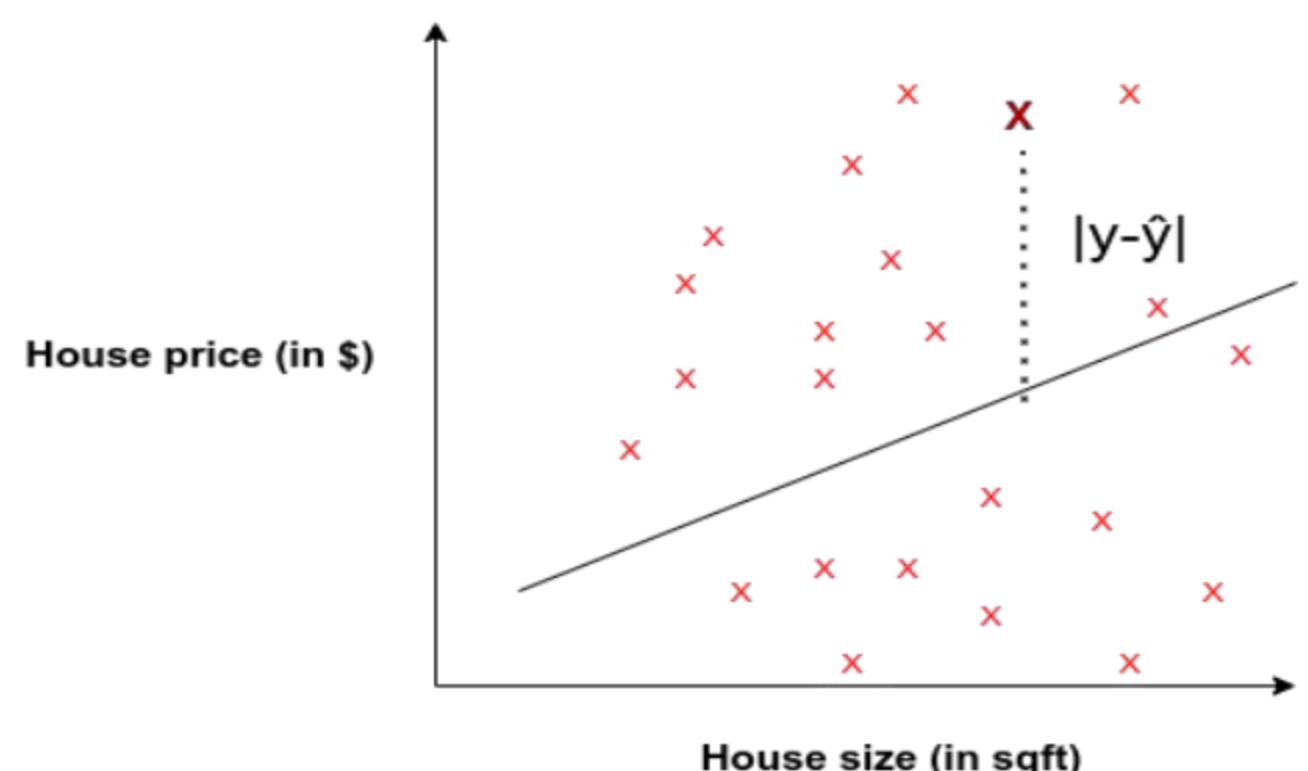


b. Mean Absolute Error

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

c. Root Mean Squared Error

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



Métricas de Performance Puntual

Percentage Errors

$$p_t = 100e_t/y_t.$$

1. Mean Absolute Percentage Error

$$\text{MAPE} = \text{mean}(|p_t|).$$

2. Symmetric Mean Absolute Percentage Error

$$\text{sMAPE} = \text{mean} \left(200|y_t - \hat{y}_t|/(y_t + \hat{y}_t) \right).$$