

Clustering subsecuencial de series de tiempo: Evidencia de patrones temporales en el tipo de cambio USD/MXN.

Juan Francisco Muñoz Elguezábal ¹ Riemann Ruíz Cruz ²

¹Msc. Ciencia de Datos - ITESO ²Departamento de Matemáticas y Física - ITESO

Hipótesis - Experimento

Hipótesis:

Existe un conjunto de condiciones bajo las cuales, la ocurrencia de evento exógeno a una serie de tiempo financiera, provoca la presencia de patrones temporales observables en la misma serie, y por lo tanto, evidencia de rechazo de la *Hipótesis del Mercado Eficiente*.

Experimento:

Comunicado de indicadores macroeconómicos de México y USA, como candidatos de eventos generadores de patrones temporales en la serie de tiempo intradía del del tipo de cambio USD/MXN.

Definiciones

Sea $\{S_T\}_{T=1}^n$, los precios $OHLC_T : \{Open_T, High_T, Low_T, Close_T\}$ de cada n minuto *bursátil* de los últimos 10 años. De los cuales se extraen ventanas de tamaño m , de tal manera que $\{OHLC_t\}_{t=1}^m$, para $m = 30$.

- micro-volatilidades: $\rightarrow HL_t = High_t - Low_t$
- micro-tendencias: $\rightarrow CO_t = Close_t - Open_t$

Sea el proceso $\{I_t\}_{t=1}^k$ como el comunicado de un indicador macroeconómico que sucede k veces, tal que, con $OHLC_k\{Open_{t=1:m}^k, High_{t=1:m}^k, Low_{t=1:m}^k, Close_{t=1:m}^k\}$ se definen las series de tiempo $CO_{t=1:m}^k$ y $HL_{1:m}^k$ como *motifs* para encontrar en $OHLC_T$

Tipos de indicadores

categoria	usa	mex	total
actividad economica	26	7	33
consumo	29	5	34
energia	4	0	4
flujos de capital	5	0	5
inflacion	0	4	4
mercado inmobiliario	11	1	12
mercado laboral	13	2	15
subasta de bonos	5	0	5
tasas de interes	1	1	2
total	94	20	114

Indicadores

$\{I_t\}_{t=1}^k = \{a_t, c_t, p_t\}$ el comunicado de un indicador económico I , con información en t de *actual* _{t} , *consenso* _{t} , *previo* _{t} .

$I_t = \{a, b, c, d\} \forall \{I_t\}_{t=1}^k$ como la categorización de un comunicado de acuerdo a 4 posibles resultados:

- a: $a_t > c_t > p_t$
- b: $a_t > c_t \leq p_t$
- c: $a_t \leq c_t > p_t$
- d: $a_t \leq c_t \leq p_t$

Similitud entre series de tiempo

Sean x_t, y_t dos series de tiempo de longitud m , se define $d(x, y)$ como la distancia euclideana z-normalizada, y $corr(x, y)$ como la correlación entre estas

$$\hat{x} = \frac{x_i - \mu_x}{\sigma_x}, \quad \hat{y} = \frac{y_i - \mu_y}{\sigma_y}, \quad d(x, y) = \sqrt{\sum_{i=1}^m (\hat{x}_i - \hat{y}_i)^2}$$

$$corr(x, y) = \frac{\sum_{i=1}^m x_i y_i - m \mu_x \mu_y}{m \sigma_x \sigma_y} \rightarrow d(\hat{x}, \hat{y}) = \sqrt{2m(1 - corr(x, y))}$$

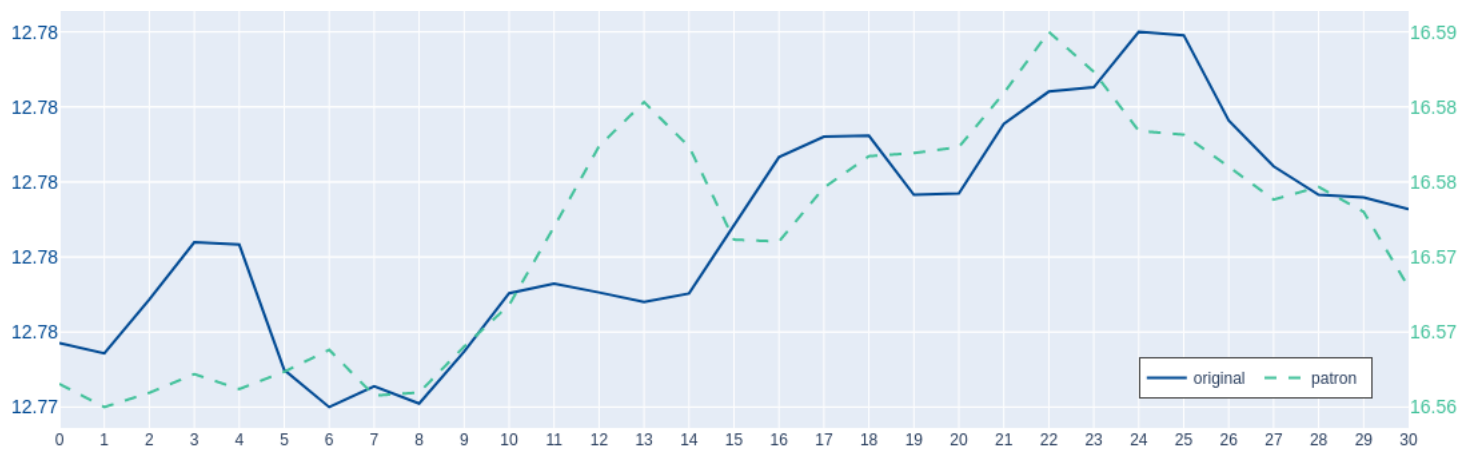
$d(\hat{x}, \hat{y}) \implies$ Si se minimiza la distancia se maximiza la correlación.

$d(\hat{x}, \hat{y}) = [0, m_p]$ donde $m_p \in \{0, \mathbb{R}^+\}$.

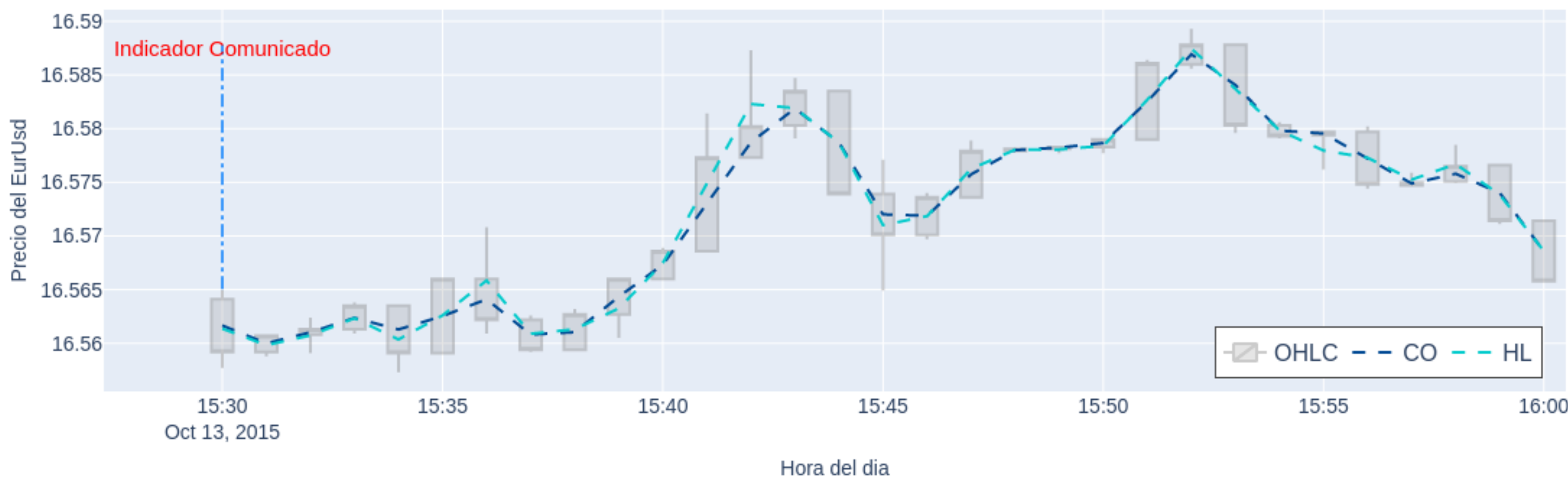
Resultados 1

id_esc	tipo_1	tipo_2	tipo_3	total_ind_esc	total_ind
APIWeelStock_USA_A	0	0	16	115	235
APIWeelStock_USA_D	2	3	4	120	235
InitiaClaims_USA_A	0	1	40	200	520
InitiaClaims_USA_B	0	0	9	53	520
InitiaClaims_USA_C	0	0	5	64	520
InitiaClaims_USA_D	1	1	30	203	520

Resultados 1



e.g: Ventana $k = 1$ de $m = 30$ Precios OHLC



e.g. Ocurrencias e invarianza en reacciones por escenario

indicador	A	B	C	D	T
Initial Jobless Claims	200	53	64	203	520
API Weekly Crude Oil Stock	115	0	0	120	235

ANOVA $\rightarrow h_0$: varianza constante

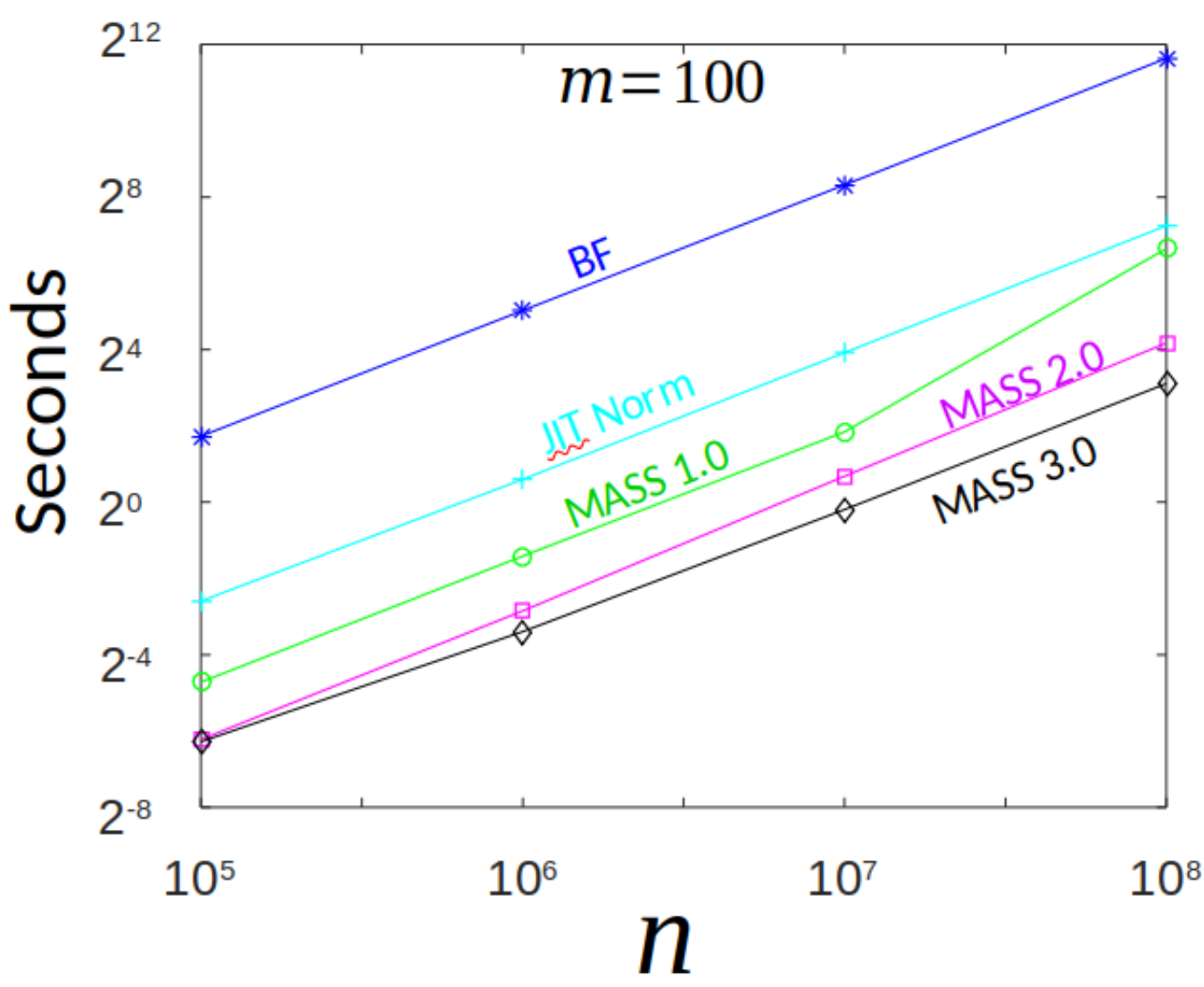
name	esc	obs	anova_hl	anova_co
Initial Jobless Claims	A	200	1	1
Initial Jobless Claims	B	53	1	0
Initial Jobless Claims	C	64	1	1
Initial Jobless Claims	D	203	1	1
API Weekly Crude Oil Stock	A	115	1	1
API Weekly Crude Oil Stock	D	120	1	1

Algoritmo MASS

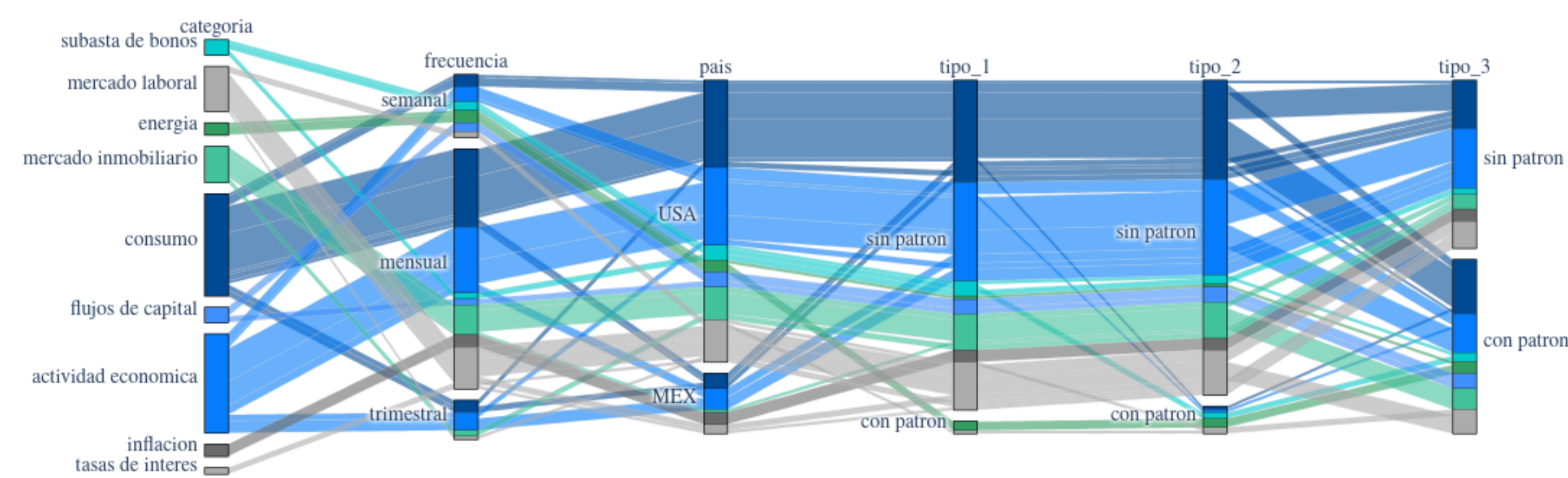
Busca una serie *query* de tamaño m en una serie de tiempo *principal* de tamaño n .

- Implementa una convolución de productos punto móviles con $O(n \log n)$.
- Normalización *just in time*.
- Transformada Rápida de Fourier para convoluciones eficientes.
- Costo computacional no depende de m (libre de efectos de dimensionalidad)

Complejidad



Resultados 3



Resultados 1

Texto breve con algunas conclusiones.