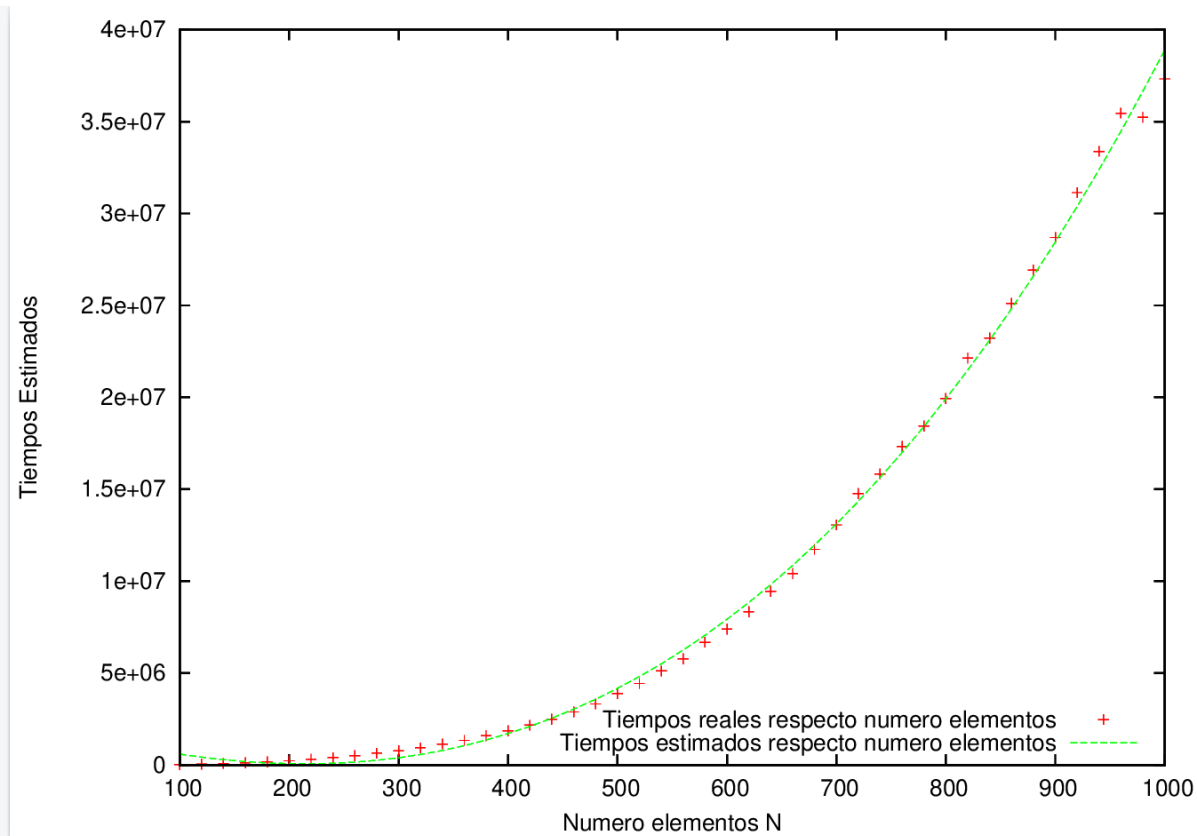


Estimación de la complejidad computacional mediante un enfoque híbrido

- A priori sabemos la complejidad computacional del algoritmo. Por ejemplo, la ordenación por inserción sería $O(n^2)$, donde n es el tamaño del conjunto a ordenar.
- Disponemos de los tiempos de ejecución de la implementación del algoritmo, para distintos valores de n (tamaño del ejemplar).
- Se pretende obtener una función que aproxime el tiempo de ejecución en función de n , es decir, **$t = f(n)$. n es la variable independiente y t la variable dependiente**
- En el caso de de la ordenación por inserción, tendríamos
 - $t = a_0 + a_1n + a_2n^2$.
- La función se obtendrá usando el método de mínimos cuadrados.
- Esta función sirve para hacer estimaciones de tiempo de esa implementación, para cualquier valor de n .

- En la siguiente gráfica se ve un ejemplo de aplicación:



- Los puntos representan los tiempos medidos, con respecto al tamaño del conjunto a ordenar y la curva representa la función que se ajusta a esos puntos, que en este caso será un polinomio de segundo grado.

¿Cómo se obtiene la función?

- Para obtener la función se utiliza el método de los mínimos cuadrados.
- Mediante dicho método podemos ajustar un conjunto de puntos a un polinomio de cualquier grado.
- Usando cambios de variable, podemos ajustar funciones que no sean polinómicas. Por ejemplo:
 - $t(n) = a_0 + a_1 * n!$
 - $t(n) = a_0 + a_1 \log(n)$
 - $t(n) = a_0 + a_1 * 2^n$

¿Cómo se obtiene la función?

- Para obtener los coeficientes de la función, hay que resolver un sistema de ecuaciones, cuyos coeficientes se obtienen a partir de los puntos que hay que ajustar

$$\begin{array}{ccccccccccc} a_0 n & + & a_1 \sum x_i & + & a_2 \sum x_i^2 & + & \dots & + & a_m \sum x_i^m & = & \sum y_i \\ a_0 \sum x_i & + & a_1 \sum x_i^2 & + & a_2 \sum x_i^3 & + & \dots & + & a_m \sum x_i^{m+1} & = & \sum x_i y_i \\ a_0 \sum x_i^2 & + & a_1 \sum x_i^3 & + & a_2 \sum x_i^4 & + & \dots & + & a_m \sum x_i^{m+2} & = & \sum x_i^2 y_i \\ \vdots & & \vdots & & \vdots & & & & \vdots & & \vdots \\ a_0 \sum x_i^m & + & a_1 \sum x_i^{m+1} & + & a_2 \sum x_i^{m+2} & + & \dots & + & a_m \sum x_i^{2m} & = & \sum x_i^m y_i \end{array}$$

- Los x_i son los distintos valores del tamaño del ejemplar, que para el caso de la ordenación por inserción serían los distintos tamaños del conjunto a ordenar, en las pruebas que se hacen.
- Los y_i son los tiempos medidos para los valores del tamaño del conjunto.
- n sería el número de pruebas que se realizan

Ejemplo de ajuste de función polinómica

- Supongamos que tenemos los tiempos de ejecución de la ordenación por inserción, para los siguientes tamaños del conjunto: 100, 110, 120, 130, 140, ..., 200, en microsegundos.
- Como la función es del tipo: $t = a_0 + a_1 n + a_2 n^2$, hay que obtener a_0 , a_1 y a_2 .
- Por tanto, tendremos que resolver un sistema de tres ecuaciones con tres incógnitas. La tabla siguiente indica como obtener el sistema

	n	t(tiempo real)	n ²	n ³	n ⁴	n*t	t*n ²	t(estimado)
	100	900	10000	1000000	100000000	90000	9000000	892,1688
	110	1080	12100	1331000	146410000	118800	13068000	1086,1228
	120	1310	14400	1728000	207360000	157200	18864000	1299,1368
	130	1525	16900	2197000	285610000	198250	25772500	1531,2108
	140	1785	19600	2744000	384160000	249900	34986000	1782,3448
	150	2030	22500	3375000	506250000	304500	45675000	2052,5388
	160	2360	25600	4096000	655360000	377600	60416000	2341,7928
	170	2635	28900	4913000	835210000	447950	76151500	2650,1068
	180	2995	32400	5832000	1049760000	539100	97038000	2977,4808
	190	3320	36100	6859000	1303210000	630800	119852000	3323,9148
	200	3710	40000	8000000	1600000000	742000	148400000	3689,4088
Sumatorios	1450	19940	218500	34075000	5473330000	3114100	500823000	19936,818
Varianza		874710						868493,111

$$11*a_0 + 1450*a_1 + 218500*a_2 = 19940$$

$$1450*a_0 + 218500*a_1 + 34075000*a_2 = 3114100$$

$$218500*a_0 + 34075000*a_2 + 5473330000*a_2 = 500823000$$

Ejemplo de ajuste de función polinómica (II)

(Nota: El coeficiente de a_0 en la primera ecuación es 11 porque la muestra tiene 11 elementos.)

De donde $a_0 = 0.9288$, $a_1 = -0.6176$, $a_2 = 0.0953$.

De esta forma, la ecuación de la curva para obtener los tiempos estimados es:

$$t(n) = 0.9288 - 0.6176 * n + 0.0953 * n^2.$$

La última columna de la tabla muestra los tiempos estimados por medio de la ecuación de la curva, que como se puede apreciar son muy similares a los reales.

Esta función permitiría estimar el tiempo para cualquier valor de n .

Ejemplo de ajuste de función $n\log(n)$

- Para ajustar curvas no polinómicas se pueden emplear cambios de variable, de forma tal que la función pasa a ser polinómica.
- En este ejemplo se va a ajustar una curva del tipo $n\log(n)$ a una curva que representa tiempos frente a valores de n . Este ejemplo podría ser el de la ordenación de un vector por el quicksort.
- $t(n) = a_0 + a_1 n\log(n)$.
- Para linealizarla, se realiza el cambio $z = n\log(n)$ y de esta forma la función es $t(n) = a_0 + a_1 z$ que es un polinomio de primer grado.
- La tabla con los sumatorios sería:

	n	$t(\text{tiempo real})$	$z=n\log(n)$	z^2	$z \cdot t$	$t(\text{estimado})$
	100,000	50,500	200,000	40000,000	10100,000	49,813
	110,000	58,411	224,553	50424,138	13116,296	55,288
	120,000	58,400	249,502	62251,123	14570,989	60,852
	130,000	66,463	274,813	75521,985	18264,742	66,496
	140,000	67,592	300,458	90274,965	20308,427	72,215
	150,000	80,783	326,414	106545,896	26368,591	78,003
	160,000	83,032	352,659	124368,509	29281,942	83,856
	170,000	89,335	379,176	143774,679	33873,816	89,769
	180,000	93,690	405,949	164794,632	38033,290	95,740
	190,000	102,093	432,963	187457,119	44202,353	101,764
	200,000	109,541	460,206	211789,562	50411,517	107,839
Sumatorios	1450,000	859,839	3606,693	1257202,607	298531,964	861,636
Varianza		374,830				371,157

Ejemplo de ajuste de función $n\log(n)$ (II)

El sistema resultante sería

$$11*a_0 + 3606,693*a_1 = 859.84$$

$$3606,693*a_0 + 1257202,607*a_1 = 298531,964$$

De donde $a_0 = 5.213$ y $a_1 = 0.223$. La ecuación de la curva para obtener los tiempos estimados sería:

$$t(n) = 5.213 + 0.223*n*\log(n).$$

La última columna de la tabla muestra los tiempos estimados por medio de la ecuación de la curva, que como se puede apreciar son muy similares a los reales.

Coeficiente de determinación

- ¿Cómo podemos saber si la curva se ajusta correctamente a los datos medidos?
- Para ello, se calcula el coeficiente de determinación, cuyo valor está comprendido entre 0 y 1. Para que el ajuste sea bueno, el coeficiente de determinación ha de ser superior a 0.9.
- Se calcula dividiendo la varianza de los tiempos estimados por la varianza de los tiempos observados (tiempos reales).
- Para el primer ejemplo el coeficiente es: $868493.111/874710 = 0.992$ (ajuste del 99.2 %).
- Para el segundo ejemplo el coeficiente es: $371,157/374,830 = 0.99$ (ajuste del 99%).