



MÉTRICAS DE EVALUACIÓN

Matriz de confusión, métricas y sobreajuste.

TABLA DE CONTENIDOS

01. INTRODUCCIÓN

02. MATRIZ DE
CONFUSIÓN

03. MÉTRICAS

04. SOBREAJUSTE

01

INTRODUCCIÓN

INTRODUCCIÓN

Tenemos el error que comete el método, pero :

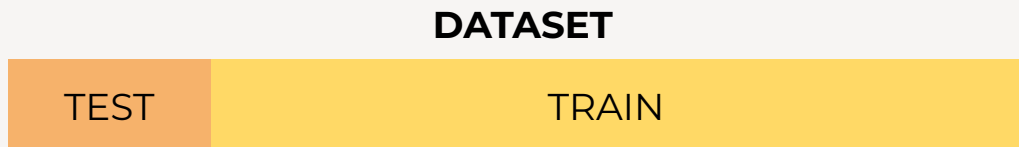
- ¿Cuántos ejemplos del conjunto de entrenamiento **clasifica correctamente**?
- ¿Cuántos ejemplos que **NO** pertenecen al conjunto de entrenamiento **clasifica correctamente**?
- ¿Qué significa “clasificar correctamente”?

OBJETIVO

- Evaluar un método de clasificación con **métricas estándar**.
- Comparar varios **métodos de clasificación** aplicados a un mismo conjunto de datos.
- Comparar un mismo método de clasificación aplicado a **diferentes conjuntos de datos**.
- Evaluar el efecto que produce el **cambio de parámetros** en un método.

EVALUACIÓN

- Dividir el conjunto de datos disponible en
 - Conjunto de Entrenamiento (training set)
 - Conjunto de Prueba (testing set)



Luego, entrenar y evaluar

1. **Entrenar** la red con el conjunto de entrenamiento.
2. **Evaluar** el método con el conjunto de **entrenamiento**.
3. **Evaluar** el método con el conjunto de **prueba**

EVALUACIÓN

1. ¿Cómo sabemos si esta **división es apropiada**?
2. ¿Cómo evaluamos cuantitativamente la **capacidad de clasificación**?
3. ¿Es verdad que si $\mathbf{E}(\mathbf{w}) \equiv \mathbf{0}$, entonces es un clasificador perfecto?

METODOLOGÍAS

- Matriz de Confusión
- Métricas para la evaluación del rendimiento
- Diseño del experimento para la evaluación

02

MATRIZ DE CONFUSIÓN

MATRIZ DE CONFUSIÓN

Tabla que permite evaluar el desempeño de un algoritmo de **clasificación**.

- **Columna** → cantidad de instancias en la clase dada por el método. (Predicción)
- **Fila** → cantidad de instancias en la clase verdadera. (Real/Actual)

		Prediction	
		Positive	Negative
Actual	Positive	TP	FN
	Negative	FP	TN

EJEMPLO

Consideremos un modelo de clasificación que fue entrenado para distinguir entre: perros y gatos.

Se toma una muestra de 27 animales:

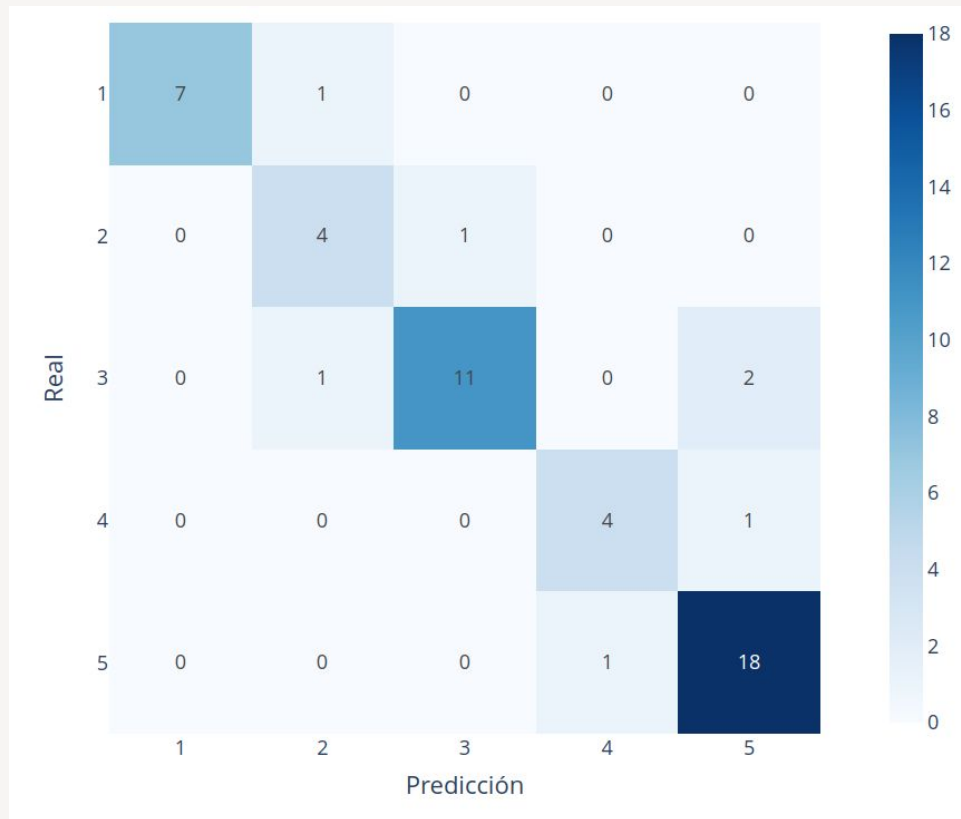
- 15 perros
- 12 gatos

Después de clasificar se obtiene la siguiente matriz de confusión:

	Perro	Gato
Perro	11	4
Gato	2	10

MATRIZ DE CONFUSIÓN MULTICLASE

Supongamos
que tenemos
5 clases



03

MÉTRICAS

MÉTRICAS

Criterios para evaluar la clasificación

Métricas estándar:

- Accuracy
- Precision
- Recall
- F1-Score
- Tasa de TP
- Tasa de FP

MÉTRICAS

ACCURACY

$$\frac{TP + TN}{TP + TN + FN + FP}$$

PRECISION

$$\frac{TP}{TP + FP}$$

RECALL

$$\frac{TP}{TP + FN}$$

F1-SCORE

$$\frac{2 * precision * recall}{precision + recall}$$

TASAS

Tasa de Verdaderos Positivos

$$\frac{TP}{TP + FN}$$

Tasa de Falsos Positivos

$$\frac{FP}{FP + TN}$$

PROCEDIMIENTO

Teniendo los dos conjuntos de datos, el de entrenamiento y el de prueba:

Dada una cantidad de épocas: epoch

1. Calcular los **w** utilizando el conjunto de entrenamiento.
2. **Clasificar** los datos del conjunto de prueba utilizando los w encontrados en el paso 1
3. Calcular el valor de las **métricas** para los ejemplos del conjunto de entrenamiento y luego para los del conjunto de prueba.

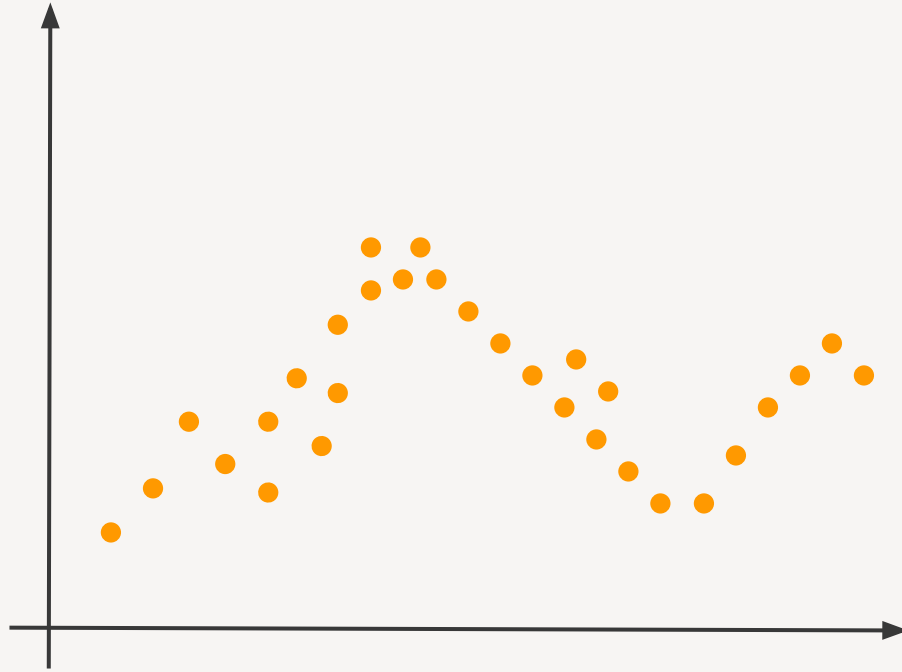
Realizar el mismo experimento para epoch = 1, 10, 20, . . . , 300

04

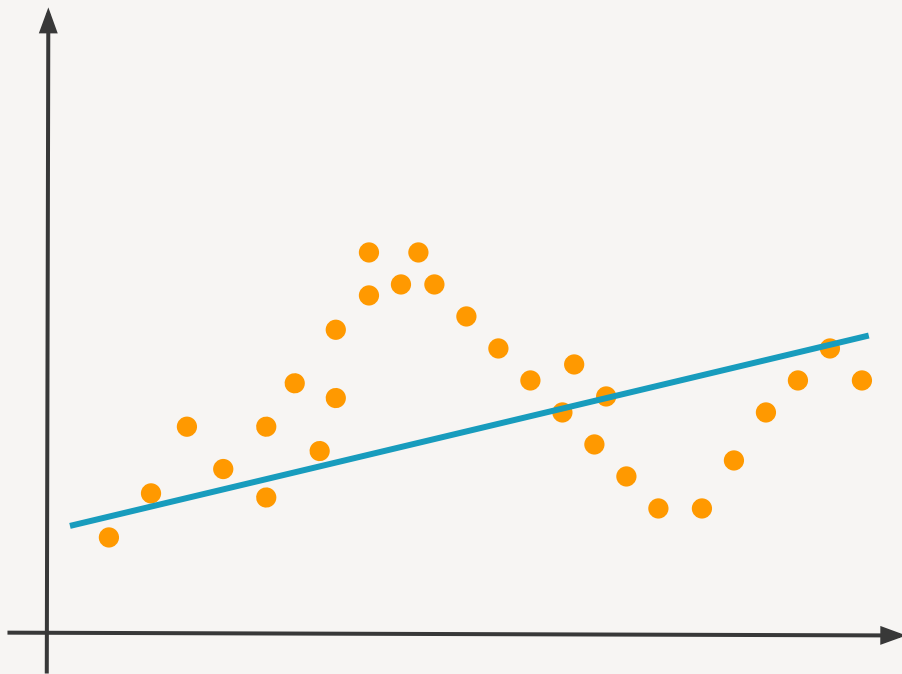
SOBREAJUSTE

Overfitting

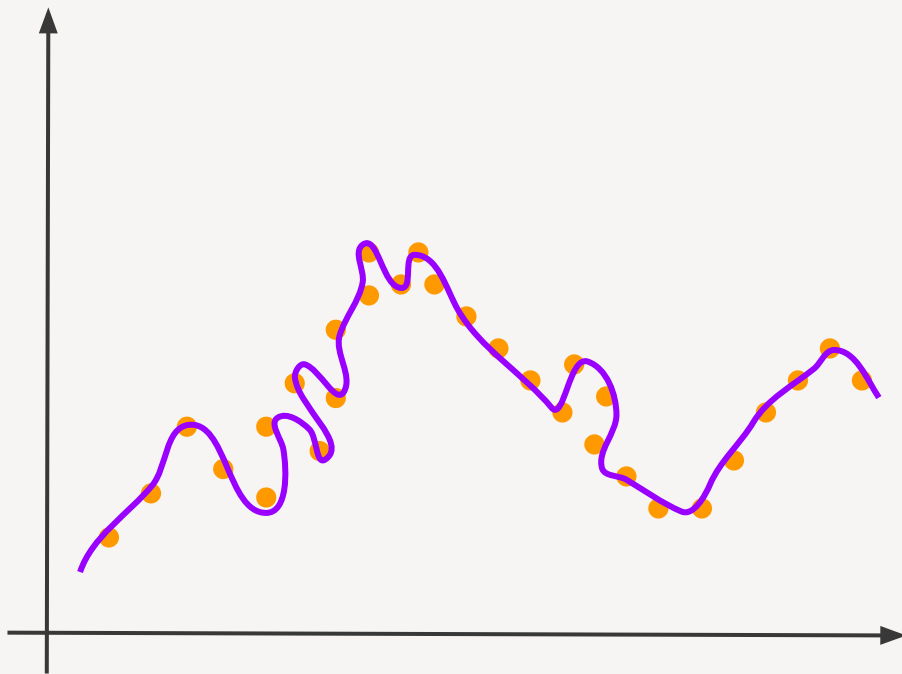
DATASET



UNDERFITTING



OVERFITTING



OVERFITTING / SOBREAJUSTE

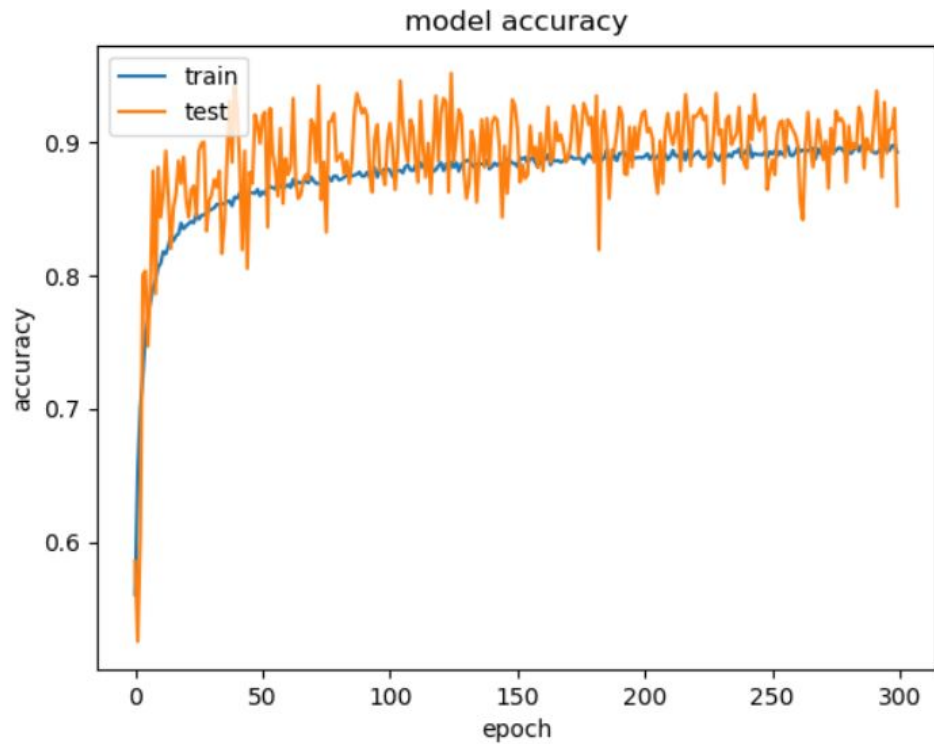
En cualquier método de aprendizaje:

Es el efecto de **sobreentrenar** un algoritmo de aprendizaje con datos para los que se conoce el resultado deseado.

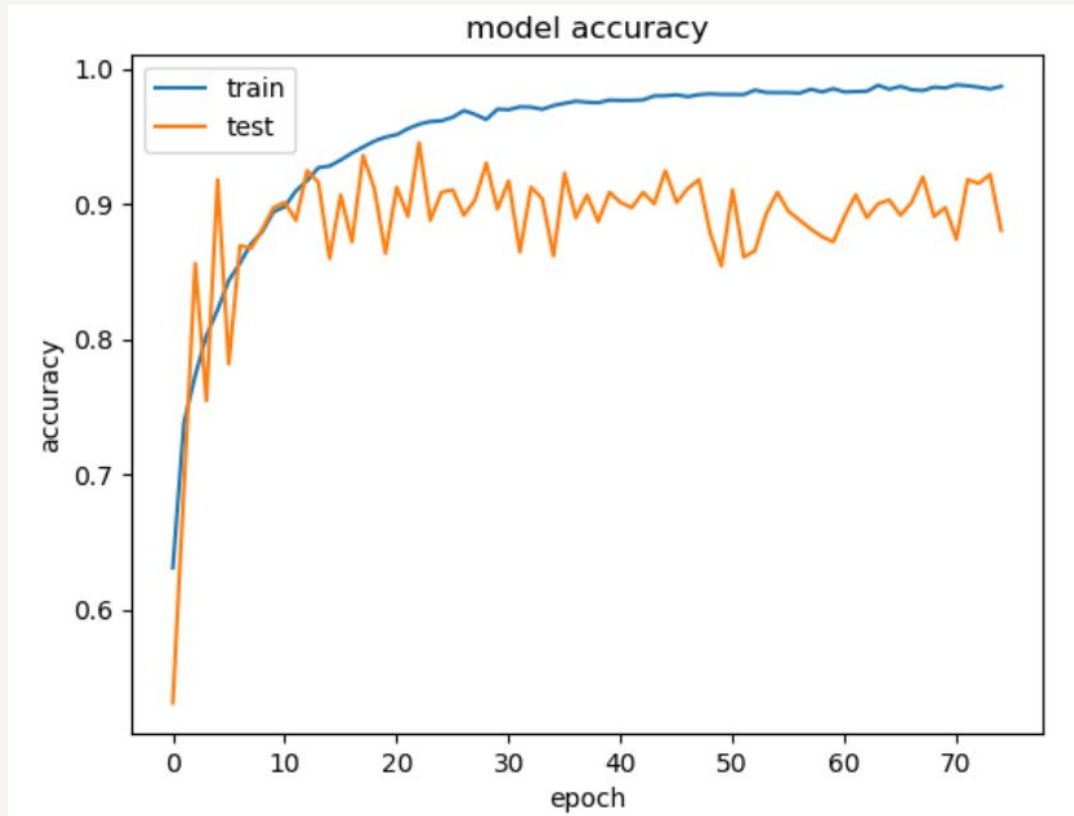
El método clasifica con gran precisión los datos del conjunto de entrenamiento

Pero... **no puede generalizar**

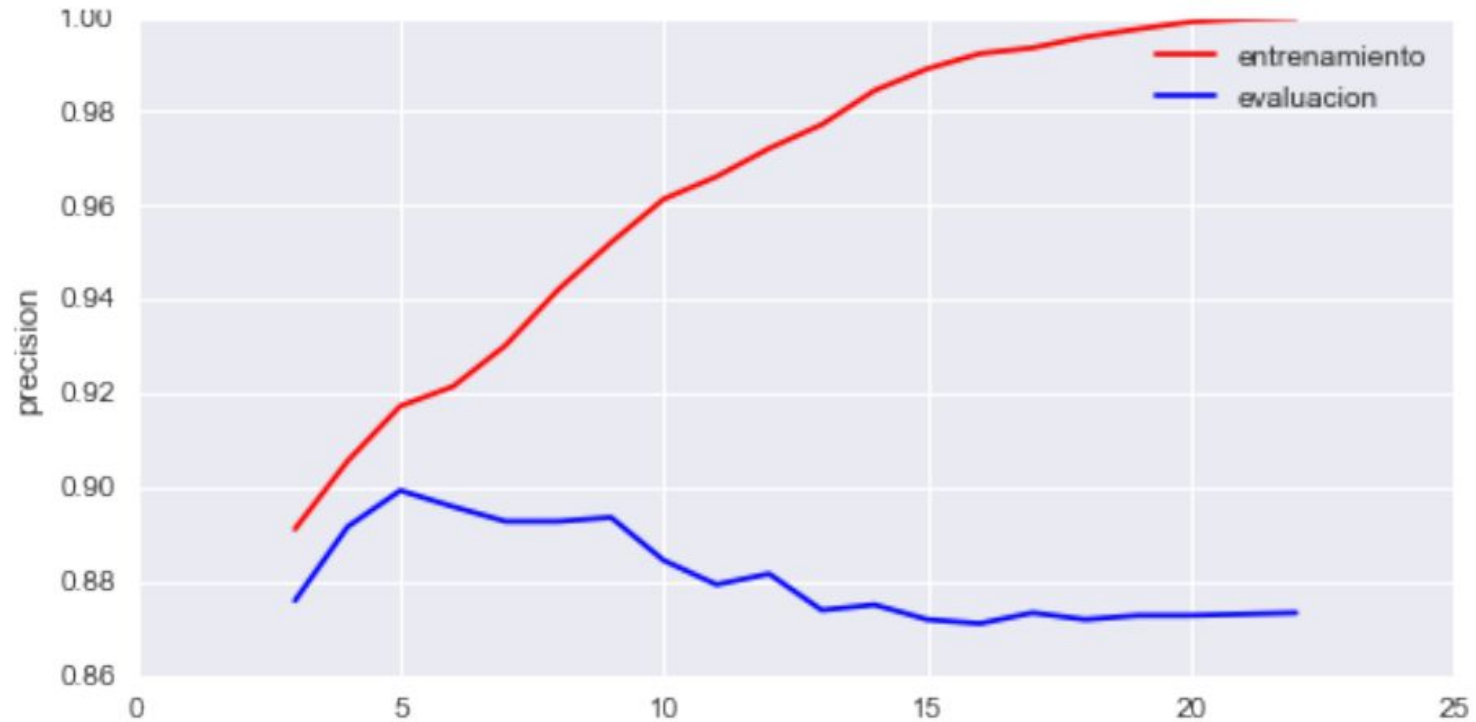
EJEMPLO ACCURACY



EJEMPLO SOBREAJUSTE



EJEMPLO SOBREAJUSTE



CAUSAS

El sobreajuste puede deberse a diversos factores:

- Conjunto de datos de entrenamiento **no balanceado**.
- **Pocos registros** en el conjunto de entrenamiento.
- Conjunto de datos de entrenamiento con mucho **ruido**.

¿SE PUEDE EVITAR?

¿Cómo sabemos si la partición train/test es apropiada?

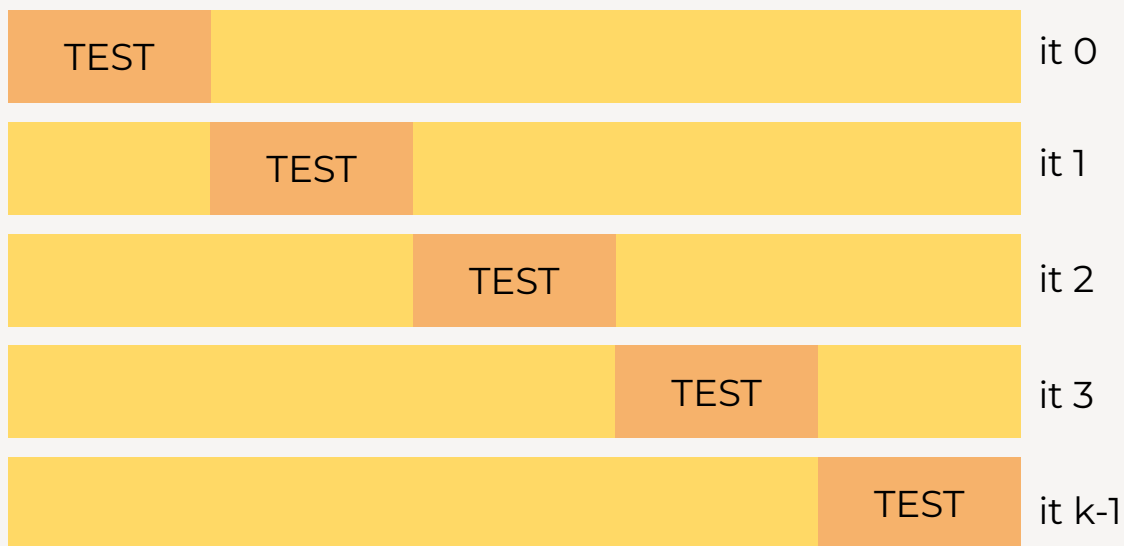
Utilizamos métodos de experimentación

Validación Cruzada (K-Fold Cross Validation)

Nota: NO es la única técnica

K-FOLD CROSS-VALIDATION

1. Se **divide** aleatoriamente el dataset en **k partes** “iguales”.
2. Tomar $k - 1$ partes y usarlas para entrenar (training_set) y el remanente como testing_set



K-FOLD CROSS-VALIDATION

for $j=1\dots T$ (generalmente $T=k$)

1. Entrenamiento con el j -ésimo conjunto
2. Evaluación del método usando como testing set la parte separada para tal efecto.
3. Se calcula alguna medida de la precisión del método.

Finalmente se obtienen parámetros de desempeño del promedio de las T iteraciones.

RESUMEN

