

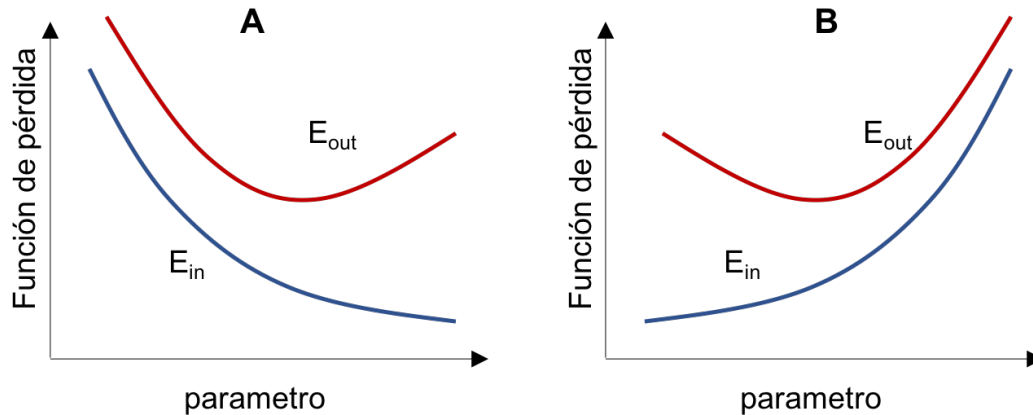
Examen rápido No. 4

Reconocimiento de Patrones (2023-2)

Julio Waissman Vilanova

Nombre: _____

1. Considera las siguientes figuras A y B

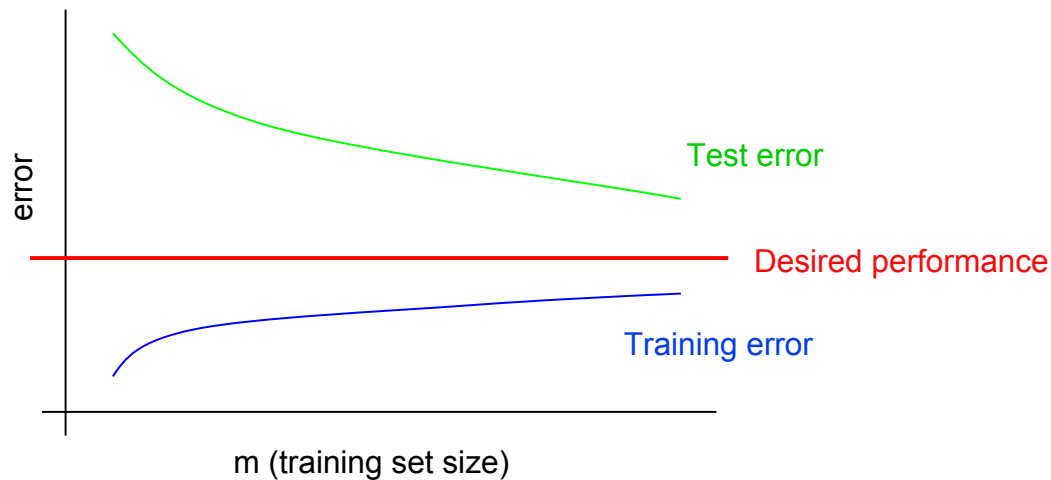


Asigna cual es la curva sobre los errores en muestra E_{in} y fuera de muestra E_{out} que debería salir teóricamente para los siguientes parámetros de ajuste de métodos de aprendizaje:

- A B Número de neuronas en la capa oculta en una red neuronal.
- A B Parámetro λ de regularización en regresión lineal.
- A B Umbral γ entre 0 y 1 que es el valor por el cual se considera que un objeto pertenece a la clase 1 en regresión logística (por default $\gamma = 0.5$).
- A B Valor de C es una SVM con kernel lineal.
- A B Número de capas ocultas en una red neuronal.
- A B Parámetro σ utilizado en el kernel gaussiano para una SVM.

2. Supongamos que se esta resolviendo un problema de análisis de sentimientos en documentos utilizando como método para clasificar los diferentes sentimientos que se pueden obtener de un *twit* utilizando un algoritmo basado en regresión logística. Si tenemos al rededor de 5000 palabras diferentes en nuestra *bolsa de palabras*, 4 sentimientos básicos («satisfecho», «molesto», «triste», «otro») y al rededor de 20,000 *twits* previamente clasificados.

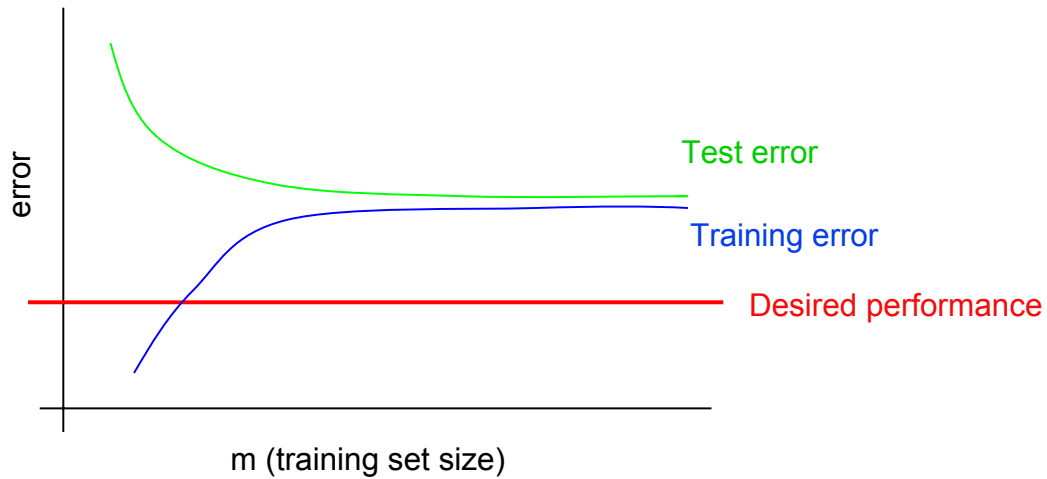
Si realizamos una curva de aprendizaje y obtenemos algo similar a la curva siguiente:



subraya las acciones que podrían mejorar al sistema de aprendizaje.

- Tratar de generar más *twits* clasificados.
 - Tratar de reducir el número de palabras de la *bolsa de palabras*.
 - Tratar de aumentar las palabras de la *bolsa de palabras*.
 - Agregar otras características como la longitud del *twit* o la hora a la que fue enviado.
 - Aumentar el número máximo de iteraciones del algoritmo de optimización.
 - Aumentar el valor de λ (parámetro de regularización).
 - Disminuir el valor de λ (parámetro de regularización).
 - Utilizar una SVM con kernel gaussiano.
3. Supongamos que estamos estimando la demanda de energía eléctrica doméstica en la Cd. de Hermosillo para el próximo día, utilizando como información el consumo de energía eléctrica de los 30 días anteriores, la temperatura máxima en Hermosillo de los 30 días anteriores, la temperatura mínima en Hermosillo de los 30 días anteriores, el día de la semana, una variable que indica si el día es festivo o no y una variable que indica la estación del año (invierno, primavera, verano y otoño). Se aplica un método de regresión lineal con la información de los últimos 5 años.

Para analizar el desempeño del algoritmo de regresión lineal, se realiza una curva de aprendizaje la cual resulta ser de la forma siguiente:



subraya las acciones que podrían mejorar al sistema de aprendizaje.

- Solicitarle a CFE información de otros 5 años anteriores.
- Disminuir el valor de λ (parámetro de regularización).
- Aumentar el valor de λ (parámetro de regularización).
- Utilizar solo la información histórica de los últimos 15 días y no de los 30 días anteriores.
- Utilizar una red neuronal en lugar de la regresión lineal.
- Agregar como atributos la raíz cuadrada de la demanda de energía eléctrica de los 30 días anteriores y la raíz cuadrada de los valores máximos y mínimos de temperatura de los 30 días anteriores.
- Agregar la humedad relativa de los 30 días anteriores.

4. Sea la siguiente matriz de confusión, resuelta después de utilizar un método de aprendizaje para clasificar datos de un problema real:

		y	
		0	1
$h_{\theta}(x)$	0	300	5
	1	10	30

Responde a las siguientes preguntas:

- a) ¿Cual es el error de clasificación? _____
- b) ¿Cual es la precisión del clasificador? _____
- c) ¿Cual es el *recall* del clasificador? _____
- d) ¿Cual es el F_1 -score del clasificador? _____