

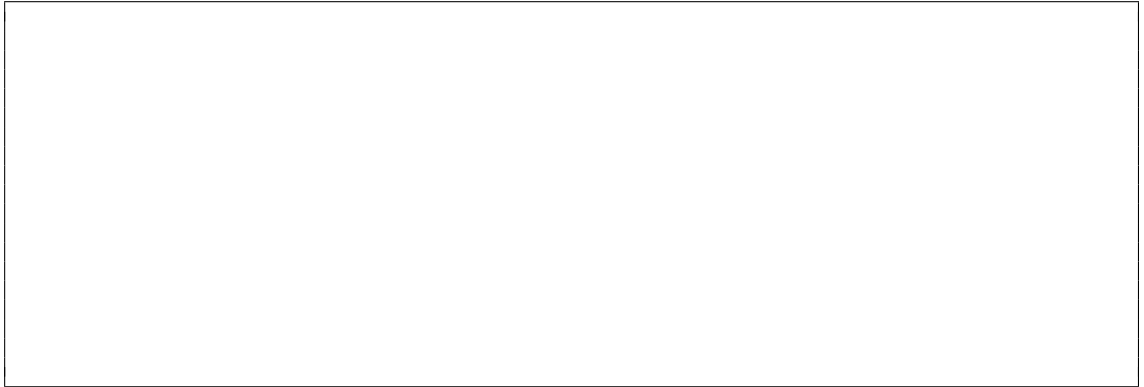
**Examen intermedio**  
**Redes neuronales, período 2017–2.**  
Profesor: Julio Waissman Vilanova.

Nombre: \_\_\_\_\_

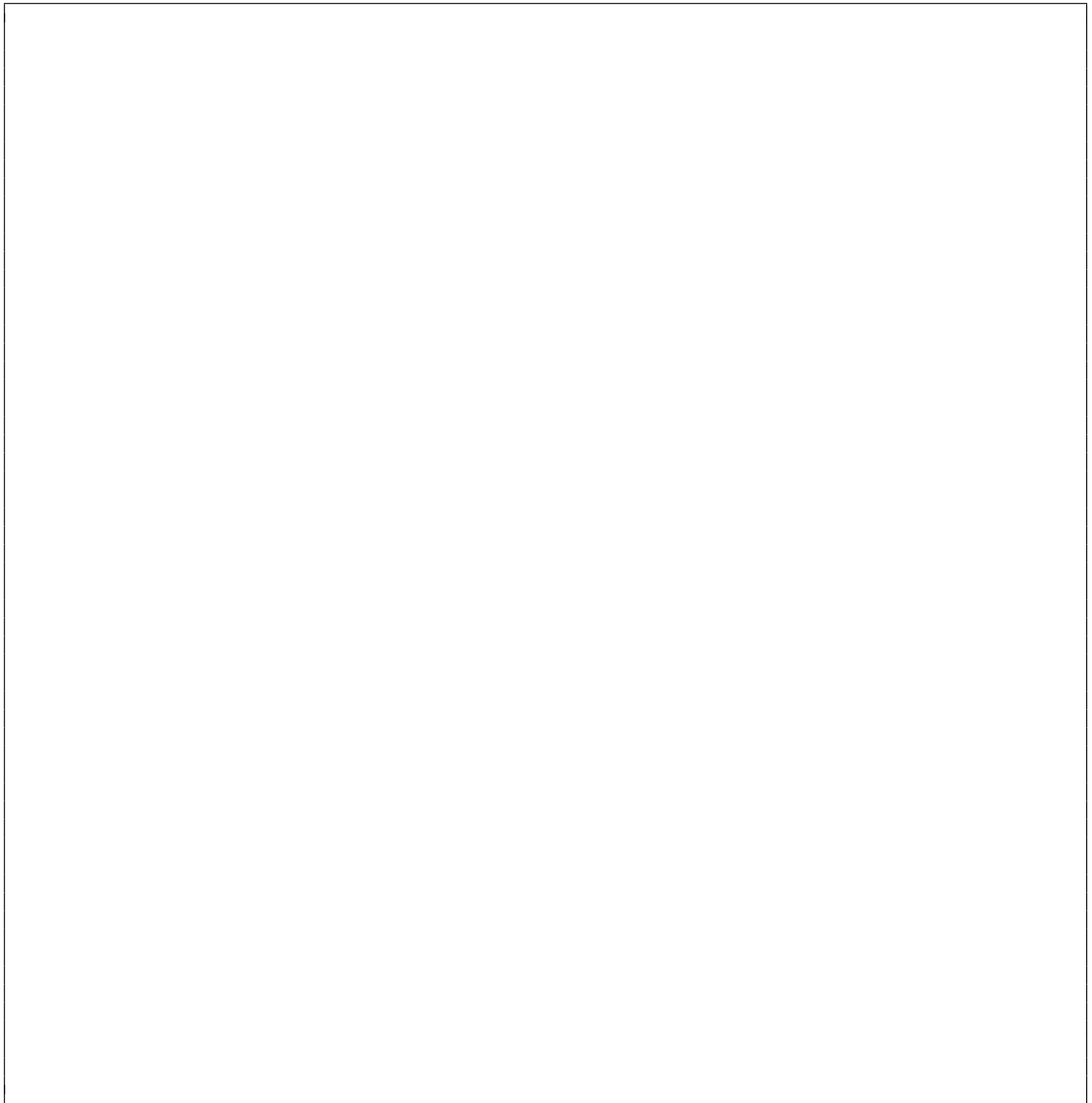
1. (5 puntos) La empresa *Necrosoft, Inc.* vende software y películas piratas por internet. Para esto tienen un sitio atractivo, una página en *facebook* y un sistema de pagos por tarjeta de crédito y *Paypal*. Dado que es una empresa de punta utiliza métodos de reconocimiento de patrones para analizar y administrar la información empresarial y como ayuda a la toma de decisiones (una cosa llamada por los gringos como *business intelligence*. Para las siguientes tareas selecciona si la tarea es de regresión (R), clasificación (C), aprendizaje no supervisado (N) u otro tipo de aprendizaje (O).
- (a) \_\_\_\_ Predecir el tráfico en la página web para el día de mañana, conociendo el tráfico en los días pasados.
  - (b) \_\_\_\_ Determinar si un usuario es cliente potencial de películas o de software para enviarle publicidad orientada a su correo electrónico.
  - (c) \_\_\_\_ Estimar el ancho de banda necesario para el próximo año de la página del sistema.
  - (d) \_\_\_\_ Encontrar, a partir de la página de *facebook* grupos de personas con características comunes para seleccionar el software a piratear.
  - (e) \_\_\_\_ Predecir las ventas a finales de mes sobre hipotéticas ofertas, con el fin de realizar planeación estratégica en la empresa.
2. (10 puntos) Consideremos una red neuronal con 2 neuronas en la capa de entrada, dos neuronas en la primera capa oculta, dos neuronas en la segunda capa oculta y una neurona en la capa de salida. Todas las neuronas utilizan una función de activación logística. Las matrices de peso de la red neuronal están dadas por:

$$W^{(1)} = \begin{bmatrix} -1.0 & 2.0 & -1.0 \\ 1.0 & 0.5 & 1.0 \end{bmatrix} \quad W^{(2)} = \begin{bmatrix} -1.0 & -2.0 & 4.0 \\ 0.1 & 0.2 & 0.3 \end{bmatrix} \quad W^{(3)} = \begin{bmatrix} 1.0 & 1.0 & 1.0 \end{bmatrix}$$

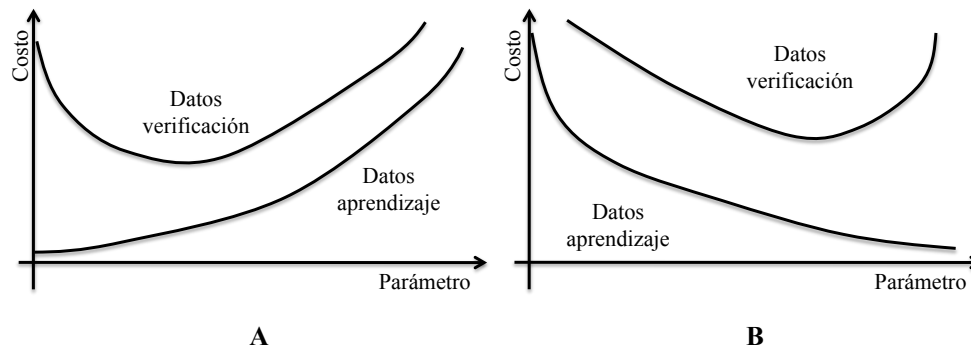
- (a) (3 puntos) La salida para una entrada  $x = [1, 1]^T$  es \_\_\_\_\_.
- (b) (1 punto) El criterio de pérdida para este tipo de red neuronal es: \_\_\_\_\_.
- (c) (2 puntos) El costo para este único dato de entrada es \_\_\_\_\_.
- (d) (4 puntos) Si para  $x = [1, 1]^T$  sabemos que le corresponde la clase 1, entonces calcula, utilizando *backpropagation*, las matrices de gradientes  $\nabla_{W^l} J(W)$  para  $l = 1, 2, 3$ .



3. (15 puntos) Ahora vamos a revisar el costo computacional de el aprendizaje de una red neuronal. Consideremos una estructura de red neuronal en alimentación hacia adelante, en la cual todas las neuronas de la capa anterior se encuentran completamente conectadas con las neuronas de la capa siguiente.
- (a) (5 puntos) Considera una red neuronal con una sola capa oculta, con 5 neuronas de entrada, 3 neuronas en la capa oculta, y una neurona en la capa de salida. Solo vamos a contar como operaciones los productos de la forma  $w_{i,j}^{(l)} a_j^{(l-1)}$ ,  $w_{i,j}^{(l)} a_i^{(l)}$  y  $a_i^{(l-1)} \delta_j^{(l)}$  manteniendo la notación utilizada en el curso. ¿Cual es el número total de operaciones requeridas para realizar un solo *epoch* de una operación del algoritmo de *backpropagation*, si contamos con un solo dato en el conjunto de aprendizaje? \_\_\_\_\_.
- (b) (5 puntos) Vamos a llamar *nodo* a cualquier unidad de una red neuronal, independientemente si es entrada, salida, o neuronas de la capa oculta (incluidas las neuronas para  $a_0^{(l)}$ ). Consideremos una red con 10 unidades de entrada ( $a_0^{(0)}$  se incluye en las unidades de entrada), una sola unidad de salida, y 36 unidades ocultas. Las neuronas ocultas se pueden acomodar en tantas capas ocultas como se quiera, siempre que las neuronas de la capa  $l - 1$  se encuentren completamente conectadas con la capa  $l$ .
1. ¿Cuál es el menor número de pesos que se puede lograr con esta red neuronal?  
\_\_\_\_\_.
  2. ¿Cuál es el mayor número de pesos que se puede lograr con esta red neuronal?  
\_\_\_\_\_.
- (c) (5 puntos) Supongamos ahora que la función de activación de cada unidad es lineal, el lugar de una función logística. Esto es  $g(z_i^{(l)}) = z_i^{(l)}$ . Supongamos ahora que tenemos una red con 10 neuronas de entrada, 20 neuronas en la primer capa oculta, 20 neuronas en una segunda capa oculta y 1 neurona de salida. Demuestre que con este tipo de funciones de activación, dicha red neuronal es equivalente a solamente tener las 10 neuronas de la capa de entrada y la neurona de la capa de salida directamente.



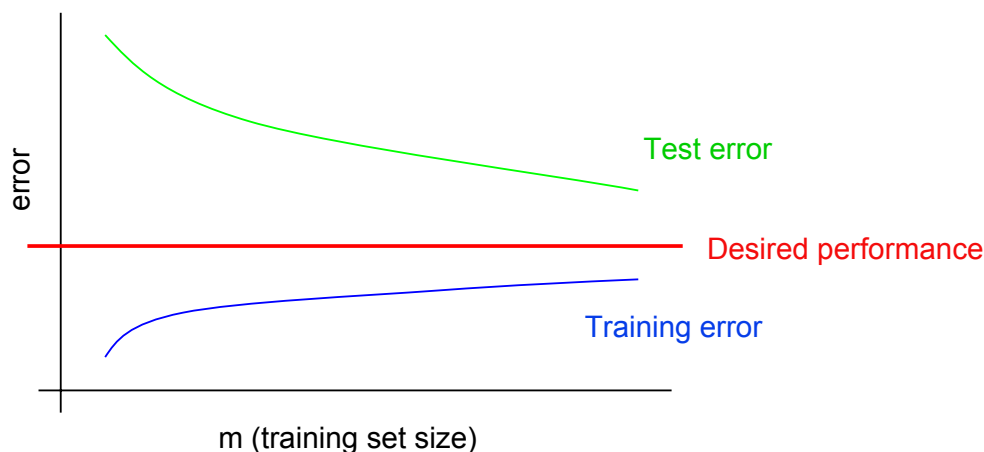
4. (10 puntos) Considera las siguientes figuras A y B donde el costo en los datos de validación y verificación (después del aprendizaje), se grafican respecto a el valor de un parámetro. Cada punto de la curva representa el valor final de un proceso de aprendizaje completo.



Asigna cual es la curva (A o B) que podría corresponder al variar cada uno de los siguientes parámetros:

- (a) \_\_\_ Número de neuronas en la capa oculta (asumiendo una sola capa oculta).
  - (b) \_\_\_ Número de capas ocultas en una red neuronal.
  - (c) \_\_\_ Valor de la tasa de aprendizaje  $\alpha$ .
  - (d) \_\_\_ Valor del factor de olvido  $\lambda$  del método de momento (inercia).
  - (e) \_\_\_ Umbral  $\gamma$  entre 0 y 1 por el cual se considera que un objeto pertenece a la clase 1, si la salida de la red neuronal es logística (por default  $\gamma = 0.5$ ).
  - (f) \_\_\_ Número de epochs utilizados en el aprendizaje.
5. (8 puntos) 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 una red neuronal hacia adelante. Si tenemos al rededor de 5000 palabras diferentes en nuestra *bolsa de palabras*, 4 sentimientos básicos («satisfecho», «molesto», «triste», «otro») y al rededor de 200,000 *twits* previamente clasificados.

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

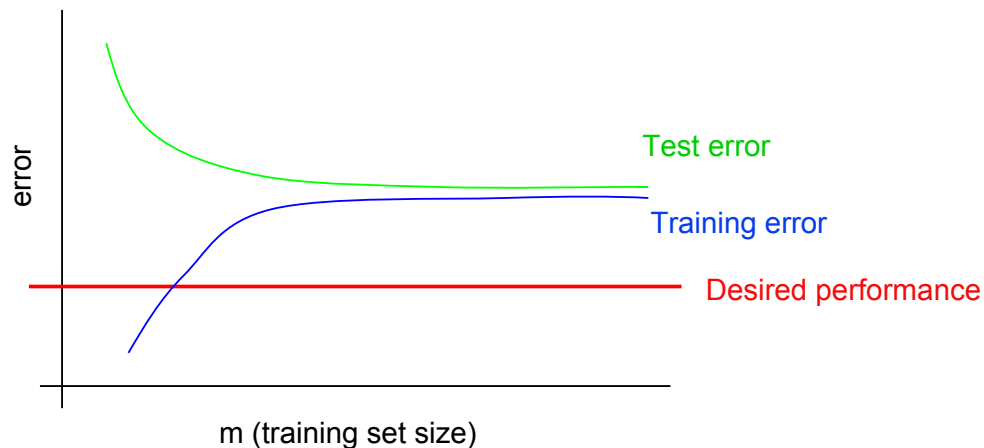


selecciona las acciones que podrían mejorar al sistema de aprendizaje.

- ☐ 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 *tweet* o la hora a la que fue enviado.
  - ☐ Aumentar el número máximo de epochs.
  - ☐ Aumentar el valor de  $\alpha$  (tasa de aprendizaje).
  - ☐ Disminuir el valor de  $\alpha$ .
  - ☐ Utilizar mas neuronas en la capa oculta.
6. (7 puntos) 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 una red neuronal con una sola capa oculta 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:



selecciona las acciones que podrían mejorar al sistema de aprendizaje.

- ☐ Solicitarle a CFE información de otros 5 años anteriores.
- ☐ Disminuir el valor de  $\alpha$  (tasa de aprendizaje).
- ☐ Aumentar el valor de  $\alpha$ .
- ☐ Utilizar solo la información histórica de los últimos 15 días y no de los 30 días anteriores.
- ☐ Utilizar dos capas ocultas.
- ☐ 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.