Examen de Teoría de Redes Neuronales

Máster MIARFID, Universitat Politècnica de València, Febrero de 2019

Apellidos: | | **Nombre:** | |

Cuestiones (2 puntos, 30 minutos, sin apuntes)

1 $\boxed{C}$ En el marco de las funciones discriminates lineales, indicar qué afirmación es cierta:

A) El valor de la función discriminante en un punto es igual a la distancia del punto al hiperplano separador.

B) El valor de la función discriminante en un punto es igual a la distancia del punto al hiperplano separador dividido por el módulo de vector de pesos.

C) El valor de la función discriminante en un punto es proporcional a la distancia del punto al hiperplano separador.

D) El valor de la función discriminante en un punto siempre es mayor que la distancia del punto al hiperplano separador.

2 $\boxed{B}$ En un problema de clasificación en 10 clases se a utilizar una red neuronal donde a partir de diferentes variables de entrada y con 10 neuronas en la capa de salida pueda resolver dicho problema. En este sentido, indicar cuál de las siguientes funciones objetivo es el más adecuado para entrenar la red:

A) Error cuadrático medio.

B) La entropía cruzada.

C) Una combinación del error cuadrático y la entropía.

D) Softmax

3 $\boxed{C}$ Dada una neurona definimos su valor pre-activación (o entrada total) como x y su valor post-activación como y, por lo tanto $y = f(x)$ siendo $f()$ la función de activación. Si en un punto x_0 $f(x_0) = 0.3$ y $f()$ una sigmoid, ¿cuál es la derivada de $f()$ en x_0?

A) 1

B) 0.7

C) 0.21

D) -0.3

4 $\boxed{A}$ En un problema de regresión se va a utilizar un perceptrón multicapa y el algoritmo BackPropagation para entrenar los pesos del perceptrón multicapa. Para decidir cuándo parar el entrenamiento, indicar qué criterio se debe usar en general:

A) Se para cuando el error cuadrático calculado en un conjunto distinto del de entrenamiento alcanza un mínimo.

B) Se para cuando el error cuadrático del conjunto de entrenamiento alcanza un mínimo.

C) Se para cuando el error cuadrático del conjunto de entrenamiento empieza a oscilar.

D) Se para cuando el error cuadrático calculado en un conjunto distinto del de entrenamiento alcanza un máximo.

5 $\boxed{A}$ En el algoritmo Contrastive Divergence de 1 paso, el gradiente de la matriz de pesos se obtiene como:

A) $(h_0 x_0^t - h_1 x_1^t)$

B) $(x_0 x_1^t + h_0 h_1^t)$

C) $(h_0 x_0^t + h_1 x_1^t)$

D) $(x_0 x_1^t - h_0 h_1^t)$

6 $\boxed{\text{D}}$ Dado un conjunto de muestras de aprendizaje $X = \{(\boldsymbol{x}_1, c_1), (\boldsymbol{x}_2, c_2), \cdots, (\boldsymbol{x}_{10}, c_{10})\}$, donde $\boldsymbol{x} \in \mathcal{R}^d$ y $c_i \in \{-1, +1\}$ se calculan los multiplicadores de Lagrange óptimos para una SVM con márgenes blandos ($C = 20$), obteniéndose $0, 0, 0, 0, 20, 3, 2, 20, 0, 0$, respectivamente. Indicar la afirmación correcta:

A) Los vectores soporte son el 1, 2, 3, 4, 9 y 10.

B) El umbral se calcula a partir de la muestra 1.

C) El umbral se calcula a partir de la muestra 8.

D) Los vectores soporte son el 5, 6, 7 y 8.

7 $\boxed{\text{C}}$ Siendo R la recompensa total (total reward), γ el factor de descuento, a una acción, s un estado y Q la función de valor estado-accion. El objetivo del Reinforcement Learning es encontrar un "optimal policy" tal que:

A) $R = \sum_{t=0}^{T} \gamma^t r_{t+1}$

B) $Q^\pi(s, a) = \mathbb{E}[R \mid s, a, \pi^*]$

C) $\pi^* = \text{argmax}_\pi \, \mathbb{E}[R \mid \pi]$

D) $Q^\pi(s, a) = \mathbb{E}[R \mid s, a, \pi]$

8 $\boxed{\text{C}}$ Una GAN, Generative Adversarial Network, está compuesta principalmente por:

A) Una red neuronal encoder-decoder, normalmente son denoising autoenconders.

B) Una red neuronal encoder-decoder, normalmente son variational autoenconders.

C) Una función G derivable, normalmente una red neuronal, y una función D derivable, normalmente una red neuronal.

D) Una variable latente z y una variable real x como entrada y salida respectivamente de un autoencoder.

9 $\boxed{\text{A}}$ El objetivo principal de añadir ruido gaussiano al resultado de las funciones de activación, o en la entrada de la red, es:

A) Aportar generalización

B) Mejorar la convergencia del algoritmo SGD

C) Alcanzar un mínimo global de la función de coste

D) Emplear Batch Normalization

10 $\boxed{\text{A}}$ Hay diversos algoritmos para entrenar una red recurrente simple a partir de un conjunto de datos de entrenamiento. Idicar cuál es la afirmación correcta

A) El algoritmo basado en el gradiente exacto es más lento que el "Back-propagation through time" en cada iteración.

B) El algoritmo basado en el gradiente exacto es más rápido que el "Back-propagation through time" en cada iteración.

C) El algoritmo basado en el gradiente exacto tiene la misma complejidad temporal que el "Back-propagation through time" en cada iteración.

D) El algoritmo basado en el gradiente exacto necesita la misma memoria que el "Back-propagation through time".

Examen de Teoría de Redes Neuronales

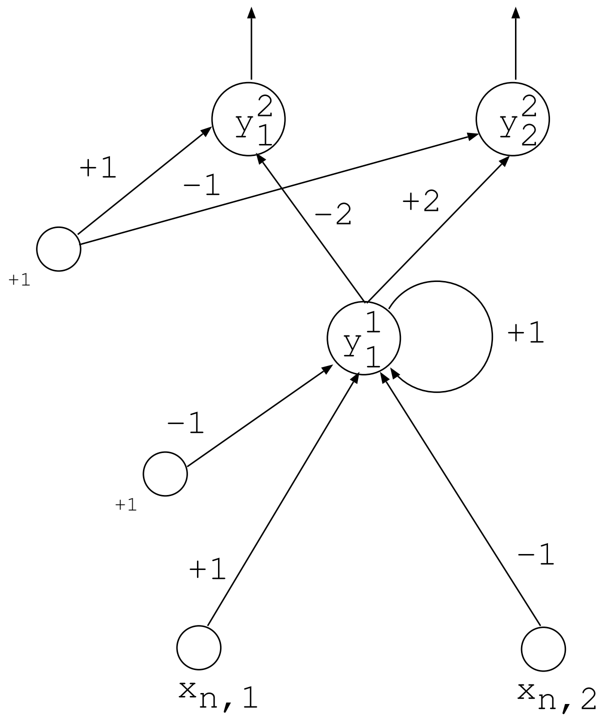
Máster MIARFID, Universitat Politècnica de València, Febrero de 2019

Apellidos: **Nombre:**

Problemas (4 puntos, 60 minutos, con apuntes)

Problema 1 - (2 puntos)

Las funciones de activación en los nodos de la figura son del tipo ReLU: $f(z) = \begin{cases} z & z \geq 0 \\ 0 & z < 0 \end{cases}$.



Dada una secuencia de entrada $(+3, +1), (+1, 0), (-1, +2)$

a) Calcular la secuencia de los valores del nodo oculto Y_2^1. (1 punto)

b) Calcular la secuencia de los valores del nodo de salida Y_1^2 e Y_2^2. (1 punto)

a)

$$y_{n,1}^1 = f((-1) + (+1) \cdot x_{n,1} + (-1) \cdot x_{n,2} + (+1) \cdot y_{n-1,1}^1)$$

$$y_{1,1}^1 = f((-1) + (+1) \cdot (+3) + (-1) \cdot (+1) + (+1) \cdot (0)) = f(+1) = +1$$
$$y_{2,1}^1 = f((-1) + (+1) \cdot (+1) + (-1) \cdot (0) + (+1) \cdot (+1)) = f(+1) = +1$$
$$y_{3,1}^1 = f((-1) + (+1) \cdot (-1) + (-1) \cdot (+2) + (+1) \cdot (+1)) = f(-3) = 0$$

b)

$$y_{n,1}^2 = f((+1) + (-2) \cdot y_{n,1}^1)$$
$$y_{n,2}^2 = f((-1) + (+2) \cdot y_{n,1}^1)$$

$$y_{1,1}^2 = f((+1) + (-2) \cdot (+1)) = f(-1) = 0$$
$$y_{1,2}^2 = f((-1) + (+2) \cdot (+1)) = f(+1) = +1$$
$$y_{2,1}^2 = f((+1) + (-2) \cdot (+1)) = f(-1) = 0$$
$$y_{2,2}^2 = f((-1) + (+2) \cdot (+1)) = f(+1) = +1$$
$$y_{3,1}^2 = f((+1) + (-2) \cdot (0)) = f(+1) = +1$$
$$y_{3,2}^2 = f((-1) + (+2) \cdot (0)) = f(-1) = 0$$

Problema 2 - (2 puntos)

Siguiendo el estandar de notación para los mapas 4D en una red convolucional: {batch,canales,filas,columnas}, se tiene un mapa con las siguientes dimensiones $Z = 100 \times 32 \times 16 \times 16$. A dicho mapa Z se le aplican dos capas convolucionales diferentes:

X :

- $N = 64$
- $K_R = K_c = 3$
- $stride = 2$
- $padding = 1$

Y :

- $N = 128$
- $K_R = K_c = 3$
- $stride = 4$
- $padding = 1$

Indica:

a) Mapa de salida después de X: (**0.25 puntos**)

Batch=100, Canales=64, filas=columnas=lower((16+(2*1)-3)/2+1)=8, por lo tanto:

$100 \times 64 \times 8 \times 8$

b) Mapa de salida después de Y: (**0.25 puntos**)

Batch=100, Canales=128, filas=columnas=lower((16+(2*1)-3)/4+1)=8, por lo tanto:

$100 \times 128 \times 4 \times 4$

Siguiendo la idea de un modelo Inception quisiéramos poder combinar ambos mapas de salida concatenándolos. Sin embargo los tamaños no son compatibles.

c) Indica qué habría que hacer para poder concatenar X e Y: (**0.5 puntos**)

Lo más normal sería tener que reducir la dimensionalidad de los mapas de X para adecuarla a 4×4. Para ello se podrían hacer dos cosas: Un pooling de 2×2 o bien una nueva convolución con $stride = 2$.

d) Indica qué tamaño final se obtendría de dicha concatenación. (**0.5 puntos**)

En el caso de haber hecho un pooling se obtendría el siguiete mapa 4D:

$100 \times 192 \times 4 \times 4$

En el caso de haber hecho una nueva convolución con stride=2, dependería del número de filtros M de esta nueva convolución. En todo caso sería:

$100 \times (M + 64) \times 4 \times 4$

Partiendo del mapa original Z

e) ¿ Qué tipo de convolución necesitaríamos para obtener un mapa como este: $100 \times 128 \times 1 \times 1$? (**0.25 puntos**)

- $N = 128$
- $K_R = K_c = 16$
- $stride = 1$
- $padding = 0$

f) Si a continuación realizáramos un flatten (reshape lineal), qué tamaño de tensor obtendríamos? (**0.25 puntos**)

100×128