Examen de Teoría de Redes Neuronales

Máster MIARFID, Universitat Politècnica de València, Febrero de 2020

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

Cuestiones (2 puntos, 30 minutos, sin apuntes)

1 $\boxed{C}$ En una tarea de clasificación donde dos o más objetos se representan mediante el mismo vector aunque sean de clases distintas, indicar qué afirmación es cierta:

A) Un buen clasificador debe suministar la misma probabilidad para todas las clases a todos los objetos que se representan con el mismo vector.

B) Dos o más objetos se pueden representar con el mismo vector solo si son de la misma clase.

C) La representación de un objeto mediante un vector se realiza mediante una serie de transformaciones a partir de las cuales se obtienen las características que permiten distinguir los objetos de una clase con respecto a los de otra clase, pero suponen pérdidas de información del objeto.

D) La representación de un objeto mediante un vector se realiza mediante una serie de transformaciones que permiten distinguir los objetos de una clase con respecto a los de otra clase, pero no suponen pérdidas de información del objeto.

2 $\boxed{C}$ En un problema de clasificación en 10 clases se utiliza una red neuronal con 10 unidades de entrada y con 10 neuronas en la capa de salida. En este sentido, indicar cuál de las siguientes afirmaciones es falsa:

A) Los pesos de la red neuronal se obtienen minimizando la entropía cruzada

B) Los pesos de la red neurnal se pueden obtener minimizando el error cuadrático medio.

C) La red neuronal está mal diseñada porque el número de clases, el vector de características de entrada y el tamaño de la capa de salida no siguen los criterios básicos que debe seguir un clasificador basado en una red neuronal.

D) Es conveniente que las funciones de activación de la capa de salida sean softmax.

3 $\boxed{A}$ Supongamos que en la capa de salida de un perceptrón multicapa para un problema de clasificación de 5 clases se utiliza la función softmax y los valores de la capa de salida antes de aplicar la función de activación son: $1,\ 0.5,\ 5.5,\ 2.1, 2.0$. Indicar cuál de las siguientes alternativas se obtendrá al aplicar la función de activación:

A) $0.0103,\ 0.0062,\ 0.9247,\ 0.0309,\ 0.0279$

B) $-0.0103,\ 0.0062,\ -0.9247,\ 0.0309,\ -0.0279$

C) $0.7310,\ 0.6224,\ 0.9959,\ 0.8909,\ 0.8808$

D) $0.0900,\ 0.0450,\ 0.4955,\ 0.1892,\ 0.1802$

4 $\boxed{C}$ Cuando nos enfrentamos a un problema dónde se debe emplear una topología con muchas capas podemos considerar que las Deep Belief Networks son:

A) Una forma de evitar el problema del desvanecimiento de gradiente "vanishing gradient"

B) Una forma de poder entrenar un auto-encoder

C) Una forma de inicializar los pesos de manera adecuada

D) Una forma de poder entrenar de manera supervisada

5 $\boxed{D}$ En un MLP con una capa oculta de H neuronas y funciones de activaciónn Leaky ReLU con parámetro 0.01. Cuál de las siguientes afirmaciones con respecto a la activación de dichas neuronas de la capa oculta es correcta

A) Ninguna activación será mayor que 0.01

B) Ninguna activación será negativa

C) Ninguna activación será menor que -0.01

D) Ninguna de las anteriores

6 $\boxed{C}$ ¿Cuál de las afirmaciones siguientes es falsa?

A) Con una función de activación Max-out de al menos 2 matrices de pesos W_k podríamos obtener un comportamiento lineal

B) Con una función de activación Max-out de al menos 2 matrices de pesos W_k podríamos obtener una ReLu

C) Con una función de activación Max-out de al menos 2 matrices de pesos W_k podríamos obtener una activación con la función sigmoide

D) Con una función de activación Max-out de al menos 2 matrices de pesos W_k podríamos obtener un comportamiento no-lineal

7 $\boxed{B}$ En relación a las ResidualNets, su principal ventaja es:

A) Son adecuadas para emplear en capas convolucionales

B) Son adecuadas para evitar el desvanecimiento de gadiente “vanishing gradient” en redes muy profundas

C) Son adecuadas para obtener convergencias más rápidas

D) Son adecuadas para evitar el “covariance shift” que introduce batch-norm

8 $\boxed{B}$ En Deep Q-Learning se entiende por “non-stationary targets” el proceso por el cual:

A) Se repite el entrenamiento varias veces para las mismas muestras compuestas por tuplas $\{s_t, a_t, r_{t+1}, s_{t+1}\}$

B) Se dispone de una versión “congelada” de la red Q para estimar $maxQ(s_{t+1}, a)$ en el instante t que se actualiza cada C iteraciones

C) Se dispone una versión “congelada” de la red Q para estimar $Q(s_t, a)$ en el instante t que se actualiza cada C iteraciones

D) Se repite el entrenamiento varias veces para las mismas muestras compuestas por tuplas $\{s_t, a_t, r_{t+1}, a_{t+1}\}$

9 $\boxed{A}$ Los algoritmos básicos para el aprendizaje de los pesos de una red recurrente simple son el “forward gradient” (FG) y el “back-propagation through time” (BPTT), indicar que afirmación es correcta:

A) El coste temporal del algoritmo FG es mayor que el BPTT, pero el coste espacial suele ser mejor el que presenta el FG.

B) El coste temporal y espacial del algoritmo FG es siempre menor que el BPTT.

C) El coste temporal del algoritmo FG es menor que el BPTT, pero el coste espacial suele ser mejor el que presenta el BPTT.

D) El coste temporal y espacial del algoritmo FG es siempre mayor que el BPTT.

10 $\boxed{A}$ ¿ Cuál de las siguientes afirmaciones respecto a las técnicas de Budget en Online Learning es correcta ?

A) Limita el conjunto de muestras con pérdida (con error) anteriores que se emplean para clasificar la nueva muestra (x_t, y_t)

B) Se aplica principalmente para problemas linealmente separables donde se quiere limitar el conjunto de muestras con pérdida (con error) anteriores

C) Se aplica principalmente para problemas no online de tipo batch donde se quire limitar el conjunto de muestras con pérdida (con error) anteriores

D) Limita el número de muestras con pérdida (con error) anteriores que se emplean para clasificar todas las anteriores muestras (x_t, y_t), $t = 1 \ldots T$

Examen de Teoría de Redes Neuronales

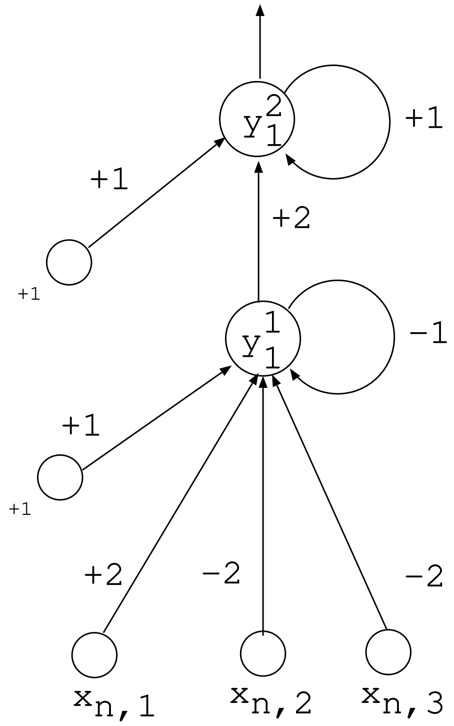
Máster MIARFID, Universitat Politècnica de València, Febrero de 2020

Apellidos: **Nombre:**

Problemas (3 puntos, 60 minutos, con apuntes)

Problema 1, (1.5 puntos)

Las funciones de activación en los nodos de la figura son del tipo sigmoid



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

a) Calcular la secuencia de los valores del nodo oculto Y_1^1 (**0.75 puntos**)

b) Calcular la secuencia de los valores del nodo de salida Y_1^2 (**0.75 puntos**)

a)

$$y^1_{n,1} = f((+1) + (+2) \cdot x_{n,1} + (-2) \cdot x_{n,2} + (-2) \cdot x_{n,3} + (-1) \cdot y^1_{n-1,1})$$

$$y^1_{1,1} = f((+1) + (+2) \cdot (+3) + (-2) \cdot (+1) + (-2) \cdot (-1) + (-1) \cdot 0) = f(+7) = \frac{1}{1+\exp(-7)} = +0.9991$$

$$y^1_{2,1} = f((+1) + (+2) \cdot (+1) + (-2) \cdot (0) + (-2) \cdot (0) + (-1) \cdot (+0.9991)) = f(+2.0009) = \frac{1}{1+\exp(-2.0009)} = +0.8809$$

$$y^1_{3,1} = f((+1) + (+2) \cdot (-1) + (-2) \cdot (+2) + (-2) \cdot (-1) + (-1) \cdot 0.8809) = f(+-3.8809) = \frac{1}{1+\exp(+3.8809)} = +0.0202$$

b)

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

$$y^2_{1,1} = f((+1) + (+2) \cdot (+0.9991) + (+1) \cdot 0) = f(2.9982) = \frac{1}{1+\exp(-2.9982)} = +0.9525$$

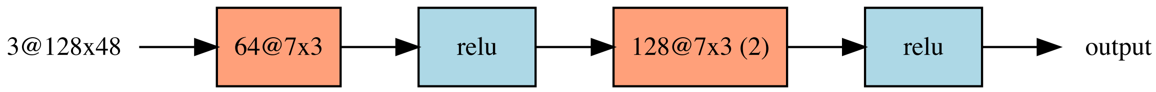
$$y^2_{2,1} = f((+1) + (+2) \cdot (+0.8809) + (+1) \cdot (+0.9525)) = f(3.7143) = \frac{1}{1+\exp(-3.7143)} = +0.9762$$

$$y^2_{3,1} = f((+1) + (+2) \cdot (+0.0201) + (+1) \cdot (+0.9766)) = f(2.0166) = \frac{1}{1+\exp(-2.0166)} = +0.8825$$

Problema 2 (1.5 puntos)

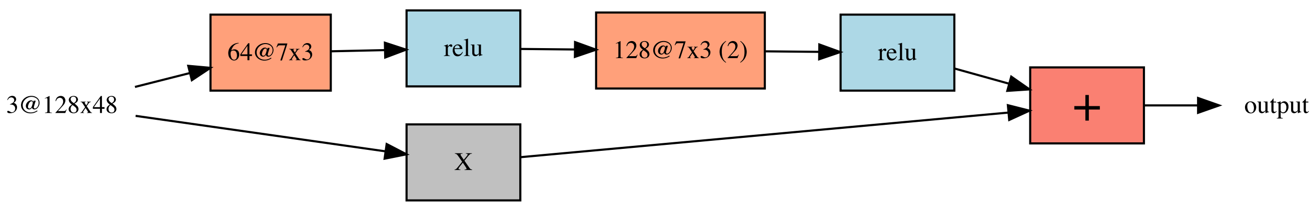
Resolver para cada una de las tres topologías teniendo en cuenta que las convoluciones son **con** padding y que "(2)" significa stride=2.

a) Dado el tensor de entrada calcular el tamaño del tensor de salida **output** (**0.5 puntos**)



output=128@64x24

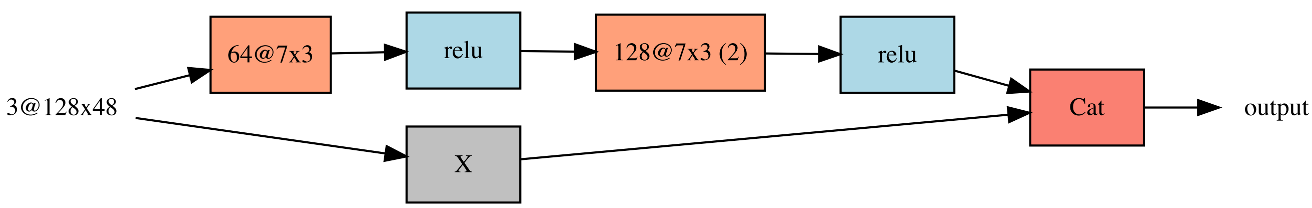
b) ¿ Qué operador sería necesario poner en **X** para tener una topología correcta ? Calcular el tensor de salida **output** que se obtendría. (**0.5 puntos**)



La única restricción es que para sumar dos tensores deben tener las mismas dimensiones. El tensor resultado también tiene las mismas dimensiones, obviamente. En este sentido tenemos que conseguir que X sea una capa que pueda ir del tamaño del input al del output del apartado anterior, o sea 128@64x24. Para ello no nos vale un Max-Pool por ejemplo, dado que no conseguiremos la profundida de 128. Pero por ejemplo X=128@1x1 con padding y stride=2 sería una potencial solución.

Por el hecho de emplear padding ("same") cualquier capa convolucional del tipo X=128@MxN con stride=2 sería solución (M<=128 y N<=48). Y el output seguirá siendo output=128@64x24.

c) ¿ Qué operador sería necesario poner en **X** para tener una topología correcta? Calcular el tensor de salida **output** que se obtendría. (**0.5 puntos**)



La única restricción es que para concatenar dos tensores deben tener las mismas dimensiones espaciales 2D pero pueden tener diferentes profundidades. De hecho el Cat concatena dichas profundidades. El tensor resultado también tiene las mismas dimensiones 2D y la profundidad sería la concatenación de las profundidades de entrada. En este sentido tenemos que conseguir que X sea una capa que pueda ir del tamano del input al del output del apartado anterior pero con cualquier profundidad, o sea Z@64x24.

Para ello podría valer un Max-Pool o cualquier convolución del tipo X=Z@MxN (M<=128 y N<=48) con padding y stride=2. Y el output será output=(128+Z)@64x24.