

**Técnicas de Inteligencia Artificial**  
**Examen evaluación continua - 16 de enero de 2023**

**1.- Aprendizaje Supervisado (1.5 puntos)**

**1.1- Responde a las siguientes preguntas (0.4 puntos):**

**Nota Importante:** Razona tu respuesta, si el razonamiento es correcto se contabilizará el valor de la pregunta pero si la respondes sin razonamiento o con razonamiento incorrecto se te descontará la puntuación que esta pregunta valga.

1.- Teniendo un perceptrón simple, en cada actualización de los pesos, ¿la instancia que ha condicionado esa actualización sería bien etiquetada tras la actualización que acabamos de realizar?.

☐ Siempre ☐ Algunas Veces con perceptrón y siempre con MIRA ☐ Nunca, solo sucederá si empleamos MIRA con el perceptrón no

2.- Dadas la red-1 y la red-2:

¿Hay alguna que represente un aprendizaje multiclase?.

☐ No, ninguna ☐ Una ☐ Ambas

3.- Dadas la red-1 y la red-2:

¿Para qué se emplean las ReLUs?.

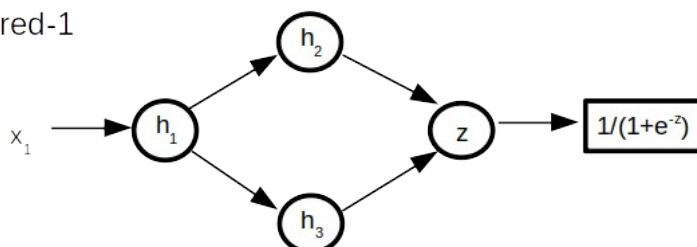
☐ Para introducir una capa más y mejorar la predicción ☐ Para introducir no linealidad  
☐ Para convertir una regresión en clasificación

4.- Dadas la red-1 y la red-2:

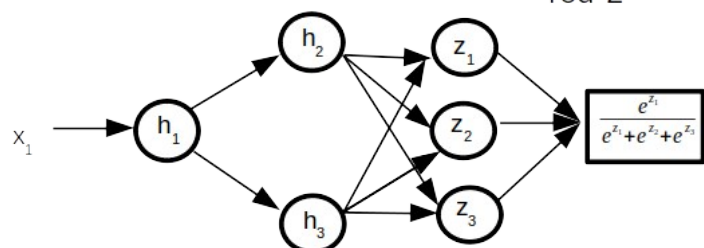
En la red-1 tras el sigmoide, ¿cuál será el punto de corte para determinar si la instancia pertenece o no a cada clase?. Quizás te sirva para razonar recordar que en el perceptrón el valor de corte es 0, intenta hacer un paralelismo.

☐ 1 ☐ 0.5 ☐ 0

red-1



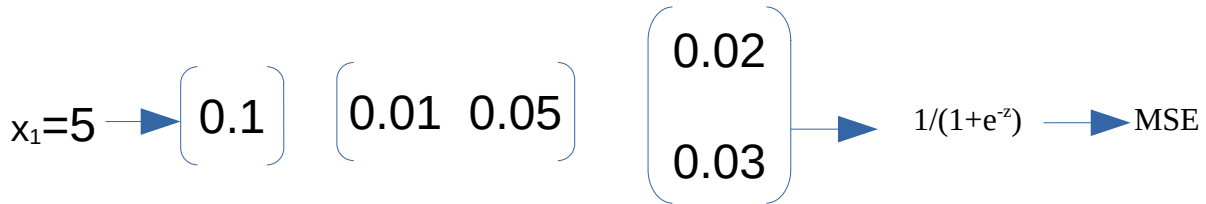
red-2



**Técnicas de Inteligencia Artificial**  
**Examen evaluación continua - 16 de enero de 2023**

**1.2- Soluciona el siguiente problema (1.1 puntos)**

Dada la siguiente red, sabiendo que los valores son los siguientes:



$$y_{real}=1$$

La función de activación  $a$  es en este caso es el sigmoide, es decir,  $\text{sigmoid}(z) = \frac{1}{1+e^{-z}}$  (sabemos que  $\frac{\partial \text{sigmoid}(z)}{\partial z} = \text{sigmoid}(z) * (1 - \text{sigmoid}(z))$ )




La función de error es el Mean Square Error  $MSE: (y_{real} - y_{predicha})^2$  y su derivada  $\frac{\partial MSE}{\partial a} = 2 * (y_{real} - y_{predicha})$

Calcula la actualización que se producirá con el descenso del gradiente para el peso  $w_1^1$  (el peso 0.1 en la representación matricial) . Para ello deberás calcular el error (0.3 puntos), calcular el gradiente asociado a  $w_1^1$  (0.7 puntos) y realizar su actualización con  $lr=0.01$  (0.1 puntos)


**2.- Aprendizaje por Refuerzo (1.5 puntos)**

**2.1-Iteración del Valor (0.75 puntos)**

Suponiendo que disponemos de la siguiente cuadrícula (Grid):

		
1,3	2,3	3,3
1,2	2,2 	3,2

**Técnicas de Inteligencia Artificial**  
**Examen evaluación continua - 16 de enero de 2023**

 <p>1,1</p>	2,1	3,1
--	-----	-----

Encontrar **un solo diamante** supone una recompensa de **+0.06**, encontrar **varios diamantes** de **+0.2**, **los lingotes** de **+0.5**, encontrar **el tesoro** supone una recompensa de **+10** y **la muerte** **-10** con la consecuente finalización del juego.

Sabemos que **los diamantes que están solos** están protegidos por el **espíritu de pirata solitario** que **aparece con una probabilidad de 0.5** y cuando aparece esconde el diamante, así que no obtenemos ninguna recompensa. En los lugares **con muchos diamantes** hay un dragón protegiendo la mitad de ellos. Este suele **estar dormido el 0.75** de las veces, y si el caso **te deja cogerlos sin problema**, sin embargo **si está despierto** debes contar que **tu recompensa no será de 0.2** sino de la mitad porque no podrás hacerte con todos los diamantes. **El tesoro no está protegido** y por lo tanto podrás recibir la recompensa asociada a él en cuanto caigas en el lugar donde se encuentra. Por último, **la dama de la muerte** si caes en su casilla, te hará una pregunta que normalmente **responde correctamente el 0.6 de la gente**. Si la **aciertas el juego acaba, pero no te penalizan con -10**, sin embargo, **si no la aciertas, no solo finalizará el juego, sino que morirás y recibirás la penalización esperada de -10**.

Suponiendo una  $\gamma = 0.1$  y sabiendo que la fórmula de Bellman para calcular el valor de un estado es:

$$\forall s \in S, V_{k+1}(s) \leftarrow \max_a \sum_{s'} T(s, a, s') [R(s, a, s') + \gamma V_k(s')]$$

Calcular el MDP, es decir:

- 1.- las probabilidades de transición y de premio
- 2.- una vez que las hemos identificado y disponemos de ellas, realizar 1 iteración (la 0) y determinar el valor de los estados tras esa iteración. Aplicaremos la versión sincrónica.

## 2.2- Q-learning

En este caso supongamos que en el mismo escenario no disponemos de información sobre el entorno, es decir no conocemos ni la función de transición, ni la de recompensas, y por lo tanto no nos queda más remedio que aplicar aprendizaje-Q (Q-learning), exponiendo al agente al entorno para que éste a través de la experimentación determine cuál es la política que habrá de seguir.

Sabiendo que la fórmula de Bellman para calcular los valores Q es la siguiente y sabiendo que  $\alpha=0.1$  y  $\gamma = 0.5$ .

**Técnicas de Inteligencia Artificial**  
**Examen evaluación continua - 16 de enero de 2023**

$$NewQ(s, a) = \underbrace{Q(s, a)}_{\text{Current Q value}} + \underbrace{\alpha}_{\text{Learning Rate}} [\underbrace{R(s, a)}_{\text{Reward for taking that action at that state}} + \underbrace{\gamma}_{\text{Discount rate}} \underbrace{\max_{a'} Q'(s', a')}_{\text{Maximum expected future reward given the new s' and all possible actions at that new state}} - Q(s, a)]$$

Se dispone de los siguientes valores-Q (Q-values) tras X episodios,

0.00	0.00	0.00
0.00	0.05	0.00
0.00	0.00	0.00
0.01	0.00	0.25
0.00	0.00	0.00
0.94	0.00	-0.9
0.01	-0.9	0.00
0.00	0.01	0.15
0.00	0.00	0.00
0.00	0.00	0.00

Q Valores actuales

**2.2.1 (0.1 puntos)** Generar la Tabla-Q (Q-table) asociada a estos valores transformando el dibujo Q Valores actuales en la Tabla-Q.

**2.2.2 (0.4 puntos)** Suponiendo que el siguiente episodio sea partiendo del {1,2}: ({Arriba, Derecha, Abajo}) actualizar los valores-Q (Q-values) asociados empleando la tabla-Q (Q-table) y escribir las operaciones realizadas a tal efecto.

**2.2.3 (0.1 punto)** Explica cuando se emplea  $\epsilon$ -greedy porque se suelen modificar a la vez la  $\alpha$  y  $\epsilon$ .

**2.2.4 (0.15 punto)** Escribe el algoritmo Q-learning empleando  $\epsilon$ -greedy e identifica en qué paso del algoritmo se emplea el valor  $\epsilon$  y que supondría tener un  $\epsilon$  de 0.8.