

# Técnicas de Inteligencia Artificial – Segunda Parte del Examen Final

## 1.- Clasificación multiclase (5 puntos)

### *1.1 Dibujar la red (1 punto)*

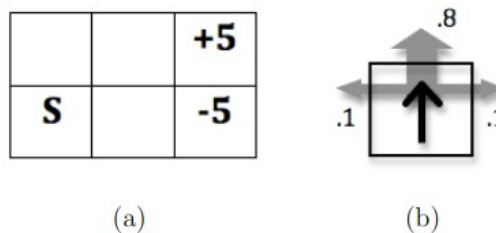
Dados items de entrada que vienen representados por 4 rasgos (features) es decir, que cada item del entrenamiento viene representado por un vector de dimensión  $[1 \times 4]$  dibuja una red neuronal de 2 capas intermedias, donde la primera capa consta de 4 neuronas y la segunda de 3, sabiendo que esta red se empleará para realizar clasificación multiclase consistente en 3 clases. Decide tu la dimensión de la última capa. Inicializa los pesos con valores al azar, pero dibuja dichos valores también en el dibujo, por ejemplo  $w_1=0.2$ ,  $w_2=0.15$ , etc.

### *1.2 Calcular el error y la actualización por descenso del gradiente (4 puntos)*

Una vez hecho el dibujo, calcula la salida y el error empleando cross-entropy sabiendo que la clase real es la 2, es decir el vector  $y_{\text{real}}$  será  $0,1,0$ . Presenta las operaciones (no hace falta que las resuelvas) para actualizar un peso de la segunda capa. Consulta la hoja de fórmulas.

## 2. Value Iteration (2 puntos)

Dada la fórmula de Bellman adaptada al Value Iteration que encontrarás en la hoja de fórmulas, considera el siguiente MDP. Los estados son cuadrados de cuadrícula, identificados por su número de fila y columna (primera fila). El agente siempre comienza en el estado (1,1), marcado con la letra S. Hay dos estados objetivo terminales, (2,3) con recompensa +5 y (1,3) con recompensa -5. Las recompensas son 0 en estados no terminales. La función de transición para las acciones Norte, Sur, Oeste o Este aparece representada en el dibujo, pero si ocurre una colisión con una de las paredes que delimitan las cuadrículas, el agente permanece en el mismo estado.



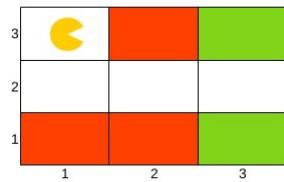
Dibuja gráficamente el MDP correspondiente **(1 punto)**.

Calcula el valor de los estados dadas dos iteraciones empleando el algoritmo de Value Iteration o Iteración por valor, en su versión síncrona, suponiendo que todos los estados se inicialicen a valor 0.0 y que el valor de descuento gamma es de 0.9 **(1 punto)**

## 3. Q-learning (3 puntos)

Dada la cuadrícula que se muestra a continuación y un agente que está tratando de aprender la política óptima, sabemos que las recompensas se consiguen solamente por realizar la acción **salir**

desde uno de los estados coloreados. Al realizar esta acción, el agente pasa al estado **Final (D)** que aunque no aparece en la cuadrícula es un estado al que se puede transitar desde cualquiera de los estados coloreados aunque con distinto premio, al salir desde los estados verdes se obtiene un premio de +30 y desde los estados rojos un premio de -100. En cualquier caso, al salir el episodio terminaría. Supongamos  $\gamma = 1$  y  $\alpha = 0.5$  para todos los cálculos.



Sabiendo que la fórmula de Bellman adaptada para Q-learning es la siguiente:

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

Dados los episodios que aparecen a continuación, completa el momento en el que los siguientes valores de Q se vuelven distintos de cero por primera vez. Tu respuesta debe tener el formato (# de episodio, # de iter) donde iter # es la iteración de actualización de Q-learning en ese episodio. Si el valor Q especificado nunca pasa a ser distinto de cero, escribe nunca.

Q ((1,2), E) = \_\_\_\_\_ Q ((2,2), E) = \_\_\_\_\_ Q ((3,2), S) = \_\_\_\_\_

#### Episodio 1

(1,3), S, (1,2), 0  
(1,2), E, (2,2), 0  
(2,2), E, (3,2), 0  
(3,2), N, (3,3), 0  
(3,3), Exit, D, +50

#### Episodio 2

(1,3), S, (1,2), 0  
(1,2), E, (2,2), 0  
(2,2), S, (2,1), 0  
(2,1), Exit, D, -100

#### Episodio 3

(1,3), S, (1,2), 0  
(1,2), E, (2,2), 0  
(2,2), E, (3,2), 0  
(3,2), S, (3,1), 0  
(3,1), Exit, D, +30

#### Episodio 4

(1,3), S, (1,2), 0  
(1,2), E, (2,2), 0  
(2,2), E, (3,2), 0  
(3,2), N, (3,3), 0  
(3,3), Exit, D, +50

#### Episodio 5

(1,3), S, (1,2), 0  
(1,2), E, (2,2), 0  
(2,2), E, (3,2), 0  
(3,2), S, (3,1), 0  
(3,1), Exit, D, +30