

BLOQUE I

Control intermedio de Inteligencia Artificial (EPS – UAM) 2020/05/05

Calificaciones: 2020/05/20

DECLARACIÓN DE AUTORÍA

Instrucciones:

- Redactar a mano, bien en papel, bien en formato electrónico.
- Fechar y rubricar (firma manuscrita)
- Incluir esta cara como página [B1.0 de **x**] del pdf entregado para este bloque

BLOQUE I

[B1.0 de **3**]

Nº de páginas en bloque (excluyendo esta página):

Fecha y hora de comienzo real del examen (igual o posterior a las 11:00):

Hora de finalización (no la de entrega) real del examen (anterior a las 13:00).

Yo, D. / Da. **<Nombre y apellidos>** con pasaporte/NIE/DNI **<número de identificación>** declaro que he realizado el examen en el periodo indicado, de manera individual, sin colaborar, prestar o recibir ayuda de otras personas.

Y para que así conste, lo rubrico en **<lugar de realización del examen>** con fecha 2020/05/05

Fdo: <Nombre y apellidos>

INSTRUCCIONES:

1. El examen tiene una duración de **dos horas**.
2. Responde de manera clara, completa y concisa detallando toda la información que se solicita de manera ordenada.
3. Las respuestas deben ser justificadas para recibir calificación. Es decir, una respuesta correcta sin explicación será considerada inválida.
4. Escanea las respuestas a las preguntas de este bloque en un solo pdf de nombre

examen_IA_2020_05_05_<apellido1>_<apellido2>_<nombre>_B1.pdf

5. Realiza la entrega en Moodle en los siguientes 30 minutos de la realización del examen.
 6. En caso de incidencias o dudas, por favor ponte en contacto con el foro habilitado a tal efecto en MS Teams o en **alberto.suarez@uam.es**
-

1. [5 puntos en Bloque I] Búsqueda A*

Un supermercado quiere instalar un sistema para reconocer las matrículas de las furgonetas de los proveedores que acceden a su parking. El sistema tiene fallos por el efecto de sombras en la imagen, matrículas que están sucias, etc. Hay caracteres que tiende a confundir más, como el 1 con el 7. Por ejemplo, la matrícula 1234ABC puede confundirla con 7234ABC. A veces también un carácter puede no ser leído. Por ejemplo la matrícula 1234ABC puede confundirla con 234ABC. El supermercado tiene la lista de matrículas de los proveedores, por lo que en el futuro quiere implementar un sistema que tome la matrícula leída y devuelva la matrícula del listado más cercana.

Para ello, vamos a enfocarnos en el primer problema, que es definir la distancia entre dos matrículas:

- La distancia entre una matrícula M y otra Mg se define como el **coste mínimo** de convertir M en Mg
- Para convertir M en Mg se realiza una secuencia de acciones individuales. Cada una de ellas puede ser:
 - **Convertir un carácter en otro.** El coste de la acción depende del carácter convertido:
 - Convertir un “1” en un “7”: **coste 1**
 - Convertir un “7” en un “1”: **coste 1.5**
 - **Añadir “1” a la izquierda de M:** **coste 1.** Se puede realizar sólo si la longitud de M es un carácter
 - **Añadir “7” a la izquierda de M:** **coste 3.** Se puede realizar sólo si la longitud de M es un carácter
 - **Eliminar último carácter de M** (sólo en caso de que la longitud de M es dos caracteres): **coste 1**

Con estas definiciones, la distancia entre “17” y “77” es 1, ya que la forma menos costosa de convertir “17” en “77” es simplemente reemplazar el 1 por 7.

En nuestro primer prototipo consideramos sólo consideraremos los caracteres 1 y 7. Si M es el estado actual de la matrícula, y Mg una matrícula objetivo:

- a) Propón una heurística monótona para este problema (convertir M en Mg con el menor coste total posible).
- b) Considerando que la matrícula leída es “7”, y la matrícula objetivo es “71”, desarrolla el árbol de búsqueda empleando el algoritmo A* **con** eliminación de estados repetidos. Para cada nodo, indicar el orden de expansión y los valores de coste $g+h = f$. El orden de las acciones aplicables a cada estado es el mismo que el orden en el que están definidas más arriba
- c) ¿Cuál es la distancia entre “7” y “71”?
- d) ¿Es A* **con** eliminación de estados repetidos con la heurística propuesta óptima?
- e) ¿Sería A* **sin** eliminación de estados repetidos con la heurística propuesta óptima?

2. [5 puntos en Bloque I] Árboles de decisión

Considera la siguiente base de datos:

ID ejemplo	Edad	X2	X3	Clase
1	20	0	0	Sí
2	30	0	0	No
3	50	0	0	Sí
4	20	0	1	No
5	40	1	1	Sí
6	60	1	1	Sí
7	30	1	2	No
8	20	1	2	No
9	40	1	2	No

Responde razonadamente e incluye todos los cálculos intermedios que has realizado. No se darán por válidas respuestas sin los cálculos intermedios:

1. ¿Cuál es la entropía inicial de la clase?
2. ¿Cuáles son las preguntas candidatas a nodo raíz que va a evaluar C4.5?
3. ¿Cuál es la ganancia de información de cada una de esas preguntas candidatas?
4. ¿Qué pregunta elige finalmente C4.5 como nodo raíz?