[Área personal](https://evaulas.inf.uva.es/my/) / Mis cursos / [MINERÍA DE DATOS](https://evaulas.inf.uva.es/course/view.php?id=94) / [EXAMEN 1º CONVOCATORIA - 16 DE ENERO DE 2023](https://evaulas.inf.uva.es/course/view.php?id=94&section-1)

/ [EXAMEN TEST 1º CONVOCATORIA.](https://evaulas.inf.uva.es/mod/quiz/view.php?id=249)

**Comenzado el** lunes, 16 de enero de 2023, 16:06

**Estado** Finalizado

**Finalizado en** lunes, 16 de enero de 2023, 16:36

Tiempo empleado

30 minutos

**Puntos** 7,20/20,00

**Calificación 3,60** de 10,00 (**36**%)



La respuesta correcta es:

Perfectamente. Se obtendría la función de coste acumulada tras el procesamiento de la ultima secuencia del lote y, a continuación, se aplicaría el optimizador escogido al efecto.

1. Sí, siempre que se normalice las entradas, tras cada procesamiento del lote.
2. Perfectamente. Se obtendría la función de coste acumulada tras el procesamiento de la ultima secuencia del lote y, a continuación, se aplicaría el optimizador escogido al efecto.
3. No tendría sentido, porque el lote ya está implícito en el conjunto de muestras formado por cada secuencia.
4. No, debido a que los optimizadores basado en el gradiente, en general, no admiten procesamiento de lotes.

¿Es posible el procesamiento por lotes en el aprendizaje de una red neuronal recurrente?

Pregunta **1**

Sin contestar

Puntúa como 1,00



La respuesta correcta es:

Porque la definición de la heterogenidad va buscando los clusters más compactos (conexos y convexos) y lo suficientemente distanciados unos de otros.

1. Porque es una mezcla de mínimo (single link) y del máximo (complete link)
2. Porque es claramente el método aglomerativo que más complejidad conlleva en su cálculo y, por tanto, en su cómputo.
3. Porque es el que más se parece al de las k-medias, que tan ampliamente es usado en la práctica
4. Porque la definición de la heterogenidad va buscando los clusters más compactos (conexos y convexos) y lo suficientemente distanciados unos de otros.

¿Por qué el método de Ward es uno de los métodos jerárquicos más utilizados?

Pregunta **2**

Correcta

Se puntúa 1,00 sobre 1,00



La respuesta correcta es:

Discretizando el conjuto de entrenamiento y aplicando esa discretizazión al conjunto de test.

1. Discretizando todo el conjunto de datos.
2. Discretizando por separado el conjunto de entrenamiento y el conjunto de test.
3. Discretizando el conjunto de test y aplicando esa discretización al conjunto de entrenamiento.
4. Discretizando el conjuto de entrenamiento y aplicando esa discretizazión al conjunto de test.

¿Cómo hay que aplicar el método de discretización de atributos basado en la entropia?

Pregunta **3**

Incorrecta

Se puntúa -0,20 sobre 1,00



La respuesta correcta es:

Evitar que la varianza disminuya con el aumento de las repeticiones sobre conjuntos de prueba que no son independientes.

1. Evitar que la varianza aumente por utilizar conjuntos de entrenamiento diferentes que no son independientes.
2. Reducir los problemas de redondeo numérico al realizar muchas repeticiones.
3. Evitar que la varianza disminuya con el aumento de las repeticiones sobre conjuntos de prueba que no son independientes.
4. Evitar que la varianza disminuya con el aumento de las repeticiones sobre conjuntos de prueba independientes.

¿Cuál es le efecto de la corrección que se introduce en el test de student pareado corregido, empleado para la

comparación de dos métodos de aprendizaje sobre el mismo problema?

Pregunta **4**

Incorrecta

Se puntúa -0,20 sobre 1,00



La respuesta correcta es:

se forma un vector con la salida de todas las neuronas y se escala por neuronas, es decir, como si cada neurona del mapa auto-organizado fuera un atributo.

1. se escoge sólo el valor que ofrece la neurona ganadora de cada muestra
2. se escoge la componente del vector de entrada más cercana a la media de dicho vector
3. se fija un radio de vecindad y se recolecta, en un vector, la salidas de la neurona ganadora, junto con las que caen dentro de dicho radio de vecindad
4. se forma un vector con la salida de todas las neuronas y se escala por neuronas, es decir, como si cada neurona del mapa auto-organizado fuera un atributo.

Cuando se usa la salida de un mapa auto-organizado como entrada a un perceptrón multicapa:

Pregunta **5**

Sin contestar

Puntúa como 1,00



La respuesta correcta es:

340

a. 340

b. 4000

c. 40

d. 5040

e. 400

Las instancias de un conjunto de datos están descritas por 10 atributos. Se emplea un método envolvente con selección hacia adelante para seleccionar el mejor conjunto de 4 atributos. Sabiendo que la tasa de error se estima mediante validación cruzada de 10 particiones, indicar cuántos clasificadores hay que crear par realziar la selección

de atributos.

Pregunta **6**

Correcta

Se puntúa 1,00 sobre 1,00



La respuesta correcta es:

No, porque no cumpliría la propriedad reflexiva

1. No, porque no cumpliría la propriedad reflexiva
2. No habría problema, porque sabemos que existe una biyección entre el intervalo (0,1) y el conjunto de los números reales.
3. No, porque no cumpliría la propiedad simétrica.
4. Sí, porque esta dependencia funcional traduce perfectamente la definición de una similaridad mediante al distancia.

Dado que distancia y similaridad representan conceptos antagónicos, se podría definir una similaridad a partir de

una distancia mediante la inversa (similaridad=1/distancia).

Pregunta **7**

Sin contestar

Puntúa como 1,00



La respuesta correcta es:

la jerarquía indexada de cada agrupación y los clusters involucrados para ello.

1. al tratarse de un árbol, con una estructura de lista recurrente como en Prolog, valdría
2. la jerarquía indexada de cada agrupación
3. solamente las muestras que conforman los clusters finales
4. la jerarquía indexada de cada agrupación y los clusters involucrados para ello.

El dendograma de un método de clustering aglomerativo precisa de una estructura de datos que almacene implícita

o explícitamente:

Pregunta **8**

Correcta

Se puntúa 1,00 sobre 1,00



La respuesta correcta es:

con dos mecanismos: decremento del radio de vecindad con el paso de cada época y reducción progresiva del factor de aprendizaje en cada iteración.

1. sólo a través del decremento del factor de aprendizaje conforme crece el número de iteraciones
2. mediante una tabla, en la que se especifica de antemano, lo que va a valer el factor de aprendizaje a lo largo de toda una época.
3. sólo a través de la reducción del radio de vecindad al aumentar las épocas
4. con dos mecanismos: decremento del radio de vecindad con el paso de cada época y reducción progresiva del factor de aprendizaje en cada iteración.

En el algoritmo de aprendizaje de un mapa auto-organizado, ¿cómo se implementan las fases de ajuste de grano

grueso y las de más fino?.

Pregunta **9**

Correcta

Se puntúa 1,00 sobre 1,00



La respuesta correcta es:

se implementa mediante la interacción vecinal, en particular, por la modificación de los pesos de las neuronas topológicamente cercanas a la ganadora.

1. se implementa mediante la interacción vecinal, en particular, por la modificación de los pesos de las neuronas topológicamente cercanas a la ganadora.
2. no existe como tal, porque sólo la neurona ganadora es la que actualiza sus pesos en cada iteración.
3. solamente aparece entre las neuronas de la capa de entrada y la de salida.
4. se materializa a través de los pesos que las unen todas entre sí.

En el mapa auto-organizado, la interacción entre neuronas:

Pregunta **10**

Correcta

Se puntúa 1,00 sobre 1,00



La respuesta correcta es:

es una variable que enumera de manera ordenada en el tiempo las entradas que conforman una secuencia de ellas.

1. en una variable contínua
2. es una variable que enumera de manera ordenada en el tiempo las entradas que conforman una secuencia de ellas.
3. en una variable discreta controlada por la señal de reloj interna del sistema de cómputo.
4. es directamente el número de época en el que se llega el algoritmo correspondiente

En las redes neuronales recurrentes, la aparición de la variable tiempo de manera explícita se traduce:

Pregunta **11**

Incorrecta

Se puntúa -0,20 sobre 1,00

Pregunta **12**



Sin contestar Puntúa como 1,00

En un problema de predicción de series temporales, tratado con redes neuronales recurrentes, frente a uno de clasificación:

1. supondría formular funciones de pérdida / coste diferentes, pero su optimización podría hacerse aplicando los mismos métodos basado en el gradiente
2. las funciones de pérdida / coste son idénticas, tanto en la predicción, como en la clasificación, por lo tanto, los métodos de optimización no variarán
3. las funciones de pérdida / coste no varían, pero sí los métodos de optimización para adaptarse a cada propósito: predicción o clasificación.
4. como las funciones de pérdida / coste ya no son las mismas, su optimización también debería abandonar los métodos de optimización basados en el gradiente

La respuesta correcta es:

supondría formular funciones de pérdida / coste diferentes, pero su optimización podría hacerse aplicando los mismos métodos basado en el gradiente



La respuesta correcta es:

las capas más profundas

1. las capas más profundas
2. al procesamiento llevado a cabo en la capa de salida
3. en las intermedias, ya que todavía no se ha dejado notar el efecto de la evanescencia del gradiente
4. es una tarea que lleva a cabo pro igual en todas las capas de la red.

En una red convolutiva empleada para la clasificación de imágenes, las formas más simples de cada imagen son

atribuibles a:

Pregunta **13**

Sin contestar

Puntúa como 1,00



La respuesta correcta es:

La estimación tiene poco sesgo, pues compensa la ausencia de instancias en el conjunto de entrenamiento con la estimación optimista del error de resubstitución.

1. La estimación tiene poco sesgo, pues ambos errores permiten estimar el error verdadero.
2. Realiza una estimación optimista por incluir el error de resubstitución.
3. Realiza una estimación pesimista, pues el clasificador no se entrena con todas las instancias disponibles.
4. La estimación tiene poco sesgo, pues compensa la ausencia de instancias en el conjunto de entrenamiento con la estimación optimista del error de resubstitución.

El método de estimación de la tasa de error Bootstrap realiza una media ponderada de la tasa de error sobre el

conjunto de prueba y el error de resubtitución. Seleccione la opción correcta:

Pregunta **14**

Sin contestar Puntúa como 1,00



Las respuestas correctas son: Eliminar atributos redundantes., Eliminar atributos irrelevantes.

1. Eliminar los atributos que tienen una distribución uniforme.
2. Eliminar los atributos que tienen menor varianza.
3. Eliminar atributos redundantes.
4. Crear los atributos que se consideran relevantes para identificar la clase.
5. Eliminar atributos irrelevantes.

Los métodos de selección de atributos intentan:

Pregunta **15**

Correcta

Se puntúa 1,00 sobre 1,00



La respuesta correcta es:

Estandarización.

1. Depende de si es un problema de clasificación o de regresión.
2. Lo mejor es no normalizarlos debido a la presencia de outliers.
3. Estandarización.
4. Escalado.

Un conjunto de datos contiene outliers y se ha optado por no eliminarlos. ¿Cuál es la mejor forma de normalizar el

conjunto de datos?

Pregunta **16**

Incorrecta

Se puntúa -0,20 sobre 1,00



La respuesta correcta es:

Porque las tasas de error cometidas sobre distintos conjuntos de datos de distintos dominios no son directamente comparables.

1. Porque al promediar las tasa de error sobre distintos conjuntos de datos hay que asumir la distribución t- Student para estimar los intervalos de confianza del error, y es menos precisa que la distribución normal.
2. Porque obtener los rankings de cada método es más sencillo.
3. Porque utilizar la distribución de Fisher para comprobar la hipotesis nula es preferible a utilizar la t-Student.
4. Porque las tasas de error cometidas sobre distintos conjuntos de datos de distintos dominios no son directamente comparables.

Cuando se quiere comparar el comportamiento de distintos métodos de aprendizaje sobre varios problemas, se

recurre a contar el número de veces que un algoritmo se comporta mejor que los demás, eleborando tablas de victorias o de rankings. ¿Por qué no se usa directamente la tasa de error de los clasificadores inducidos?

Pregunta **17**

Correcta

Se puntúa 1,00 sobre 1,00



La respuesta correcta es:

(0.76, 0.76)

a. (0.4, 0.6)

b. (0.76, 0.76)

c. (0.7, 0.2)

d. (0.3, 0,8)

e. (0.6, 0.6)

En un problema de clasificación binario, un clasificador asigna el 40% de las veces la clase mayoritaria con probabilidad 70% y el 60% de la veces la clase mayoritaria con una probabilidad del 80%. Indicar qué punto le

representa en el espacio ROC.

Pregunta **18**

Sin contestar Puntúa como 1,00



La respuesta correcta es: 1.0

a. 0.9

b. 0.8

c. 0.2

d. 1.0

Se dispone de 5 instancias para crear la curva ROC de un clasificador para la clase +. La información de cada instancia, con el formato <número de instancia, clase real, clase predicha, puntuación>, se incluye en la lista: (<1, +,

+, 0.9>, <2, +, +, 0.7>, < 3, +, -, 0.6>, < 4, +, -, 0.4>, < 5, -, -, 0.2>). Determinar el área bajo la curva ROC del

clasificador:

Pregunta **19**

Correcta

Se puntúa 1,00 sobre 1,00



La respuesta correcta es:

Asumiendo una distribución binomial.

1. No se pueden calcular, porque el conjunto de prueba tiene menos de 30 instancias.
2. Asumiento una distribución normal.
3. Suponiendo una distribución t-student.
4. Asumiendo una distribución binomial.

¿Cómo se calculan los intevalos de confianza de la tasa de error de un clasificador si el conjunto de prueba tiene 20

muestras?

Pregunta **20**

Sin contestar Puntúa como 1,00

[◄ Avisos](https://evaulas.inf.uva.es/mod/forum/view.php?id=93&forceview=1)

Ir a...

[Conjunto de datos weatherAUS-examen ►](https://evaulas.inf.uva.es/mod/resource/view.php?id=246&forceview=1)