

--En el contexto de comparar varios clasificadores sobre varios métodos, el test de Bonferroni-Dunn  
Seleccione una o más de una:

- a. Tiene validez si el test de Iman y Davenport no rechaza la hipótesis nula.
- b. Permite comparar todos los métodos entre sí.
- c. Discrimina más que el test de Nemenyi.
- d. Permite comparar un método frente a los demás.
- e. Tiene validez si el test de Iman y Davenport rechaza la hipótesis nula.
- f. Se realiza después del test de Nemenyi.

Las respuestas correctas son: Discrimina más que el test de Nemenyi., Permite comparar un método frente a los demás., Tiene validez si el test de Iman y Davenport rechaza la hipótesis nula.

--El algoritmo de las k-medias, a pesar de su antigüedad, sigue siendo un de los métodos de clustering más populares, ¿por qué?:

- a. por su robustez al ruido y a los puntos aislados
- b. por su complejidad (lineal) y su eficiencia
- c. por ser capaz de formar clusters no convexos
- d. por su complejidad (polinómica de grado 2)

La respuesta correcta es:

por su complejidad (lineal) y su eficiencia

--¿Por qué se utilizan los rankings o los test de victorias para comparar algoritmos sobre varios conjuntos de datos?

- a. Porque el algoritmo que mejor se comporta siempre obtiene la menor tasa de error.
- b. Porque los errores cometidos por los algoritmos en distintos conjuntos de datos no son directamente comparables.
- c. Porque ningún algoritmo obtiene siempre la menor tasa de error.
- d. Porque los signos y los rankings no dependen del error sobre cada conjunto de datos.

La respuesta correcta es: Porque los errores cometidos por los algoritmos en distintos conjuntos de datos no son directamente comparables.

--En el algoritmo AC de Clustering, comparando el criterio del mínimo, máximo y media, en cuál se calculan más distancias:

a. Igual para todos ellos: hay que calcular la distancia de todas las muestras de un cluster con todas las del otro que se quiere agrupar.

b. en el máximo.

c. usando la media.

d. en el mínimo

La respuesta correcta es: Igual para todos ellos: hay que calcular la distancia de todas las muestras de un cluster con todas las del otro que se quiere agrupar.

--En una red recurrente implementada a base de celdas LSTM, GRU o SimpleNN, ¿tendría sentido emplear la entropía cruzada como una función de coste o de pérdida?

a. No, debido a que siempre lleva a cabo una predicción y salida tiene que ser un valor continuo

b. Sí, si lo que se está llevando a cabo es una clasificación de secuencias

c. No, porque lo que hay a la salida es siempre un escalar

d. Sí, pero sólo admitiría como optimizador el SGD (Descenso del Gradiente Estocástico) en su versión más simple, es decir, sin término momento y sin posibilidad del procesamiento por lotes de secuencias

La respuesta correcta es:

Sí, si lo que se está llevando a cabo es una clasificación de secuencias

--En un mlp, se puede intercambiar el orden de dos atributos, pero ¿en una convolutiva?:

a. Sí, porque la convolución es una transformación isométrica.

b. No, ya que la convolución no es el producto escalar de todo el vector de entrada, sino a trozos, según un núcleo.

c. Sí, porque de lo contrario, la solución dependería del orden de los atributos y esto no es admisible.

d. No, porque el escalado de los atributos ha de hacerse teniendo en cuenta todos los valores dentro de una misma muestra.

La respuesta correcta es:

No, ya que la convolución no es el producto escalar de todo el vector de entrada, sino a trozos, según un núcleo.

--Una imagen almacenada en canales RGB, tras pasar por una capa convolutiva de 16 neuronas, ¿en cuántos canales pasará a la siguiente capa? Nota: considérese, por ejemplo, que el almacenamiento es WHCS (width, high, channel, Sample)

- a.En uno.
- b.En 16.
- c.En 3.
- d.En  $16 \times 3 = 48$ .

La respuesta correcta es:  
En  $16 \times 3 = 48$ .

--Las instancias de un conjunto de datos están descritas por 10 atributos. Se emplea un método de filtro con selección hacia adelante para seleccionar el mejor conjunto de 3 atributos. Indicar cuántos subconjuntos de atributos hay que evaluar.

- a.100
- b.30
- c.720
- d.27

La respuesta correcta es: 27

--¿Qué sucede en el borde, cuando la imagen tiene un número impar de pixel en, al menos, una dimensión, y el Polling es de  $2 \times 2$ ?

- a.Se rellenan de unos hasta completar una cuadrícula  $2 \times 2$ .
- b.Se reduce a un único pixel igual a cero.
- c.Depende de la operación,es decir, si corresponde con el máximo, mínimo o media.
- d.Se lleva acabo la operación, sólo que con menos celdas.

La respuesta correcta es:  
Se lleva acabo la operación, sólo que con menos celdas.

--Se dice que un test de hipótesis es muy exigente cuando:  
Seleccione una o más de una:

- a.El error de tipo II es muy bajo
- b.El error de tipo I es muy alto
- c.El error de tipo II es muy alto
- d.El error de tipo I es muy bajo

Las respuestas correctas son: El error de tipo I es muy bajo, El error de tipo II es muy alto

--Dentro de los diferentes paradigmas de aprendizaje, el Clustering o Análisis de Grupos, en general, dónde lo encuadraría:

- a. Con refuerzo
- b. Competitivo
- c. Supervisado
- d. No supervisado

La respuesta correcta es:  
No supervisado

--Desde el punto de vista tensorial, el procesamiento de imágenes, respecto al vídeo, qué implica:

- a. Nada, se trata como una colección de imágenes estáticas.
- b. Normalmente, en el vídeo se reducen los canales a uno, con lo que el tensor resultante, tendría la misma dimensionalidad que una imagen.
- c. Una dimensión más para las imágenes.
- d. Una dimensión más para el vídeo.

La respuesta correcta es:  
Una dimensión más para el vídeo

--En un dendograma, la altura en la que aparece un nodo no terminal dentro de la jerarquía:

- a. tiene su importancia, porque determina el número de muestras que finalmente va a tener esa rama del árbol.
- b. marca de manera cronológica, cuándo se debe agrupar o dividir esa rama y, por tanto, cuando se corte por un número prefijado de clusters, determinará si el nodo correspondiente se incluirá, o no, en la agrupación final.
- c. es un parámetro que depende de la distribución de probabilidad asociada al espacio muestral
- d. es un dato irrelevante, sirve sólo para organizar el dibujo de tal manera, que no se pisen las diferentes ramas del árbol.

La respuesta correcta es:  
marca de manera cronológica, cuándo se debe agrupar o dividir esa rama y, por tanto, cuando se corte por un número prefijado de clusters, determinará si el nodo correspondiente se incluirá, o no, en la agrupación final.

--¿Es posible que un clasificador tenga una tasa de error de 0.99 y el área bajo la curva ROC sea 0.5, sabiendo que, como es habitual, la clase positiva es minoritaria?

a. Solo si la distribución de clases es muy desequilibrada y el clasificador predice siempre la clase negativa.

b. Solo si la distribución de clases es muy desequilibrada y el clasificador predice siempre la clase positiva.

c. Solo si el clasificador es aleatorio.

d. No.

La respuesta correcta es: Solo si la distribución de clases es muy desequilibrada y el clasificador predice siempre la clase positiva.

--Sea  $a(i, j)$  el elemento de la fila "i", columna "j" de la matriz de confusión de un clasificador binario. Sea la siguiente matriz de confusión:  $a(1, 1)=80$ ,  $a(1, 2)=20$ ,  $a(2, 1)=30$  y  $a(2, 2)=170$ . La tasa de falsos positivos del clasificador es:

a. 0.72

b. 0.27

c. 0.15

d. 0.80

La respuesta correcta es: 0.15

--El número de clusters que surjan de aplicar un algoritmo concreto ¿se fija de antemano?

a. Depende del algoritmo.

b. Sólo se hace en el caso de las k-medias, en el resto de los vistos en clase, se obtiene de manera automática

c. Siempre. A posteriori se comprobará si esta hipótesis es correcta o no.

d. Nunca, son los algoritmos los que conducen a un número óptimo.

La respuesta correcta es:  
Depende del algoritmo

--Indicar que características son ciertas del método de Bagging:  
Seleccione una o más de una:

a. Introduce diversidad usando particiones del conjunto de entrenamiento.

b. El tamaño del ensamble se fija de antemano.

c. Es un método Universal, aplicable a cualquier clasificador base.

d. Construcción del ensamble: entrenamiento independiente.

e. La estrategia de combinación es "Entrenable".

Las respuestas correctas son: Construcción del ensamble: entrenamiento independiente., El tamaño del ensamble se fija de antemano., Es un método Universal, aplicable a cualquier clasificador base.

--Al comparar una GRU y una LSTM, qué es lo que justifica que la primera sea más rápida que la segunda:

- a.La GRU no admite entradas vectoriales, pero sí la LSTM.
- b.La GRU dispone de menos entradas
- c.Las GRU no se pueden distribuir más que en una capa, mientras que las LSTM sí pueden disponerse en varias.
- d.La GRU internamente presenta menos puertas

La respuesta correcta es:

La GRU internamente presenta menos puertas

--La aparición explícita de la variable tiempo en las redes neuronales se debe a:

- a.La necesidad de medir tiempos para obtener las prestaciones del sistema.
- b.El procesamiento de secuencias, lo que conlleva a que en un mismo canal de entrada, aparezcan valores diferentes en distintos instantes de tiempo.
- c.El ahorro de canales de entrada, ya que por uno sólo, se admite multitud de entradas.
- d.Que es una variable básica de la Física.

La respuesta correcta es:

El procesamiento de secuencias, lo que conlleva a que en un mismo canal de entrada, aparezcan valores diferentes en distintos instantes de tiempo.

--De los siguientes métodos de discretización, indicar cuál de ellos es local y basado en la entropía:

- a.Discretización mediante un árbol de decisión sobre el atributo a discretizar y criterio de parada MDL.
- b.Discretización por frecuencias.
- c.Discretización mediante un árbol de decisión.
- d.Discretización por intervalos.

La respuesta correcta es: Discretización mediante un árbol de decisión.

--El uso del test de Student para comparar dos algoritmos sobre un mismo problema de aprendizaje en base a la tasa de error requiere estimar la misma mediante validación cruzada. Indicar por qué.

- a. Porque la estimación de la tasa de error es más precisa.
- b. Porque los conjuntos de prueba no se solapan.
- c. Porque los conjuntos de entrenamiento se solapan.
- d. Porque permite realizar test pareados.

La respuesta correcta es: Porque los conjuntos de prueba no se solapan.

--Indicar que características son ciertas del método de Boosting:

Seleccione una o más de una:

- a. Introduce diversidad usando particiones del conjunto de entrenamiento.
  - b. Construcción del ensamble: entrenamiento independiente.
  - c. El tamaño del ensamble se fija de antemano.
  - d. Es un método Universal, aplicable a cualquier clasificador base.
  - e. La estrategia de combinación es "Entrenable".
- Retroalimentación

Las respuestas correctas son: La estrategia de combinación es "Entrenable"., Es un método Universal, aplicable a cualquier clasificador base.

--¿Qué distingue una red neuronal parcialmente recurrente, de una totalmente recurrente?

- a. En la parcial hay neuronas aisladas.
- b. Las parcialmente recurrentes se emplean para predicción y, las totalmente recurrentes, para clasificación.
- c. No hay diferencia en la arquitectura, sino que su diferencia está en el algoritmo de aprendizaje
- d. En la total, todas las neuronas están conectadas con todas, mientras que en la parcial algunos nodos pueden no estar conectados con todos.

La respuesta correcta es:

En la total, todas las neuronas están conectadas con todas, mientras que en la parcial algunos nodos pueden no estar conectados con todos.

--Indicar qué algoritmos de aprendizaje son sensibles a la normalización de los datos:

Seleccione una o más de una:

- a.K-vecinos más próximos.
- b.SVM.
- c.Naive Bayes.
- d.Regresión logística.
- e.Inducción de reglas de clasificación.
- f.Inducción de árboles de decisión.

Las respuestas correctas son: K-vecinos más próximos., SVM., Regresión logística.

--Rotation Forest:

Seleccione una o más de una:

- a.Introduce diversidad usando particiones del espacio de características.
- b.El tamaño del ensamble se fija de antemano.
- c.La estrategia de combinación es "Entrenable".
- d.Es un método Universal, aplicable a cualquier clasificador base.
- e.Construcción del ensamble: entrenamiento independiente.

Las respuestas correctas son: Construcción del ensamble: entrenamiento independiente., Introduce diversidad usando particiones del espacio de características., El tamaño del ensamble se fija de antemano.

--¿Por qué razón los métodos de clustering divisivos apenas se usan?

- a.No se puede usar con muestras categóricas.
- b.Dan lugar a agrupamientos que raramente satisfacen los requerimientos del problema.
- c.Su lentitud: el número de posibilidades que hay que analizar antes de pasar al siguiente nivel, crece de manera exponencial.
- d.Dado que el número prefijado de clusters obliga a cortar por abajo, resulta poco eficiente empezar por arriba.

La respuesta correcta es:

Su lentitud: el número de posibilidades que hay que analizar antes de pasar al siguiente nivel, crece de manera exponencial.



---Los métodos de selección de atributos intentan:  
Seleccione una o más de una:

- a. Eliminar atributos redundantes.
- b. Eliminar atributos que siguen una distribución uniforme.
- c. Eliminar atributos irrelevantes.
- d. Eliminar atributos con ruido.

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

--Los métodos de selección de atributos de la familia de envoltorio se caracterizan por:

- a. Realizar la selección de atributos con el algoritmo de aprendizaje con el que se crea el modelo final.
- b. Realizar la selección de atributos con un algoritmo de aprendizaje.
- c. Realizar la selección de atributos en base a la tasa de error del modelo final.
- d. Realizar la selección de atributos mediante selección hacia adelante.

La respuesta correcta es: Realizar la selección de atributos con el algoritmo de aprendizaje con el que se crea el modelo final.

--En el algoritmo de k-medias por lotes:

- a. La elección de las muestras se hace de manera aleatoria.
- b. Las n más cercanas al baricentro del experimento.
- c. Se pueden emplear las n-primeras, si éstas están ordenadas por algún atributo.
- d. Se escoge siempre de manera estratificada.

La respuesta correcta es:

La elección de las muestras se hace de manera aleatoria.

--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 de la clase +>, 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:

- a. 0.9
- b. 0.8
- c. 0.6
- d. Ninguna de las anteriores.

La respuesta correcta es: Ninguna de las anteriores.