Tratamiendo de datos Masivos: Cuestionario

|  |  |  |  |
| --- | --- | --- | --- |
| Apellidos: |  | Nombre: |  |

Instrucciones:

Seleccionar la respuesta adecuada, escribiéndola al lado de la pregunta. Solo hay una respuesta válida por pregunta

|  |  |  |
| --- | --- | --- |
| 1) |  | ¿Cuál de las siguientes afirmaciones es correcta? |
|  | a. | MongoDB incluye SQL de forma natural |
|  | b. | MongoDB es orientado a documentos |
|  | c. | Las colecciones tienen un esquema común. |

|  |  |  |
| --- | --- | --- |
| 2) |  | ¿Cuál de las siguientes afirmaciones es cierta para la clasificación basada en el algoritmo de NaïveBayes? |
|  | a. | Solo es aplicable a dimensiones tomando valores en el intervalo [0,1], por eso es un enfoque probabilista |
|  | b. | Las dimensiones deben ser independientes entre sí |
|  | c. | Es especialmente útil para datos con valores perdidos, ya que es capaz de ‘rellenar huecos’ |

|  |  |  |
| --- | --- | --- |
| 3) |  | Tenemos unos datos para los que conocemos sus dimensiones, pero no el valor de sus etiquetas. Entonces: |
|  | a. | No importa, podemos aplicar técnicas de clustering para agrupar valores similares |
|  | b. | Podemos aplicar árboles de decisión, que nos indicarán posibles valores para las etiquetas |
|  | c. | No se puede hacer nada, en machine learning necesitamos las etiquetas para trabajar |

|  |  |  |
| --- | --- | --- |
| 4) |  | En Mongo queremos listar la colección books ordenada por título (title) empezando desde la A. ¿Cuál de las siguientes consultas debemos utilizar? |
|  | a. | db.books.sort( { title : 1 } ).find( { } ) |
|  | b. | db.books.find( { } ).sort( { title : -1 } ) |
|  | c. | db.books.find( { } ).sort( { title : 1 } ) |
|  | d. | db.books.find( { } , { title : "$sort" } ) |

|  |  |  |
| --- | --- | --- |
| 5) |  | ¿Cuál de las siguientes afirmaciones es cierta? |
|  | a. | El aprendizaje supervisado se llama así porque requiere la atención de un experto durante el proceso |
|  | b. | Los árboles de decisión ofrecen una tasa de acierto mayor que los random forest |
|  | c. | La regresión logística es útil para clasificar datasets con etiquetas binarias |

|  |  |  |
| --- | --- | --- |
| 6) |  | En una matriz de confusión… |
|  | a. | Es deseable que la mayor parte de los valores estén en las diagonales |
|  | b. | Debe suceder que el determinante es no nulo. |
|  | c. | Interesa que las diagonales tengan valores lo más próximos a cero posible. |
|  |

|  |  |  |
| --- | --- | --- |
| 7) |  | ¿Qué devolverá la consulta db.users.find( { age: { $gt : 50 }, age : { $lt : 60 } } ); |
|  | a. | Usuarios con edad mayor que 50 y menor que 60 |
|  | b. | Usuarios con edad mayor o igual a 50 y menor o igual a 60 |
|  | c. | Usuarios con edad menor que 60 |
|  | d. | Usuarios con edad menor a 50 |

|  |  |  |
| --- | --- | --- |
| 8) |  | ¿Cuál de las siguientes repuestas es incorrecta con respecto a k-means? |
|  | a. | El valor k debe ser proporcionado por el usuario |
|  | b. | Es independiente del escalado de las dimensiones |
|  | c. | Pueden producirse grupos vacíos |
|  | d. | El algoritmo depende de un valor aleatorio, por lo que conviene repetirlo varias veces |

|  |  |  |
| --- | --- | --- |
| 9) |  | Queremos evaluar una regresión lineal con algún valor que no dependa de las magnitudes de los datos y que permita comparar varias regresiones. ¿Qué valor aconsejarías? |
|  | a. | El RMSE |
|  | b. | El error absoluto |
|  | c. | El coeficiente r |
|  | d. | Ninguno de los anteriores |

|  |  |  |
| --- | --- | --- |
| 10) |  | ¿Cuál de las siguientes afirmaciones es cierta en programa con RDDs? |
|  | a. | Las transformaciones modifican un RDD |
|  | b. | Las transformaciones crean una copia modificada de un RDD |
|  | c. | La acción filter sirve para eliminar elementos de un RDD |
|  | d. | flatMap es similar a Map pero se aplica a RDDs con elementos de tipo lista |