



DECSAI

Departamento de Ciencias de la Computación e I.A.

Universidad de Granada

INTELIGENCIA DE NEGOCIO

7 EXÁMENES 2014-2016



DECSAI

Departamento de Ciencias de la Computación e I.A.

Universidad de Granada

INTELIGENCIA DE NEGOCIO

Convocatoria Ordinaria de Febrero

27 de enero de 2014

1. a) Aunque por abuso del lenguaje hemos hablado de minería de datos y de KDD como sinónimos, indica las diferencias entre ambos términos.
b) Identifica y describe muy brevemente las etapas dentro del proceso de KDD. Haz un gráfico con las etapas.
2. Enumera y describe dos problemas abordados en minería de datos. Pon un ejemplo de aplicación real (que no sea el de prácticas) y menciona un algoritmo clásico para cada uno de ellos.
3. ¿Por qué es interesante realizar selección de atributos antes de construir un clasificador? Enumera brevemente al menos dos justificaciones para su uso.
4. Describe el significado de las medidas de soporte y confianza en reglas de asociación.
5. Suponed un conjunto de datos de clasificación que tiene 4 atributos de entrada, 500 ejemplos y 3 clases. Tres de los atributos de entrada son numéricos en el dominio [1.0, 5.0] y el cuarto es categórico con 4 valores diferentes. ¿Qué preprocesamiento aplicarías a los datos para emplear las técnicas de vecino más cercano?

Todas las preguntas tienen la misma puntuación
Tiempo: 100 minutos



DECSAI

Departamento de Ciencias de la Computación e I.A.

Universidad de Granada

INTELIGENCIA DE NEGOCIO

Convocatoria Extraordinaria de Septiembre

1 de Septiembre de 2014

1. (1 pto.) a) Aunque por abuso del lenguaje hemos hablado de minería de datos y de KDD como sinónimos, indica las diferencias entre ambos términos.

b) Identifica y describe muy brevemente las etapas dentro del proceso de KDD. Haz un gráfico con las etapas.
2. (2 ptos.) Explica brevemente los modelos básicos de multclasificador: Bagging y Boosting. Aspectos a destacar de cada uno de ellos. Sus diferencias como multclasificadores.
3. (2 ptos.) a) ¿Por qué (y en qué situaciones) es interesante realizar selección de variables antes de construir un clasificador?
b) ¿Por qué (y en qué situaciones) es interesante realizar selección de instancias antes de construir un clasificador?
4. (3 ptos.) En las siguientes situaciones, determinar si estamos ante un problema de minería de datos y qué tipo de técnicas se podrían aplicar. Razonarlo (3 puntos).
 - a. Dividir los clientes de una compañía de acuerdo a su género
 - b. Dividir los clientes de una compañía de acuerdo a su fiabilidad
 - c. Calcular las ventas totales de una compañía
 - d. Predecir el precio futuro del almacén de una compañía en base a registros históricos
 - e. Monitorizar el latido de corazón de un paciente para detectar situaciones anormales
 - f. Extraer frecuencias de una señal de sonido
5. (2 ptos.) Suponed un conjunto de datos de clasificación que tiene 4 atributos de entrada, 500 ejemplos y 2 clases. El 15% de los ejemplos están en la segunda clase, y el 85 en la primera. ¿Qué preprocesamiento se puede aplicar a los datos para emplear las técnicas árboles de decisión?

Todas las preguntas tienen la misma puntuación

Tiempo: 120 minutos



DECSAI

Departamento de Ciencias de la Computación e I.A.

Universidad de Granada

INTELIGENCIA DE NEGOCIO

Convocatoria Extraordinaria de Septiembre

4 de Diciembre de 2014

6. (1 pto.) Explica las ventajas e inconvenientes de un selector de características de tipo envolvente y de tipo filtro.
7. (1 pto.) Describe brevemente las diferencias de C4.5 respecto a ID3
8. (1 pto.) Ventajas de Bagging y Boosting frente a otros clasificadores denominados fuertes.
9. (1 pto.) Enumera las ventajas y los inconvenientes que presenta el método de agrupamiento k-medias.
10. (1 pto.) Suponed un conjunto de datos de clasificación que tiene 5 atributos de entrada, 500 ejemplos y 4 clases. Tres de los atributos de entrada son numéricos en el dominio $[1.0, 5.0]$, el cuarto es categórico con 5 valores diferentes y el quinto es binario (2 valores categóricos). Queremos aplicar k-NN y árboles de decisión. ¿Qué preprocesamiento harías para cada una de las dos técnicas de aprendizaje?
11. (1 pto.) ¿Por qué (y en qué situaciones) es interesante realizar selección de instancias antes de construir un clasificador?
12. (1 pto.) Suponed un conjunto de datos de clasificación que tiene 4 atributos de entrada, 500 ejemplos y 2 clases. El 25% de los ejemplos están en la segunda clase, y el 75 en la primera. ¿Cómo abordarías el problema para aplicar árboles de decisión?
13. (1 pto.) Describe la técnica de validación bootstrapping.
14. (1 pto.) ¿Cómo abordarías un problema de clasificación con 6 clases? Explica formas de resolver el problema, y sus ventajas e inconvenientes.
15. (1 pto.) ¿Cómo resolverías un problema de clasificación con 30 millones de instancias y 300 características?

Tiempo: 120 minutos



DECSAI

Departamento de Ciencias de la Computación e I.A.

Universidad de Granada

INTELIGENCIA DE NEGOCIO

Convocatoria Ordinaria de Febrero

27 de Enero de 2015

1. (1 pto.) ¿Por qué (y en qué situaciones) es interesante realizar selección de variables antes de construir un clasificador?
2. (1 pto.) Define un árbol de decisión y describe el algoritmo básico de construcción de un árbol de decisión.
3. (1 pto.) Describe la técnica de validación bootstrap y discute qué aporta en el diseño de clasificadores.
4. (2 ptos) Disponemos de la siguiente base de datos conteniendo 4 transacciones:

TID artículos comprados

t1	K,A,D,B
t2	D,A,C,E,B
t3	C,A,B,E
t4	B,A,D

Suponiendo los umbrales mínimos de soporte y confianza al 50% y 90% respectivamente, se pide obtener todos los conjuntos frecuentes fijado dicho soporte usando el algoritmo A priori y las reglas asociadas al nivel indicado de confianza.

Tiempo: 90 minutos



DECSAI

Departamento de Ciencias de la Computación e I.A.

Universidad de Granada

INTELIGENCIA DE NEGOCIO

Convocatoria Extraordinaria de Septiembre

10 de Septiembre de 2015

1. (1 pto.) Explica las ventajas e inconvenientes de un selector de características de tipo envolvente y de tipo filtro.
2. (1 pto.) Describe brevemente las diferencias de C4.5 respecto a ID3
3. (1 pto.) Ventajas de Bagging y Boosting frente a otros clasificadores denominados fuertes.
4. (1 pto.) Enumera las ventajas y los inconvenientes que presenta el método de agrupamiento k-medias.
5. (1 pto.) Suponed un conjunto de datos de clasificación que tiene 5 atributos de entrada, 500 ejemplos y 4 clases. Tres de los atributos de entrada son numéricos en el dominio $[1.0, 5.0]$, el cuarto es categórico con 5 valores diferentes y el quinto es binario (2 valores categóricos). Queremos aplicar k-NN y árboles de decisión. ¿Qué preprocesamiento harías para cada una de las dos técnicas de aprendizaje?
6. (1 pto.) ¿Por qué (y en qué situaciones) es interesante realizar selección de instancias antes de construir un clasificador?
7. (1 pto.) Suponed un conjunto de datos de clasificación que tiene 4 atributos de entrada, 500 ejemplos y 2 clases. El 25% de los ejemplos están en la segunda clase, y el 75 en la primera. ¿Cómo abordarías el problema para aplicar árboles de de decisión?
8. (1 pto.) Describe la técnica de validación bootstrapping.
9. (1 pto.) ¿Cómo abordarías un problema de clasificación con 6 clases? Explica formas de resolver el problema, y sus ventajas e inconvenientes.
10. (1 pto.) ¿Cómo resolverías un problema de clasificación con 30 millones de instancias y 300 características?

Tiempo: 120 minutos



DECSAI

Departamento de Ciencias de la Computación e I.A.

Universidad de Granada

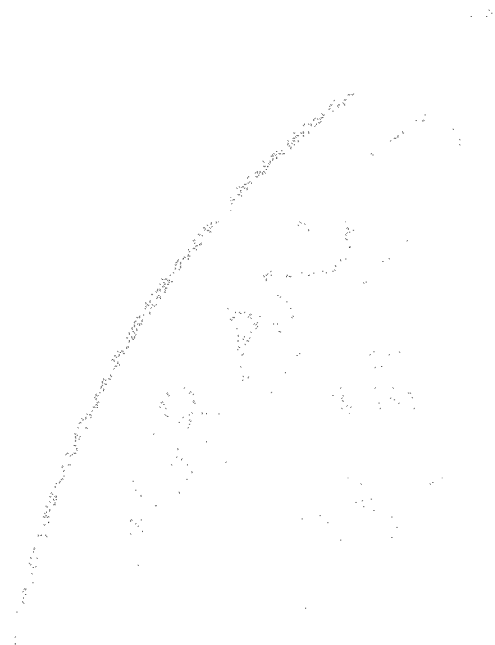
INTELIGENCIA DE NEGOCIO

Convocatoria Ordinaria de Febrero

5 de Febrero de 2016

5. (1 ptos.) ¿Por qué (y en qué situaciones) es interesante utilizar las técnicas de imputación de valores perdidos?
Enumera, describe y justifica el funcionamiento de 2 técnicas de estimación de valores perdidos.
6. (1 ptos.) Explica brevemente los modelos básicos de multclasificador: Bagging y Boosting. Aspectos positivos a destacar de cada uno de ellos. Sus diferencias como multclasificadores.
2. (1 ptos.) Enumera las ventajas y los inconvenientes que presenta el método de agrupamiento k-medias.
7. (2 ptos) Enumera y describe cinco problemas abordados en minería de datos. Pon un ejemplo de aplicación real y menciona un algoritmo clásico para cada uno de ellos.

Tiempo: 100 minutos





DECSAI

Departamento de Ciencias de la Computación e I.A.

Universidad de Granada

INTELIGENCIA DE NEGOCIO

Convocatoria Extraordinaria de Septiembre

15 de Septiembre de 2016

1. (1 pto.) Explica las ventajas e inconvenientes de un selector de características de tipo envolvente y de tipo filtro.
2. (1 pto.) Describe brevemente las diferencias de C4.5 respecto a ID3
3. (1 pto.) Ventajas de Bagging y Boosting frente a otros clasificadores denominados fuertes. Diferencias entre Bagging y Boosting
4. (1 pto.) Enumera las ventajas y los inconvenientes que presenta el método de agrupamiento k-medias.
5. (1 pto.) Suponed un conjunto de datos de clasificación que tiene 5 atributos de entrada, 500 ejemplos y 3 clases (cada clase entre 30 y 40% de tamaño). Tres de los atributos de entrada son numéricos en el dominio $[1.0, 5.0]$, el cuarto es categórico con 5 valores diferentes y el quinto es binario (2 valores categóricos). Queremos un algoritmo de clasificación. ¿Qué algoritmo aplicarías y por qué? ¿Qué preprocesamiento h sería interesante para aplicar dicho algoritmo?
6. (1 pto.) ¿Por qué (y en qué situaciones) es interesante realizar selección de instancias antes de construir un clasificador?
7. (1 pto.) Suponed un conjunto de datos de clasificación que tiene 4 atributos de entrada, 500 ejemplos y 2 clases. El 25% de los ejemplos están en la segunda clase, y el 75 en la primera. ¿Cómo abordarías el problema para aplicar árboles de de decisión?
8. (1 pto.) Describe la técnica de validación bootstrapping.
9. (1 pto.) ¿Cómo abordarías un problema de clasificación con 6 clases? Explica formas de resolver el problema, y sus ventajas e inconvenientes.
10. (1 pto.) ¿Cómo resolverías un problema de clasificación con 5 millones de instancias y 5 millones características?

Tiempo: 120 minutos