

Tercera práctica de Inteligencia Artificial

Curso 2019-2020

Departament d'Informàtica i Enginyeria Industrial
Universitat de Lleida
carlos@diei.udl.cat
jojeda@diei.udl.cat
eduard.torres@udl.cat

1. Enunciado

El objetivo de esta práctica es evaluar el conocimiento del alumno sobre algoritmos de aprendizaje supervisado y no supervisado. Estos algoritmos están descritos en las diapositivas de la asignatura, junto a las tareas que componen esta práctica.

1.1. Árboles de decisión

1.1.1. T9 - Construcción del árbol de forma recursiva. 1 punto

Implementación de la función `buildtree` que depende de una medida de impureza `scoref=(entropy|gini_impurity)` y el criterio β para limitar la construcción.

1.1.2. T10 - Construcción del árbol de forma iterativa. 1.75 puntos

Implementación de la función `buildtree_ite` que depende de una medida de impureza y el criterio β para limitar la construcción. El árbol resultante debe ser igual que el de la construcción recursiva.

1.1.3. T12 - Función de clasificación. 0.75 puntos

Una vez tenemos un árbol construido podemos evaluar nuevos datos contra el árbol. Implementar la función `classify` que devuelve la partición del nodo hoja donde acaba cayendo un nuevo dato.

1.1.4. T13/14 - Evaluación del árbol. 1.5 puntos

Implementar la evaluación del árbol (función `test_performance`) y aplicarla a un nuevo dataset¹ con diferentes ratios training-test. Demostrar el incremento en la calidad de clasificación (accuracy), incrementando el training set.

1.1.5. T15 - Missing data. 0.5 puntos

Sugerir alternativas a como tratar ejemplos en los datasets a los que les falta uno o más datos (valores para sus atributos).

1.1.6. T16 - Poda del árbol. 1 punto

Implementar la poda del árbol siguiendo la estrategia bottom-up descrita en las slides.

¹Tenéis nuevos datasets disponibles en <http://archive.ics.uci.edu/ml/index.php>

1.2. Clustering

1.2.1. T11 - Restarting policies. 1.5 puntos

Queremos que k-means asigne los k clusters tal que se minimizen la suma de las distancias al cuadrado de todos los ítems a sus respectivos centroides. Dado que el resultado de k-means depende de la inicialización de los k centroides, deberíamos ejecutarlo varias veces para conseguir la mejor asignación de clusters, aquella que minimiza esa suma de distancias.

Mejorad la función de `kcluster` para que tenga en cuenta ese número de veces de ejecución (parametrizable) y devuelva la mejor configuración.

2. Implementación - 1 punto

Se valorará la calidad y eficiencia de los algoritmos implementados. Es necesario tener en cuenta los siguientes aspectos de la implementación:

- El lenguaje de programación es Python.
- La utilización de un diseño orientado a objetos.
- La simplicidad y legibilidad del código.
- Se recomienda seguir la guía de estilo ².

3. Documentación - 1 punto

Documento en pdf (máximo ≈ 4 páginas) que incluya una descripción de las decisiones que se han tomado en la implementación, así como los posibles resultados experimentales y comentarios teóricos.

La página inicial ha de contener el nombre de los integrantes del grupo. La práctica puede realizarse en grupos de dos personas o individualmente.

Se valorará la redacción y presentación del documento.

4. Material a entregar

El material evaluable de esta práctica es:

- Todos los archivos de código fuente, modificados o añadidos.
- Documento de la práctica.

Todo el material requerido se entregará en el paquete de nombre `ia-prac3.[tgz|tar.gz|zip]`

²<https://www.python.org/dev/peps/pep-0008/>