

Universitat Oberta de Catalunya

Estudis d'Informàtica i Multimèdia

ASSIGNATURA: Intel·ligència Artificial II

Pràctica: **Pràctica**

Data de publicació: 10/12/2007 **Data de lliurament:** ≤ 08/01/2008

- Observacions:**
- Envieu la solució en un sol fitxer amb format .zip o .rar amb l'infome i el programa a realitzar
 - Dipositeu l'exercici a la bústia de Lliurament de Pràctiques, SENSE còpia a la bústia del consultor
 - La prova és individual. No s'acceptaran proves similars, copiades o realitzades en grups

Cognoms, nom:

Enunciat

Disposem d'un conjunt de dades de 8124 bolets¹, disponible en l'arxiu Excel adjunt. Ens donen, per a cada bolet, 22 atributs i la classe. En l'arxiu de text adjunt teniu una descripció dels mateixos. Tots els atributs són nominals. Molts dels exemples contenen valors absents o nuls codificats amb el símbol "?". Per tal de no realitzar cap tractament en les dades, utilitzarem la **distància de Hamming**. La distància de Hamming entre dos exemples serà el número de valors diferents dels seus atributs².

Volem comprovar si, a partir d'aquestes dades, es pot construir un model que sàpiga predir la classe de l'exemple (edible o poisonous). Per a construir el model predictiu s'utilitzarà una variant del **vector quantization**. Es vol que el sistema afegixi **penalització per les classes no guanyadores**³.

Es demana que realitzeu el programa indicat i un informe que ha de contenir:

1. Una explicació de com heu generat els conjunt de training i test.
2. Una explicació de com heu tractat la funció del factor d'aprenentatge.
3. Una petita explicació general de l'algorisme que implementeu, explicant tots aquells detalls que considereu rellevants i les decisions de disseny preses.
4. Una explicació de com heu tractat els valors nuls.
5. Una taula amb almenys la precisió i el temps de càlcul de l'algorisme.
6. Un apèndix amb el llistat del codi font del vostre programa.
7. I, en general, una justificació de tot el que estigueu fent.

Heu de lliurar el programa que hagueu implementat per realitzar la pràctica (es recomana que sigui en Java, però la podeu realitzar en Pascal, Delphi, Python, C, C++, Visual Basic o "similar"⁴). La qualitat del codi (estructura, comentaris...) és un dels criteris importants de la correcció de la pràctica.

¹ S'ha extret del problema "mushroom" del repositori del UCI: <http://www.ics.uci.edu/~mllearn/MLRepository.html>

² Podeu consultar la wikipedia per a més informació: http://ca.wikipedia.org/wiki/Dist%C3%A0ncia_de_Hamming

³ Aquesta variant la teniu explicada a la plana 48 de la documentació d'aprenentatge de l'assignatura.

⁴ Consulteu si no és un d'aquests llenguatges. Els consultors no respondran a preguntes sobre detalls de programació.