

## PRÀCTICA

### Presentació

Pràctica sobre el desenvolupament i aplicació d'algoritmes d'aprenentatge automàtic.

### Competències

#### Competències de grau

- Capacitat per utilitzar els fonaments matemàtics, estadístics i físics i comprendre els sistemes TIC.
- Capacitat per analitzar un problema en el nivell d'abstracció adequat a cada situació i aplicar les habilitats i coneixements adquirits per abordar-lo i resoldre'l.
- Capacitat per conèixer les tecnologies de comunicacions actuals i emergents i saber-les aplicar, convenientment, per dissenyar i desenvolupar solucions basades en sistemes i tecnologies de la informació
- Capacitat per proposar i avaluar diferents alternatives tecnològiques i resoldre un problema concret

#### Competències específiques

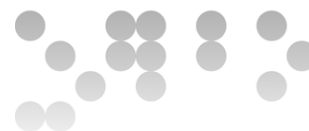
- Capacitat per utilitzar la tecnologia d'aprenentatge automàtic més adequada per a un determinat problema.
- Capacitat per avaluar el rendiment dels diferents algorismes de resolució de problemes mitjançant tècniques de validació creuada.

### Objectius

Seguint amb les dades utilitzades en les altres PACs, disposem d'un conjunt de dades de persones extretes del cens (arxius "adult.data.csv" i "adult.names.txt"). Agafarem com a classe l'última columna. Es vol desenvolupar un categoritzador a partir de les diferents variables (veure la descripció en el fitxer "adult.names.txt") per veure si podem agrupar les persones utilitzant únicament els atributs nominals sense binaritzar-los tot obviant els atributs numèrics.

Volem construir un model capaç de categoritzar els exemples utilitzant l'algorisme de categorització *Partition Around Medoids* (PAM). L'objectiu és implementar l'algorisme i avaluar els resultats amb les dades esmentades.

La pràctica està formada per 3 exercicis amb diferent pes. La memòria de la pràctica ha de contenir les respostes a cadascun dels exercicis.



## Descripció de la PRÀCTICA

### Exercici 1. PAM (30%)

El Partition Around Medoids (PAM) és un algorisme del tipus k-medoids. Un algorisme de clustering en el que s'utilitza el *medoid* en comptes del promig per generar els *clústers*. Això ens permet fer categorització amb atributs nominals directament.

Cerqueu informació sobre aquest algorisme i descriuiu-lo en detall, posant l'èmfasi en com es podria implementar.

Cerqueu informació per mirar com podríem avaluar els resultats d'aquest tipus de categorització.

### Exercici 2. Preparació de les dades (10%)

Seleccioneu aleatòriament 100 exemples de cada classe del conjunt de dades de l'arxiu `adult.data.csv` tot eliminant els atributs numèrics.

### Exercici 3. Implementació del PAM (60%)

Implementeu l'algorisme del PAM i l'avaluació dels resultats tal i com ho heu descrit a l'exercici 1. Apliqueu la implementació per categoritzar l'arxiu generat a l'exercici 2.

Cal que tingueu en compte:

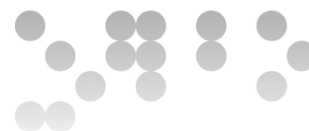
1. Cal explicar l'algorisme implementat, explicant tots aquells detalls que considereu rellevants i les decisions de disseny preses. Feu especial esment en els passos de la implementació de l'algorisme.
2. Una taula amb almenys la precisió, la matriu de confusió i el temps de càlcul de l'algorisme.
4. Un apèndix amb el llistat del codi font del vostre programa.
5. I, en general, una justificació de tot el que estigueu fent.

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

## Recursos

### Bàsics

Per a realitzar aquesta PRÀCTICA disposeu d'uns fitxers adjunts ("adult.names.txt", "adult.data.csv") on trobareu les dades corresponents.



## Criteris de valoració

Els tres exercicis d'aquesta PRÀCTICA es valoraran amb 3, 1 i 6 punts respectivament.

**Exercici 1:** aquest exercici val 2 punts on es valorarà significativament la claredat, simplicitat i estructura del discurs; així com l'enfoc de cara a la implementació.

**Exercici 2:** es valorarà 1 punt. S'ha d'incloure la descripció general del procés.

**Exercici 3:** la valoració d'aquest exercici va per fases:

- El funcionament del programa té una valoració binària de 0 o 4 punts en funció de si funciona o no.
- La qualitat del codi i la presentació es porten un punt. Com a qualitat entenen: l'estructura, comentaris, disseny...; i com a presentació: la sortida del programa, l'ús d'entorns de finestres, l'entrada de dades, la possible parametrització del programa...
- El punt restant valorarà: les conclusions generals, l'anàlisi de resultats...

## Format i data de lliurament

Cal lliurar la PRÀCTICA en un únic fitxer comprimit (.zip) que contingui la memòria i els codi font dels diferents exercicis al registre d'activitats d'avaluació continuada.

El nom del fitxer ha de ser CognomsNom\_AC\_Practica amb l'extensió .zip (ZIP).

**Data Límit: 29/05/2017 a les 24 hores.**

Per a dubtes i aclariments sobre l'enunciat, adreceu-vos al consultor responsable de la vostra aula.

### Nota: Propietat intel·lectual

Sovint és inevitable, en produir una obra multimèdia, fer ús de recursos creats per terceres persones. És per tant comprensible fer-ho en el marc d'una pràctica dels estudis d'Enginyeria Informàtica, sempre i això es documenti clarament i no suposi plagi en la pràctica.

Per tant, en presentar una pràctica que faci ús de recursos aliens, s'ha de presentar juntament amb ella un document en què es detallin tots ells, especificant el nom de cada recurs, el seu autor, el lloc on es va obtenir i el seu estatus legal: si l'obra està protegida pel copyright o s'acull a alguna altra llicència d'ús (Creative Commons, llicència GNU, GPL ...). L'estudiant haurà d'assegurar-se que la llicència que sigui no impedeix específicament seu ús en el marc de la pràctica. En cas de no trobar la informació corresponent haurà d'assumir que l'obra està protegida pel copyright.

Hauran, a més, adjuntar els fitxers originals quan les obres utilitzades siguin digitals, i el seu codi font si correspon.