



## 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 referent a l'origen de distints vins a partir de la seva composició i color (arxius "wine3c\_pract.csv" i "wine.names.txt"). Els arxius de dades "csv" tenen un format tipus taula, on cada fila correspon a un exemple. L'última columna és la classe (noteu que ara tenim 3 classes) i la resta corresponen als atributs de l'exemple. L'arxiu adjunt "txt" conté la descripció d'aquests atributs. Es vol desenvolupar un classificador a partir dels diferents atributs per veure si podem predir l'origen del vi segons suggereix la descripció del problema.

Volem construir un model capaç de predir la classe de l'exemple utilitzant una variant de l'algorisme de classificació k veïns més propers ( $kNN$ ). L'objectiu és implementar l'algoritme 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. Random kNN (RkNN) (20%)

El kNN és un algorisme basat en exemples molt utilitzat. Moltes vegades no s'utilitza en la seva forma bàsica, existeixen múltiples variants. Una proposta de modificació recent consisteix en utilitzar varis kNN aplicats en subconjunts aleatoris d'atributs, semblant a la idea del classificador Random Forest (RF) però en comptes d'utilitzar arbres de decisió utilitzar varis classificadors kNN. El resultat del classificador RkNN en una dada de test serà el vot majoritari de tots els kNN utilitzats. La implementació RkNN també pot servir per seleccionar els atributs més significatius (vindran del subconjunt de kNN que tinguin millor precisió de classificació).

Cerqueu informació sobre aquesta modificació i descriviu-la en detall, posant l'èmfasi en com es podria implementar tant la classificació utilitzant RkNN, com es pot implementar la selecció de característiques (ajuda: podeu cercar el següent article "Li S, Harner EJ, Adjeroh DA. Random KNN feature selection - a fast and stable alternative to Random Forests. BMC Bioinformatics. 2011;12:450. doi:10.1186/1471-2105-12-450").

### Exercici 2. kNN (20%)

Dividiu les dades amb conjunts de train i test i utilitzeu l'algorisme kNN original per classificar les dades de test. Doneu els resultats mitjançant precisió i matriu de confusió. Podeu utilitzar la implementació kNN de les PACs anteriors en Python (i scikit-learn, vegeu KNeighborsClassifier).

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

Implementeu el kNN amb la modificació de l'exercici 1. Apliqueu la implementació per classificar els mateixos casos de test obtinguts en l'exercici 2. .

Heu de proporcionar:

1. Una explicació de 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, comparant-lo amb el kNN original (exercici 2).
3. Utilitzant la selecció de característiques, obteniu un llistat ordenat dels atributs més rellevants.
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 amb les llibreries scikit-learn. Contacteu amb el consultor per utilitzar altres llenguatges. 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 ("wine3c\_pract.csv", i "wine.names.txt") on trobareu les dades corresponents.

## Criteris de valoració

Els tres exercicis d'aquesta PRÀCTICA es valoraran amb 2, 2 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 la inclusió de la taula de resultats amb 1,5 punts. Els resultats han de contenir com a mínim la precisió (*accuracy*), el nombre d'exemples erronis i la matriu de confusió. Els 0,5 punts restants s'adjudiquen als comentaris, valoracions i justificacions de l'exercici.

**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, les comparacions amb l'exercici 2...

## Format i data de lliurament

Cal lliurar la PRÀCTICA en un únic fitxer comprimit (ex.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: 28/05/2018 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.