



## PAC3: Classificació

### Presentació

En aquesta prova d'avaluació estudiarem com fer un classificador per a detectar la fallida de projectes empresarials.

### Competències

En aquest enunciat es treballaran en un determinat grau les competències generals de màster següents:

- Capacitat per a projectar, calcular i dissenyar productes, processos i instal·lacions en tots els àmbits de l'enginyeria en informàtica
- Capacitat per al modelat matemàtic, càlcul i simulació en centres tecnològics i d'enginyeria d'empresa, particularment en tasques de recerca, desenvolupament i innovació en tots els àmbits relacionats amb l'enginyeria en informàtica
- Capacitat per a l'aplicació dels coneixements adquirits i per solucionar problemes en entorns nous o poc coneguts dins de contextos més amplis i multidisciplinaris, sent capaços d'integrar aquest coneixements
- Posseir habilitats per a l'aprenentatge continuat, autodirigit i autònom
- Capacitat per modelar, dissenyar, definir l'arquitectura, implementar, gestionar, operar, administrar i mantenir aplicacions, xarxes, sistemes, serveis i continguts informàtics
- Capacitat per assegurar, gestionar, auditar i certificar la qualitat dels desenvolupaments, processos, sistemes, serveis, aplicacions i productes informàtics

Les competències específiques d'aquesta assignatura que es treballaran són:

- Entendre que és l'aprenentatge automàtic en el context de la Intel·ligència Artificial
- Distingir entre els diferents tipus i mètodes d'aprenentatge
- Aplicar les tècniques estudiades a un cas concret



## Objectius

L'objectiu d'aquesta prova d'avaluació és l'estudi per a la implementació en python classificador de clients d'una distribuïdora d'aliments.

## Descripció de la PAC

L'objectiu d'aquesta prova d'avaluació és classificar les dades dels arxius adjunts relacionats amb dades de clients d'una distribuïdora d'aliments. Volem veure si es pot detectar el tipus de client (al detall o a l'engròs) i la seva zona de procedència (Lisboa, Oporto o d'altres).

L'arxiu de dades "Wholesale customers.csv" té un format tipus taula, on cada fila correspon a un dels 440 exemples. La primera columna correspon al tipus, la segona a la zona i la resta corresponen als atributs de l'exemple. Així, les dues primers columnes corresponen a la classe. L'arxiu adjunt "Wholesale customers.txt" conté la descripció d'aquests atributs.

Aquests arxius pertanyen al problema "Wholesale customers" del repositori d'aprenentatge de l'UCI:

<http://archive.ics.uci.edu/ml/>

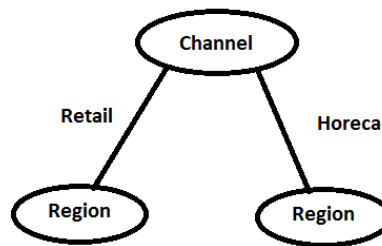
### Exercici 1

L'objectiu d'aquest exercici és la implementació d'un programa python que apliqui el *Naïve Bayes*, els *arbres de decisió*, el *kNN* i les *màquines de vectors de suport* (amb diferents kernels) a l'arxiu de dades "Wholesale customers.csv"; utilitzant la validació simple com a protocol de validació i l'*error* com a mesura d'avaluació.

La definició de classes serà la combinació de les dues primeres columnes de l'arxiu. És a dir, tindrem un problema amb sis classes que surten de totes les combinacions possibles de les dues classes: {Lisbon, Horeca}, {Lisbon, Retail}, {Oporto, Horeca}...

### Exercici 2

Una aproximació diferent a la de l'exercici anterior és fer servir una classificació en dues fases. És a dir, primer muntem un classificació per a distingir una de les classe i després un classificador per cadascun dels valors d'aquesta classe en un segon nivell (sabent que ja són de la classe); tal i com mostra la figura següent:



O a la inversa, tenint un classificador per Region en el primer nivell i tres per Channel en el segon.

Repetiu l'exercici anterior seguint ara aquest model de tractament jeràrquic de les classes.

### Exercici 3

Doneu una taula amb tots els resultats dels exercicis anteriors. Realitzeu una valoració global comparant els mètodes i redacteu unes conclusions globals sobre l'aplicació dels mètodes a aquest conjunt de dades. Els criteris de correcció de la PAC invaliden una A si tots els processos no estan ben justificats i comentats.

## Recursos

Aquesta pràctica requereix els recursos següents:

### Bàsics:

Per a realitzar aquesta PAC disposeu d'uns fitxers adjunts:

- *Wholesale customers.csv*
- *Wholesale customers.txt*
- Vídeo Classification + Naïve Bayes
- Vídeo Support Vector Machines i sklearn
- Codi Arbres de decisió
- Codi kNN

**Complementaris:** manual de teoria de l'assignatura

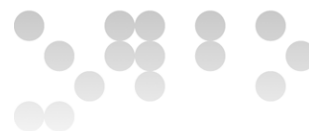
## Criteris de valoració

Els exercicis tindran la valoració següent associada:

Exercici 1: 4 punts

Exercici 2: 4 punts

Exercici 3: 2 punts



**S'han de raonar les respostes de tots els exercicis. Les respostes sense justificació no rebran puntuació.**

## Format i data de lliurament

La pràctica s'ha de lliurar abans del **proper 14 de maig** (abans de les 24h).

La solució ha de consistir en un arxiu zip que contingui un informe en format pdf i els arxius en format python (\*.py) que corresponguin a la solució adoptada.

Adjunteu l'arxiu a un missatge en el apartat de **Lliurament i Registre de AC (RAC)**. El nom de l'arxiu ha de ser CognomsNom\_IAA\_PAC3 amb extensió zip.

Per a dubtes i aclaracions sobre l'enunciat, dirigiu-vos al consultor responsable de l'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 del Grau Multimèdia, 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.

Un altre punt a considerar és que qualsevol pràctica que faci ús de recursos protegits pel copyright no podrà en cap cas publicar-se en Mosaic, la revista del Graduat en Multimèdia a la UOC, a no ser que els propietaris dels drets intel·lectuals donin la seva autorització explícita.