

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}...

Per a realitzar l'exercici he fet servir el codi proporcionat per els materials de l'assignatura, per a cadascun dels mètodes mencionats a l'enunciat. Ha fet falta modificar el codi per tal de fer-lo més llegible i modular, refactoritzant la manera en que es llegeixen les dades del fitxer i els paràmetres que fan servir alguns mètodes. Un cop llegit el fitxer, es fa el càlcul corresponent per a assignar la classe que toca a cada entrada, tenint en compte la combinació entre les dues primeres columnes i es modifica la estructura de les dades de manera que ens és més fàcil treballar amb elles: després de calcular la classe esborrem les dues primeres columnes i afegim al final de cada entrada la nova classe.

La implementació de la execució la trobem en el fitxer **activitat1.py** i la implementació de cadascun dels mètodes la trobem en els fitxers **utils.py**, **knn.py**, **naive_bayes.py**, **decisionTree.py** i **vectorMachine.py**.

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)

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.

He tingut diversos entrebancs a l'hora de fer aquesta pac, tant per la dificultat de la mateixa, per problemes amb el fòrum i complicacions a la feina que m'han impedit realitzar aquest exercici.

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.

A continuació presento una taula amb els resultats obtinguts de la execució de la implementació de cadascun dels mètodes.

	Primer exercici	Segon exercici
Knn	0.6802721088435374 %	-
Naïve Bayes	0.0 %	-
Decision Tree	2.0408163265306123 %	-
Vector machine	-	-

He de dir que no estic gaire content amb aquests resultats, no m'aporten confiança, ja que no són els que jo m'esperava. És possible que la refactorització del codi no hagi sigut completament correcta i s'hagi perdut exactitud en la execució dels càlculs.