

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

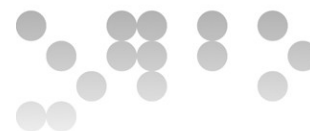
- Capacitat per utilitzar la tecnologia d'aprenentatge automàtic més adequada per a un determinat problema.
- Capacitat per avaluar el rendiment dels diferents algorismes de resolució de problemes mitjançant tècniques de validació creuada.

Objectius

Seguint amb les dades utilitzades en les altres PACs, disposem d'un conjunt de dades de clients de targetes de crèdit (arxius "credit_card.csv" i "credit_card.txt"). Agafarem com a classe l'última columna (*default*). Es vol desenvolupar un classificador a partir de les diferents variables (veure la descripció en el fitxer "credit_card.txt") per veure si podem predir el tipus i donar la seva probabilitat. Noteu que hem d'ignorar la primera columna.

Volem construir un model capaç de predir la classe de l'exemple utilitzant l'algorisme de classificació basat en exemples *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. kNN i valor de confiança (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 de les modificacions interessants consisteix en fer que la votació dels diferents exemples més propers estigui ponderada per la distància a l'exemple de test. És a dir, que els exemples més propers tinguin més pes que els que estan més lluny.

Cerqueu informació sobre aquesta modificació i descriu-la en detall, posant l'èmfasi en com es podria implementar. Descriu com tractar els valors nuls o infinits en el cas de que puguin aparèixer.

Si tenim la modificació comentada anteriorment, podem generar un valor de confiança (que no de probabilitat) tenint en compte el tant per cent de vot que va a parar a «default» de entre tots els vots ponderats.

Descriu com modifiquem l'algorisme anterior per a que doni aquest valor de confiança.

Exercici 2. Weka (20%)

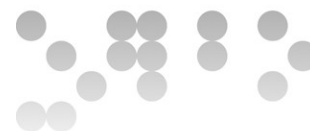
Partiu el conjunt de dades en dos: un per a l'entrenament i un per al test. Expliciteu a l'informe com heu realitzat aquesta divisió. Utilitzant el Weka apliqueu el kNN descrit en l'apartat anterior (el Weka té implementada la modificació esmentada) per predir la classe de l'arxiu de test. Discussiu els resultats.

Exercici 3. Implementació del kNN ponderat i de la confiança (60%)

Selecioneu aleatòriament 5 exemples de cada classe del conjunt de test generat en l'apartat anterior. Implementeu el kNN amb la modificació de l'exercici 1 i la generació de la confiança. Apliqueu la implementació per classificar tot l'arxiu de test a partir del d'entrenament. Doneu el valor de la confiança per als 10 exemples anteriors.

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, comparant els resultats amb els obtinguts al Weka (exercici 2).
3. Una taula amb els valors de confiança i la classe dels 10 exemples seleccionats a l'inici de l'exercici.
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++, 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 (“credit_card.txt”, “credit_card.csv”) 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 les matrius de confusió de tots els mètodes especificats a l'enunciat. 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 el Weka...

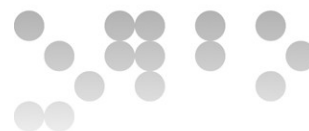
Format i data de lliurament

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: 27/05/2016 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.