

KONPUTAZIO-ZIENTZIEN METODO MATEMATIKOAK
2010.eko iraila

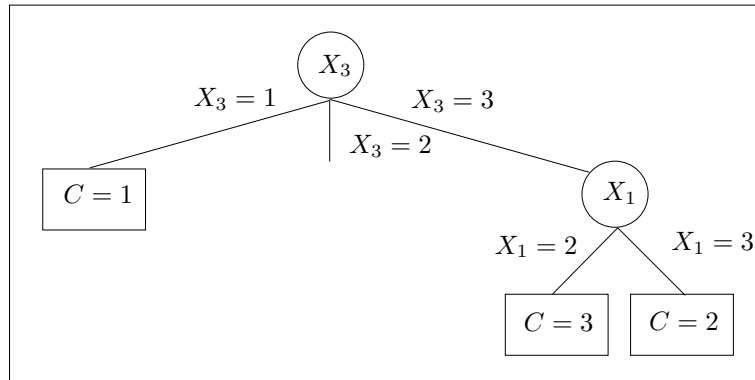
(Azterketa honetan gehienez 5 puntu lortuko dira. Praktikan ere gehienez 5. Biak gainditu behar dira irakasgaia bere osotasunean gainditzeko.)

1. (Puntu 1) Ondorengo taulan entrenamendurako 10 kasuz osatutako datu-basea agertzen da.

Kasua	X	Y	Klasea
$\mathbf{x}_1$	2	2	1
$\mathbf{x}_2$	2	4	1
$\mathbf{x}_3$	4	2	2
$\mathbf{x}_4$	2	6	1
$\mathbf{x}_5$	2	8	2
$\mathbf{x}_6$	2	10	1
$\mathbf{x}_7$	6	2	2
$\mathbf{x}_8$	8	2	2
$\mathbf{x}_9$	10	2	1
$\mathbf{x}_{10}$	6	4	1

- 1.1 Distantzia Euklidestarrak kalkulatu eta eman ezazu $\mathbf{x}_{11} = (7, 10)$ kasu berriaren klasearen iragarpen bat oinarritzko 3-NN sailkatzailea erabiliz.
- 1.2 Egin ezazu 1-NN sailkatzailearen holdout bidezko ebaluazioa. Har itzazu datubaseko lehenengo % 70 kasuak entrenamendurako eta azkeneko % 30 kasuak testerako. Kalkula ezazu asmatze-tasa, eraiki ezazu erroreen matrizea eta irudika ezazu ROC kurba.
2. (1.25 puntu) 12 kasu dituen honako datubasetik abiatuz *ID3* algoritmoa aplikatu da sailkapen-zuhaitza eraikitzeko, baina prozesua ez da amaitu (ikus irudia). Jarrai ezazu algoritmoarekin sailkapen-zuhaitza guztiz osatu arte. Ez egin inausketarik.

Kasuak	X_1	X_2	X_3	X_4	C
$\mathbf{x}_1$	1	2	1	2	1
$\mathbf{x}_2$	2	2	1	1	1
$\mathbf{x}_3$	1	1	1	2	1
$\mathbf{x}_4$	2	1	2	1	1
$\mathbf{x}_5$	1	2	2	3	2
$\mathbf{x}_6$	1	2	2	3	2
$\mathbf{x}_7$	2	2	2	2	2
$\mathbf{x}_8$	3	2	3	2	2
$\mathbf{x}_9$	3	2	3	2	3
$\mathbf{x}_{10}$	2	2	3	2	3
$\mathbf{x}_{11}$	3	1	2	1	3
$\mathbf{x}_{12}$	3	2	2	1	3



3. (1.25 puntu) Aurreko ikasturteetako datuetatik abiatuz dakigu karrera bukaerako proiektuan (KBP) matrikulatzen diren ikasleen artean % 40ak besterik ez duela defentsa ikasturte horretako ekainean eginen. Gainera, KBParen defentsa ikasturte horretako ekainean egin dutenen artean % 70ak 25 kreditu baino gutxiagoren matrikula egin du. KBParen defentsa ikasturte horretako ekainean egin ez dutenen artean % 10ak egin du 25 kreditu baino gutxiagoren matrikula.

Bestalde, KBParen defentsa ikasturte horretako ekainean egin dutenen artean lehenengo lauhilabetekoan KBParen garapenean emandako lanordu kopuruak probabilitate-banaketa normalari jarraitzen dio, 100 orduko itxaropen matematikoarekin eta 10 orduko desbideratze tipikoarekin. KBParen defentsa ikasturte horretako ekainean egin ez dutenen artean lehenengo lauhilabetekoan KBParen garapenean emandako lanordu kopuruak ere probabilitate-banaketa normalari jarraitzen dio, baina 50 orduko itxaropen matematikoarekin eta 20 orduko desbideratze tipikoarekin.

- (a) Ikasle batek 20 kredituren matrikula egin du eta KBParen garapenean 80 orduz egin du lan. KBParen defentsa ikasturte horretako ekainean egingo duela esango al zenuke?
 - (b) Erabakia hartzeko erabili duzun sailkapenerako paradigma azal ezaizu, eta esan aldez aurretiko suposizioak egin behar izan ote duzun.
4. (0.5 puntu) Demagun 3. ariketan emandako datubasea dugula eta demagun 12 kasu horien klasea (taulako azken zutabea) ez dela ezaguna. Hasierako partiketan kasu bakoitza konglomeratu batean dago. Esan ezaizu pausu batean zein konglomeratu bilduko diren distantziaren arabera.
5. (1 puntu) Idatzi Algoritmo Genetikoaren sasikodea eta azaldu Algoritmo Genetikoei buruzko hurrengo kontzeptuak: egokitzen funtzioa, gurasoen hautapenaren metodoa, gurutzaketa-eragilea, mutazio eragilea eta Konbergentzia irizpidea