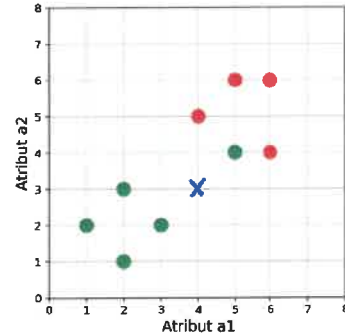


Odgovarjajte kratko in jedrnat, točno na zastavljena vprašanja. Vse odgovore pišite na črto pod vprašanji in **izključno na ta list**, ki ga edinega oddate na koncu! Čas pisanja: 45 minut.

(1)

Na sliki je podanih devet primerov, ki so opisani z dvema atributoma, a_1 in a_2 , barva pa označuje njihov razred. Za klasifikacijo teh primerov uporabljamo algoritem k-NN, pri čemer parameter k nastavimo na 1, uporabljamo pa evklidsko razdaljo.



(a) Za oceno klasifikacijske točnosti uporabljamo 3-kratno prečno preverjanje. Ob tem so podatki razdeljeni v tri podmnožice takole: $\{(a_1=1, a_2=2), (3,2), (6,6)\}$, $\{(2,1), (5,4), (5,6)\}$ in $\{(2,3), (4,5), (6,4)\}$. Kakšna je tako ocenjena klasifikacijska točnost, če se primer $(a_1=4, a_2=5)$ označi za rdečega?

100% (2/3), 2/3, 2/3; odg: 7/9 = 78%

(b) Zdaj pa za oceno kvalitete napovedi uporabljamo metodo izpusti enega (leave-one-out). Pri tem se zgodi, da se primer $(a_1=4, a_2=5)$ označi za rdečega. Kakšna je tako ocenjena klasifikacijska točnost? Kaj pa priklic za zeleni razred?

klasifikacijska točnost: 7/9 = 78% priklic (zelen): 4/5 = 80%

(c) Kako bi ta algoritem klasificiral primer $(a_1=4, a_2=3)$ pri $k=1$? Kaj pa pri $k=5$ in navadnem glasovanju?

k=1: zelen k=5: zelen

(2)

Sedaj na sliki iz prve naloge zgoraj zanemarimo barvo, torej za naše primere nimamo podatka o razredu. Tako se soočimo s problemom nenadzorovanega učenja. Lotimo se ga s pomočjo algoritma k-voditeljev (k-means). Tudi tu uporabljamo evklidsko razdaljo.

(a) Začetni točki dveh ($k=2$) centroidov sta $(1,3)$ in $(5,4)$. Kakšni sta njuni končni koordinati?

$c_1: (2,2)$ $c_2: (5\frac{1}{5}, 5)$

(b) V katero gručo bi model zgrajen v nalogi (a) uvrstil primer $(a_1=4, a_2=3)$?

V gručo C_1 (bližje je centroid s koordinato $(2,2)$)

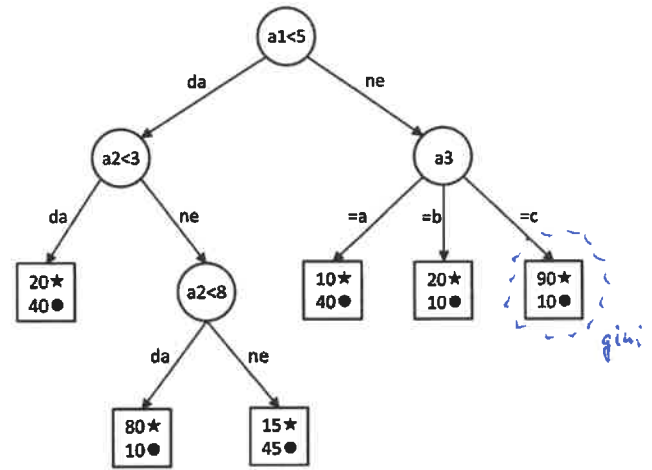
(c) Recimo, da nam domenski ekspert pove, da primer $(6,4)$ spada v rdeč razred, primer $(2,3)$ pa v zelen razred. Ostale primere iz iste gruče (vsak je v eni) kot sta označena primera po principu delno nadzorovanega učenja (semi-supervised learning) »pobarvamo« z isto barvo kot dana primera. Tako imamo pred seboj spet problem nadzorovanega učenja. A bodimo pozorni, *pravilne* barve razredov so takšne kot so na sliki zgoraj!

Kam bi algoritem 5-NN uvrstil primer $(a_1=4, a_2=3)$? Kakšna bi bila ocenjena klasifikacijska točnost tega algoritma (5-NN) ob uporabi metode izpusti enega? Algoritem uporablja navadno glasovanje.

$(4,3)$ se uvrsti kot: rdeč klasifikacijska točnost: 8/9

(3)

Podano je odločitveno drevo, ki ločuje črne luknje (●) od (ostalih) zvezd (★) na podlagi treh atributov. V listih drevesa na desni so podane tudi frekvence učnih primerov, ki pripadajo dani veji. List napove tisti razred, ki ima v njem večino.



(a) Izračunajte gini nečistočo v listu veje $a_1 \geq 5$ in $a_3 = 'c'$.

1
$$G = 1 - \left(\frac{10}{100} \right)^2 - \left(\frac{90}{100} \right)^2 = 0,18$$

(b) Kateri izmed treh atributov je v tem modelu najpomembnejši?

1 a_1 , atribut v korenu drevesa

(c) Kako bi podano odločitveno drevo klasificiralo naslednjih osem primerov 1: $\langle a_1=3, a_2=2, a_3=c \rangle$, 2: $\langle 4,5,a \rangle$, 3: $\langle 6,5,b \rangle$, 4: $\langle 6,2,a \rangle$, 5: $\langle 1,9,b \rangle$, 6: $\langle 8,8,a \rangle$, 7: $\langle 8,3,c \rangle$, 8: $\langle 7,1,b \rangle$?

2 Po vrsti (1-8): 0, ★, ★, 0, 0, 0, ★, ★

(d) Resnične vrednosti razredov za primere iz naloge (c) so po vrsti [●, ●, ★, ●, ★, ●, ★, ★]. Kakšna je vrednost AUROC, torej ploščine pod ROC krivuljo gledano s strani razreda ★?

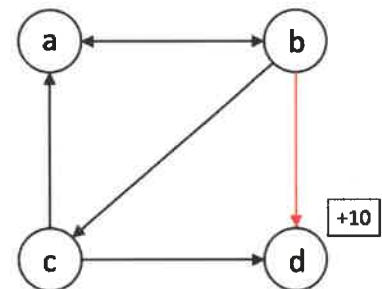
1 Napovedi vredimo po verjetnosti, da je ★ (primer 1 je npr. 33%, da je ★). $AUROC = \frac{12}{16} = 75\%$

(e) V kateri točki koordinatnega sistema (x,y) smo, ko narišemo prvih pet primerov na ROC krivuljo? Koordinati x in y izrazite v odstotkih!

1 $x = 50\%$, $y = 75\%$

(4)

Na sliki je podan preprost svet, ki se ga naš agent uči s pomočjo spodbujevanega učenja (reinforcement learning). Stanja so označena s črkami a-d, možne akcije v stanjih predstavljajo puščice. Tako je npr. v stanju b možno izbrati akcije »pojdi v a«, »pojdi v c« ali »pojdi v d«. Stanje d je končno, tam agent dobi nagrado +10 točk. Vse akcije so deterministične, edina izjema je označena z rdečo puščico. Tukaj se v tretjini primerov res zgodi akcija »pojdi v d«, v preostalih primerih pa se preprosto ne zgodi nič. Pri spodnjih vprašanjih konkretno napišite kako bi agent ravnal v stanjih b in c (pri a in d seveda nima izbire).



(a) Katere (optimalne) strategije se nauči naš agent, če ob vsaki akciji dobi še nagrado -1, tudi ob prehodih v končno stanje in tudi, če se po akciji ne zgodi nič?

1 b: pojdi v c, c: pojdi v d (strategija: čim hitroje na cilj)

(b) Kaj pa, če ob vsakem prehodu dobi nagrado +1?

1 b: a ali c (uxeno), c: pojdi v a (strategija: nikoli ne končaj)

(c) Sedaj pa agent ob nobeni akciji ne dobi nagrade (izjema je +10, ko doseže stanje d). Ob tem se uči tako, da ima parameter γ (discount factor) nastavljen na 0,99. Katere strategije se agent nauči?

1 Enake strategije kot v nalogi (a).