

Začeto dne	petek, 15. november 2019, 21:19
Stanje	Zaključeno
Dokončano dne	ponedeljek, 25. november 2019, 00:00
Porabljeni čas	9 dni 2 ure
Točke	9,72/15,00
Ocena	64,81 od možno največ 100,00

Vprašanje 1

Pravilno

Ocena 1,00 od 1,00

Podana je naslednja učna množica primerov, ki imajo podan atribut A in razred:

Primer	A	Razred
1	70	Yes
2	90	No
3	85	No
4	95	No
5	70	Yes
6	90	Yes
7	78	Yes
8	65	Yes
9	75	Yes
10	80	No
11	70	No
12	80	Yes
13	80	Yes
14	96	Yes

Katera od naslednjih diskretizacij atributa A zagotovi največji informacijski prispevek (najbolj minimizira nečistost)?

- ☒ a. [$\leq 65, > 65$] ✓
- ☐ b. [$\leq 70, > 70$]
- ☐ c. [$\leq 95, > 95$]
- ☐ d. [$\leq 90, > 90$]
- ☐ e. [$\leq 85, > 85$]

Vaš odgovor je pravilen.

Pravilni odgovori so: [$\leq 65, > 65$], [$\leq 95, > 95$]

Vprašanje 2

NEpravilno

Ocena 0,00 od 1,00

Izberi pravilne trditve glede regresijskih dreves.

Izberite enega ali več odgovorov:

- ☒ Regresijska drevesa uporabljamo, kadar imamo zvezne attribute. ✗
- ☐ Linearna regresija je poseben primer regresijskega drevesa.
- ☐ Za mero nečistoče pri regresijskih drevesih lahko uporabimo entropijo.
- ☒ V listih regresijskega drevesa včasih napovemo kar povprečno vrednost. ✓

Pravilni odgovori so: Linearna regresija je poseben primer regresijskega drevesa., V listih regresijskega drevesa včasih napovemo kar povprečno vrednost.

Vprašanje 3

NEpravilno

Ocena 0,00 od 1,00

Kateri od naslednjih učnih algoritmov lahko zgradijo hipotezo tudi če imajo učni primeri manjkajoče vrednosti atributov (ki jih umetno ne nadomeščamo z izbranimi vrednostmi ali pri predprocesiranju prvotno ne odstranimo nepopolnih primerov)?

- 1. klasifikacijsko drevo
- 2. naivni Bayes
- 3. kNN
- 4. linearna regresija

- ☒ a. 2, 3 ✖
- ☐ b. 1, 2, 3
- ☐ c. 2, 3, 4
- ☐ d. 1, 2, 3, 4
- ☐ e. 1, 2
- ☐ f. 2, 4

Vaš odgovor je napačen.

Pravilen odgovor je: 1, 2

Vprašanje 4

Pravilno

Ocena 1,00 od 1,00

V tabeli so predstavljeni podatki za potočne in soške postrvi. Želimo se naučiti model (hipotezo), s katerim bi lahko na podlagi genskih testov razlikovali med vrstami. A, B, C in D so geni (podatki so izmišljeni).

A	B	C	D	Vrsta
0	0	0	0	P
0	0	0	1	P
0	1	1	0	P
1	1	1	0	P
0	2	1	0	P
0	2	0	1	S
1	0	0	0	P
1	0	0	1	P
1	1	1	0	S
1	2	0	0	P
1	2	1	1	S
1	1	1	0	P

Izračunaj verjetnost $p(\text{vrsta}=\text{P}|\text{A}=1,\text{B}=0,\text{C}=1,\text{D}=0)$ z naivnim Bayesovim klasifikatorjem. Vse verjetnosti izračunaj z relativno frekvenco.

Odgovor: ✔

Pravilen odgovor je: 0,988

Vprašanje 5

Pravilno

Ocena 1,00 od 1,00

Spet izračunaj verjetnost $p(\text{vrsta}=\text{P}|\text{A}=1,\text{B}=0,\text{C}=1,\text{D}=0)$ z naivnim Bayesovim klasifikatorjem, a tokrat vse verjetnosti izračunaj z Laplaceovo oceno verjetnosti.

Odgovor: ✔

Pravilen odgovor je: 0,8867

Vprašanje 6

Delno pravilno

Ocena 0,33 od 1,00

Z metodo za izračun nomograma pojasni, kako k odločitvi, da gre za potočno postrv (razred P) prispevata različni vrednosti atributa C.

Povedano drugače, izračunaj število točk, ki ju k odločitvi prispevata različni vrednosti. Pri izračunu uporabi naravni logaritem, rezultat pa napiši s tremi decimalkami in brez presledkov (npr. "-0,123").

točke($P|C=0$) = ✖

točke($P|C=1$) = ✖

!! DISCLAIMER: Za naslednje rodove !! Rezultata sta pravilna, le zamenjana :(LP in lep pozdrav ... Odgovori še na naslednje vprašanje:

K odločitvi, da gre za soško (razred S) postrv več prispeva vrednost atributa ✔

Vprašanje 7

Pravilno

Ocena 1,00 od 1,00

Katere izmed naslednjih primerov bi 3NN klasificiral kot vrsta=S? Uporabi manhattansko razdaljo.

Izberite enega ali več odgovorov:

- ☒ a. A=1,B=2,C=0,D=1 ✔
- ☐ b. A=1,B=1,C=1,D=0
- ☒ c. A=0,B=2,C=1,D=1 ✔
- ☐ d. A=1,B=0,C=0,D=0

Vaš odgovor je pravilen.

Pravilni odgovori so: A=0,B=2,C=1,D=1, A=1,B=2,C=0,D=1

Vprašanje 8

Pravilno

Ocena 1,00 od 1,00

Kaj je to prekletstvo dimenzionalnosti?

- ☐ a. pove nam, da z večanjem števila dimenzij postajajo problemi nadzorovanega učenja nerešljivi
- ☐ b. pove, da pri dodajanju števila dimenzij potrebujemo tudi več učnih primerov
- ☐ c. pove, da moramo vrednosti atributov normalizirati na isto skalo
- ☐ d. pove, da se z večanjem števila dimenzij povečujeta prostorska in časovna zahtevnost algoritmov za učenje
- ☒ e. pove nam, da se z večanjem števila dimenzij povečuje tudi prazen prostor okoli primerov ✔

Vaš odgovor je pravilen.

Pravilen odgovor je: pove nam, da se z večanjem števila dimenzij povečuje tudi prazen prostor okoli primerov

Vprašanje 9

Pravilno

Ocena 1,00 od 1,00

Koliko koeficientov hiperravnine v n-dimenzionalnem prostoru moramo določiti pri učenju z linearno regresijo?

- ☐ a. $\frac{n}{2}$
- ☐ b. $n - 1$
- ☐ c. $\frac{n \times (n-1)}{2}$
- ☒ d. n ✔
- ☐ e. $n + 1$

Vaš odgovor je pravilen.

Pravilen odgovor je: n

Vprašanje 10

Delno pravilno

Ocena 0,17 od 1,00

Kaj velja za učenje linearne regresije?

Izberite enega ali več odgovorov:

- ☐ a. stohastični gradientni spust vedno najde optimalno rešitev
- ☒ b. pri klasifikaciji ne obstaja eksaktna analitična rešitev ✓
- ☐ c. pri gradientnem spustu lahko s parametrom nadzorujemo hitrost učenja in s tem vedno prej zaključimo z optimizacijo
- ☒ d. za učenje v več kot 2 dimenzijah obstaja samo optimizacijski algoritem ✗
- ☐ e. metoda je primerna samo za regresijske probleme

Vaš odgovor je delno pravilen.

Pravilno ste izbrali 1.

Pravilni odgovori so: pri klasifikaciji ne obstaja eksaktna analitična rešitev, pri gradientnem spustu lahko s parametrom nadzorujemo hitrost učenja in s tem vedno prej zaključimo z optimizacijo

Vprašanje 11

NEpravilno

Ocena 0,00 od 1,00

V tabeli so predstavljeni podatki za postrvi, ki so opsani z atributi A, B, C (različna genska testiranja) in dolžino postrvi, ki je ciljna spremenljivka.

A	B	C	dolžina
0	0	0	9
0	0	0	10
0	1	1	9
0	2	0	12
0	2	1	12
1	0	0	12
1	0	0	15
1	1	1	11
1	1	1	15
1	1	1	9
1	2	0	9
1	2	1	12

Z lokalno uteženo regresijo želimo napovedati dolžino postrvi z atributi $x_j = \{A = 1, B = 2, C = 1\}$. Pri izračunu uporabi:

- Manhattansko razdaljo za merjenje razdalj,
- jedrno funkcijo $w_i = \frac{1}{1+d_{ij}}$.

Odgovor vnesi s tremi decimalnimi mesti, decimalno vejico in brez presledkov (npr. "8.123").

Odgovor: ✗

Pravilen odgovor je: 11,359

Vprašanje **12**

Pravilno

Ocena 1,00 od 1,00

Podana je učna množica naslednjih točk (podana je oznaka točke in koordinati (x,y)): A(3,1), B(1,2), C(3,4), D(5,2), E(1,1).

Na množici izvedemo hierarhično združevalno gručenje, pri katerem uporabimo manhattansko razdaljo in pristop popolne povezanosti (complete linkage) merjenja razdalj med gručami. V primeru, če je med gručami/točkami razdalja enaka, združimo najprej tiste, ki vsebujejo točke, ki se v abecednem redu nahajajo prej (npr. izmed gruč točk AD in BC bomo s točko E naprej združili gručo AD, ker vsebuje točko A).

Navedi gruče, ki jih dobimo, če dendrogram porežemo tako, da dobimo tri gruče. Odgovor podaj kot oznake točk, urejene po abecedi, brez presledkov, gruče pa loči z vejicami (npr.: "A,BC,DE").

Odgovor: ABE,C,D



Pravilen odgovor je: ABE,C,D

Vprašanje **13**

Pravilno

Ocena 1,00 od 1,00

Uporabimo podatke iz prejšnje naloge. Navedi gruči, ki ju dobimo, če dendrogram porežemo tako, da dobimo dve gruči. Odgovor podaj kot oznake točk, urejene po abecedi, brez presledkov, gruče pa loči z vejicami (npr.: "A,BC,DE").

Odgovor: ABE,CD



Pravilen odgovor je: ABE,CD

Vprašanje **14**

Pravilno

Ocena 1,00 od 1,00

Ali bi pri pristopu z uporabo enojne povezanosti (angl. single linkage) v prejšnjih nalogah dobili drugačen rezultat? (Da/Ne)

Odgovor: Da



Pravilen odgovor je: Da

V množici točk A(3,1), B(1,2), C(3,4), D(5,2), E(1,1) izvedemo gručenje z algoritmom k-means, ki uporablja manhattansko razdaljo.

Začetni naključno določeni vrednosti centroidov sta C1(4,4) in C2(5,4).

Začetna pripadnost točk centroidom je naslednja:

- A pripada centroidu

C2

✖
- B pripada centroidu

C2

✖
- C pripada centroidu

C1

✔
- D pripada centroidu

C2

✔
- E pripada centroidu

C2

✖

 .

V naslednji iteraciji sta koordinati centroidov naslednji:

- C1(

3

✖

 ,

4

✖

)
- C2(

2,5

✖

 ,

1,5

✖

).

Pripadnost točk centroidom v tej (naslednji iteraciji) je naslednja:

- A pripada centroidu

C2

✖
- B pripada centroidu

C2

✖
- C pripada centroidu

C1

✔
- D pripada centroidu

C2

✔
- E pripada centroidu

C2

✖

 .