Osnove umetne inteligence, izpit 1. sept. 2011

Navodila:
Čas: 70 min. Uporaba literature in zapiskov ni dovoljena.
Vse naloge so enakovredne.
Ustno: ponedeljek, 5. sept. 2011 ob 14h

1. Naj bo prostor stanj drevo s konstantnim vejanjem 3. Cene vseh povezav v drevesu naj bodo enake 1. Uporabljamo hevristično funkcijo h, ki je za vsako vozlišče *V* v drevesu definirana takole:

$$h(V) = g(V)^2$$

pri čemer je g(*V*) globina vozlišča *V* v drevesu (za koren drevesa štejemo, da je na globini 0). Začetno vozlišče preiskovanja je koren drevesa. Preiskujemo z algoritmom IDA*. Na začetku je meja Bound enaka 0. Ciljno vozlišče je na globini 3.
(a) Kako se spreminja meja Bound med iteracijami algoritma IDA*? Koliko je cena rešitvene poti?
(b) Najmanj koliko in največ koliko vozlišč generira IDA*, preden najde rešitev? Med generirana vozlišča štejemo tudi morebitna ponovno generirana vozlišča vključno z začetnim vozliščem.

2.
(a) Kaj v strojnem učenju pomeni izraz »pristranskost« (angl. bias)? Navedi dve vrsti pristranskosti in ju na kratko razloži.
(b) Kaj meri Ginijev indeks, za kaj se uporablja v strojnem učenju? Zapiši formulo za izračun Ginijevega indeksa in razloži pomen simbolov v formuli.
(c) Zapiši vsaj tri različne formule za ocenjevanje verjetnosti dogodka na osnovi eksperimentalnih podatkov. Izvedli smo vsega *N* poskusov, pri čemer se je dogodek zgodil v *n* poskusih. Razloži uporabljane simbole v formulah in za vsako od formul na kratko podaj njihove prednosti in slabosti.
(d) Opiši na kratko bayesovski postopek ocenjevanja verjetnosti.
(e) Kaj pomeni v preiskovalnih algoritmih, da je ocenitvena funkcija *monotona*? V katerem algoritmu je ta lastnost pomembna in zakaj?

3. V verjetnostnem grafu imamo dogodke a, b, c in d. Struktura grafa je naslednja: c je naslednik od a in b, d je naslednik od b. Pripadajoče verjetnosti so:

p(a) = 0.3
p(b) = 0.1
p(c | a b) = 0.9, p(c | ~a b] = 0.6, p(c | a ~b) = 0.8, p(c | ~a ~b) = 0.1
p(d | b) = 0.9, p(d, ~b) = 0.1

(a) Oceni brez računanja in na kratko utemelji, katera verjetnost je večja:

p(a | c) ali p(a | b c d) ?
(b) Izračunaj pogojno verjetnost p(b | c d).
(c) Zapiši vse množice vozlišč, ki d-ločujejo vozlišči a in b.

4. Imamo problem atributnega strojnega učenja, v katerem nastopata binarna atributa A in B ter binarni razred C. Možni vrednosti atributov in razreda so 0 in 1.

Podanih je naslednjih 10 učnih primerov:

A	B	C
0	0	1
0	1	0
1	0	0
1	1	1
0	0	1
0	1	0
1	0	0
1	1	1
0	0	1
0	1	0

(a) Koliko je informacijski dobitek atributa A? Koliko je relativni informacijski dobitek atributa A? Brez računanja ugotovi, kateri atribut A ali B ima večji relativni informacijski dobitek, oz. ali imata oba enakega? Utemelji odgovor.

(b) Oceni iz gornjih učnih primerov z Laplacovo oceno verjetnosti, koliko je klasifikacijska točnost P1 klasifikacijskega pravila (upoštevaj seveda samo primere, ki jih pravilo pokriva):

if B = 1 then C = 0

(c) Oceni iz gornjih učnih primerov z Laplacovo oceno verjetnosti, koliko je klasifikacijska točnost P2 klasifikatorja:

if A = 1 & B = 1 then C = 0 else C = 1

(d) Če vemo, da so znane verjetnosti vrednosti atributov:

P(A=1) = 0.5
P(B=1 | A=1) = 0.6
P(B=1 | A=0) = 0.4,

kako se spremeni ocena klasifikacijske točnosti P2 klasifikatorja iz vprašanja c?

Odgovori na izbrana vprašanja

1.
(a) Bound = 0, 2, 6, 12; cena rešitve = 3
(b) min # = 4+13+40+4 = 61, max # = 1+4+13+40+(40+81-3) = 175

2.
(b) Meri (ne)čistost množice primerov glede na razred, Gini = 1 - SUM[i=1..N_c] p_i^2
(c)

- Relativna frekvenca: p = n/N, prednost = enostavnost, OK za velik N, slabo za majhen N
- Laplacova ocena: p = (n+1)/(N+2),   prednost = ne potrebuje dodatnih parametrov, bolj zanesljivo kot rel. frekv. za majhen N; pomanjkljivost: ne upošteva morebitnega predznanja o verjetnosti (verj. porazdelitvi) možnih izidov
- m-ocena: p = (n+m*P_a)/(N+m); prednost: upošteva apriorno verjetnostno porazdelitev, slabost: zahteva parametra m in Pa, ki sta lahko povsem neznana in sta stvar ugibanja.

(e) Monotona ocenitvena funkcija narašča vzdolž katerekoli poti v prostoru stanj, to je: Če je vozlišče v1 predhodnik v2, potem je f(v1) ≤ f(v2). Ta lastnost je pomembna pri IDA*. Če je f monotona, potem je zagotovljeno, da IDA* preiskuje prostor stanj v priorotetnem zaporedju

3.
(a) p(a|c) > p(a|bcd), ker sta a in b oba razloga za c; če se b zgodi, je a manj potreben za c
(b) p(b|cd) = p(bcd)/p(cd) = p(b)p(d|b)p(c|b)/[p(d)p(c|d)] = 0.1*0.9*0.69/[0.18*0.5] = 0.69
(c) {}, {d}

4.
(a) Gain(A) = 0, GainRatio(A) = 0
GainRatio(B) > GainRatio(A), ker sta v celotni učni množici verjetnosti P(C=0) = P(C=1) = 0.5, medtem ko se po uporabi atributa B spremenita v P(C=0|B=0)=0.4 in P(C=0|B=1)=0.6. Zato se entropija po uporabi atributa B zmanjša, po uporabi atr. A pa ostane enaka.
(b) P1 = $P_{Laplace}$(C=0|B=1) = (3+1)/(5+2) = 4/7 = 0.5714
(c) P2 = P(A=1 & B=1) * (0+1)/(2+2) + (1 - P(A=1 & B=1)) * (3+1)/(8+2) =
¼ * ¼ + ¾ * 4/10 = 0.3625
(d) P2 = P(A=1)*P(B=1|A=1)*1/4 + (1- P(A=1)*P(B=1|A=1))*0.4 =
0.5*0.6 * 0.25 + 0.7 * 0.4 = 0.355