① Imamo problem strojnega učenja
ATRIBUTI: X in Y
RAZRED: R

$H(C) = -\sum_k p_k \log_2 p_k$ NEDOLOČENOST RAZREDA

$Gain(A) = H(C) - I_{RES}(A)$ INFO GAIN ATRIBUTA A

Koliko je (Gain(Y))?

$X = \{0, 1\}$ 8 primerov
$Y = \{a, b\}$

$H(R) = -\frac{4}{8} \log \frac{4}{8} - \frac{4}{8} \log \frac{4}{8}$

$= \underline{1}$

Nedoločenost razreda R je $\underline{1}$.

$I_{RES}(Y) = \overset{Y}{\underset{\underset{[2,2]\ [2,2]}{+a\quad +b}}{\diagup\diagdown}} = 2 \cdot \left(\frac{4}{8} \left(\frac{2}{4} \cdot \log_2 \frac{2}{4} + \frac{2}{4} \log_2 \frac{2}{4} \right) \right)$

$= \underline{1}$

$Gain(Y) = H(R) - I_{RES}(Y) = 1 - 1 = \underline{0}$

Od Y ne dobimo 0 informacij.

② Program za strojno
učenje je dobil:

if X=0 AND Y=a
    R = +
ELSE
    R = −

→ Hipoteza

Oceni klasifikacijsko
napako Hipoteze!
↳ Laplacetova

$P = \frac{h+1}{N+K}$

k - število vseh razredov

N - število vseh primerov

h - št. primerov ki pripadajo raz. C

	X	Y	R
1.	0	a	+
2.	0	b	−
3.	1	a	−
4.	1	b	+
5.	0	a	+
6.	0	b	−
7.	1	a	−
8.	1	b	+

	REZULTAT	H	
1.	0 a +	+	1
2.	0 b −	−	
3.	1 a −	−	\
4.	1 b +	÷	\
5.	0 a +	+	2
6.	0 b −	−	
7.	1 a −	−	\
8.	1 b +	−	\

Dva razreda

$k = 2 \Leftarrow \{+, -\}$

$P = \frac{h+1}{N+K} = \frac{h+1}{N+2}$

h... število pravilno opredeljenih v listu

N... število vseh opredeljenih v listu
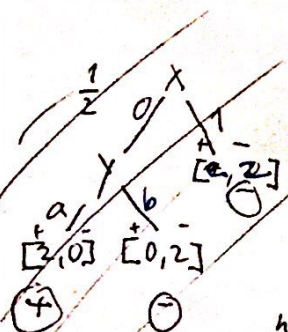
$L_1: \frac{2+1}{2+2} = \frac{3}{4}$

$L_2: \frac{3}{4}$

$L_3: \frac{4+1}{4+2} = \frac{5}{6}$

KLASIF.
TOČNOST → $\frac{2}{8} \cdot \frac{3}{4} + \frac{2}{8} \cdot \frac{3}{4} + \frac{4}{8} \cdot \frac{5}{6} = 0.79$

NAPAKA ↳ $1 - 0.79 = 0.208$

DN1.1

- Poglej kako algoritem naredi drevo!
- v listk napiši kako se klasificirajo posamezni testni → KLASIFICIRA ~~testni~~ SE GLEDE NA
  primeri   VEČINO
- Laplace glede na List! → $\frac{n+1}{N+k}$ n ... pravilno klas. v listu (oz. glede na večinski razred)
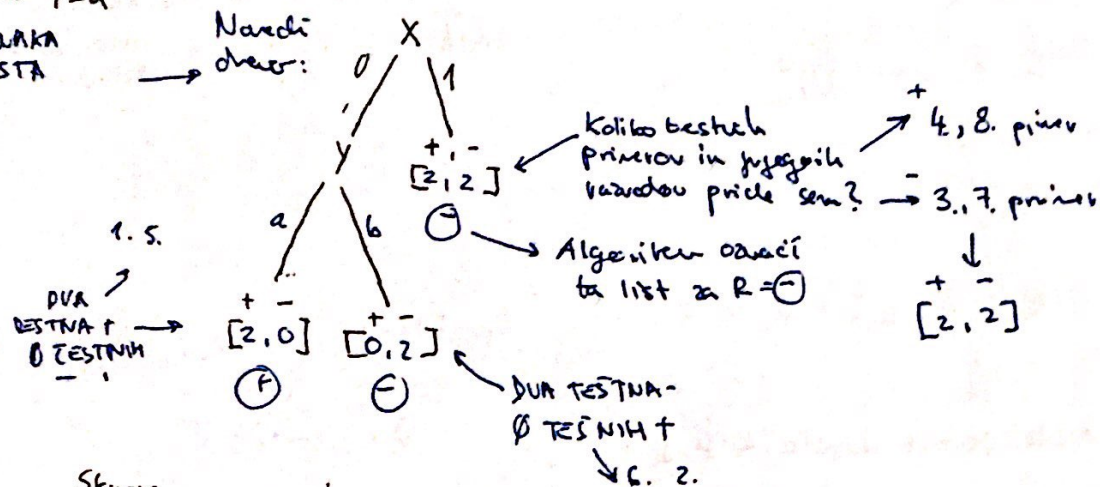   k ... število razl. razredov

② if $X=0$ ~~TEN~~ AND $Y=a$
       R = + ← OZNAKA LISTA   Naredi drevo:
   ELSE
       R = -

OZNAKA LISTA

1. 5.

DVA TESTNA + → [2,0]   [0,2]
0 TESTNIH -   ⊕   ⊖

Koliko testnih primerov in njihovih razredov pride sem? → 4, 8. primer → 3, 7. primer

Algoritem označi ta list za $R=⊖$

$^+_{[2,2]}$

DVA TESTNA -
0 TESTNIH +
↓ 6. 2.

$L_1: \frac{2+1}{2+2} = \frac{3}{4}$

$L_2: \frac{2+1}{2+2} = \frac{3}{4}$

$L_3: \frac{2+1}{4+2} = \frac{3}{6} = \frac{1}{2}$

Skupna točnost: $\frac{primerov}{v~listu} \sum_i primerov \cdot laplace$
                                        v listu $_i$ lista i

$= \frac{2}{8} \cdot \frac{3}{4} + \frac{2}{8} \cdot \frac{3}{4} + \frac{1}{2} \cdot \frac{1}{2}$

$t(d) = 0.6~\cancel{8}25$

~~točnost drevesa~~ klasif točnost

klasif Napaka ~~drevesa~~ $= \cancel{1-t} 1 - t(d)$
$= \boxed{0.375}$

③ Atributa X in Y neodvisna
Velja: $P(X=0)=0.5$
       $P(Y=a)=0.6$ $\longmapsto$ $P(X=1)=0.5$ Neodvisnost
                                    $P(Y=b)=0.4$ $P(A \wedge B) = P(A) P(B)$
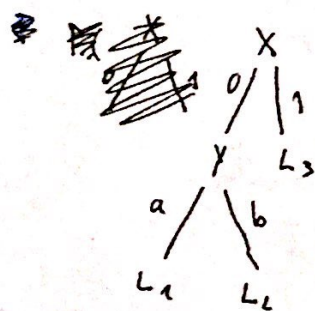
Od prej:
$L_1 = 3/4$
$L_2 = 3/4$
$L_3 = 1/2$

Utež izračunaw:
~~$P(X=0)$...~~
$t(d) = P(X=0)\left(P(Y=a) \cdot \frac{3}{4} + P(Y=b) \cdot \frac{3}{4} \right) + P(X=1) \cdot \frac{1}{2}$
$t(d) = 0.5 \left(0.6 \cdot \frac{3}{4} + 0.4 \cdot \frac{3}{4} \right) + 0.5 \cdot \frac{1}{2}$
$\boxed{= 0.625}$

~~$t(d) = 1 - t(d) = 0.375$~~

KLASIF. TOČNOST ZA VSAK LIST:
• LAPLACE ali
• M-OCENA
  krat
VERJETNOST TEGA LISTA

DVA FAKTORJA

⑧ PROBLEM: Atributa A in B
Razred C

KOLIKO JE REL. INFO
PRISPEVEK ATR. A
↳ GainRatio (A)?

	A	B	C	
X_1	0	0	1	01
X_2	0	1	0	00
X_3	1	0	0	01
X_4	1	1	1	00
X_5	0	0	1	01
X_6	0	1	0	
X_7	1	0	0	
X_8	1	1	1	
X_9	0	0	1	
X_{10}	0	1	0	

$I(A) = H(A)$
$I = I(C) = H(C)$

$GainRatio(A) = \dfrac{Gain(A)}{I(A)}$

$\neq \quad = \dfrac{I - I_{RES}(A)}{I(A)}$

$I_{RES}(A) = $ A ⟋⟍ ≈ ....

[ ] [ ]

↓
Po delitvi
uL. klasif

$A = \langle 0,1 \rangle$
$[6,4] \longmapsto I(A) = \dfrac{6}{10} \log_2 \dfrac{6}{10} + \dfrac{4}{10} \log_2 \dfrac{4}{10}$

$= 0{\cdot}970950594$

$I_{RES}(A) = $ A $\;{}^{0}\!\diagup \diagdown\!{}^{1}\; = \dfrac{6}{10}\cdot 1 + \dfrac{4}{10}\cdot 1 = \underline{\underline{1}}$

$C = 0\ 1 \quad 0\ 1$
$[3,3]\ [2,2]$

$I(C) = \dfrac{5}{10} \log_2 \dfrac{5}{10} \cdot 2 = 1$

$Gain(A) = I(C) - I_{RES}(A) = \cancel{0}$

$GainRatio(A) = \dfrac{\cancel{\;}0}{0{\cdot}970950} = 0$

⑩ TEST~~IRI~~ H ⑧
~~LAPLACE~~ ~~OCENA~~
Na nalogo ⑧ oceni klas. točnost spodnjega
pravila z Laplacovo oceno verjetnosti!

IF A = 1 AND B = 1
    C = 0
ELSE
    C = 1

1. zgradimo drevo
2. Poglejmo v katere liste padejo učni primeri!
3. Izrač. LAPLACE $L_1, L_2 \ldots$
4. Izrač. TOČNOST

①

$[0,2]$ $[2,0]$
$(C=0)$ $(C=1)$
then   else

Ⓐ — 1 / 0 — $[3,3]$ $(C=1)$ else
Ⓑ
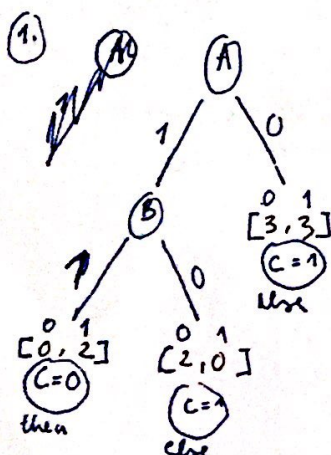
② Obravnavano
črtno → vstavi
v $[x_{i,n}]$ pod
listi

④ $E(d) = \left(\dfrac{2}{10}\cdot\dfrac{3}{4}\right) +$
$\left(\dfrac{2}{10}\cdot\dfrac{3}{4}\right) +$
$\dfrac{6}{10}\cdot\dfrac{1}{2}$

$= \underline{0{\cdot}6}$ ⑩
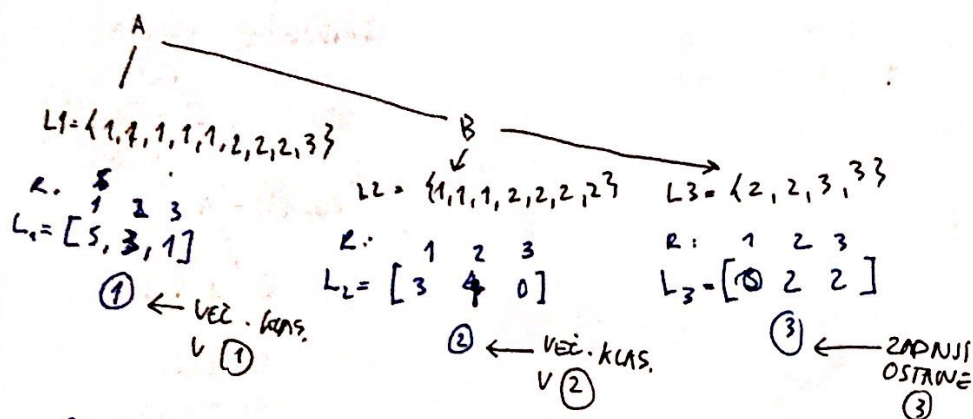
③ $P_{LAP} = \dfrac{h+1}{N+k}$ $k=2 \Leftarrow \langle 0,1\rangle \ni C$

$L_1 = \dfrac{2+1}{2+2} = \dfrac{3}{4}$

$L_2 = \dfrac{3}{4}$

$L_3 = \dfrac{3+1}{6+2} = \dfrac{4}{8} = \dfrac{1}{2}$

DN1. 2

(5) Imamo problem SU s $R = \{1,2,3\}$, ALG za V je iz 20 primerov izgradilo spodnje drevo:

$L_1 = \{1,1,1,1,1,2,2,2,3\}$

R. $\begin{array}{ccc} 5 & 3 & 1 \\ 1 & 2 & 3 \end{array}$

$L_1 = [5, 3, 1]$

① ← več. klas. V ①

$L_2 = \{1,1,1,2,2,2,2\}$

R: $\begin{array}{ccc} 1 & 2 & 3 \\ 3 & 4 & 0 \end{array}$

$L_2 = [3 \ 4 \ 0]$

② ← več. klas. V ②

$L_3 = \{2,2,3,3\}$

R: $\begin{array}{ccc} 1 & 2 & 3 \\ 0 & 2 & 2 \end{array}$

$L_3 = [0 \ 2 \ 2]$

③ ← ZADNJI OSTANE ③

OCENI KLAS. TOČNOST V L_1

$k = 3$

$N = 5 + 3 + 1 = 9$

$h = 5$

$P_{LAP}(L_1) = \frac{5 + 1}{9 + 3} = \frac{6}{12} = \frac{1}{2}$

OCENI KLAS. TOČNOS V L_1 po LAPLACE

$k = 3$

$N = 7$

$LAPLACE = \frac{4 + 1}{7 + 3} = \frac{5}{10} = \frac{1}{2}$ ⑤

---

(6) SPADA POD ⑤

OCENI KLAS. TOČNOST DRELESA

$L_1: \frac{1}{2}$

$L_2: \frac{4+1}{7+3} = \frac{5}{10} = \frac{1}{2}$

$L_3: \frac{2+1}{4+3} = \frac{3}{7}$

P do listov

$$t(d) = \frac{9}{20} \cdot \frac{1}{2} + \frac{7}{20} \cdot \frac{1}{2} + \frac{4}{20} \cdot \frac{3}{7} = 0.486 \quad 6$$

LAPLACE LISTOV

---

(7) Isto drevo kot pri 5. nalogi!

↳ m ocena
↳ $m = 6$
↳ APRIORNE VERJETNOSTI:

$P_0(1) = 0.5$

$P_0(2) = 0.35$

$P_0(3) = 0.15$

? V katri razred se klasificirajo novi primeri ki padejo v L3?

↳ Izračunaj m oceno L3 za vsak razred posebej!
Tisti ki ima maksimalno, v tistega se klasificira!

$\begin{array}{ccc} & 1 & 2 & 3 \end{array}$

$L_3 = [0 \ 2 \ 2] \quad \begin{array}{l} N = 4 \\ m = 6 \end{array}$

$t_m 3(1) = \frac{0 + 0.5 \cdot 6}{4 + 6} = \frac{3}{10}$

$t_m 3(2) = \frac{2 + 0.35 \times 6}{10} = \frac{4.1}{10}$ → List se klasificira v razred ②

$t_m 3(3) = \frac{2 + 0.15 \times 6}{10} = \frac{2.9}{10}$

M OCENA

$P = \frac{h + P_a \cdot m}{N + m}$

Scanned with CamScanner

⑨ Testni primeri iz ⑧ naloge
↳ Točnost spodnjega pravila
po Laplaceovi !
↳ Klasif. točnost !

$t(d)$, Laplace ?

IF B = 1 THEN
    C = 0
ELSE
    C = 1

①   Ⓑ
1 /   \ 0
0,1   0,1
[~~5,2~~]   [2,3]
Ⓒ=0   Ⓒ=1
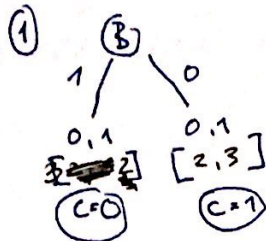
② vstavi UP v
pravilni list
[×,×]

③ LAPLACE
$L_1: \frac{3+1}{5+2} = \frac{4}{7}$

$L_2: \frac{3+1}{5+2} = \frac{4}{7}$

③ $t(d) = \frac{5}{10} \cdot \frac{4}{7} + \frac{5}{10} \cdot \frac{4}{7}$

$= \frac{1}{2} \cdot \frac{4}{7} + \frac{1}{2} \cdot \frac{4}{7}$

$= \frac{4}{7} = \boxed{0.8571}$ ⑨

⑪ • Isti problem kot ⑧   UP.
• Laplace ocena
• Oceni $t(d) = ?$
• Poznamo: $P(A=1) = 0.5$
$P(B=1|A=1) = 0.6$ →
$P(B=1|A=0) = 0.4$
↳ Neporabno?

IF A = 1 AND B = 1
    C = 0
ELSE
    C = 1

$P=0.5$ Ⓐ
1 /   \ 0 $P=0.5$
Ⓑ L_3
$1.0.6$ 1 / \ 0 $P=0.4$
L_1 L_2

$P(B=y|A=1) = P(B=1|A=1)$
     $+ P(B=0|A=1)$
     $= 1$
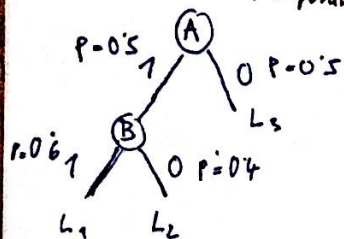
L_1 in L_2 in L_3
Izračunani v ⑩
nalogi !

$L_1 = \frac{3}{4}$

$L_2 = \frac{3}{4}$

$L_3 = \frac{1}{2}$

→ $t(d) = P(A=1) \cdot P(B=1|A=1) \cdot L_1$
     $+ P(A=1) \cdot P(B=0|A=1) \cdot L_2$
     $+ P(A=0) \cdot L_3$
     $= 0.5 \times 0.6 \times \frac{3}{4} + 0.5 \times 0.4 \times \frac{3}{4} + 0.5 \cdot \frac{1}{2}$
     $= \boxed{0.625}$ ⑪

DN1.3

(14) REZANJE NAZAJ
- MEP z Minimizacijo napake
- Ocena z Laplacetovo
- Uporablja Matrico za gradnjo drevesa
- Minimizacija klas. napake/max klas. točnosti

A [14,15,18]
B [6,10,8]   C [0,6,10]
D [9,8,10]
E [5,3,3]   F [1,7,5]
G [4,0,4]   H [4,6,3]   I [1,2,3]
J [2,2,2]   K [3,1,1]   L [1,1,1]   M [0,5,1]   N [0,1,3]
O [2,2,2]   P [2,4,1]
R [1,1,1]   S [2,0,0]

STATIČNA NAPAKA
V LISTU
$E(T) = e(V)$

VZVRATNA NAPAKA
$\sum_i p_i E(T_i) = p_1 \cdot E(T_1) + \dots$

Probujemo ksit   Lahko z točnostjo ali napako → max → min

$E(R) = \dfrac{1+1}{3+3} = \dfrac{2}{6} = \dfrac{1}{3}$

$E(S) = \dfrac{2+1}{2+3} = \dfrac{3}{5}$

k? → VZVRATNA NAPAKA V K

$E(k) = \dfrac{3}{5} \cdot \dfrac{1}{3} + \dfrac{2}{5} \cdot \dfrac{3}{5} = \dfrac{6}{15} + \dfrac{4}{25}$ → VZVRATNA NAPAKA VK JE VEČJA OD STATIČNE → POD k POREŽEMO

STATIČNA NAPAKA V K

$p(k) = \dfrac{3+1}{5+3} = \dfrac{4}{8} = \dfrac{1}{2}$

$e(k) = \dfrac{1}{2}$

$t(k) = \dfrac{1}{2} = T(k)$

$e(k) < E(k)$
POREŽEMO

$t(O) = \dfrac{2+1}{6+3} = \dfrac{3}{9} = \dfrac{1}{3}$

$t(P) = \dfrac{4+1}{7+3} = \dfrac{5}{10} = \dfrac{1}{2}$

H? → VZVR. TOČNOST

$T(H) = \dfrac{6}{13} \cdot \dfrac{1}{3} + \dfrac{7}{13} \cdot \dfrac{1}{2} = \dfrac{6}{39} + \dfrac{7}{26} = 0.423$

$T(H) < E(H)$ → POREŽEMO POD H

STAT. TOČNOST

$E(H) = \dfrac{6+1}{13+3} = \dfrac{7}{16} = 0.4375$

$t(J) = \dfrac{2+1}{6+3} = \dfrac{3}{9} = \dfrac{1}{3}$

$T(k) = \dfrac{1}{2}$

Sva zračunali ko sva povezali
vzeli statičen.

E? → $T(E) = \dfrac{6}{11} \cdot \dfrac{1}{3} + \dfrac{5}{11} \cdot \dfrac{1}{2}$

$= \dfrac{6}{33} + \dfrac{5}{22} = 0.409$

$E(E) = \dfrac{5+1}{11+3} = \dfrac{6}{14} = \dfrac{3}{7}$

$= 0.428$

$T(E) < E(E)$ → POREŽEMO POD E

$$t(L) = \frac{1+1}{3+3} = \frac{2}{6} = \frac{1}{3}$$

$$\longrightarrow T(F) = \frac{3}{13} \cdot \frac{1}{3} + \frac{6}{13} \cdot \frac{2}{3} + \frac{4}{13} \cdot \frac{4}{7} \quad \left| \begin{array}{l} T(F) > t(F) \;\; POD \\ F \\ NE\text{-}REŽEMO \end{array} \right.$$

$$t(M) = \frac{5+1}{6+3} = \frac{6}{9} = \frac{2}{3}$$

$$= \frac{3}{39} + \frac{12}{39} + \frac{16}{91} = 0.560$$

$$t(N) = \frac{3+1}{4+3} = \frac{4}{7}$$

$$t(F) = \frac{7+1}{13 + 3} = \frac{8}{16} = \frac{1}{2} = 0.500$$

$$t(G) = \frac{4+1}{8+3} = \frac{5}{11}$$

$$\longrightarrow T(D) = \frac{8}{27} \cdot \frac{5}{11} + \frac{13}{27} \cdot 0.4375 + \frac{6}{27} \cdot \frac{4}{9} = \quad \longrightarrow \begin{array}{l} T(F) > t(F) \;\; POD\ D \\ NE\ OBREŽEMO \end{array}$$

$$t(H) = 0.4375$$

$$= \frac{40}{297} + 0.2106 + 0.0988 =$$

$$t(I) = \frac{3+1}{6+3} = \frac{4}{9}$$

$$= 0.444$$

$$t(D) = \frac{10+1}{27+3} = 0.366$$

$$T(E) = 0.428$$

$$T(F) = 0.560$$

$$\longrightarrow T(B) = \frac{11}{24} \cdot 0.428 + \frac{13}{24} \cdot 0.560 = \quad \longrightarrow \begin{array}{l} Pod\ B\ ne \\ obrežemo! \end{array}$$

$$= 0.4995 = 0.500$$

$$t(B) = \frac{10+1}{24+3} = \frac{11}{27} = 0.467$$
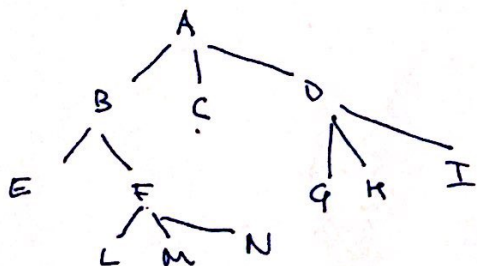
$$T(B) = 0.500$$

$$t(C) = \frac{6+1}{6+3} = \frac{7}{9} \longrightarrow T(A) = \frac{24 \cdot 0.500 + 6 \cdot \frac{7}{9} + 27 \cdot 0.444}{57} = \quad \longrightarrow \begin{array}{l} POD\ A\ NE \\ OBREŽEMO \end{array}$$

$$T(D) = 0.444$$

$$= 0.503$$

$$t(A) = \frac{25+1}{57+3} = 0.433$$

KONČNO DREVO

Oduzali smo po MEP vendar smo maksimizirali KT.
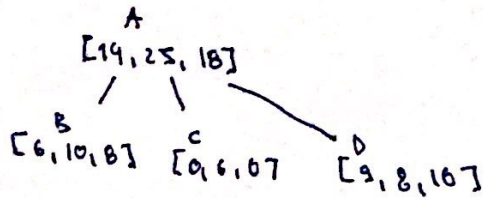
↳ ODREZALI: J, K, R, S, O, P

Odg, to drevo sedaj vsebuje 8 listov!

(16) Kar iz 1? naloge vzememo rezultat. Kolikšna je Vsuratna napaka v vozlišču B? Uporabi Laplaceovo oceno verjetnosti!

Računali smo 'z klas. točnostje zato:
$E(B) =$

(15) klasif. točnost pri vezaju u neprei č drevo dodrino do 1. Globine; uporabi Laplačtovo oceno verjetnost!

A
[14, 25, 18]

B   C   D
[6,10,8]   [9,6,07]   [9,8,10]

$t(B) = \dfrac{11}{27}$

$t(C) = \dfrac{7}{9}$ $\longrightarrow$ $T(A) = \dfrac{24 \cdot \frac{11}{27} + 6 \cdot \frac{7}{9} + 27 \cdot \frac{11}{30}}{57} = 0.427$

$t(D) = \dfrac{11}{30}$

(17) Kako bomo klas. novi primer u vozl. F če so apriorne verjetnosti:

$P(R=1) = 0.1$
$P(R=2) = 0.3$
$P(R=3) = 0.6$

in $m = 10$!

$F = [1, 7, 5]$

$t_m F(1) = \dfrac{1 + 0.1 \cdot 10}{13 + 10} = \dfrac{2}{23}$

$t_m F(2) = \dfrac{7 + 0.3 \cdot 10}{13 + 10} = \dfrac{10}{23}$

$t_m F(3) = \dfrac{5 + 0.6 \cdot 10}{13} = \dfrac{11}{23}$ $\leftarrow$ MAX

V VOZL. F SE KLASIFICIRAJO NOVI PRIMERI U 3. RAZRED

Scanned with CamScanner