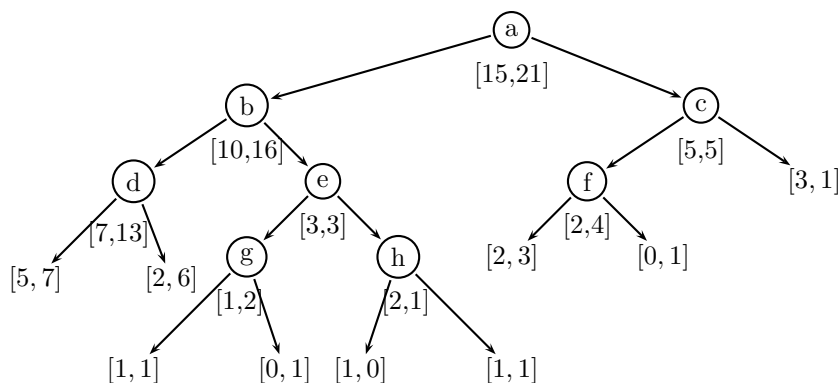


Naloga: rezanje drevesa s postopkom MEP

Porežite drevo na sliki 1 s postopkom MEP z uporabo Laplaceove ocene. V oglatih oklepajih je navedeno število primerov obeh razredov, npr. vozlišče b vsebuje 10 primerov prvega in 16 primerov drugega razreda. V vozlišču b je večinski drugi razred.



Slika 1

Rešitev:

V vsakem notranjem vozlišču v (govoriti o rezanju v listih ni smiselno) primerjamo statično napako e in vzvratno napako E . Statična napaka $e(v)$ je napaka drevesa v primeru, če v postane list (torej, če bi drevo porezali tik pod vozliščem v). Vzvratna napaka $E(v)$ je napaka, če drevesa ne porežemo. Kdaj naj torej drevo porežemo pod v ? Napako hočemo minimizirati, torej ga porežemo takrat, ko je statična napaka manjša ali enaka od vzvratne napake. Zakaj tudi, če je enaka? Z vidika točnosti je vseeno ali drevo tu porežemo ali ne, ker pa je manjši model ponavadi tudi bolj razumljiv, ga porežemo.

Ponovimo najprej formulo za izračun Laplaceove ocene verjetnosti,

$$p = \frac{n + 1}{N + k},$$

kjer je:

- n število primerov večinskega razreda v vozlišču
- N število vseh primerov v vozlišču
- k število razredov

V našem primeru je $k = 2$.

Vozlišče d :

Ker računamo napako, moramo ocenjeno verjetnost odšteti od 1 (glej nalogo 3). Statična napaka v vozlišču d je $e(d) = 1 - \frac{n+1}{N+2} = 1 - \frac{13+1}{20+2} = 8/22 = 0.363$. Za izračun vzvratne napake potrebujemo napako v levem in desnem listu d , ki ju bomo označili kar z d_L in d_D .

$$e(d_L) = 1 - \frac{7+1}{12+2} = 6/14 = 0.428$$

$$e(d_D) = 1 - \frac{6+1}{8+2} = 3/10 = 0.3$$

Vzvratna napaka je utežena vsota napak v obeh vozliščih. Utežimo ju z deležem primerov, ki pripadajo vsakemu listu. V levi list vozlišča d gre 12 od 20 primerov iz d , v desnega pa preostalih 8: $E(d) = \frac{12}{20} 0.428 + \frac{8}{20} 0.3 = 0.376$. Izračunali smo, da je statična napaka $e(d)$ manjša od vzvratne napake $E(d)$, torej je boljše, da drevo **porežemo** pod d .

Podobno izračunamo obe napaki za ostala vozlišča, pri čemer je pomembno, da gremo po drevesu od spodaj navzgor:

Vozlišče g :

$$e(g) = 1 - \frac{2+1}{3+2} = 2/5 = 0.4$$

$$e(g_L) = 1 - \frac{1+1}{2+2} = 0.5$$

$$e(g_D) = 1 - \frac{1+1}{1+2} = 1/3 = 0.333$$

$$E(g) = \frac{2}{3} 0.5 + \frac{1}{3} 0.333 = 0.444$$

$$e(g) \leq E(g), \text{ torej režemo pod } g.$$

Vozlišče h :

$$e(h) = 1 - \frac{2+1}{3+2} = 2/5 = 0.4$$

$$e(h_L) = 1 - \frac{1+1}{1+2} = 1/3 = 0.333$$

$$e(h_D) = 1 - \frac{1+1}{2+2} = 0.5$$

$$E(h) = \frac{1}{3} 0.333 + \frac{2}{3} 0.5 = 0.444$$

$$e(h) \leq E(h), \text{ torej režemo pod } h.$$

Vozlišče e :

$$e(e) = 1 - \frac{3+1}{6+2} = 0.5$$

$$e(e_L) = e(g) = 0.4$$

$$e(e_D) = e(h) = 0.4$$

$$E(e) = \frac{1}{2} 0.4 + \frac{1}{2} 0.4 = 0.4$$

$e(e) > E(e)$, torej **NE** režemo pod e .

Vozlišče f :

$$e(f) = 1 - \frac{4+1}{6+2} = 3/8 = 0.375$$

$$e(f_L) = 1 - \frac{3+1}{5+2} = 3/7 = 0.429$$

$$e(f_D) = 1 - \frac{1+1}{1+2} = 1/3 = 0.333$$

$$E(f) = \frac{5}{6} 0.429 + \frac{1}{6} 0.333 = 0.412$$

$e(f) \leq E(f)$, torej režemo pod f .

Vozlišče c :

$$e(c) = 1 - \frac{5+1}{10+2} = 0.5$$

$$e(c_L) = e(f) = 0.375$$

$$e(c_D) = 1 - \frac{3+1}{4+2} = 1/3 = 0.333$$

$$E(c) = \frac{6}{10} 0.375 + \frac{4}{10} 0.333 = 0.358$$

$e(c) > E(c)$, torej **NE** režemo pod c .

Vozlišče b :

$$e(b) = 1 - \frac{16+1}{26+2} = 0.393$$

$e(b_L) = e(d) = 0.363$ (ker smo pri d porezali, vzamemo tu statično napako vozlišča d)

$e(b_D) = E(e) = 0.4$ (pri e nismo rezali, zato vzamemo tu vzvratno napako)

$$E(b) = \frac{20}{26} 0.363 + \frac{6}{26} 0.4 = 0.371$$

$e(b) > E(b)$, torej **NE** režemo pod b .

Vozlišče a :

$$e(a) = 1 - \frac{21+1}{36+2} = 0.421$$

$e(a_L) = E(b) = 0.371$ (ker pri b nismo rezali, vzamemo tu vzvratno napako vozlišča b)

$e(a_D) = E(c) = 0.358$ (vzamemo vzvratno napako pri c , ker pod c nismo rezali)

$$E(a) = \frac{26}{36} 0.371 + \frac{10}{36} 0.358 = 0.367$$

$e(a) > E(a)$, torej **NE** režemo pod a .