Opće napomene

1. Nezavisnost, uvjetna nezavisnost – [redacted] produkt i "brisanje" argumenta:

Ⓐ Bezuvjetna nezavisnost slučajnih varijabli X i Y:

produkt: $$P(X,Y) = P(X)\cdot P(Y)$$

brisanje drugog argumenta (jedan uvjet):

$$P(X|Y) = P(X)$$
$$P(Y|X) = P(Y)$$

Ⓑ Uvjetna nezavisnost sl. varijabli X, Y, za dani Z

produkt (u prvom argumentu):

$$P(X,Y|Z) = P(X|Z)\cdot P(Y|Z)$$

brisanje drugog argumenta – prema uvjetnoj nezav.

$$P(X|Y,Z) = P(X|Z) \quad Y \text{ se briše}$$
$$P(Y|X,Z) = P(Y|Z) \quad X \text{ se briše}$$

Sve ostalo se prvo mora svesti na nešto od ovog, a to ide po Bayesovom pravilu ("okreće" argumente!).

2. Bayesova mreža smije imati višak grana, sve dok graf ostaje usmjeren i aciklički.

Višak grana se kasnije može eliminavati i to računanjem odgovarajućih distribucija i provjerom (uvjetne) nezavisnosti (v. gore). To NE IDE topološkim ili grafovskim algoritmom – poput d-separacije!

Uočiti – ako čvorove numeriramo "topološki", tj. tako da roditelji dolaze prije djece, onda lančano pravilo

$$P(X_1,\dots,X_n) = \prod_{i=1}^{n} P(X_i|X_1,\dots,X_{i-1})$$

koje uvijek vrijedi, odgovara Bayesovoj mreži koja je potpuni usmjereni aciklički graf!

$$\text{parents}(X_i) = \{X_1,\dots,X_{i-1}\}$$

Primjena na Zad. 1 (a)

(ii) Ova mreža je potpun graf

$$X_1 = \text{Omot}, \quad X_2 = \text{Oblik}, \quad X_3 = \text{Okus}$$

i sigurno je korektna, tj. što god pisalo u tablicama uvjetnih vjerojatnosti (CPT) u čvorovima, dobivamo korektnu združenu distribuciju.

Problem = mreža je neprirodna, ima previše grana i iz zadanog teksta treba ozbiljno računati da se nađu sve 3 CPT.

(iii) je očito korektna – prema tekstu i ima jednu granu manje – koja odražava uvjetnu nezavisnost Omot i Oblik za dani Okus, a CPT se trivijalno napišu.

(i) nije korektna, jer graf (prema topološkoj semantici) tvrdi da su Omot i Oblik nezavisne, tj.

$$P(\text{Omot} \mid \text{Oblik}) = P(\text{Omot})$$

ili

$$P(\text{Oblik} \mid \text{Omot}) = P(\text{Oblik})$$

ili

$$P(\text{Omot}, \text{Oblik}) = P(\text{Omot}, \text{Oblik})$$

No, ako iz teksta ili mreže (iii) izračunamo distribucije na lijevoj i desnoj strani, izlazi da ove jednakosti NE vrijede.

1. zad. – rješenje

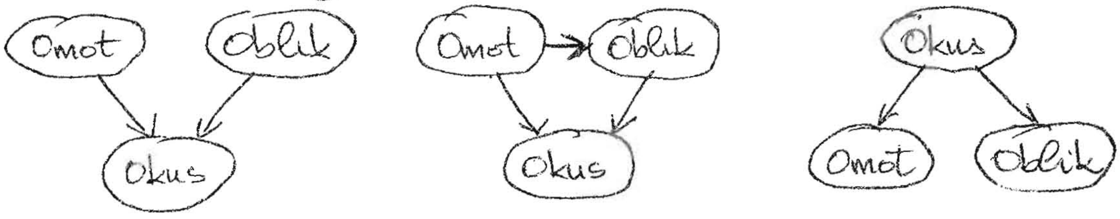
Sintaksa Bayesove mreže za n slučajnih varijabli $X_1, -, X_n$

- skup od n čvorova, svaki čvor odgovara jednoj slučajnoj varijabli X_i (varijabla = ime čvora)
- usmjerena grana između dva čvora opisuje izravni utjecaj polaznog čvora = roditelja, na dolazni čvor = dijete.
  Cijeli graf mora biti usmjereni aciklički graf.
- u svakom čvoru X_i zadana je uvjetna distribucija vjerojatnosti
  $$P(X_i \mid Parents(X_i))$$
  koja zadaje vjerojatnosti čvora X_i za dane (poznate) sve njegove roditelje $Parents(X_i)$.

- Globalna ili numerička semantika mreže:
  potpuna združena distribucija vjerojatnosti (konjunkcije) svih varijabli $X_1, -, X_n$ je produkt (prema Bayesovom lančanom pravilu) lokalnih uvjetnih distribucija vjerojatnosti
  $$P(x_1, -, x_n) = \prod_{i=1}^{n} P(x_i \mid parents(X_i))$$
  gdje su $x_1, -, x_n$ konkretne vrijednosti varijabli, a $parents(X_i)$ su konkretne vrijednosti roditelja od X_i.
  Na razini distribucija
  $$P(X_1, -, X_n) = \prod_{i=1}^{n} P(X_i \mid Parents(X_i))$$
  (produkt = produkt po točkama)

- Lokalna ili topološka semantika iz strukture grafa:
  svaki čvor X_i je uvjetno nezavisan od svih svojih nesljedbenika (nepotomaka), ako su dani (kao uvjet/dokaz) njegovi roditelji $Parents(X_i)$.

"Bombovi iznenađenja" - pomućene mreže:



(i) (ii) (iii)

(a) Iz teksta zadatka je očito da Okus ne ovisi ni o čemu (tj. nema roditelja), dok Omot i Oblik izravno ovise o Okusu.

Dakle, Omot i Oblik su međusobno zavisne - porast vjerojatnosti jedne povlači porast vjerojatnosti druge varijable - preko Okus.

(i) NE VALJA - jer nema veze između Omot i Oblik. Po lokalnoj semantici, to bi značilo da su Omot i Oblik nezavisne - v. odgovor na (c).

(ii) MOŽE reprezentirati bilo koju distribuciju jer je mreža potpuna (između svaka dva čvora ima veza u jednom smjeru). Naravno, to nije prirodna mreža (po smjeru veza) i zato ima previše veza.

(iii) MOŽE, i to je prirodna mreža za ovaj problem, s najmanjim brojem veza (veze su kauzalne).

(b) Mreža (iii) je najbolja - ima najmanje veza i veze imaju prirodan = kauzalan smjer. Dodatno, v. odgovor na (d).

(c) U mreži (i) nema veze od Omot prema Oblik. Po lokalnoj (topološkoj) semantici, čvor Omot je uvjetno nezavisan od svih svojih nesljedbenika za dane njegove roditelje.

No, Omot nema roditelja ⇒ uvjetna nezavisnost prelazi u bezuvjetnu, a Oblik nije sljedbenik od Omot (nema te veze). Dakle, Omot i Oblik su nezavisne, tj. slijedi P(Omot|Oblik) = P(Omot). Odgovor je DA - i baš zato mreža (i) NE VALJA.

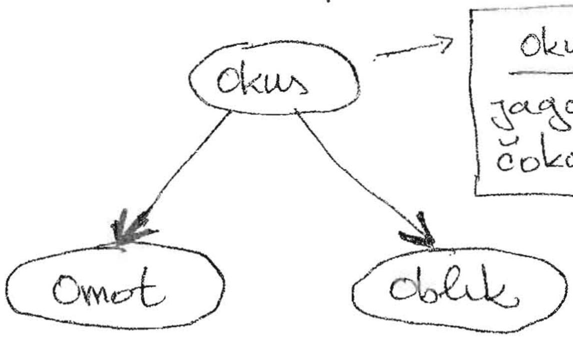
(d) Iz lokalne (topološke) semantike u (iii) slijedi:

- Za varijablu Okus → ništa, jer su joj preostale dvije varijable sljedbenici.
- Za Omot i Oblik → njih dvije su uvjetno nezavisne za dani Okus (roditelj za obje).

Točno ova nezavisnost izlazi iz teksta zadatka: kad znamo Okus, odgovarajući bomboni se nezavisno oblikuju i pakiraju (prema opisanim pravilima).

Zato mreža (iii) korektno/najbolje prikazuje problem.

(e) U najboljoj mreži (iii), pripadne tablice uvjetnih distribucija vjerojatnosti (CPT) izlaze izravno iz vrijednosti u opisu problema:



Okus	P(Okus)
jagoda	0.7
čokolada	0.3

← ne treba pamtiti, jer $\Sigma = 1$

Okus	Omot	P(Omot|Okus)
jagoda	crveni	0.8
jagoda	smeđi	0.2
čokolada	crveni	0.1
čokolada	smeđi	0.9

Okus	Oblik	P(Oblik|Okus)
jagoda	okrugli	0.8
jagoda	kockasti	0.2
čokolada	okrugli	0.1
čokolada	kockasti	0.9

Dovoljno je pamtiti po jedan red od prva i od zadnja dva reda, jer je Σ(prvi, drugi) = Σ(treći, četvrti) = 1

Napomena: Za mrežu (ii), za nalaženje odgovarajućih CPT imamo što za računati – hrpa posla, jer je mreža neprirodna!

(f) Traži se vjerojatnost P(Omot=crveni).
To ovisi samo o Okusu (ne o Obliku – možemo ga eliminirati kao varijablu).
Po Bayesovom pravilu – za distribucije vjerojatnosti

$$\mathbf{P}(Omot) = \mathbf{P}(Omot \mid Okus) \cdot \mathbf{P}(Okus)$$

pa uvrstimo Omot = crveni i marginaliziramo = zbrojimo po svim vrijednostima Okusa:

Okus=...

$$P(Omot=crveni) = P(Omot=crveni \mid Okus=jagoda) \cdot P(jagoda) + P(Omot=crveni \mid Okus=čokolada) \cdot P(čokolada)$$

$$= 0.8 \cdot 0.7 + 0.1 \cdot 0.3 = 0.56 + 0.03$$

$$= \underline{0.59}$$

Usput, to znači da je P(Omot = smeđi) = 0.41.

(g) Zadani su "dokazi" Omot = crveni i Oblik = okrugli, a traži se vjerojatnost

$$P(Okus=jagoda \mid Omot=crveni, Oblik=okrugli).$$

Možemo izračunati potpunu združenu distribuciju i onda očitati pripadni redak.
Međutim, lakše je stvar "okrenuti" po Bayesovom pravilu i onda izravno koristiti poznate podatke iz mreže

$$\mathbf{P}(Okus \mid Omot=crveni, Oblik=okrugli) = \{\text{okret po Bayesu}\}$$

$$= \alpha \cdot \mathbf{P}(Omot=crveni, Oblik=okrugli \mid Okus) \cdot \mathbf{P}(Okus)$$

$$= \{\underline{\text{uvjetna nezavisnost}} \text{ Omot i Oblik za dani Okus}\}$$

$$= \alpha \cdot \mathbf{P}(Omot=crveni \mid Okus) \cdot \mathbf{P}(Oblik=okrugli \mid Okus) \cdot \mathbf{P}(Okus)$$

Računamo distribuciju za obа okusa, tako da na kraju možemo naći normalizacijsku konstantu α da suma bude jednaka 1.

Potrebni vektor duljine 2 (za oba okusa), s tim da je poredak prvo jagoda, pa onda čokolada:

$\mathbf{P}(\text{Omot} = \text{crveni} \mid \text{Okus}) = \langle 0.8, 0.1 \rangle$ iz CPT u Omot

$\mathbf{P}(\text{Oblik} = \text{okrugli} \mid \text{Okus}) = \langle 0.8, 0.1 \rangle$ iz CPT u Oblik

$\mathbf{P}(\text{Okus}) = \langle 0.7, 0.3 \rangle$ iz CPT u Okus

Traženi produkt po točkama je

$$\mathbf{P}(\text{Okus} \mid \text{Omot} = \text{crveni}, \text{Oblik} = \text{okrugli})$$
$$= \alpha \cdot \langle 0.8 \cdot 0.8 \cdot 0.7,\ 0.1 \cdot 0.1 \cdot 0.3 \rangle$$
$$= \alpha \cdot \langle 0.448,\ 0.003 \rangle$$

$$\Rightarrow \alpha = \frac{1}{0.451} = 2.21729\,49002\ldots$$

$$\mathbf{P}(\text{Okus} \mid \text{Omot} = \text{crveni}, \text{Oblik} = \text{okrugli}) = \left\langle \frac{448}{451}, \frac{3}{451} \right\rangle$$
$$= \langle 0.99334\,81153\ldots,\ 0.00665\,18847\ldots \rangle$$

Dakle, odgovor za okus jagode je

$$\boxed{P(\text{jagoda} \mid \text{crveni}, \text{okrugli}) = 0.99334\,81153\ldots}$$

2. zad. – rješenje

Bayesova mreža za "model" jedenja sladoleda

V_1	$P(V_1)$
k	0.4
s	0.6

$$\begin{tikzcd} V_1 \arrow[r] \arrow[d] & V_2 \arrow[d] \\ S_1 & S_2 \end{tikzcd}$$

V_1	V_2	$P(V_2 \mid V_1)$	
k	k	0.5	$\Sigma = 1$
k	s	0.5	
s	k	0.3	$\Sigma = 1$
s	s	0.7	

V_i	S_i	$P(S_i \mid V_i)$	
k	d	0.2	$\Sigma = 1$
k	n	0.8	
s	d	0.9	$\Sigma = 1$
s	n	0.1	

(a) Traži se empirijska vjerojatnost $\hat{P}(V_2 = k)$ na bazi uzorka

(k, n, k, n) (k, n, k, n) (s, n, s, d) (s, d, s, d) (s, d, k, n)
(k, n, k, d) (s, d, s, d) (s, d, s, d) (s, d, k, n) (k, n, s, d)

(poredak varijabli u uzorku je (V_1, S_1, V_2, S_2)).

$$\hat{P}(V_2 = k) = \frac{\text{broj uzoraka s } V_2 = k}{\text{broj svih uzoraka}} = \frac{5 \text{ (uzorci } 1, 2, 5, 6, 9)}{10}$$

tj. $\boxed{\hat{P}(V_2 = k) = 0.5}$

Napomena: pripadna "stvarna" vjerojatnost iz mreže je

$$\begin{aligned} P(V_2 = k) &= P(V_2 = k \mid V_1 = k) \cdot P(V_1 = k) + P(V_2 = k \mid V_1 = s) \cdot P(V_1 = s) \\ &= 0.5 \cdot 0.4 + 0.3 \cdot 0.6 = 0.20 + 0.18 \\ &= 0.38 \end{aligned}$$

(b) Kad su zadani dokazi $E = e$ u mreži, <u>odbacivanje uzoraka</u> preskače = odbacuje sve uzorke koji se ne slažu s dokazima, tj. u <u>reduciranom</u> uzorku ostaju samo oni koji se slažu s dokazima. Zatim se zaključuje "empirijski" = broj povoljnih / svih na temelju reduciranog uzorka.

Za dane dokaze $S_1 = d$ i $S_2 = n$, nakon odbacivanja u reduciranom uzorku ostaju samo 5. i 9. uzorak

$$(s, d, k, n) \qquad (s, d, k, n)$$

tj. imamo samo 2 uzorka.

U oba je $V_2 = k$, tj. nema uzoraka s $V_2 = s$. Onda je

$$\boxed{\hat{\mathbf{P}}(V_2 | S_1 = d, S_2 = n) = \langle \underset{k}{\tfrac{2}{2}}, \underset{s}{\tfrac{0}{2}} \rangle = \langle 1, 0 \rangle}$$

tj. $\boxed{\hat{P}(V_2 = k \,|\, S_1 = d, S_2 = n) = 1}$

(c) Metoda težinske izglednosti (vjerodostojnosti) uzorkuje samo nedokazne varijable (uključivo i varijablu upita), a dokazi se smatraju fiksnima.

Ako su $\mathbf{Z}$ sve nedokazne varijable, onda je uzoračka vjerojatnost za svaku moguću vrijednost $\mathbf{Z} = \mathbf{z}$, uz fiksne dokaze $\mathbf{E} = \mathbf{e}$, jednaka (po pravilu za Bayesove mreže)

$$S_{WS}(\mathbf{z}, \mathbf{e}) = \prod_{i=1}^{\ell} P(z_i \,|\, parents(z_i))$$

gdje su $Z_1, \ldots, Z_\ell$ sve nedokazne varijable.

- Iz globalne semantike Bayesove mreže, odgovarajuća vjerojatnost za "združeni" događaj $(\mathbf{z}, \mathbf{e})$ (koji pokriva sve varijable u problemu je

$$P(\mathbf{z}, \mathbf{e}) = \prod_{j=1}^{n} P(x_j \,|\, parents(X_i))$$

gdje je $\{X_1, \ldots, X_n\} = \{Z_1, \ldots, Z_\ell\} \cup \{E_1, \ldots, E_m\}$, a $E_1, \ldots, E_m$ su dokazne varijable.

Da bismo iz $S_{WS}(\mathbf{z}, \mathbf{e})$ dobili konzistentnu procjenu s pravom distribucijom, svaki uzorak $(\mathbf{z}, \mathbf{e})$ mora dobiti faktor = težinu

$$\boxed{w(\mathbf{z}, \mathbf{e}) = \prod_{k=1}^{m} P(e_k \,|\, parents(E_k))}$$

prema odgovarajućim dokazima u mreži.

Onda je

$$S_{ws}(z,e) \cdot w(z,e) = P(z,e).$$

Umjesto "broj povoljnih" i "broj svih", za procjenu vjerojatnosti koristimo zbroj odgovarajućih težina

Ako je X upitna varijabla, a x_p neka od njezinih vrijednosti onda je procjena vjerojatnosti $P(X=x_p|e)$ dana ovako:

$$\boxed{P(X=x_p|e) = \frac{\text{zbroj težina uzoraka u kojima je } X=x_p}{\text{zbroj težina svih uzoraka}}}.$$

(d) Fiksne vrijednosti dokaznih varijabli S_1, S_2 su

$$S_1 = d,\ S_2 = n.$$

(Svih 6 uzoraka ima te fiksne vrijednosti, variraju samo vrijednosti za V_1, V_2).

Prvo treba odrediti težine svih 6 uzoraka. Težina w pojedinog uzorka

$$(V_1=v_1,\ \underbrace{S_1=d}_{\text{dokaz}},\ V_2=v_2,\ \underbrace{S_2=n}_{\text{dokaz}})$$

računa se iz konkretnih vrijednosti v_1 i v_2

$$w = P(S_1=d \mid V_1=v_1) \cdot P(S_2=n \mid V_2=v_2)$$

Tablica težina za zadanih 6 uzoraka je

(v_1, d, v_2, n)	$P(d\|v_1)$	$P(n,v_2)$	$w=P(d\|v_1)\cdot P(n\|v_2)$
(s, d, k, n)	0.9	0.8	0.72
(k, d, k, n)	0.2	0.8	0.16
(s, d, k, n)	0.9	0.8	0.72
(s, d, s, n)	0.9	0.1	0.09
(s, d, s, n)	0.9	0.1	0.09
(k, d, s, n)	0.2	0.1	0.02

U svih 6 uzoraka, prva tri imaju $V_2=k$ (tzv. povoljni) pa je procjena vjerojatnosti

$$\hat{P}(V_2=k \mid S_1=d, S_2=n) = \frac{0.72+0.16+0.72}{0.72+0.16+0.72+0.09+0.09+0.02} = \frac{1.6}{1.8}$$

$$= \frac{8}{9} = 0.88889$$

3. zad. - rješenje

(a) Entropija slučajne varijable X s diskretnim vrijednostima x_i koje imaju vjerojatnosti $p_n = P(X=x_n)$ je

$$H(X) = -\sum_n p_n \log_2 p_n$$

(broj bitova potreban za kodiranje "odgovora").

U zadanom primjeru "ciljna" varijabla je Klasa s dvije moguće vrijednosti + i −.

Cilj kod konstrukcije stabla odlučivanja je što brže smanjiti njezinu entropiju, na osnovu odgovarajućeg uzorka u danom čvoru, koji još nije potpuno "klasificiran".

Skup (preostalih) podataka D u danom čvoru ima polaznu entropiju $H(D)$ – koju računamo empirijski, na temelju tog skupa podataka.

Za sve (preostale) atribute, koji još nisu iskorišteni treba izračunati pripadnu uvjetnu entropiju, kao srednju vrijednost (očekivanje na bazi uzorka), po svim mogućim vrijednostima tog atributa.

Konkretno, dani atribut A_i s k vrijednosti $a_{i1}, \dots, a_{ik}$ particionira sve primjere D u k disjunktnih podskupova

$$D_1, \dots, D_k$$

D_j = svi primjeri u kojima je $A_i = a_{ij}$, $j = 1, \dots, k$.

Entropija ciljne varijable X na skupu D_j je

$$\underbrace{H(X \mid A_i = a_{ij})}_{\text{entropija}(D_j)} = -\sum_n P(X=x_n \mid A_i = a_{ij}) \cdot \log_2 P(X=x_n \mid A_i = a_{ij})$$

a pripadno očekivanje je

$$\underbrace{H(X \mid A_i)}_{\text{entropija}_{A_i}(D)} = \sum_{j=1}^{k} P(A_i = a_{ij}) \cdot H(X \mid A_i = a_{ij})$$

$\text{entropija}_{A_i}(D)$ = (ako su vrijednosti jednako vjerojatne)

$$= \sum_{j=1}^{k} \frac{|D_j|}{|D|} \cdot \underbrace{H(X \mid A_i = a_{ij})}_{\text{entropija}(D_j)}$$

Dobitak informacije odabirom atributa A_i za grananje u tom čvoru stabla je

$$I(D, A_i) = \text{dobitak}(D, A_i) = \cancel{\text{kol}(X)}$$
$$= H(X) - H(X|A_i)$$
$$= \text{entropija}(D) - \text{entropija}_{A_i}(D)$$

Za grananje biramo onaj atribut A_i (među preostalima) koji daje <u>maksimalni</u> <u>dobitak informacije</u>, tj. onaj koji ima <u>najmanju</u> entropiju ($\text{entropija}_{A_i}(D)$).

- U našem skupu primjera, sve vrijednosti atributa su jednako vjerojatne

1. Na početku skup D sadrži <u>svih</u> 6 primjera.
Imamo 3 primjera klase + i 3 primjera klase −, pa za ciljnu varijablu Klasa vrijedi
$$p_+ = p_- = \frac{1}{2}.$$
Entropija polaznog skupa D je
$$\boxed{\text{entropija}(D) = H(\text{Klasa}) = -(\underset{\frac{1}{2}}{p_+}\underset{-1}{\log_2 p_+} + \underset{\frac{1}{2}}{p_-}\underset{-1}{\log_2 p_-}) = 1}$$
tj. nema informacije = maksimalni nered.

2. Tražimo <u>najbolji</u> atribut za korijen (prvi čvor) stabla.

2a. Izbor A_i = Boja. Imamo 3 vrijednosti: $\underset{1}{\text{crvena}}$, $\underset{2}{\text{plava}}$, $\underset{3}{\text{zelena}}$

D_1 = (Boja = crvena) – ima 3 primjera {1,3,5}
D_2 = (Boja = plava) – ima 1 primjer {2}
D_3 = (Boja = zelena) – ima 2 primjera {4,6}

Treba naći entropije svakog od ovih skupova prema Klasi.

$$\text{entropija}(D_1) = \{2+, 1-\} = -\left(\frac{2}{3}\log_2\frac{2}{3} + \frac{1}{3}\log_2\frac{1}{3}\right)$$
$$= -\left(\frac{2}{3}\underbrace{\log_2 2}_{1} - \frac{2}{3}\log_2 3 + \frac{1}{3}\underbrace{\log_2 1}_{0} - \frac{1}{3}\log_2 3\right)$$
$$= \log_2 3 - \frac{2}{3} = 0.91829\ 58341$$

$\text{entropija}(D_2) = \{1+\} = -1 \cdot \log_2 1 = 0$

$\text{entropija}(D_3) = \{2-\} = -0 \cdot \log_2 0 - 1 \cdot \log_2 1 = 0$

Pripadna srednja vrijednost

$$\text{entropija}_{\text{Boja}}(D) = \frac{3}{6} \cdot \text{entropija}(D_1) + \frac{1}{6}\,\text{entropija}(D_2) + \frac{2}{6} \cdot \text{entropija}(D_3)$$
$$= \frac{1}{2} \cdot \left(\log_2 3 - \frac{2}{3}\right)$$
$$= \underline{0.45914\ 79170}$$

2b. Izbor A_i = Oblik. Imamo 2 vrijednosti: kvadrat$_1$, krug$_2$

D_1 = (Oblik = kvadrat) – ima 4 primjera $\{1,2,4,6\}$

pripadna entropija

$\text{entropija}(D_1) = \{2+, 2-\} = -\frac{2}{4}\log_2\frac{2}{4} - \frac{2}{4}\log_2\frac{2}{4} = 1$

D_2 = (Oblik = krug) – ima 2 primjera $\{3,5\}$

$\text{entropija}(D_2) = \{1+, 1-\} = 1$

Pripadna srednja vrijednost

$$\text{entropija}_{\text{oblik}}(D) = \frac{4}{6} \cdot \text{entropija}(D_1) + \frac{2}{6} \cdot \text{entropija}(D_2)$$
$$\underline{= 1}$$

2c. Izbor A_i = Veličina. Imamo 2 vrijednosti: velika$_1$, mala$_2$

D_1 = (Veličina = velika) – ima 4 primjera $\{1,2,5,6\}$

$$\text{entropija}(D_1) = \{3+, 1-\} = -\frac{3}{4}\log_2\frac{3}{4} - \frac{1}{4}\log_2\frac{1}{4}$$
$$= -\frac{3}{4}\log_2 3 + \log_2 4 = 2 - \frac{3}{4}\log_2 3$$
$$= 0.81127\ 81245$$

D_2 = (Veličina = mala) – ima 2 primjera $\{3,4\}$

$\text{entropija}(D_2) = \{0+, 2-\} = 0$

Srednja vrijednost

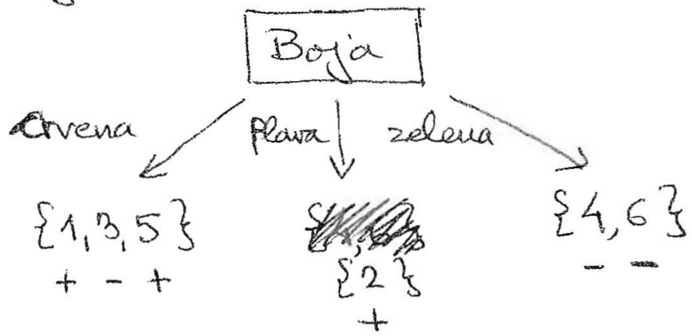
$$\text{entropija}_{\text{Veličina}}(D) = \frac{4}{6} \cdot \text{entropija}(D_1) + \frac{2}{6} \cdot \text{entropija}(D_2) = \frac{4}{3} - \frac{1}{2}\log_2 3$$
$$= \underline{0.54085\ 20830}$$

Tablica entropija i dobitaka informacija

Atribut A_i	entropija$_{A_i}(D)$	dobitak (D, A_i)
Boja	0.45914 79170	0.54085 20830
Oblik	1	0
Veličina	0.54085 20830	0.45914 79170

Najmanju entropiju / najveći dobitak informacija daje Boja, pa nju treba uzeti kao korijen stabla.

Particijom svih primjera prema vrijednosti boje dobivamo sljedeće stablo



Čvorovi "djece" za plavu i zelenu sadrže samo primjere iste klase ⟹ postaju listovi završnog stabla odlučivanja s klasama +, odn. −.

3. U čvoru za crvenu boju imamo 3 primjera, $D = \{1,3,5\}$ od čega su 2 klase + (1,5) i 1 klase − (3).

U tom čvoru nastavljamo konstrukciju (rekurzivno), sa skupom

$$D = \{1, 3, 5\},$$

na kojem je

$$p_+ = \frac{2}{3}, \quad p_- = \frac{1}{3}$$

pa je njegova "polazna" entropija

$$\boxed{\begin{aligned} \text{entropija}(D) &= -\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3} \qquad (\text{v. ranije } D_1) \\ &= \log_2 3 - \frac{2}{3} = 0.91829\ 58341 \end{aligned}}$$

Preostali atributi za grananje su Oblik i Veličina.

4. Traži se bolji od ta dva atributa - kao novi čvor grananja u stablu.

4a. Izbor A_i = Oblik - imamo 2 vrijednosti: kvadrat$_1$, krug$_2$.

D_1 = (Oblik=kvadrat) - ima 1 primjer {1}

entropija(D_1) = {1+, 0-} = 0

D_2 = (Oblik = krug) - ima 2 primjera {3,5}

entropija (D_2) = {1+, 1-} = 1

Srednja vrijednost

$$\text{entropija}_{Oblik}(D) = \frac{1}{3}\cdot \text{entropija}(D_1) + \frac{2}{3}\cdot \text{entropija}(D_2)$$

$$= \frac{2}{3} = 0.66666\ 66667$$

4b. Izbor A_i = Veličina - imamo 2 vrijednosti: velika$_1$, mala$_2$

D_1 = (Veličina = velika) - ima 2 primjera {1,5}

entropija(D_1) = {2+, 0-} = 0

D_2 = (Veličina = mala) - ima 1 primjer {3}

entropija (D_2) = {1-, 0+} = 0

Srednja vrijednost

$$\text{entropija}_{Veličina}(D) = \frac{2}{3}\cdot \text{entropija}(D_1) + \frac{1}{3}\,\text{entropija}(D_2)$$

$$= 0$$

Tablica entropija i dobitaka informacija (nepotrebna, sve je očito)

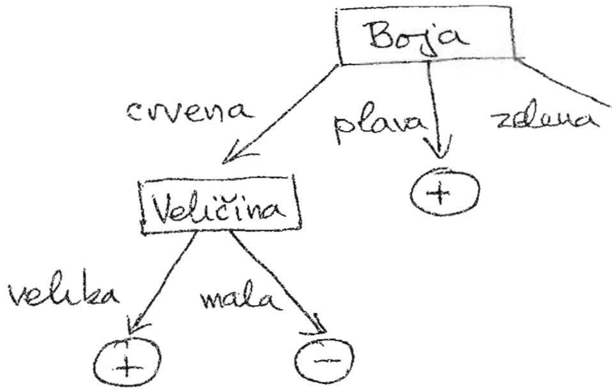
Atribut A_i	entropija A_i(D)	dobitak (D, Ai)
Oblik	0.66666 66667	0.25162 91674
Veličina	0	0.91829 58341

⇒ Treba uzeti Veličinu kao "dijete" za crvenu boju

Oba dobivena djeteta imaju samo primjere iste klase, pa postaju listovi završnog ~~st~~ stabla, a pripadne klase su + za "velika", - za "mala"

Svi primjeri su klasificirani ⇒ GOTOVO.

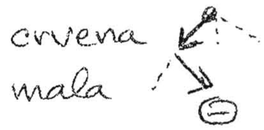
Pripadno ~~s~~tablo ima oblik



Uočiti da Oblik nema utjecaj na klasifikaciju

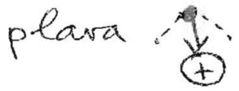
Klasifikacija novih primjera

1. (crvena, kvadrat, mala)



$\Rightarrow$ Klasa = −

2. (plava, krug, mala)



$\Rightarrow$ Klasa = +

3-6A

Grafička ilustracija konstrukcije stabla odlučivanja u pojedinim čvorovima:

1. korak: čvor = korijen, pripadni skup primjera $D=\{1,..,6\}=\{3+,3-\}$

Klasifikacija prema vrijednostima preostalih atributa (a to su sva tri) daje

Boja: crvena, plava, zelena

$D_1=\{1,3,5\} = \{2+,1-\}$ $\quad D_2=\{2\} = \{1+,0-\} \Rightarrow H_2=0$ $\quad D_3=\{4,6\} = \{0+,2-\} \Rightarrow H_3=0$

Oblik: kvadrat, krug

$D_1=\{1,2,4,6\} = \{2+,2-\} \Rightarrow H_1=1$ $\quad D_2=\{3,5\} = \{1+,1-\} \Rightarrow H_2=1$

$H_{oblik}=1$

Veličina: velika, mala

$D_1=\{1,2,5,6\} = \{3+,1-\}$ $\quad D_2=\{3,4\} = \{0+,2-\} \Rightarrow H_2=0$

Dakle, Oblik ne daje nikakvu informaciju.

2. korak – već imamo:

Boja: crvena, plava, zelena

crvena: $D=\{1,3,5\} = \{2+,1-\}$; plava: +; zelena: −

čistoi

u čvoru "crvena" imamo $D=\{1,3,5\}=\{2+,1-\}$. Klasifikacija po preostalim atributima (Oblik, Veličina) daje

2+, 1−

Oblik: kvadrat, krug

$D_1=\{1\} = \{1+,0-\} \Rightarrow H_1=\emptyset$ $\quad D_2=\{3,5\} = \{1+,1-\} \Rightarrow H_2=1$

Veličina: velika, mala

$D_1=\{1,5\} = \{2+,0-\} \Rightarrow H_1=0$ $\quad D_2=\{3\} = \{0+,1-\} \Rightarrow H_2=0$

$H_{veličina}=0$

(b) Zadan je skup ili niz T od N primjera za učenje

$$x^{(i)}, \quad i = 1, \dots, N.$$

Svaki primjer $x^{(i)}$ je "vektor" koji se sastoji od n vrijednosti tzv. atributa $x_{i,j}$, za $j = 1, \dots, n$ i od jedne dodatne vrijednosti $c^{(i)}$ – tzv. klase

$$x^{(i)} = (x_{i,1}, x_{i,2}, \dots, x_{i,n}, c^{(i)}).$$

Svaki pojedini atribut $x_{i,j}$ možemo interpretirati kao vrijednost odgovarajuće slučajne varijable A_j (za taj atribut na j-tom mjestu). Klasu $c^{(i)}$, također, možemo interpretirati kao vrijednost odgovarajuće slučajne varijable C, s tim da pretpostavljamo da klasa C ovisi o atributima $A_1, \dots, A_n$.

Bilo koji zadani "novi" primjer x ima poznate <u>samo</u> vrijednosti atributa, a pripadnu vrijednost klase $c = c(x)$ treba odrediti na temelju skupa podataka za učenje T,

$$x = (x_1, \dots, x_n), \quad c = c(x) = ?$$

Za svaki novi primjer x, metoda k najbližih susjeda (k-NN) se sastoji iz 2 dijela:

(1) Pronaći indekse $i_1, \dots, i_k$ za k <u>najbližih</u> susjeda novom primjeru x, među svim primjerima iz T, pri čemu se udaljenost $d(x, x^{(i)})$, $x^{(i)} \in T$ mjeri na sljedeći način

$$d(x, x^{(i)}) = \sqrt{\sum_{j=1}^{n} d_j^2(x_j, x_{i,j})}$$

– Euklidska udalj. na kartezijevom produktu metričkih prostora

gdje je d_j udaljenost na prostoru atributa za varijablu A_j. U našem slučaju, svi atributi su "kategoričkog" tipa, pa koristimo trivijalnu diskretnu 0-1 metriku

$$d_j(x_j, x_{i,j}) = \begin{cases} 0, & \text{za } x_j = x_{i,j} \\ 1, & \text{za } x_j \neq x_{i,j} \end{cases}, \quad j = 1, \dots, n$$

(2) Na osnovu k najbližih susjeda iz T, određujemo klasu novog primjera, prema klasama najbližih susjeda. Kod težinskog određivanja klase, koristimo i težine za klasu svakog susjeda:

$$c(x) = \arg\max_{c} \sqrt{\sum_{\ell=1}^{k} w_{i_\ell} \cdot \delta(c, c^{(i_\ell)})}$$

Gornji maksimum ide po svim mogućim vrijednostima klase $C = c$.

Funkcija δ je karakteristična funkcija na prostoru vrijednosti za klase

$$\delta(c_p, c_q) = \begin{cases} 1, & \text{za } c_p = c_q \\ 0, & \text{za } c_p \neq c_q \end{cases}$$

(obratno od diskretne 0-1 metrike).

Prema zahtjevu iz zadatka, pripadne težine w_{i_ℓ} su obrnuto proporcionalne kvadratima udaljenosti između novog i susjednog primjera iz T

$$w_{i_\ell} = \frac{1}{(d(x, x^{(i_\ell)}))^2}.$$

Zato se kod nalaženja najbližih susjeda isplati računati

$$d^2(x, x^{(i)})$$

bez korijena (najbliži su isti!) = broj različitih atributa

– Za novi primjer $x = $ (crvena, kvadrat, mala) tražimo 3 najbliža susjeda

i	atributi od $x^{(i)}$			$d^2(x, x^{(i)})$ = broj različitih atributa		
1	crvena	kvadrat	velika	1	*	3 najbliža susjeda
2	plava	kvadrat	velika	2		
3	crvena	krug	mala	1	*	
4	zelena	kvadrat	mala	1	*	
5	crvena	krug	velika	2		
6	zelena	kvadrat	velika	2		

Najbliža 3 susjeda su primjeri 1, 3, 4.

- Sva tri primjera najbližih susjeda imaju ISTU udaljenost (i kvadrat)

$$d^2(x, x^{(i)}) = 1 \quad , i = 1,3,4$$

Dakle, težine w_1, w_3, w_4 sva tri najbliža susjeda su jednake 1

$$w_1 = w_3 = w_4 = 1$$

pa možemo uzeti kao da nema težina, za određivanje klase.

~~-~~,

- Primjeri 3,4 imaju klasu "–", a primjer 1 ima klasu "+" pa odmah vidimo da "–" dominira među 3 susjeda, tj.

$$\boxed{c(\text{crvena}, \text{kvadrat}, \text{mala}) = - .}$$

- Kompliciranije = gledamo po svim mjednostima za c :

$$\sqrt{\sum_{i=1,3,4} \delta(c=+, c^{(i)})} = \sqrt{\underset{1}{\delta(+,+)} + \underset{0}{\delta(+,-)} + \underset{0}{\delta(+,-)}} = \sqrt{1}$$

$$\sqrt{\sum_{i=1,3,4} \delta(c=-, c^{(i)})} = \sqrt{\underset{0}{\delta(-,+)} + \underset{1}{\delta(-,-)} + \underset{1}{\delta(-,-)}} = \sqrt{2}$$

$$\Rightarrow \boxed{c(x) = -}$$

(c) Uzmimo iste pretpostavke kao u (b), tj. da imamo poznat skup T primjera za učenje, zajedno s pripadnim klasama

$$x^{(i)} = (x_{i,1}, \dots, x_{i,n}, c^{(i)}), \quad i = 1, \dots, N$$

i treba klasificirati novi primjer

$$x = (x_1, \dots, x_n), \quad c = c(x) = ?$$

Uzmimo da varijabla klase C ima m mogućih vrijednosti

$$c_\ell, \quad \ell = 1, \dots, m.$$

Pripadnu (a priori) distribuciju vjerojatnosti $\mathbf{P}(C)$, naravno, NE znamo, ali ju možemo "naučiti" iz zadanih podataka T za učenje.

Ideja za klasifikaciju novog primjera x je primjena Bayesovog pravila

$$P(C=c_\ell \mid x) = \frac{P(x \mid C=c_\ell) \cdot P(C=c_\ell)}{P(x)} = \alpha \cdot P(x \mid C=c_\ell) \cdot P(C=c_\ell)$$

gdje je α normalizacijska konstanta za nepoznatu vjerojatnost $P(x)$, tj. $\alpha = 1/P(x)$. Na razini distribucije vjerojatnosti

$$\mathbf{P}(C \mid x) = \alpha \cdot \mathbf{P}(x \mid C) \cdot \mathbf{P}(C)$$

s tim da se traži maksimalna a posteriori vrijednost c_{MAP} u ovoj distribuciji, a $\mathbf{P}(x \mid C)$ i $\mathbf{P}(C)$ se procjenjuju na bazi skupa za učenje T, tj.

$$c(x) = c_{MAP} = \arg\max_{c} P(x \mid c) \cdot P(c)$$

po svim mogućim vrijednostima $c = c_\ell$, $\ell = 1, \dots, m$.

– Bayesova klasifikacija procjenjuje $\mathbf{P}(c)$ i $\mathbf{P}(x \mid C)$ na osnovu skupa za učenje T.

$$\mathbf{P}(C) = \text{lagano}: \quad \boxed{\frac{\text{broj primjera s klasom } C = c_\ell}{\text{broj svih primjera}} = P(c_\ell)}$$

Međutim, slična procjena za $P(x_1, \dots, x_n \mid c_\ell)$ je gotovo ~~nemoguću~~ nemoguća, jer bismo morali imati ogroman broj primjera za učenje. Ukupan broj mogućih događaja koje treba brojati (= kardinalitet prostora mogućih događaja) je

$$N_{BC} = \underbrace{m}_{\text{za } C} \cdot |A_1| \cdot \ldots \cdot |A_n|, \quad |A_j| = \text{broj mogućih vrijednosti atributa } A_j$$

Kod <u>naivne</u> Bayesove klasifikacije pretpostavljamo da su atributne varijable $A_1, -, A_n$ <u>uvjetno nezavisne</u>, uz danu klasu C tj.

$$\mathbb{P}(A_1,-,A_n|C) = \prod_{j=1}^{n} \mathbb{P}(A_j|C).$$

Za zadani primjer atributa $x=(x_1,-,x_n)$ to znači da pretpostavljamo da je

$$\boxed{P(x_1,-,x_n|c_e) = \prod_{j=1}^{n} P(x_j|c_e), \quad \forall e=1,-,m}$$

pa nam trebaju samo uvjetne vrijednosti <u>pojedinačnih</u> atributa x_j, za danu klasu c_e.

Bitna prednost ovog modela je što je ukupan broj mogućih događaja za model

$$N_{NBC} = m\cdot(|A_1|+\ldots+|A_n|)$$

tj. trebamo mnogo <u>manji</u> skup podataka za učenje da bismo (na uzorku) procijenili

$$\boxed{P(x_j|c_e) = \frac{\text{broj primjera s klasom } c_e \text{ u kojima je } A_j=x_j}{\text{broj primjera s klasom } c_e}}$$

- <u>Algoritam</u>: Na bazi skupa T računamo procjene za $P(c_e)$. Za zadane atribute x_j računamo procjene za $P(x_j|c_e)$, iskoristimo naivno Bayesovo pravilo za računanje

$$P(x_1,-,x_n|c_e) = \prod_{j=1}^{n} P(x_j|c_e)$$

i zatim tražimo vrijednost c_{MAP} za koju se dostiže maksimum

$$P(x_1,-,x_n|c_e)\cdot P(c_e), \quad e=1,-,m.$$

U našem primjeru, moguće klase su $+$ i $-$, a na 6 primjera iz skupa T dobivamo

$$\boxed{P(C=+) = P(C=-) = \frac{1}{2}}$$

(imamo po 3 primjera svake klase). Onda je još lakše, treba naći

$$c_{MAP} = c_{ML} = \max_c \arg P(x|c) = \max_c \arg \prod_{j=1}^{n} P(x_j|c).$$

Novi primjer je $x=$ (crvena, kvadrat, mala)

Treba naći uvjetne vjerojatnosti svakog od ovih atributa, za sve moguće klase.

Za $C=+$: imamo $N_+ = 3$ primjera.

$$P(\text{crvena} \mid +) = (\text{imamo } 2 \text{ primjera}) = \frac{2}{3}$$
$$P(\text{kvadrat} \mid +) = (\quad - \text{"} - \quad) = \frac{2}{3}$$
$$P(\text{mala} \mid +) = (\text{imamo } \emptyset \text{ primjera}) = 0$$

$$\Rightarrow \underline{P(x \mid +) = \frac{2}{3} \cdot \frac{2}{3} \cdot 0 = 0}$$

Za $C=-$: imamo $N_- = 3$ primjera.

$$P(\text{crvena} \mid -) = (\text{imamo } 1 \text{ primjer}) = \frac{1}{3}$$
$$P(\text{kvadrat} \mid -) = (\text{imamo } 2 \text{ primjera}) = \frac{2}{3}$$
$$P(\text{mala} \mid -) = (\text{imamo } 2 \text{ primjera}) = \frac{2}{3}$$

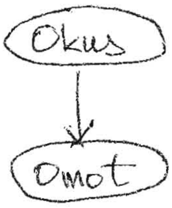
$$\Rightarrow \underline{P(x \mid -) = \frac{1}{3} \cdot \frac{2}{3} \cdot \frac{2}{3} = \frac{4}{27}}$$

Za $C=-$ dobivamo veću vrijednost $\Rightarrow$

$$\boxed{c(x) = -}$$

4. zad. - rješenje

U ovom problemu, svi bomboni imaju isti (okrugli) oblik, pa ostaju samo dvije slučajne varijable - to su Okus i Omot. Pripadna trivijalna Bayesova mreža je



Okus	**P**(Okus)
jagoda	0.7
čokolada	0.3

Okus	Omot	**P**(Omot | Okus)	
jagoda	crveni	0.8	$\Sigma=1$
jagoda	smeđi	0.2	
čokolada	crveni	0.1	$\Sigma=1$
čokolada	smeđi	0.9	

Podatke iz ove mreže koristimo za računanje potrebnih podataka za problem učenja u zadatku.

Stvarno zaključivanje u zadatku nema izravne veze s ovom mrežom - jer još imamo i hipoteze h_1, h_2, te skup podataka za učenje.

Uočiti: hipoteze su formulirane preko distribucije vjerojatnosti omota (a ne okusa).

Poznati podaci o sadržaju vreća u hipotezama mogu se zapisati u obliku tablice uvjetne distribucije **P**(Omot | H) gdje H ima vrijednosti h_1 i h_2

H = hipoteza	Omot	**P**(Omot | H)
h_1	crveni	0.93
h_1	smeđi	0.07
h_2	crveni	0.25
h_2	smeđi	0.75

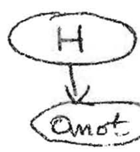
Skraćene oznake za kasnije su:

$$P(\text{crveni} \mid h_1) = p_1 = 0.93$$
$$P(\text{smeđi} \mid h_1) = 1 - p_1 = 0.07$$
$$P(\text{crveni} \mid h_2) = p_2 = 0.25$$
$$P(\text{smeđi} \mid h_2) = 1 - p_2 = 0.75$$

Znamo i a priori distribuciju vjerojatnosti za H - obje hipoteze su jednako vjerojatne

H	**P**(H)
h_1	0.5
h_2	0.5

Digresija: ovi podaci odgovarali bi Bayesovoj mreži



samo fali učenje!

Podaci za učenje zadani su vrijednostima omota:

d_1 : Omot = crveni , d_2 : Omot = smeđi

Skraćeni zapis

d_1 = crveni , d_2 = smeđi .

Cijeli skup podataka za učenje označavamo s $d = (d_1, d_2)$.

(a) A posteriori distribuciju vjerojatnosti hipoteza uz dane podatke za učenje računamo prema Bayesovom pravilu

– u terminima distribucija vjerojatnosti

$$\mathbf{P}(H|d) = \frac{\mathbf{P}(d|H) \cdot \mathbf{P}(H)}{\mathbf{P}(d)} = \alpha \cdot \mathbf{P}(d|H) \cdot \mathbf{P}(H)$$

gdje je α normalizacijski faktor – da ne moramo računati $P(d)$ – što je teško/nepoznato;

– u terminima vjerojatnosti pojedine hipoteze h_i

$$P(h_i|d) = \frac{P(d|h_i) \cdot P(h_i)}{P(d)} = \alpha \cdot P(d|h_i) \cdot P(h_i)$$

Hipoteza maksimalne a posteriori vjerojatnosti h_{MAP} je

$$h_{MAP} = \arg\max_{h_i} P(h_i|d) = \arg\max_{h_i} P(d|h_i) \cdot P(h_i)$$

– A priori vjerojatnosti hipoteza $P(h_i)$ znamo, a vjerojatnosti $P(d|h_i)$ računamo iz dobivenog uzorka za učenje, koristeći pretpostavku da su izvlačenja "omota" nezavisna i jednako distribuirana:

$$P(d|h_i) = \prod_{j=1}^{2} P(d_j|h_i).$$

U našem primjeru, vjerojatnost izvlačenja crvenog omota iz vreće "h_i" je p_i, a za smeđi omot je $1-p_i$.

Onda je:

$$\boxed{P(d_1, d_2|h_i) = p_i \cdot (1-p_i)} \quad i=1,2$$

Sad iskoristimo poznate vrijednosti za p_i, pa je

$$P(d|h_1) = 0.93 \cdot 0.07 = 0.0651$$
$$P(d|h_2) = 0.25 \cdot 0.75 = 0.1875$$

Obzirom na to da je $P(h_1) = P(h_2)$, odmah vidimo da hipoteza h_2 ima veću a posteriori vjerojatnost

$$\boxed{h_{MAP} = h_2}$$

Usput, zbog jednake a priori vjerojatnosti hipoteza, slijedi i da je hipoteza maksimalne izglednosti h_{ML} jednaka h_{MAP}

$$\boxed{h_{MAP} = h_{ML} = h_2}$$

A posteriori distribuciju $\mathbb{P}(H|d)$ dobivamo normalizacijom

$$\mathbb{P}(H|d) = \alpha \cdot \mathbb{P}(d|H) \cdot \mathbb{P}(H) = \{ P(h_1) = P(h_2) = \tfrac{1}{2} \}$$
$$= \alpha' \cdot \mathbb{P}(d|H)$$
$$= \alpha' \cdot \langle 0.0651, 0.1875 \rangle = \{ \alpha = \frac{1}{0.2526} = 3.9588281869 \}$$
$$= \langle 0.2577197150, 0.7422802850 \rangle$$

(b) Na temelju podataka za učenje, treba "predvidjeti" distribuciju vjerojatnosti "upitne" varijable X, uz dane podatke d.

– Optimalno Bayesovo predviđanje (OBP):

$$\mathbb{P}(X|d) = \sum_i \mathbb{P}(X|h_i) \cdot P(h_i|d)$$

(marginalizacija po svim vrijednostima hipoteza, h_i i još pretpostavljamo uvjetnu nezavisnost X i podataka d uz danu hipotezu h_i, tj. $\mathbb{P}(X|d, h_i) = \mathbb{P}(X|d)$).

– Maksimalno a posteriori (MAP) predviđanje

$$\mathbb{P}(X|d) \approx \mathbb{P}(X|h_{MAP})$$

(jer h_{MAP} maksimizira faktor $P(h_i|d)$, pa aproksimacija uzima samo pripadni član, kao da mu je težina $P(h_{MAP}|d) = 1$)

U našem primjeru, traži se predviđanje za X=Okus, i dovoljno je naći vjerojatnost za X=Okus=jagoda (ne treba cijela distribucija vjerojatnosti – dva člana, već samo prvi – za jagodu).

<u>Uočiti</u>: predviđamo okus na temelju omota (njihov omjer ili distribucija vjerojatnosti je hipoteza).
To je <u>obratno</u> od <u>polazne</u> Bayesove mreže i zato trebamo dodatne podatke iz nje.

Traži se $P(X=\text{jagoda} \mid d)$, a za OBP nam treba (skraćeni zapis)

$$P(\text{jagoda} \mid h_i) \quad , i=1,2.$$

(već znamo da je $h_{MAP}=h_2$ pa tamo trebamo samo $P(\text{jagoda} \mid h_2)$)

Ove vjerojatnosti računamo "marginalizacijom" po <u>omotu</u>, jer <u>omot veže okus i hipoteze</u>:

$$P(\text{jagoda} \mid h_i) = \sum_{\substack{\text{omot} \\ =\text{crveni} \\ \text{smeđi}}} \underbrace{P(\text{jagoda} \mid \text{omot})}_{\substack{\downarrow \\ \text{treba izračunati} \\ \text{iz polazne Bayesove mreže}}} \cdot \underbrace{P(\text{omot} \mid h_i)}_{\text{znamo}}$$

- Za $P(\text{jagoda} \mid \text{omot})$ koristimo Bayesovo pravilo i vjerojatnosti iz 1. zadatka (CPT na početku ovog zadatka):

$$P(\text{jagoda} \mid \text{omot}) = \frac{P(\text{omot} \mid \text{jagoda}) \cdot P(\text{jagoda})}{P(\text{omot})}$$

za omot = crveni, smeđi.
Vrijednosti za brojnik imamo, a nazivnik <u>treba</u> izračunati marginalizacijom po okusima (kao u 1. zadatku), ne smijemo samo normalizirati.

$$P(\text{Omot}=\text{crveni}) = P(\text{crveni} \mid \text{jagoda}) \cdot P(\text{jagoda}) + P(\text{crveni} \mid \text{čokolada}) \cdot P(\text{čokolada})$$
$$= 0.8 \cdot 0.7 + 0.1 \cdot 0.3 = \underline{\underline{0.59}}$$

$$P(\text{Omot}=\text{smeđi}) = \text{(može kao gore, ali lakše je ...)}$$
$$= 1 - P(\text{Omot}=\text{crveni}) = \underline{\underline{0.41}}$$

$$\mathbf{P}(\text{Omot}) = \langle 0.59, 0.41 \rangle$$

Vjerojatnosti okusa jagode za dani omot su:

$$P(\text{jagoda} \mid \text{crveni}) = \frac{P(\text{crveni} \mid \text{jagoda}) \cdot P(\text{jagoda})}{P(\text{crveni})}$$

$$= \frac{0.8 \cdot 0.7}{0.59} = \frac{0.56}{0.59} = \underline{0.94915\,25424}$$

$$P(\text{jagoda} \mid \text{smeđi}) = \frac{P(\text{smeđi} \mid \text{jagoda}) \cdot P(\text{jagoda})}{P(\text{smeđi})}$$

$$= \frac{0.2 \cdot 0.7}{0.41} = \frac{0.14}{0.41} = \underline{0.34146\,34146}$$

Vjerojatnosti okusa jagode za dane hipoteze (vreće) su:

$$P(\text{jagoda} \mid h_1) = p_1 \cdot P(\text{jagoda} \mid \text{crveni}) + (1-p_1) \cdot P(\text{jagoda} \mid \text{smeđi})$$

$$= 0.93 \cdot \frac{0.56}{0.59} + 0.07 \cdot \frac{0.14}{0.41}$$

$$= \underline{\underline{0.90661\,43034}}$$

$$P(\text{jagoda} \mid h_2) = p_2 \cdot P(\text{jagoda} \mid \text{crveni}) + (1-p_2) \cdot P(\text{jagoda} \mid \text{smeđi})$$

$$= 0.25 \cdot \frac{0.56}{0.59} + 0.75 \cdot \frac{0.14}{0.41}$$

$$= \underline{\underline{0.49338\,56966}}$$

- <u>BOP</u>!

$$P(\text{jagoda} \mid d) = P(\text{jagoda} \mid h_1) \cdot P(h_1 \mid d) + P(\text{jagoda} \mid h_2) \cdot P(h_2 \mid d)$$

$$= 0.90661\ldots \cdot 0.25771\ldots + 0.49338\ldots \cdot 0.74228\ldots$$

$$= 0.23365\,23799 + 0.36623\,04755$$

$$= \underline{\underline{0.59988\,28554}}$$

tj. vjerojatniji okus trećeg bombona je <u>jagoda</u>.

- <u>MAP</u>

$$P(\text{jagoda} \mid d) \approx P(\text{jagoda} \mid h_2) = 0.49338\,56966$$

tj. po MAP predviđanju, vjerojatniji okus trećeg bombona je <u>čokolada</u>.