Johdatus Bayes-verkoilla päättelyyn

Jouni Seppänen

2001-09-19

Todennäköisyyslaskua

- Aksiomaattinen lähestymistapa: todennäköisyysavaruus $(S, \mathcal{A}, P)$
  - S on perusjoukko
  - $\mathcal{A}$ on "sopiva" kokoelma S:n osajoukkoja, *tapahtumia*
    (tarkalleen sanottuna σ-algebra: (1) $S \in \mathcal{A}$, (2)
    $A \in \mathcal{A} \implies \bar{A} \in \mathcal{A}$,
    (3) jos J numeroituva ja $A_j \in \mathcal{A}$ kaikilla $j \in J$, niin $\cup_{j \in J} A_j \in \mathcal{A}$)
  - $P : \mathcal{A} \to \mathbb{R}$ on kuvaus, joka liittää jokaiseen tapahtumaan todennäköisyyden:
    1. $P(A) \geq 0$ kaikilla $A \in \mathcal{A}$
    2. $P(S) = 1$
    3. $P(\cup_{j \in J} A_j) = \sum_{j \in J} P(A_j)$,
       kun indeksijoukko J on numeroituva ja kaikilla $j, k \in J$
       pätee $A_j \cap A_k = \emptyset$.
- Frekventistinen tulkinta
  - Tapahtuman A todennäköisyys kertoo, kuinka usein A tapahtuu pitkään jatketussa toistokokeessa.
  - Painottamattomalla kolikolla $P(\text{kruuna}) = 0{,}5$.

- Bayeslainen tulkinta
  - Tapahtuman A todennäköisyys kertoo, kuinka paljon joku uskoo tapahtumaan.
  - Kuinka todennäköisesti huomenna sataa?
- Matemaattinen todennäköisyysteoria on tulkinnasta riippumatonta
- Ehdollinen todennäköisyys
  - Kuinka todennäköinen on A, kun tiedetään B *eikä mitään muuta olennaista*?
  - $P(A|B) = P(AB)/P(B)$
- Riippumattomuus
  - ”tieto tapahtumasta A ei kerro mitään tapahtumasta B”
  - $P(AB) = P(A)P(B)$, merkitään $A \perp B$
- Ehdollinen riippumattomuus
  - ”tieto tapahtumasta A ei kerro mitään tapahtumasta B, kun tiedetään että C tapahtuu (eikä mitään muuta olennaista)”
  - $P(AB|C) = P(A|C)P(B|C)$, $A \perp B \mid C$

Bayesin kaava

- $P(A|B) = \dfrac{P(B|A)P(A)}{P(B)}$
- todistus: $P(B|A)P(A) = P(AB) = P(A|B)P(B)$
- tulkinta: $P(A)$ on A:n *prioritodennäköisyys*, havaitaan B, $P(B|A)$ on A:n *uskottavuus*, $P(A|B)$ on A:n *posterioritodennäköisyys*

$$\boxed{P(A|B) \propto P(A)P(B|A)}$$

- Esimerkki: vakava sairaus esiintyy yhdellä 10000 ihmisestä, ja sen löytämiseksi on olemassa testi, joka antaa oikean vastauksen 99 %:ssa tapauksista. Millä todennäköisyydellä positiivisen tuloksen saanut ihminen on sairas?

$$\begin{aligned}P(\text{sairas}|\text{pos}) &= P(\text{pos}|\text{sairas})P(\text{sairas})/P(\text{pos}) \\ &= 0,99 \cdot 0,0001/\big(P(\text{pos}|\text{sairas})P(\text{sairas}) + P(\text{pos}|\text{terve})P(\text{terve})\big) \\ &= 0,000099/(0,99 \cdot 0,0001 + 0,01 \cdot 0,9999) \\ &= 0,0098.\end{aligned}$$

- Satunnaismuuttujat
  - Esim. nopan silmäluku, odotusaika pankissa
  - Muodollisesti kuvauksia $f : S \to \mathbb{R}$, joille $\{\, x \in S \mid f(x) \leq a \,\} \in \mathcal{A}$
  - Rajoitumme *diskreetteihin* satunnaismuuttujiin, joiden arvojoukko on äärellinen.
  - Merkintäkäytäntö: isot kirjaimet satunnaismuuttujia, pienet niiden arvoja, esim. $P(X = x_0) = 0{,}2$, $P(X = x_1) = 0{,}8$.
  - Muuttujan X *jakauma* $P(X)$ sisältää kaikkien tapahtumien $X = x$ todennäköisyydet. Jos esimerkiksi X on reilun nopan silmäluku, sen jakauma on seuraava.

x	$P(X = x)$
1	1/6
2	1/6
3	1/6

x	$P(X = x)$
4	1/6
5	1/6
6	1/6

- Yhteisjakauma
  - Olkoot X ja Y satunnaismuuttujia.
  - Satunnaismuuttujan (X, Y) jakauma on X:n ja Y:n yhteisjakauma $P(X, Y)$.
- Riippumattomuus
  - Jos $X = x \perp Y = y$ kaikilla x, y, niin sanotaan, että $X \perp Y$.
  - Tässä tapauksessa yhteisjakauma $P(X, Y)$ on helppo muodostaa.

- Marginalisointi
  - Yhteisjakaumasta $P(X,Y)$ päästään aina helposti marginaalijakaumiin $P(X)$ ja $P(Y)$:

$$P(X) = \sum_{Y} P(X,Y)$$

Ongelma

- Tarkastellaan kolmen eri tavoin painotetun kolikon heittojen yhteisjakaumaa.

k_1	k_2	k_3	P
kr	kr	kr	0,015
kr	kr	kl	0,015
kr	kl	kr	0,035
kr	kl	kl	0,035

k_1	k_2	k_3	P
kl	kr	kr	0,135
kl	kr	kl	0,135
kl	kl	kr	0,315
kl	kl	kl	0,315

- Neljän kolikon jakaumassa olisi 16 lukua, kymmenen kolikon jakaumassa 1024 lukua. Kuitenkin riittäisi ilmoittaa kolikoiden yksittäiset jakaumat ja riippumattomuusoletus.

- Taudin, joka on 1 %:lla väestöstä, havaitsemiseen on kaksi eri asiaa mittaavaa testiä. Ensimmäinen antaa sairaalle koehenkilölle positiivisen tuloksen todennäköisyydellä 0,99 ja terveelle negatiivisen todennäköisyydellä 0,60. Toisen vastaavat todennäköisyydet ovat 0,80 ja 0,95.

s	t_1	t_2	P
sairas	$+$	$+$	0,00792
sairas	$+$	$-$	0,00198
sairas	$-$	$+$	0,00008
sairas	$-$	$-$	0,00002

s	t_1	t_2	P
terve	$+$	$+$	0,0198
terve	$+$	$-$	0,3762
terve	$-$	$+$	0,0297
terve	$-$	$-$	0,5643

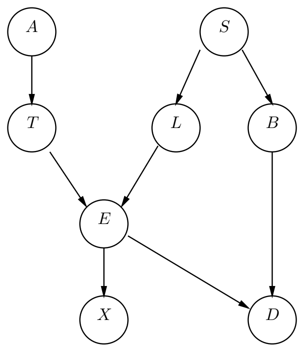
- Jakauma $P(S, T_1, T_2)$ sisältää kahdeksan lukua, kun sen määrittelemiseen riittäisi neljä lukua ja ehdollinen riippumattomuusoletus $T_1 \perp T_2 \mid S$.

Bayes-verkot

- Jakaumien esitystapa, joka käyttää hyväksi ehdollisia riippumattomuusoletuksia
  - Verkkorakenne sisältää riippumattomuusoletukset.
  - Lisäksi jokainen solmu sisältää paikallisen jakauman.
- Intuitiivinen idea: verkossa on kaari $A \to B$, jos A on B:n ”välitön syy”.
- Muodollisesti: Bayes-verkko G on suunnattu syklitön verkko, jonka solmut edustavat satunnaismuuttujia $X_1, \ldots, X_n$.
  Merkitään $\mathrm{Pa}(X_i)$:llä X_i:n välittömiä edeltäjiä (”parents”) ja $\mathrm{ND}(X_i)$:llä niitä solmuja, joihin X_i:stä ei ole suunnattua polkua (”non-descendants”). Silloin G sisältää kaikki ehdolliset riippumattomuusoletukset muotoa
  $$X_i \perp \mathrm{ND}(X_i) \mid \mathrm{Pa}(X_i).$$
- Edellisen kanssa on yhtäpitävää, että yhteisjakauma voidaan esittää hajotelmana
  $$P(U) = \prod_{i=1}^{n} P(X_i | \,\mathrm{Pa}(X_i)),$$
  missä U on kaikkien muuttujien X_i joukko.
- Mitä kaikkea seuraa G:n sisältämistä riippumattomuusoletuksista? Verkkoalgoritmi nimeltä *d-separaatio* vastaa tähän kysymykseen. Seuraavassa vain esimerkkejä.
- $X \to H \to Y$: $P(X, H, Y) = P(X)P(H|X)P(Y|H)$
  - $P(X, Y|H) = P(X)P(H|X)P(Y|H)/P(H) = P(X|H)P(Y|H)$, joten $X \perp Y \mid H$
  - Kuitenkin $X \not\perp Y \mid \emptyset$
- $X \leftarrow H \to Y$: $P(X, H, Y) = P(H)P(X|H)P(Y|H)$
  - $P(X, Y|H) = P(X|H)P(Y|H)$, joten $X \perp Y \mid H$
  - Taas $X \not\perp Y \mid \emptyset$
- $X \to H \leftarrow Y$: $P(X, H, Y) = P(X)P(Y)P(H|X, Y)$
  - Nyt $X \perp Y \mid \emptyset$ mutta $X \not\perp Y \mid H$.

Esimerkki

- Seuraavaa kuvitteellista verkkoa voitaisiin käyttää keuhkosairauksien diagnosoinnin apuna.
  - Hengitysvaikeudet (**D**yspnoea) voivat johtua **T**uberkuloosista, keuhkosyövästä (**L**ung cancer), keuhkoputkentulehduksesta (**B**ronchitis), jostain näiden yhdistelmästä tai aivan muusta syystä. Tuberkuloosin todennäköisyyttä lisää **A**asian-matka, kun taas tupakointi (**S**moking) on keuhkosyövän ja keuhkoputkentulehduksen riskitekijä. Röntgenkuva (**X**-ray) ei erota toisistaan keuhkosyöpää ja tuberkuloosia, ja myös hengitysvaikeuksien kannalta ne ovat samanlaiset.
  - Kaikki muuttujat ovat loogisia (tosi/epätosi). **E** on tosi, jos **T** tai **L** on tosi.



$$
\begin{aligned}
P(U) =& P(A)P(S) \\
& P(T|A)P(L|S) \\
& P(B|S)P(E|L,T) \\
& P(D|B,E)P(X|E)
\end{aligned}
$$

$$(U = \{A, S, T, L, B, E, X, D\})$$

Esimerkkejä päättelystä

- Yksinkertainen verkko $X \to H$ sisältää jakaumat $P(X)$ ja $P(H|X)$. Oletetaan, että muuttujan H arvo tunnetaan, "H on havaittu". Halutaan laskea $P(X|H = h)$.
  - Yhteisjakauma $P(X, H) = P(X)P(H|X)$, mistä marginalisoimalla $P(H)$.
  - Bayesin kaavalla saadaan

$$P(X|H = h) = \frac{P(H = h|X)P(X)}{P(H = h)}.$$

- Verkko $Y \leftarrow X \rightarrow H$, tunnetaan $P(X)$, $P(Y|X)$ ja $P(H|X)$, havaitaan H. Laskettava $P(Y|H=h)$.
  - Raa'alla voimalla:
    1. $P(X,Y,H)$
    2. $P(H)$
    3. $P(Y,H)$
    4. $P(Y|H=h) = P(Y,H=h)/P(H=h)$
  - Hienostuneemmin:
    1. $P(H=h) = \sum_X P(H=h|X)P(X)$
    2. $P(X|H=h) = P(H=h|X)P(X)/P(H=h)$ (Bayes)
    3. $P(Y|H=h) = \sum_X P(Y|X)P(X|H=h)$

Liittymäpuu

- Edellisen tyyppisten päättelyjen automatisoimiseksi malli kannattaa kääntää ("compile") liittymäpuuksi ("junction tree"). Vaiheet ovat pääpiirteittäin seuraavat.

  1. "Pakotetaan vanhemmat naimisiin" lisäämällä kaaret yhdistämään solmun kaikki välittömät edeltäjät, ellei niiden välillä ole valmiiksi kaarta.
  2. Muutetaan kaikki kaaret suunnattomiksi. Lopputulos on *moraalinen verkko*.
  3. *Kolmioidaan* moraalinen verkko lisäämällä niin paljon kaaria, että jokainen vähintään neljän solmun kautta kulkeva silmukka sisältää oikotien.
  4. Etsitään kolmioidun verkon klikit.
  5. Muodostetaan klikeistä liittymäpuu.

- Muodostetaan esimerkkinä keuhkotautiverkon liittymäpuu.

- Moralisointi (askeleet 1 ja 2)

- Kolmiointi (askele 3)

- Liittymäpuu (askeleet 4 ja 5)

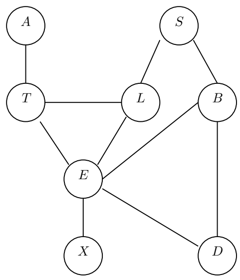
Moraalinen verkko

- Tulkitaan yhteisjakauman tekijät solmujen ja niiden edeltäjäjoukkojen funktioina:

$$P(U) = \prod_{i=1}^{n} a(X_i, \mathrm{Pa}(X_i)).$$

  Funktioita $a(\cdot, \cdot)$ kutsutaan *potentiaaleiksi.*

- Jokainen alkuperäisen verkon solmu muodostaa moraalisessa verkossa välittömien edeltäjiensä kanssa täydellisen aliverkon. Tässä yhteydessä nimitämme *klikiksi* (clique) maksimaalista täydellistä aliverkkoa. Voidaan määritellä klikkipotentiaalit a_C niin, että $P(U) = \prod a_C(X_C)$.

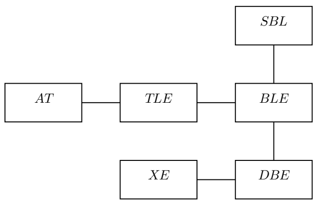


Liittymäpuun muodostaminen

- Kolmioidun verkon klikeistä muodostetaan suuntaamaton *puu* (verkko, jonka kaikkia solmupareja yhdistää yksikäsitteinen polku).

- Puulta vaaditaan seuraava ominaisuus ("running intersection property"): jos muuttuja X_i esiintyy kahdessa klikissä C ja D, se esiintyy jokaisessa C:tä ja D:tä yhdistävän (yksikäsitteisen) polun klikissä.

- Vaatimus on mahdollinen täyttää kolmioinnin ansiosta.



Liittymäpuun käyttö

- Kolmiointi ei estä potentiaalihajotelman muodostamista:

$$P(U) = \prod a_C(X_C).$$

- Lisätään puuhun klikkien väliin erotinsolmuja (separators), jotka edustavat naapuriklikkien leikkausta. Määritellään yleistetty potentiaalihajotelma

$$P(U) = \frac{\prod_C a_C(X_C)}{\prod_S b_S(X_S)}.$$

  Aluksi voidaan asettaa $b_S(X_S) = 1$ kaikilla S.

- Klikit lähettävät toisilleen viestejä erotinsolmujen kautta. Viestit muuttavat potentiaaleja niin, että edellämainittu tulojen suhde ei muutu. Lopputuloksena saadaan marginaaliesitys

$$P(U) = \frac{\prod_C P(X_C)}{\prod_S P(X_S)}.$$

- Kun tehdään havainto $\mathcal{E} : X_A = x_A^*$, määritellään funktio P^* säännöllä

$$P^*(x) = \begin{cases} P(x), & X_A = x_A^* \\ 0, & \text{muuten.} \end{cases}$$

- Nyt $P^*(U) = P(U, \mathcal{E}) = P(\mathcal{E})P(U|\mathcal{E})$, ja edelleen voidaan kirjoittaa

$$P^*(U) = P(U) \prod_{v \in A} l(v),$$

  missä $l(v) = 1$, jos $X_v = x_v^*$, 0 muuten.

- Viestinvälitysalgoritmilla voidaan laskea myös

$$P(V|\mathcal{E}) = \frac{\prod_C P(X_C|\mathcal{E})}{\prod_S P(X_S|\mathcal{E})}.$$

Lähteitä

- Michael I. Jordan (toim.): Learning in Graphical Models
  - Kjælruffin luvussa esitellään liittymäpuualgoritmi Cowellin lukua yksityiskohtaisemmin (ennen kuin aletaan kuvata sisäkkäisiä puita)
- Daphne Kollerin WWW-sivut
  - http://robotics.stanford.edu/%7Ekoller/
- Huang, Darwiche: Inference in Belief Networks: A Procedural Guide.
  - Liittymäpuun toteutus tietokoneella.
  - http://citeseer.nj.nec.com/huang94inference.html
- Myllymäki, Tirri: Bayes-verkkojen mahdollisuudet.
  - TEKES Teknologiakatsaus 58/98.