

# Bayesi võrgud

Sven Laur  
swen@math.ut.ee

19. detsember 2003. a.

# Miks eelistada Bayesi statistikat?

- Klassikaline statistika püüab olla objektiivne.
- NFL-teoreemid (David Wolpert 1995)  
Kõik õppimisalgoritmid töötavad keskmiselt sama hästi.

Probleemide klass kus algoritm A ületab algoritmi B moodustab vaid imeväikse osa kogu otsinguruumist.

- Klassikalised meetodid annavad ebastabiilseid või lubamatuid lahendeid.

Bayesi meetodid õigustavad stabiliseerivad võtteid.

## Mis on tõenäosus?

Olgu kulli ja kirja saamise tõenäosus on võrdne.

- Siis sooritades palju katseid on ligikaudu pooled neist kullid.
- Minu meelest pole mõtet kulli tulekule panustada rohkem raha kui kirja saamisele.

Olgu mündi mõlemad pooled identsed, kuid meile teadmata.

# Ratsionaalne uskumus

- Tõenäosus on subjektiivne ebakindluse mõõt, mis väljendab indiviidi ratsionaalset uskumust.
- *Dutch book* argument.
  - Iga sündmusele vastab veksel.
  - Sündmuse  $A$  tõenäosus  $p(A)$  on vekslite müügihind.
  - Ratsionaalselt käituv individ ei lase ennast petta.
- Ebakindlus on alati normeerimata tõenäosusmõõt.

# Bayesi teoreem

Iga sündmuste paari  $A$  ja  $B$  korral kehtib

$$p(A)p(B|A) = p(A, B) = p(B)p(A|B).$$

- Tegemist pole tautoloogiaga.
- Tõenäosus  $p(A|B)$  tähistab ebakindlust  $A$  suhtes, kui  $B$  on toimunud.
- *A priori* pole selge, miks ebakindlus rahuldab Bayesi teoreemi.

# Mis on õppimine?

Õppimine on Bayesi teoreemi rakendamine

$$p(X = x|D) = \frac{p(D|X = x)p(X = x)}{p(D)}$$

$$p(X = x|D) \propto p(D|X = x)p(X = x).$$

- Aposteriaalne tõenäosus  $p(X = x|D)$ .
- Mudeli tõepära  $p(D|X = x)$ .
- Aprioorne uskumus  $p(X = x)$ .

# Keskmistamine

Milline on eksamihinde  $A$  saamise tõenäosus?

- Eksamil on viisteist piletit  $T_i$ .
- Iga pileti korral saab hinnata  $p(A|T_i)$ .
- Loomulikult

$$p(A) = \sum_i p(T_i)p(A|T_i) = \mathbf{E} [p(A|T_i)] .$$

## Kuidas kirjeldada eelteadmist?

Milline on minu tõenäosus sattuda autoavariisse?

- Otse on küsimusele raske objektiivselt vastata.
- Tuleb vaadata erinevaid faktoreid.
  - Kui tihti ma kannan helkurit?
  - Kas ma olen maal või linnas?
  - Kas väljas on pime?
- Faktoreid ennast mõjutavad teised faktorid.

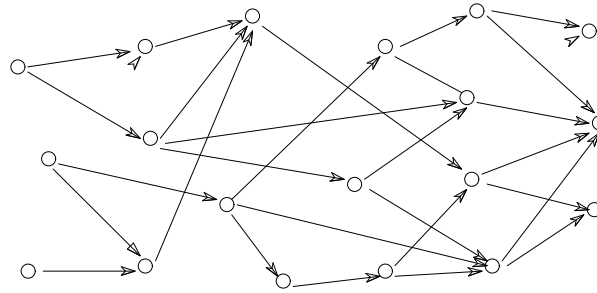


# Bayesi võrk

- Bayesi võrk (*belief network*) määrab ära juhuslike suuruste  $\mathcal{X} = (X_1, X_2, \dots, X_n)$  ühisjaotuse.
- Bayesi võrk on atsükliline orienteeritud graaf.
- Põhjuslikud seoste tugevust kirjeldatakse tinglike tõenäosustega  $p(X_i | Pa(X_i))$ . Näiteks võib see olla antud tabelina.

Helkur on	Maa on	Pime on	Tõenäosus 0.0001
on	on	ei	0.00001

# Bayesi võrgu arvutusreegel



Kuna graaf on atsükliline, siis saame avaldada

$$p(X_1, X_2, \dots, X_n) = \prod_i p(X_i | \text{Pa}(X_i)),$$

$$p(X_1) = \sum_{X_2, \dots, X_n} p(X_1, X_2, \dots, X_n).$$

Teooria lõpp!!?

## Tsirkus (*hack 1*)

Igal järgmine mänguseis on määratud eelmise poolt

$$X_1 \longrightarrow X_2 \longrightarrow \cdots \longrightarrow X_{i-1} \longrightarrow X_i$$

Teisisõnu  $p(X_i|X_1, \dots, X_{i-1}) = p(X_i|X_{i-1})$ .

On lihtne leida  $p(X_i)$  kui on teada  $p(X_{i-1})$

$$p(X_i) = \sum_{X_{i-1}} p(X_{i-1})p(X_i|X_{i-1})$$

## Tsirkus jätkub ...

Graafiliselt illustreerides

$$\boxed{X_1} \longrightarrow X_2 \longrightarrow \cdots \longrightarrow X_{i-1} \longrightarrow X_i$$

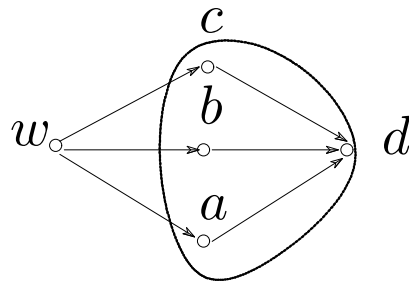
$$\boxed{X_1 \longrightarrow X_2} \longrightarrow \cdots \longrightarrow X_{i-1} \longrightarrow X_i$$

$$X_1 \longrightarrow \boxed{X_2} \longrightarrow \cdots \longrightarrow X_{i-1} \longrightarrow X_i$$

$$X_1 \longrightarrow \boxed{X_2 \longrightarrow} \cdots \longrightarrow X_{i-1} \longrightarrow X_i$$

## Keerukad sõltuvused (*hack 2*)

$$p(a, b, c, d) = p(a, b, c)p(d|a, b, c)$$



- Tipu  $d$  korral vajame muutujate  $a, b, c$  ühisjaotust.
- Selleks tuleb keskmistada üle  $w, a, b, c$  jaotuse.

## Loomulikud järeldused

- Õige arvutusskeem hoiab kokku mälu ja aega.
- Tipu  $X_i$  jaotuse leidmiseks peame leidma tipuhulkade jada  $C_1, C_2, \dots, C_n$  nii, et
  - nende lõiked rahuldaksid teatud tingimusi;
  - iga tipuhulk oleks võimalikult väike;
- Selleks piisab vaid Bayesi võrgu topoloogiast.

# Moraal

Arvutuslikust seisukohast on olulised kaks reeglit

- Tipp ja tipu järglane peavad kuuluma samasse tipuhulka.
- Tipp ja tipu eellased peavad kuuluma samasse tipuhulka.

Tulemuseks on graafi teisendusreeglid

- Kaar  $X_i \rightarrow X_j$  tuleb asendada servaga  $X_i - X_j$ .
- Kaartepaari  $X_i \rightarrow X_k \leftarrow X_i$  korral lisada serv  $X_i - X_j$ .
- Saadud graafi nimetatakse moraaliks.



## Tingimuslik sõltumatus

Süsteemaatiline arvutusskeem kasutab Bayesi võrgu sekundaarset struktuuri. Peamiseks vahendiks on tingimuslik sõltumatus.

**Definitsioon 1.** Juhuslike suuruste hulgad  $A$  ja  $B$  on tinglikult sõltumatud kui

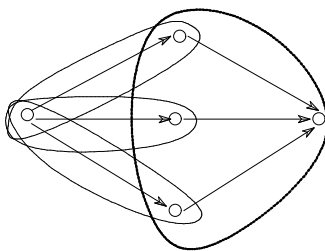
$$p(\mathcal{A}, \mathcal{B} | \mathcal{C}) = p(\mathcal{A} | \mathcal{C})p(\mathcal{B} | \mathcal{C}),$$

$$p(\mathcal{A} | \mathcal{B}, \mathcal{C}) = p(\mathcal{A} | \mathcal{C}).$$

## Moraali klikid

Vastavalt teisendusreeglitele on sobivad tipuhulgad, mille tõenäosusjaotust arvutada maksimaalsed klikid.

- On garanteeritud tipu ja tema eelaste olemine ühes klikis.
- Kliki tipud jagunevad kaheks eeldus ning järelsus.



## Maagiline tingimus

Kui saab jagada tipuhulga hulkadeks  $A$  ja  $B$  nii:

- teed hulkade  $A$  ja  $B$  vahel läbivad kindlasti  $A \cap B$ ;
- $A \cap B$  on täielik graaf;

siis saab leida

$$P(A \cup B) = \frac{P(A)P(B)}{P(A \cap B)}.$$

See on samaväärne graafi trianguleerimisega — iga tsükkel, mis omab vähemalt nelja tippu lõhutakse kõõluga.

# Klikipuu

Pärast graafi trianguleerimist leidub sellel klikipuu:

- tippudeks on maksimaalsed klikid  $C_k$ ;
- servadeks on klikkide lõiked  $S_k$ .

Kuna klikipuus puuduvad tsüklid, siis saame leida iga kliki tõenäosusjaotuse ning kogu ühisjaotuse

$$p(X_1, X_2, \dots, X_n) = \prod_{k=1}^m \frac{p(C_k)}{p(S_k)} = \prod_{k=1}^m p(R_k | S_k).$$

# Arvutuslik algoritm

Tegelikult pole vaja leida tippude kausaalset järjestust, piisab kui kasutada kolme-etapilist algoritmi:

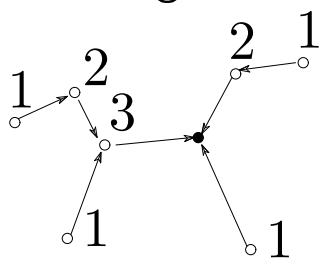
- potentsiaalide initsialiseerimine,
- tõstusmaterjali kogumine,
- tõestumaterjali laiali saatmine.

Eesmärgiks on saavutada kooskõla

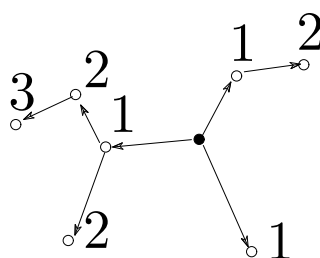
$$\int P(X_C)dX_R = P(S) = \int P(X_D)dX_{R'}$$

# Kooskõla saavutamise kampaaniad

Info kogumine



Korralduste jagamine



# Bayesi võrgu treeninime

Kuidas leida ainult andmetele toetudes Bayesi võrk?

- Puhas Bayesi printsiibi rakendamine.
- Integraali lähendmine lõpliku summaga.
- MAP ja ML lähendid — EM ja  $MC^2$  meetodid.
- Strukturaalne optimiseerimine.