

Parametrien oppiminen

- Tilastollisen mallin (Bayes-verkko) rakenne on kiinnitetty, sen numeeristen parametrien (ehdollisten todennäköisyyksien) arvot pyritään määrittämään
- Oletamme havaintojen olevan täydellisiä; s.o., jokaisen datapisteen sisältävän arvot kaikille muuttujille
- Jos makeissäkkien salmiakki-hedelmä-karkkien osuudet voivat olla mielivaltaisia, niin hypoteeseja onkin jatkumo
- Salmiakkimakeisten suhteellinen osuus säkissä θ on ainut parametri ja sitä vastaava hypoteesi on h_θ
- Bayes-verkkoon tarvitaan vain yhtä satunnaismuuttujaa (*Maku*), jolla on mahdolliset arvot *salmiakki* (tn. θ) ja *hedelmä* (tn. $1 - \theta$), vastaava solmu

- Avataan kääreestä n karkkia, joista s kpl on salmiakkia ja h kpl on hedelmää
- Oletetaan kaikkien sekoitussuhteiden olevan yhtä todennäköisiä *a priori* \Rightarrow maksimuskottavuuden menetelmä

$$P(\mathbf{d} | h_\theta) = \prod_{j=1, \dots, n} P(d_j | h_\theta) = \theta^s (1 - \theta)^h$$

- Samaan maksimaalisen uskottavuuden hypoteesiin päädytään maksimoimalla uskottavuuden logaritmia (log likelihood)

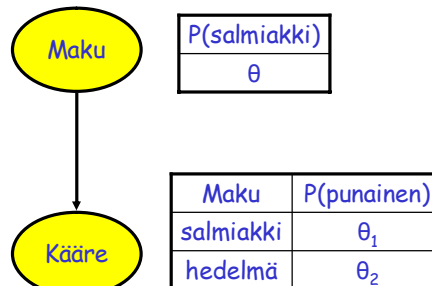
$$\begin{aligned} L(\mathbf{d} | h_\theta) &= \log P(\mathbf{d} | h_\theta) \\ &= \sum_{j=1, \dots, n} \log P(d_j | h_\theta) \\ &= s \log \theta + h \log(1 - \theta) \end{aligned}$$

- Näin tulo yli aineiston vaihtuu summaksi, joka yleensä on helpompi maksimoida
- Derivoimalla L θ :n suhteen ja etsimällä nollakohta saadaan selville θ :n maksimiuskottavuuden arvo

$$dL(d | h_{\theta})/d\theta = s/\theta - h/(1 - \theta) = 0$$

$$\Rightarrow \theta = s/(s + h) = s/n$$
- Maksimaalisen uskottavuuden hypoteesi h_{ML} sanoo säkissä olevien salmiakkien osuuden olevan sama kuin mitä avatuista karkeista on havaittu
- Edellä kuvattua menetelmää voidaan käyttää yleisesti (useammankin parametrin arvojen selvittämiseksi)

- Menetelmän merkittävä ongelma: jos jotain tapahtumaa ei ole havaittu laisinkaan (pienessä havaintoaineistossa), niin h_{ML} antaa sille todennäköisyyden nolla
- Muutetaan esim. siten, että riippuen karkin mausta se kääritään probabilistisen säännön perusteella joko punaiseen tai vihreään paperiin



- Nyt todennäköisyysmallissa on kolme parametria θ , θ_1 , ja θ_2
- Bayes-verkon standardisemantiikan perusteella voidaan laskea tapahtumien todennäköisyyksiä; esim.

$$P(\text{Maku} = \text{salmiakki}, \text{Kääre} = \text{vihreä} \mid h_{\theta, \theta_1, \theta_2})$$

$$= P(\text{Maku} = \text{salmiakki} \mid h_{\theta, \theta_1, \theta_2}) \cdot$$

$$P(\text{Kääre} = \text{vihreä} \mid \text{Maku} = \text{salmiakki}, h_{\theta, \theta_1, \theta_2})$$

$$= \theta(1 - \theta_1)$$

- Avaamme nyt n makeista, joista s on salmiakkia ja h hedelmää, joiden käärepaperien värien lkm:t ovat p_s , v_s , p_h ja v_h
- Tämän aineiston uskottavuus $P(\mathbf{d} \mid h_{\theta, \theta_1, \theta_2})$ on

$$\theta^s (1 - \theta)^h \theta_1^{p_s} (1 - \theta_1)^{v_s} \theta_2^{p_h} (1 - \theta_2)^{v_h}$$



- Otetaan logaritmi edellisestä:
- $$L = [s \log \theta + h \log(1 - \theta)] + [p_s \log \theta_1 + v_s \log(1 - \theta_1)] + [p_h \log \theta_2 + v_h \log(1 - \theta_2)]$$

- Parametrien suhteen derivoimalla ja hakemalla nollakohta saadaan

$$\partial L / \partial \theta = s / \theta - h / (1 - \theta) = 0 \Rightarrow \theta = s / (s + h)$$

$$\partial L / \partial \theta_1 = p_s / \theta_1 - v_s / (1 - \theta_1) = 0 \Rightarrow \theta_1 = p_s / (p_s + v_s)$$

$$\partial L / \partial \theta_2 = p_h / \theta_2 - v_h / (1 - \theta_2) = 0 \Rightarrow \theta_2 = p_h / (p_h + v_h)$$

- θ :n arvo on kuten ennenkin ja θ_1 :n arvo on punaisen käärepaperin osuus salmiakkikarkkien joukossa (θ_2 vast.)
- Näin oppimisongelma jakaantuu erillisiksi oppimistehtäviksi kullekin parametrille



Naiivi Bayes

- Luokkamuuttuja C on Bayes-verkon juuri ja attribuutit X_i ovat sen lehtiä
- Naiivi oletus on, että attribuutit ovat ehdollisesti riippumattomia toisistaan annettuna luokka
- Kun käytössä on Boolean muuttujat, niin parametrit ovat

$$\theta = P(C = \text{true})$$

$$\theta_{i1} = P(X_i = \text{true} \mid C = \text{true})$$

$$\theta_{i2} = P(X_i = \text{true} \mid C = \text{false})$$
- Näiden arvot löydetään kuten edellä
- Kun verkko on opetettu, niin havainto $[x_1, \dots, x_n]$, jonka luokkamuuttujan C arvo on tuntematon, voidaan luokitella

$$P(C \mid x_1, \dots, x_n) = \alpha P(C) \prod_i P(x_i \mid C)$$

Jatkuva-arvoiset muuttujat

- Tarkastellaan yhden muuttujan Gaussisen tiheysfunktion parametrien oppimista
- Aineiston siis tuottaa

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- Mallin parametrit ovat keskiarvo μ ja hajonta σ
- Olkoot havaitut arvot x_1, \dots, x_n
- Nyt uskottavuuden logaritmi on

$$L = \sum_{j=1}^n \log \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_j-\mu)^2}{2\sigma^2}}$$

$$= n(-\log \sqrt{2\pi} - \log \sigma) - \sum_{j=1}^n \frac{-(x_j - \mu)^2}{2\sigma^2}$$

- Osittaisderivaattojen nollakohdat:

$$\frac{\partial L}{\partial \mu} = -\frac{1}{\sigma^2} \sum_{j=1}^n (x_j - \mu) = 0 \quad \Rightarrow \mu = \frac{\sum_j x_j}{n}$$

$$\frac{\partial L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{j=1}^n (x_j - \mu)^2 = 0 \quad \Rightarrow \sigma = \sqrt{\frac{\sum_j (x_j - \mu)^2}{n}}$$

- Siis keskiarvon maksimaalisen uskottavuuden arvo on havaintojen keskiarvo
- ja hajonnan vastaava arvo on varianssin neliöjuuri
- Jälleen saatiin siis intuitiivisesti oikeat arvot

Yhteenveto

- Tekoäly on todella laaja tutkimuskenttä, myös metodologisesti
- Merkittävää kehitystä tutkimuksessa tapahtuu päivittäin
- Probabilistinen lähestyminen on ajanut ohi loogisen suuntauksen
- Näyttäviä demonstraatioita menetelmien mahdollisuuksista saadaan kiihtyvällä tahdilla
- Myös vähemmälle julkisuudelle jäävät arkisovellukset lisääntyvät
- Fyysisen agentin interaktio toimintaympäristönsä sekä ihmisten kanssa kaipaava vielä edistysaskelia