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 Boolen muuttujat, niin parametrit ovat

$$\theta = P(C = true)$$
$$\theta_{i1} = P(X_i = true \mid C = true)$$
$$\theta_{i2} = P(X_i = true \mid C = false)$$

- Näiden arvot löydetään kuten edellä
- Kun verkko on opetettu, niin havainto $[x_1, \ldots, x_n]$, jonka luokkamuuttujan C arvo on tuntematon, voidaan luokitella

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

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 $x_1, \ldots, x_n$
- 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:

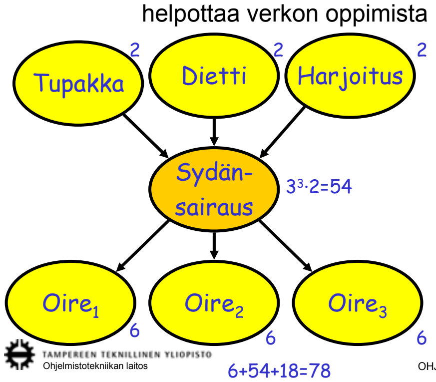
$$\frac{\partial L}{\partial \mu} = -\frac{1}{\sigma^2}\sum_{j=1}^{n}(x_j - \mu) = 0 \qquad \Rightarrow \mu = \sum_j x_j \Big/ n$$

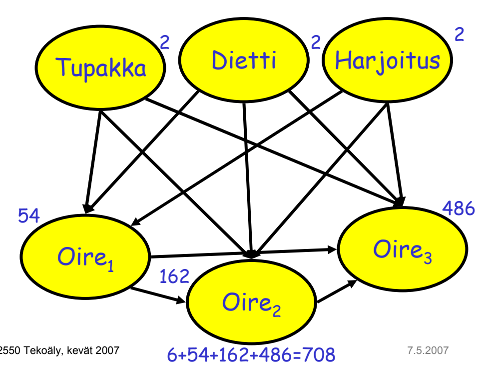
$$\frac{\partial L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{j=1}^{n}(x_j - \mu)^2 = 0 \quad \Rightarrow \sigma = \sqrt{\frac{\sum_j (x_j - \mu)^2}{n}}$$

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

EM-algoritmi

- Havainnot eivät käytännössä ole täydellisiä, kuten edellä oletettiin
- Sen sijaan on olemassa *piilomuuttujia* l. *latentteja muuttujia*, joiden arvoja ei ole mukana havainnoissa
- Latentin muuttujan ottaminen mukaan Bayes-verkkoon voi vähentää tarvittavien parametrien lukumäärää oleellisesti ja siten helpottaa verkon oppimista





- Kun edellisessä esimerkissä kullakin muuttujalla on 3 mahdollista arvoa ovat Bayes-verkkojen parametrien lukumäärät 78 ja 708
- Piilomuuttujat kuitenkin hankaloittavat oppimisongelmaa
- Kuinka esim. oppia solmun Sydänsairaus ehdolliset todennäköisyydet annettuna sen vanhemmat, koska muuttujan arvoa ei voida havainnoida?
- Sama ongelma koskee oireiden todennäköisyysjakaumien oppimista
- EM-algoritmi (expectation-maximization) on yleinen algoritmi tämän ongelman ratkaisemiseksi

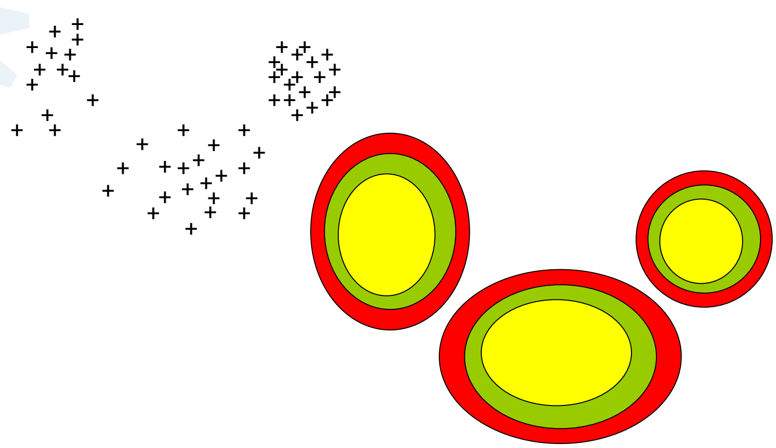
Ohjaamaton klusterointi

- Tapaukset, joita saamme nähtäväksemme, ovat luokittelemattomia
- Siitä huolimatta usein voidaan tunnistaa tapausluokkia
- Peruslähtökohta luokittelulle on, että havaintoaineiston on tuottanut *sekoitejakauma* P (mixture distribution), jossa on osana k jakaumaa
- Tapauksen generoimiseksi ensin valitaan komponenttijakauma ja sitten arvotaan satunnainen otos valitun jakauman mukaan
- Olkoon C valittua komponenttia vastaava satunnaismuuttuja, jonka mahdolliset arvot ovat $1, \ldots, k$

- Sekoitejakauma on

$$P(\mathbf{x}) = \sum_{i=1,\ldots,k} P(C = i)\, P(\mathbf{x} \mid C = i),$$

  missä $\mathbf{x}$ on tapauksen attribuuttiarvot
- Jatkuva-arvoiselle datalle luonteva valinta komponenttijakaumiksi on monimuuttujainen (multivariate) normaalijakauma
- Sekoitejakauma on siis *mixture of Gaussians*, jonka parametrit ovat
  - komponettijakaumien painot $w_i = P(C = i)$ sekä
  - kunkin komponentin keskiarvo μ_i ja kovarianssi Σ_i
- Tehtävänä on löytää aineiston perusteella sekoitemalli, joka olisi voinut tuottaa havaintoaineiston



- Jos tietäisimme mikä komponenttijakaumista tuotti minkäkin pisteen, voisimme noudattaa normaalia Gaussisen jakauman oppimista
- Jos toisaalta tietäisimme komponenttijakaumien parametriarvot, niin pisteiden (probabilistinen) jakaminen komponentteihin olisi helppoa
- Kumpaakaan ei kuitenkaan tunneta
- EM-algoritmi toimii kuin tuntisimme mallin parametrien arvot ja sen perusteella lasketaan todennäköisyydet kaikille pisteille kaikkiin komponentteihin kuulumiselle
- Sen jälkeen kukin komponentti sovitetaan uudelleen koko aineistoon, pisteet painotettuna ko. komponenttiin kuulumisen todennäköisyydellä

- Edellä mainittuja askeleita toistetaan kunnes menetelmä konvergoituu
- Menetelmä täydentää aineistoa päätellen piilomuuttujien jakaumia nykyisen mallin perusteella
- Piilomuuttajia tässä tapauksessa ovat pisteiden kuulumista komponettijakaumiin indikoivat binääriset satunnaismuuttujat Z_{ij}:

$$Z_{ij} = \begin{cases} 1 & \text{jos } \mathbf{x}_j\text{:n on tuottanut } i\text{:s komponenttijakauma} \\ 0 & \text{muuten} \end{cases}$$

- Sekoitemallin parametrit alustetaan satunnaisesti ja sen jälkeen toistetaan seuraavia askeleita

1. **E-askel**: Laske todennäköisyydet, että havainnon $\mathbf{x}_j$ on tuottanut komponenttijakauma i, $p_{ij} = P(C = i \mid \mathbf{x}_j)$
   - Bayesin kaavan perusteella $p_{ij} = \alpha\, P(\mathbf{x}_j \mid C = i)\, P(C = i)$
   - Tekijä $P(\mathbf{x}_j \mid C = i)$ on i:nnen Gaussin jakauman todennäköisyys pisteessä $\mathbf{x}_j$
   - $P(C = i)$ on i:nnen jakauman painoparametrin arvo
   - Olkoon $p_i = \Sigma_j\, p_{ij}$
2. **M-askel**: Päivitä mallin parametrit:

$$\mu_i \leftarrow \Sigma_j\, p_{ij}\, \mathbf{x}_j / p_i$$
$$\Sigma_i \leftarrow \Sigma_j\, p_{ij}\, \mathbf{x}_j \mathbf{x}_j^T / p_i$$
$$w_i \leftarrow p_i$$

- E-askel laskee piilomuuttujien Z_{ij} odotusarvot p_{ij}
- M-askel puolestaan hakee mallin parametreille uudet arvot, jotka maksimoivat aineiston log-uskottavuuden annettuna odotusarvot p_{ij}
- EM-algoritmi parantaa aineiston log-uskottavuutta jokaisella iteraatiolla
- Joissain tapauksissa EM:n voidaan todistaa saavuttavan uskottavuuden lokaali maksimi (huom. ilman askelkokoa)
- Ongelmatilanteita:
  - Komponenttijakauma supistuu kattamaan vain yhden havainnon, varianssi = 0 ⇨ uskottavuus = ∞
  - Kaksi komponenttia kattavat samat pisteet (niillä on samat keskiarvot ja varianssit)

- Bayes-verkossa piilomuuttujat ovatkin havaitsemattomien muuttujien arvot kullakin esimerkillä
- Piilo-Markov-mallissa (hidden Markov model, HMM) piilomuuttujia ovat tilasiirtymätodennäköisyydet
- EM-algoritmista saadaan eri instantiaatioita erilaisiin malleihin
- Yleisimmässä muodossaan algoritmi typistyy päivityssäännöksi

$$\theta^{(i+1)} = \arg\max_{\theta} \sum_{z} P(\mathbf{Z} = \mathbf{z} \mid \mathbf{x}, \theta^{(i)})\, L(\mathbf{x}, \mathbf{Z} = \mathbf{z} \mid \theta),$$

  $\mathbf{x}$ on kaikki havaitut arvot,

  $\mathbf{Z}$ on kaikki piilomuuttujat ja

  θ on kaikki parametrit
- Summaus vastaa E-askelta ja maksimointi M-askelta

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 kaipaa vielä edistysaskelia