1. TILASTOLLINEN HAHMONTUNNISTUS

Tilastollisissa hahmontunnistusmenetelmissä piirteitä tarkastellaan tilastollisina muuttujina

Luokittelussa käytetään hyväksi seuraavia tietoja: luokkien *a priori* tn:iä, luokkien ehdollisia tnjakaumia ja tehtyjä havaintoja eli opetusdataa

Hahmon luokka valitaan korkeimman *a posteriori* tn:n tai siitä johdetun funktion mukaisesti tai minimoimalla päätökseen liittyvän riskin odotusarvoa

1.1 Ongelman asettelu

- Hahmo esitetään piirrevektorin $\mathbf{x} \in \mathbb{R}^n$ avulla
- Hahmot halutaan jakaa M:ään luokkaan $\omega_1, \ldots, \omega_M$
- Luokkien *a priori* tn:t ovat $P(\omega_1), \ldots, P(\omega_M)$
- Luokkien ehdolliset tnjakaumat ovat $p(\mathbf{x}|\omega_1), \ldots, p(\mathbf{x}|\omega_M)$

- Jos edellä mainittuja tn:iä ja tnjakaumia ei tunneta, ne voidaan estimoida opetusdatasta:
  - Kun opetusnäytteitä yhteensä N kpl ja luokasta ω_i niitä on N_i kappaletta, $\hat{P}(\omega_i) = N_i/N$
  - Luokkien ehdollisten tnjakaumien estimointia käsitellään myöhemmin
- Mitkä ovat luokkien *a posteriori* tnjakaumat?
- Ol., että jokaisen luokittelutapahtumaan voidaan liittää kustannus: Mitkä ovat eri päätösten riskien odotusarvot?
- Jaetaan piirreavaruus osiin $R_1, \ldots, R_M$ ja tehdään luokittelupäätös seuraavasti: jos $\mathbf{x} \in R_i$, valitaan luokka ω_i
- Mitkä ovat eri luokkia vastaavat alueet piirreavaruudessa, mitkä ovat aluejaon kriteerit?

1.2 "Bayes Decision Theory"

- Optimaalinen tapa suorittaa luokittelu
- Tarkastellaan aluksi kahden luokan, ω_1 ja ω_2, tapausta
- Käyttäen Bayes-sääntöä (Bayes rule) *a posteriori* tn:t saadaan laskettua seuraavasti:

$$P(\omega_i|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_i)P(\omega_i)}{p(\mathbf{x})}, \tag{1}$$

missä

$$p(\mathbf{x}) = \sum_{i=1}^{2} p(\mathbf{x}|\omega_i)P(\omega_i) \tag{2}$$

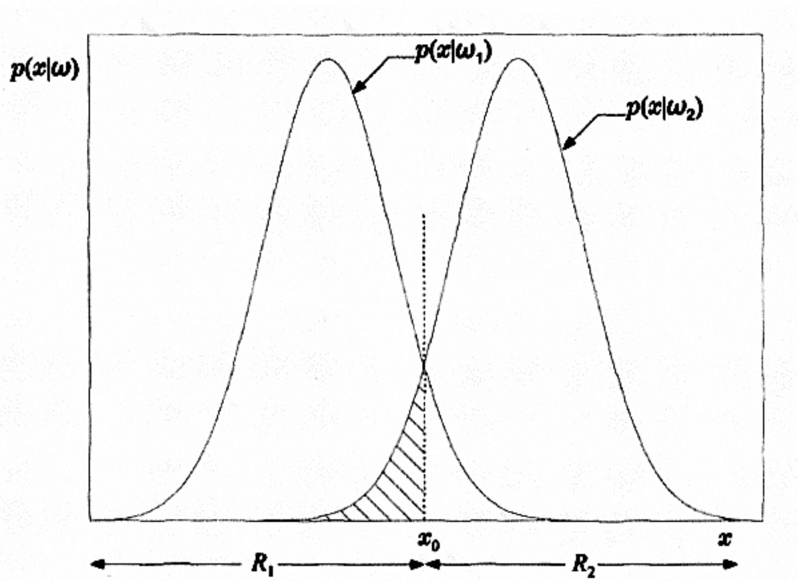
- Jaetaan piirreavaruus osiin R_1 ja R_2 siten, että valitaan aina se luokka, jonka *a posteriori* tn on korkeampi. Luokittelu tehdään siis seuraavasti:

$$\begin{aligned} &\text{Jos } P(\omega_1|\mathbf{x}) > P(\omega_2|\mathbf{x}),\ \mathbf{x} \text{ kuuluu luokkaan } \omega_1 \\ &\text{Jos } P(\omega_1|\mathbf{x}) < P(\omega_2|\mathbf{x}),\ \mathbf{x} \text{ kuuluu luokkaan } \omega_2 \end{aligned} \tag{3}$$

- Jos luokkien *a priori* tn:t ovat samat, saadaan:

$$p(\mathbf{x}|\omega_1) \gtrless p(\mathbf{x}|\omega_2) \tag{4}$$

- Kohtaa, jossa luokkien *a posteriori* tn:t ovat samat, kutsutaan päätösraja ja se jakaa piirreavaruuden alueiksi R_1 ja R_2


$p(x|\omega)$
$p(x|\omega_1)$
$p(x|\omega_2)$
x_0
x
R_1
R_2

Luokitteluvirheen tn:n minimointi

- Täydellisen, virheettömän luokittelun saavuttaminen ei ole aina edes teoriassa mahdollista
- Luokitteluvirheen tn voidaan laskea seuraavasti:

$$P_e = P(\mathbf{x} \in R_2, \omega_1) + P(\mathbf{x} \in R_1, \omega_2) \tag{5}$$

- Edellinen kaava voidaan kirjoittaa myös näin:

$$\begin{aligned} P_e &= P(\mathbf{x} \in R_2 | \omega_1) P(\omega_1) + P(\mathbf{x} \in R_1 | \omega_2) P(\omega_2) \\ &= P(\omega_1) \int_{R_2} p(\mathbf{x}|\omega_1)\, d\mathbf{x} + P(\omega_2) \int_{R_1} p(\mathbf{x}|\omega_2)\, d\mathbf{x} \\ &= \int_{R_2} P(\omega_1|\mathbf{x}) p(\mathbf{x})\, d\mathbf{x} + \int_{R_1} P(\omega_2|\mathbf{x}) p(\mathbf{x})\, d\mathbf{x} \end{aligned} \tag{6}$$

- Koska R_1:n ja R_2:n unioni kattaa koko piirreavaruuden,

$$\int_{R_1} P(\omega_1|\mathbf{x}) p(\mathbf{x})\, d\mathbf{x} + \int_{R_2} P(\omega_1|\mathbf{x}) p(\mathbf{x})\, d\mathbf{x} = P(\omega_1) \tag{7}$$

- Yhdistämällä kaavat (6) ja (7) saadaan:

$$P_e = P(\omega_1) - \int_{R_1} (P(\omega_1|\mathbf{x}) - P(\omega_2|\mathbf{x}))p(\mathbf{x})\, d\mathbf{x} \tag{8}$$

- Edellisestä nähdään helposti, että P_e minimoituu silloin, kun

$$\begin{aligned} R_1 &: \; P(\omega_1|\mathbf{x}) > P(\omega_2|\mathbf{x}) \\ R_2 &: \; P(\omega_2|\mathbf{x}) > P(\omega_1|\mathbf{x}), \end{aligned} \tag{9}$$

- Useamman kuin kahden luokan tapaukselle voidaan johtaa vastaavalla päättelyllä luokitteluvirhetn:n minimoiva päätössääntö:

$$R_i : \; P(\omega_i|\mathbf{x}) > P(\omega_j|\mathbf{x}) \; \forall j \neq i \tag{10}$$

Riskin odotusarvon minimointi

- Luokitteluvirhetn minimointi ei ole paras suunnittelukiriteeri, silloin jos erilaisiin luokittelupäätöksiin liittyy erilaiset riskit, esim. "Onko palohälytys oikea vai pelkkä testi?"

- Liitetään kaikkiin luokittelutapahtumiin ($\mathbf{x} \in R_i$, oikea luokka ω_k) kustannuskertoimet λ_{ki}, jotka voidaan koota matriisiksi $L(k,i) = \lambda_{ki}$

- Luokkaan ω_k liittyvä riski:

$$r_k = \sum_{i=1}^{M} \lambda_{ki} \int_{R_i} p(\mathbf{x}|\omega_k)\, d\mathbf{x}, \tag{11}$$

  missä M on luokkien lukumäärä

- Kuinka valita piirreavaruuden jako siten, että riskin odotusarvo r minimoituu?

$$\begin{aligned} r &= \sum_{k=1}^{M} r_k P(\omega_k) \\ &= \sum_{i=1}^{M} \int_{R_i} (\sum_{k=1}^{M} \lambda_{ki} p(\mathbf{x}|\omega_k) P(\omega_k)) d\mathbf{x} \end{aligned} \tag{12}$$

- r minimoituu seuraavalla jaolla:

$$\begin{aligned} &\mathbf{x} \in R_i, \text{ jos } l_i < l_j \ \forall j \neq i \\ &l_m = \sum_{k=1}^{M} \lambda_{km} p(\mathbf{x}|\omega_k) P(\omega_k) \end{aligned} \tag{13}$$

- Huom! Jos valitaan $\lambda_{ki} = 1 - \delta_{ki}$ (δ_{ki} on Kroneckerin delta-funktio), minimoidaan luokitteluvirhetn:ttä

- Kahden luokan tapaus:
  - Eri päätöksiin liittyvät kustannusten odotusarvot:

$$\begin{aligned} l_1 &= \lambda_{11} p(\mathbf{x}|\omega_1) P(\omega_1) + \lambda_{21} p(\mathbf{x}|\omega_2) P(\omega_2) \\ l_2 &= \lambda_{12} p(\mathbf{x}|\omega_1) P(\omega_1) + \lambda_{22} p(\mathbf{x}|\omega_2) P(\omega_2) \end{aligned} \tag{14}$$

  - Valitaan luokka ω_1, kun $l_1 < l_2$:

$$(\lambda_{21} - \lambda_{22}) p(\mathbf{x}|\omega_2) P(\omega_2) < (\lambda_{12} - \lambda_{11}) p(\mathbf{x}|\omega_1) P(\omega_1) \tag{15}$$

  - Yleensä $\lambda_{ij} \geq \lambda_{ii}$. Silloin:

$$\begin{aligned} R_1 :\ \frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} &> \frac{P(\omega_2)}{P(\omega_1)} \frac{\lambda_{21} - \lambda_{22}}{\lambda_{12} - \lambda_{11}} \\ R_2 :\ \frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} &< \frac{P(\omega_2)}{P(\omega_1)} \frac{\lambda_{21} - \lambda_{22}}{\lambda_{12} - \lambda_{11}} \end{aligned} \tag{16}$$

    (*a posteriori* tnjakaumien suhde on ns. "likelihood ratio")

- **Esimerkki 1:**
  - Tarkastellaan kahden luokan, ω_1 ja ω_2, ongelmaa ja oletetaan, että $p(x|\omega_1) \sim \mathsf{N}(0, 1/2)$ ja $p(x|\omega_2) \sim \mathsf{N}(1, 1/2)$ eli

    $$p(x|\omega_1) = \frac{1}{\sqrt{\pi}} \exp(-x^2)$$
    $$p(x|\omega_2) = \frac{1}{\sqrt{\pi}} \exp(-(x-1)^2)$$

    ja että luokkien *a priori* tn:t $P(\omega_1)$ ja $P(\omega_2)$ ovat samat
  - Silloin luokitteluvirhetn:n minimoiva päätösraja on $x_0 : \exp(-x^2) = \exp(-(x-1)^2)$ eli $x_0 = 1/2$ (katso kaava (10))

– Kun käytetään seuraava kustannusmatriisia L:

$$L = \begin{bmatrix} 0 & 0.5 \\ 1.0 & 0 \end{bmatrix},$$

$x_0: \ \exp(-x^2) = 2\exp(-(x-1)^2)$ eli $x_0 = (1-\ln(2))/2 < 1/2$ (katso kaava (15))

– Huom! Jos luokkien *a priori* tn:t eivät ole samat, $P(\omega_1) \lesseqgtr P(\omega_2)$, siirtyy päätösraja myöskin vasemmalle tai oikealle

- **Esimerkki 2:**

  – Palohälytyksen luokittelu eli juostaanko ulos (tulipalo) vai jäädäänkö sisälle ihmettelemään (väärä hälytys)?

  – Olkoot $\omega_1 =$ "tulipalo", $\omega_2 =$ "väärä hälytys"

  – Oletetaan *a priori* tn:n sille että talossa on tulipalo olevan 1 päivä 10:ssä vuodessa eli $P(\omega_1) = 1/3650$ ja $P(\omega_2) = 3649/3650$

– Piirrevektori $\mathbf{x}$ koostuu esim. seuraavista havainnoista: kuinka iso osa muista ihmisistä säntää ulos, kuinka sakeaa on savu, näkyykö liekkejä jne

– $p(\mathbf{x}|\omega_1)$ ja $p(\mathbf{x}|\omega_2)$ arvioidaan aikaisempien kokemusten pohjalta

– Liitetään eri luokittelupäätöksiin seuraavat ajassa mitatut kustannukset: $\lambda_{11} =$ "kymmenen vuoden tulot ja 1 tunti pihalla", $\lambda_{12} =$ "kymmenen vuoden tulot ja 60 vuotta loppuelämästä", $\lambda_{21} =$ "1 tunti pihalla", $\lambda_{22} = 0$

– Kaavan (15) perusteella riskin odotusarvo minimoituu,kun päätös tehdään seuraavasti:

$$\text{tulipalo}: \ \frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} > \frac{3649}{525599} \approx 0.007$$
$$\text{väärä hälytys}: \ \text{muulloin}$$

eli pelkkien havaintojen perusteella pitää olla reilusti yli 100-kertainen luottamus siihen ettei ole tulipaloa, jos aikoo jäädä sisälle

1.3 Diskriminanttifunktiot ja päätöspinnat

- Edellisten tarkastelujen pohjalta tiedetään, että kun luokittelu perustuu joko luokitteluvirhetn:n tai riskin odotusarvon minimointiin, piirreavaruus jaetaan M:ään **päätösalueeseen** $R_1, \ldots, R_M$, kun luokkia on M kappaletta

- Mikäli luokkia ω_i ja ω_j vastaavat, luokitteluvirheen tn:n minimoivat päätösalueet ovat R_i ja R_j, määritellään **päätöspinta** (decision boundary, decision surface) seuraavasti:

$$P(\omega_i|\mathbf{x}) - P(\omega_j|\mathbf{x}) = 0 \tag{17}$$

  Toisella puolella päätöspintaa erotus on positiivinen ja toisella negatiivinen

- Toisinaan on laskennallisesti kätevämpää esittää päätöspinnan yhtälö **diskriminanttifunktioiden** $g_i(\mathbf{x}) = f(P(\omega_i|\mathbf{x}))$ avulla. $f(\cdot)$ voi olla mikä tahansa monotonisesti kasvava, jatkuva funktio

- Tällöin minivirhetn:n tuottama päätössääntö (10) saa seuraavan muodon:

$$R_i: \; g_i(\mathbf{x}) > g_j(\mathbf{x}) \; \forall j \neq i \tag{18}$$

ja päätöspinnat on määritelty seuraavasti:

$$g_{ij}(\mathbf{x}) = g_i(\mathbf{x}) - g_j(\mathbf{x}) = 0 \tag{19}$$