Klasifikacija utemeljena na Bayesovoj decizijskoj teoriji
Statistički pristup raspoznavanju uzoraka

Uvod u raspoznavanje uzoraka

ak. god. 2019/2020

Prof. dr. sc. Slobodan Ribarić

Statistički pristup raspoznavanju uzoraka

Neki osnovni pojmovi:

- Uvjetne (relativne) vjerojatnosti
- Adicijski i multiplikacijski teorem
- Bayesov teorem

Statistički pristup raspoznavanju uzoraka

Formule za uvjetne vjerojatnosti mogu se jednostavno izvesti ako se pođe od relativnih frekvencija

- Promatrajmo dva događaja, A i B, kako se pojavljuju unutar n pokusa.

Pri tome mogu nastupiti ove mogućnosti:

1) A nastupi, B ne nastupi, (n_1)
2) B nastupi, A ne nastupi, (n_2)
3) Nastupe A i B, (n_3)
4) Ne nastupe ni A ni B, (n_4)

Neka su n_1, n_2, n_3 i n_4 brojevi koji nam predočuju koliko puta je nastupila prva, druga, treća , odnosno četvrta mogućnost, pri čemu vrijedi:

$$n_1 + n_2 + n_3 + n_4 = n$$

Statistički pristup raspoznavanju uzoraka

1) A nastupi, B ne nastupi, (n_1)
2) B nastupi, A ne nastupi, (n_2)
3) Nastupe A i B, (n_3)
4) Ne nastupe ni A ni B, (n_4)

Relativne frekvencije za pojedine događaje A i B su ove:

$$f(\mathrm{A}) = \frac{n_1 + n_3}{n}$$

$$f(\mathrm{B}) = \frac{n_2 + n_3}{n}$$

Relativna frekvencija da će se dogoditi ili događaj A ili događaj B ili oba:

$$f(\mathrm{A} \cup \mathrm{B}) = \frac{n_1 + n_2 + n_3}{n}$$

Relativna frekvencija da će se dogoditi događaj A i događaj B:

$$f(\mathrm{A\,B}) = \frac{n_3}{n}$$

Statistički pristup raspoznavanju uzoraka

1) A nastupi, B ne nastupi, (n_1)
2) B nastupi, A ne nastupi, (n_2)
3) Nastupe A i B, (n_3)
4) Ne nastupe ni A ni B, (n_4)

Relativna frekvencija od B pod uvjetom da je nastupio A:

$$f(\mathrm{B} \mid \mathrm{A}) = \frac{n_3}{n_1 + n_3}$$

Relativna frekvencija od A pod uvjetom da je nastupio B:

$$f(\mathrm{A} \mid \mathrm{B}) = \frac{n_3}{n_2 + n_3}$$

Statistički pristup raspoznavanju uzoraka

Pokažimo da vrijedi:

$$f(A \cup B) = f(A) + f(B) - f(AB)$$

$$\frac{n_1 + n_2 + n_3}{n} = \frac{n_1 + n_3}{n} + \frac{n_2 + n_3}{n} - \frac{n_3}{n} = \frac{n_1 + n_2 + n_3}{n}$$

$$f(A \cup B) = \frac{n_1 + n_2 + n_3}{n}$$

$$f(A) = \frac{n_1 + n_3}{n}$$

$$f(B) = \frac{n_2 + n_3}{n}$$

$$f(AB) = \frac{n_3}{n}$$

Statistički pristup raspoznavanju uzoraka

Pokažimo da vrijedi:

$$f(\mathrm{A\,B}) = f(\mathrm{A})f(\mathrm{B}|\mathrm{A}) = f(\mathrm{B})f(A|\mathrm{B})$$

$$\frac{n_3}{n} = \frac{n_1 + n_3}{n} \times \frac{n_3}{n_1 + n_3} = \frac{n_3}{n}$$

$$f(\mathrm{A\,B}) = \frac{n_3}{n}$$

$$f(\mathrm{A}) = \frac{n_1 + n_3}{n}$$

$$f(\mathrm{B} \mid \mathrm{A}) = \frac{n_3}{n_1 + n_3}$$

Statistički pristup raspoznavanju uzoraka

- Pretpostavimo da broj pokusa → ∞; relativne frekvencije → vjerojatnosti

onda se

$$f(A \cup B) = f(A) + f(B) - f(AB)$$

i

$$f(AB) = f(A)f(B|A) = f(B)f(A|B)$$

pretvaraju u:

$$P(A \cup B) = P(A) + P(B) - P(AB)$$ adicijski teorem

$$P(AB) = P(A)P(B|A) = P(B)P(B|A)$$ multiplikacijski teorem

Statistički pristup raspoznavanju uzoraka

Bayesov teorem

Polazimo od formule $P(AB) = P(A)P(B|A)$, odnosno $P(AB) = P(B)P(A|B)$

/vjerojatnost da se dogode i događaj A i događaj B jednak je produktu apsolutne vjerojatnosti događaja A i relativne vjerojatnosti događaja B, uz uvjet da se dogodi događaj A i obratno /

Iz gornjih jednadžbi sljedi: $P(B|A) = \frac{P(AB)}{P(A)} = \frac{P(B)P(A|B)}{P(A)}$ Bayesov teorem (1764. god.)

Formula pokazuje nam kolika je vjerojatnost da će se dogoditi događaj B, ako se dogodio događaj A – važno nalazimo vjerojatnost $P(B|A)$ pomoću vjerojatnosti $P(A|B)$

Statistički pristup raspoznavanju uzoraka

Bayesov teorem (nastavak)

- Pomoću relativnih frekvencija pokažimo da vrijedi izraz za Bayesov teorem:

$$\mathrm{P}(B|A) = \frac{P(AB)}{P(A)} = \frac{P(B)\mathrm{P}(A|B)}{P(A)}$$

$$f(\mathrm{B}\,|\,\mathrm{A}) = \frac{n_3}{n_1 + n_3}$$

$$f(\mathrm{A\,B}) = \frac{n_3}{n}$$

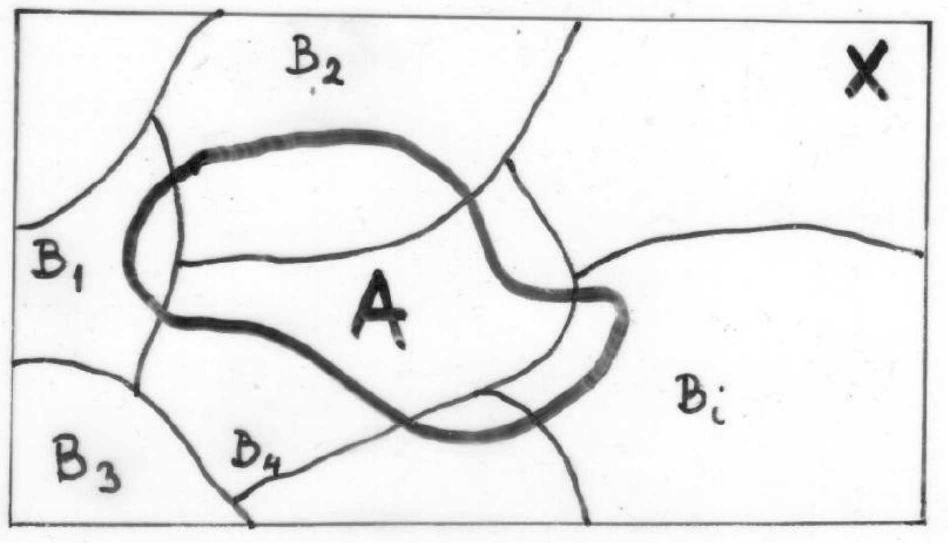
$$f(\mathrm{A}) = \frac{n_1 + n_3}{n}$$

$$\frac{n_3}{n_1 + n_3} = \frac{\frac{n_3}{n}}{\frac{n_1 + n_3}{n}} = \frac{\frac{n_2 + n_3}{n} \times \frac{n_3}{n_2 + n_3}}{\frac{n_1 + n_3}{n}} = \frac{n_3}{n_1 + n_3}$$

Statistički pristup raspoznavanju uzoraka

Bayesov teorem (nastavak)

- neka je X prostor elementarnih događaja
- $B_1, B_2, \ldots, B_n$ njegov rastav na disjunktivne dijelove (događaje koji se međusobno isključuju)
- neka je A događaj definiran na istom prostoru X i vrijedi $P(A) > 0$

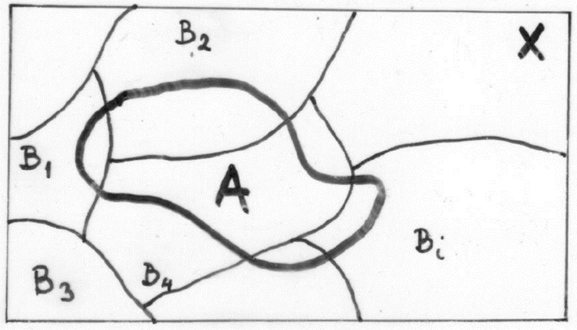


- za svaki događaj B_k; k = 1, 2, …, n vrijedi uvjetna vjerojatnost

$$f(A\,B) = f(A) f(B|A) = f(B) f(A|B)$$

$$P(B_k|A) = \frac{P(B_k A)}{P(A)}$$

Statistički pristup raspoznavanju uzoraka



- događaj A može se realizirati istodobno ili s događajem B_1, ili s događajem B_2, itd.
- to možemo zapisati:

$$A = B_1A + B_2A + \ldots + B_nA$$

- budući da se događaji B_iA, i = 1, 2, …, n međusobno isključuju vrijedi:

$$P(A) = \sum_{i=1}^{n} P(B_i\,A) = \sum_{i=1}^{n} P(B_i)\,P(A|B_i)$$

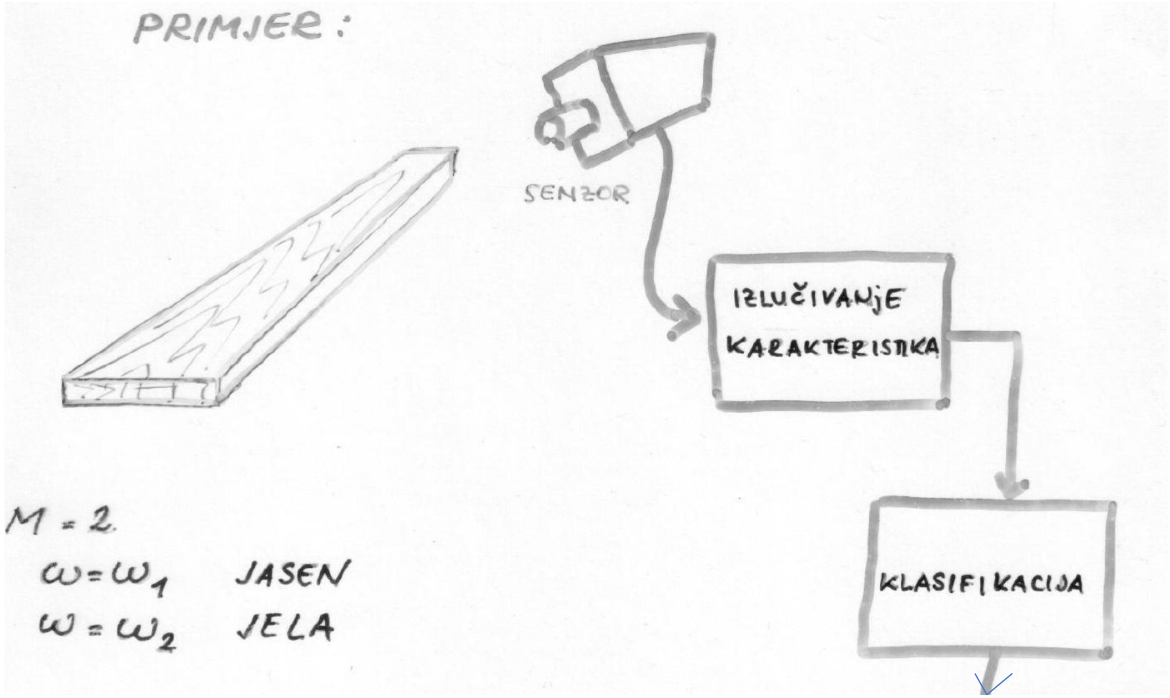
- Uvrštavanjem izraza za P(A) u $P(B_k|A) = \frac{P(B_kA)}{P(A)}$ dobivamo:

$$P(B_k|A) = \frac{P(B_k)P(A|B_k)}{\sum_{i=1}^{n} P(B_i)\,P(A|B_i)}$$

Bayesov teorem

Statistički pristup raspoznavanju uzoraka

Primjer: deterministički pristup – statistički pristup



odluka

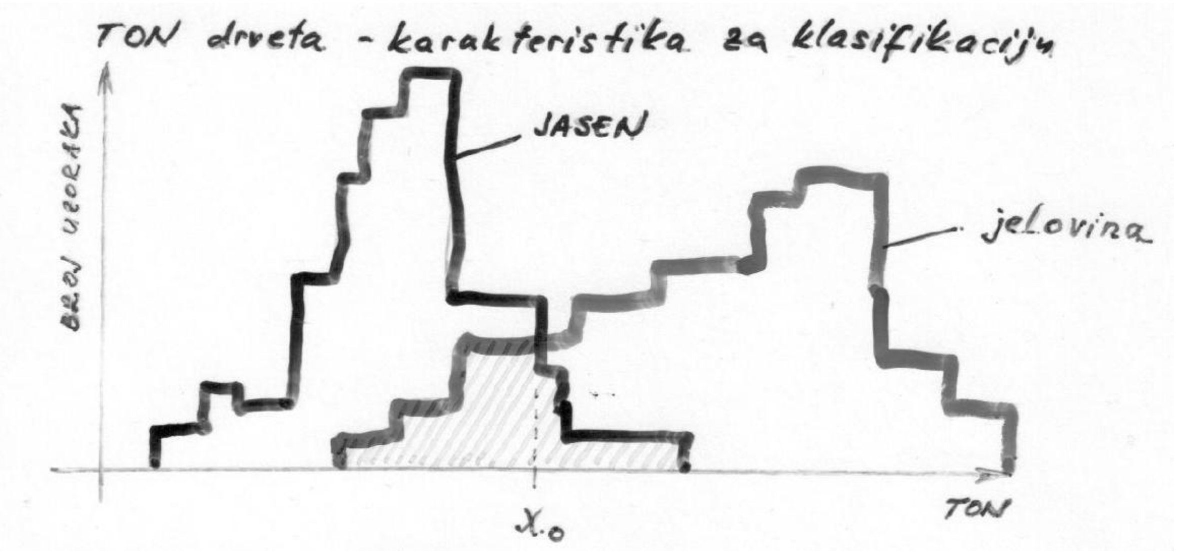
Statistički pristup raspoznavanju uzoraka

Primjer: deterministički pristup – statistički pristup (nastavak)

- znamo da je jela (daske iz razreda ω_2) svjetlija
- na temelju skupa za učenje

(daske s označenom klasifikacijom)

određujemo histogram



- ako odluku o pripadnosti razredu

donosimo samo na temelju svjetline

drveta – vrijednost praga x_o nam

određuje razrede

- prag x_o NIJE dovoljno dobar – klasifikator griješi

Statistički pristup raspoznavanju uzoraka

Primjer: determinističi pristup – statistički pristup (nastavak)

- ekspert nam je rekao da se daske mogu razlikovati i po izraženosti strukture drveta:

jasen (razred ω_1) ima izraženiju strukturu drveta

- uključujemo i tu značajku te dobivamo dvodimenzionalni vektor značajki:

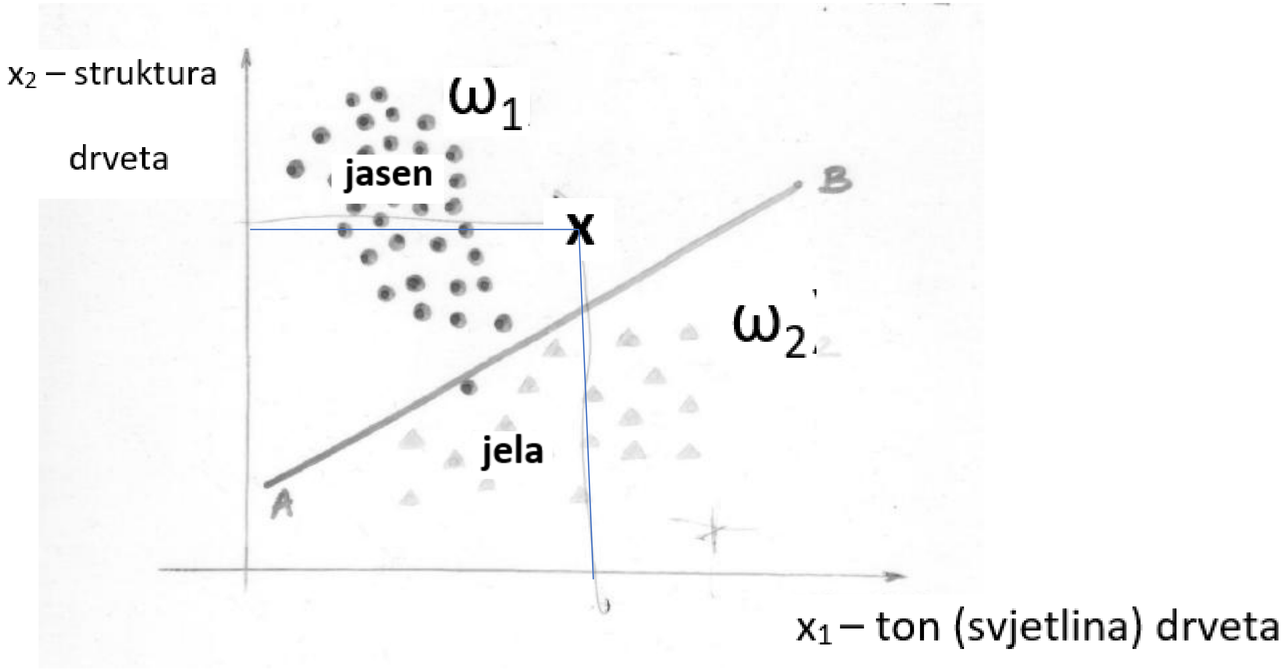
$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

x_1 – ton (svjetlina) drveta

x_2 – struktura drveta

Statistički pristup raspoznavanju uzoraka

Primjer: deterministički pristup – statistički pristup (nastavak)



Statistički pristup raspoznavanju uzoraka

Primjer: deterministički pristup – statistički pristup (nastavak)

- Promatramo pojavljivanje daske na pokretnoj vrpci: slučajno se pojavljuje vrsta daske (jasen ili jela)

- „igra prirode" – stanje prirode (ili vanjskog svijeta) neka je **ω**:

$$\boldsymbol{\omega} = \omega_1 \text{ jasen}$$

$$\boldsymbol{\omega} = \omega_2 \text{ jela}$$

- budući da se stanje prirode ne može predvidjeti **ω** se promatra kao slučajna varijabla

- pretpostavimo da pilana proizvodi godišnje jednak broj jasenovih i jelovih dasaka
  - sljedeća daska ima jednaku vjerojatnost da bude jasen ili pak jela

Statistički pristup raspoznavanju uzoraka

Primjer: deterministički pristup – statistički pristup (nastavak)

- apriorna vjerojatnost $P(\omega_1)$ da daska jasenova
- apriorna vjerojatnost $P(\omega_2)$ da daska iz jelovine
- naše „znanje o procesu”: $P(\omega_1)$ i $P(\omega_2)$

$$P(\omega_1) \geq 0 \text{ i } P(\omega_2) \geq 0 \text{ i } P(\omega_1) + P(\omega_2) = 1$$

Pretpostavimo da odluku moramo donijeti samo na temelju apriorne vjerojatnosti:

Decizijsko pravilo:

odlučujem se za ω_1 akko je $P(\omega_1) > P(\omega_2)$, inače odlučujem se za ω_2

Statistički pristup raspoznavanju uzoraka

Primjer: deterministički pristup – statistički pristup (nastavak)

Decizijsko pravilo:

odlučujem se za ω_1 akko je $P(\omega_1) > P(\omega_2)$, inače odlučujem se za ω_2

„čudno” pravilo: „klasifikator” uvijek donosi istu odluku iako znamo da se pojavljuju obje vrste dasaka!

- međutim, ako je $P(\omega_1) >> P(\omega_2)$ – možda bi i „klasifikator” dobro radio – pogreška klasifikatora bi bila jednaka $P(\omega_2)$!
- ako je $P(\omega_1) \sim P(\omega_2)$ ili $P(\omega_1) = P(\omega_2)$ pogreška 50%!

Statistički pristup raspoznavanju uzoraka

Primjer: deterministički pristup – statistički pristup (nastavak)

- uzmimo u obzir svjetlinu, odnosno ton drveta!

  - znamo da različiti uzorci dasaka imaju različitu svjetlinu:

Neka je x kontinuirana (neprekinuta) slučajna varijabla čija distribucija zavisi od stanja prirode

$p(x|\omega_j)$ - uvjetna gustoća vjerojatnosti (engl. state-conditional probability density function)

Statistički pristup raspoznavanju uzoraka

Primjer: deterministički pristup – statistički pristup (nastavak)

Općenito:

Funkcija gustoće vjerojatnosti za kontinuiranu slučajnu varijablu x je takva funkcija $f(x)$

Koja ima svojstva:

1. $f(x) \geq 0$
2. $\int_{-\infty}^{+\infty} f(x)dx = 1$
3. $\int_{x_1}^{x_2} f(x)dx = P\{x_1 < x < x_2\}$

Statistički pristup raspoznavanju uzoraka

Primjer: deterministički pristup – statistički pristup (nastavak)

Općenito:

- Za diskretnu slučajnu varijablu koja može poprimiti vrijednost $x_1, x_2, \dots, x_n$ funkcija vjerojatnosti je ona za koju vrijedi:

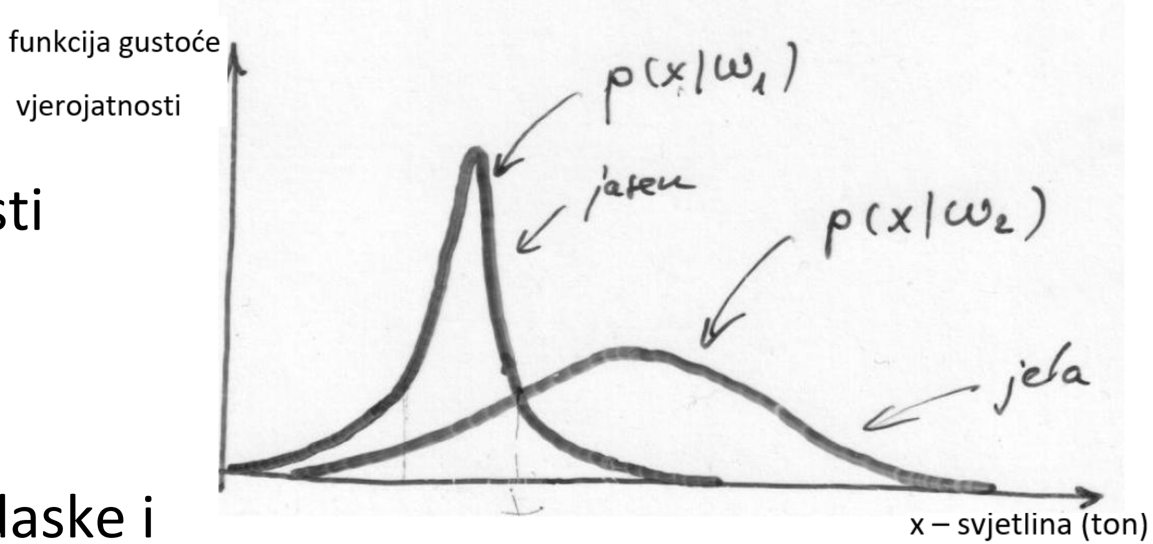
1. $f(x_i) \geq 0$
2. $\sum_{i=1}^{n} f(x_i) = 1$
3. $f(x_i) = P(X = x_i)$

Statistički pristup raspoznavanju uzoraka

Primjer: deterministički pristup – statistički pristup (nastavak)

Pretpostavka:
-znamo apriorne vjerojatnosti:
$P(\omega_1)$ i $P(\omega_2)$ i uvjetne gustoće vjerojatnosti
$p(x|\omega_j)$, j = 1, 2

vjerodostojnost (engl. Likehood of ω_j)



Pretpostavimo da smo izmjerili svjetlinu daske i dobili vrijednost *x* !

Kako to mjerenje utječe na odluku o pravom stanju prirode (odluka: daska je iz razreda ω_1 ili ω_2)?

Statistički pristup raspoznavanju uzoraka

Primjer: deterministički pristup – statistički pristup (nastavak)

Kako to mjerenje utječe na odluku o pravom stanju prirode (odluka: daska je iz razreda ω_1 ili ω_2)?

Primjenimo Bayesov teorem:

$$P(B_k|A) = \frac{P(B_k)P(A|B_k)}{\sum_{i=1}^{n} P(B_i)\, P(A|B_i)}$$

$$P(\omega_j|x) = \frac{p(x|\omega_j)P(\omega_j)}{p(x)}$$

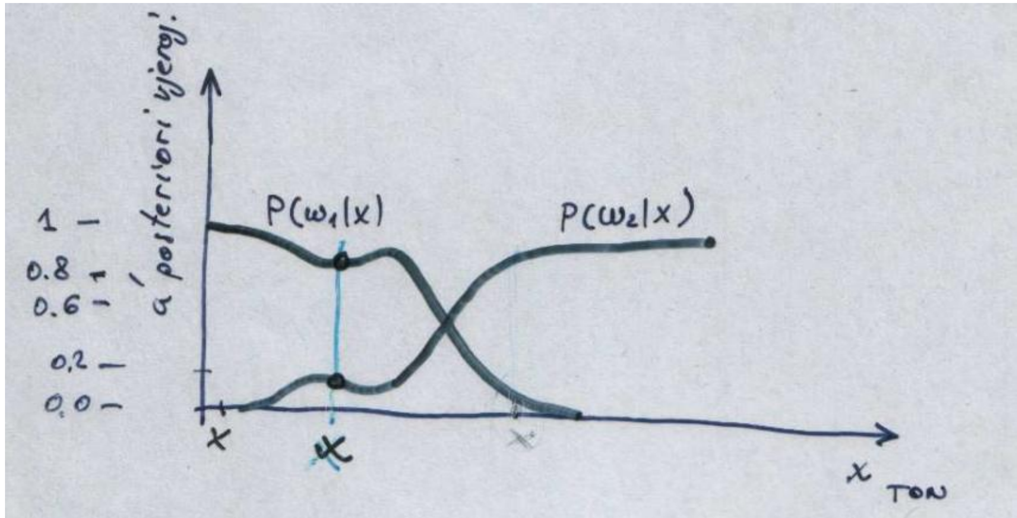
$$p(x) = \sum_{j=1}^{2} p(x|\omega_j)\, P(\omega_j)$$

Bayesov teorem (pravilo) – pokazuje nam kako promatrana (izmjerena) vrijednost *x* *mijenja* apriornu vjerojatnost $P(\omega_j)$ u aposteriornu vjerojatnost $P(\omega_j|x)$,

Statistički pristup raspoznavanju uzoraka

Primjer: deterministički pristup – statistički pristup (nastavak)

- na primjer, ako je $P(\omega_1) = 1/3$ i $P(\omega_2) = 2/3$ onda aposteriorne vjerojatnosti izgledaju ovako



Bayesovo pravilo:

- imamo promatranje (mjerenje x) za koje je

$$P(\omega_1|x) > P(\omega_2|x)$$

ODLUKA: stanje prirode je ω_1

Statistički pristup raspoznavanju uzoraka

Primjer: deterministički pristup – statistički pristup (nastavak)

Vjerojatnost pogreške?

(vjerojatnost pogreške pri svakoj odluci: P(error|x))

$$P(error|x) = \begin{cases} P(\omega_1|x) & \text{ako se odlučimo za } \omega_2 \\ P(\omega_2|x) & \text{ako se odlučimo za } \omega_1 \end{cases}$$

Minimiziramo vjerojatnost pogreške donošenjem

sljedećih odluka:

ω_1 ako je $P(\omega_1|x) > P(\omega_2|x)$

ili

ω_2 ako je $P(\omega_2|x) > P(\omega_1|x)$

Statistički pristup raspoznavanju uzoraka

Primjer: deterministički pristup – statistički pristup (nastavak)

Da li pravilo mininizira srednju vrijednost pogreške?

Da!

$$P(error) = \int_{-\infty}^{+\infty} P(error, x)dx$$

$$= \int_{-\infty}^{+\infty} P(error|x)p(x)dx$$

- jer ako je za svaki *x* vjerojatnost P(error|*x*) *najmanja* moguća vrijednost tada i integral ima najmanju vrijednost.

Statistički pristup raspoznavanju uzoraka

Bayesovo decizijsko pravilo za minimizaciju vjerojatnosti pogreške:

odluči se za ω_1 ako je $P(\omega_1|x) > P(\omega_2|x)$, inače odluči se za ω_2

Promotrimo izraze:

$$P(\omega_j|x) = \frac{p(x|\omega_j)P(\omega_j)}{p(x)}$$

nije bitno! Služi samo za skaliranje i osigurava da $P(\omega_1|x) + P(\omega_2|x) = 1$

$$p(x) = \sum_{j=1}^{2} p(x|\omega_j)P(\omega_j)$$

Modificirano pravilo:

odluči se za ω_1 ako je $p(x|\omega_1)P(\omega_1) > p(x|\omega_2)P(\omega_2)$, inače ω_2

Statistički pristup raspoznavanju uzoraka

Pretpostavimo dvije sljedeće situacije:

- ako je za neki *x*

$$p(x|\omega_1) = p(x|\omega_2)$$

odluka se temelji na apriornoj vjerojatnosti!

- ako je $P(\omega_1) = P(\omega_2)$

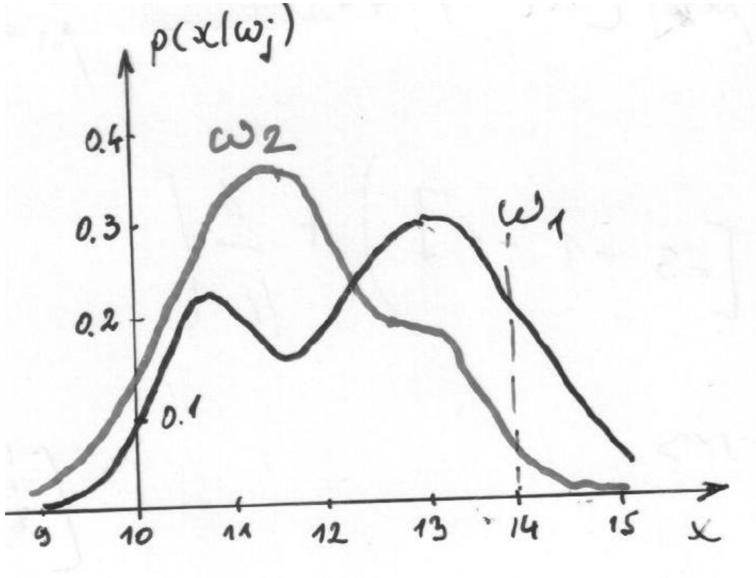
odluka se temelji na $p(x|\omega_j)$

Statistički pristup raspoznavanju uzoraka

Primjer:

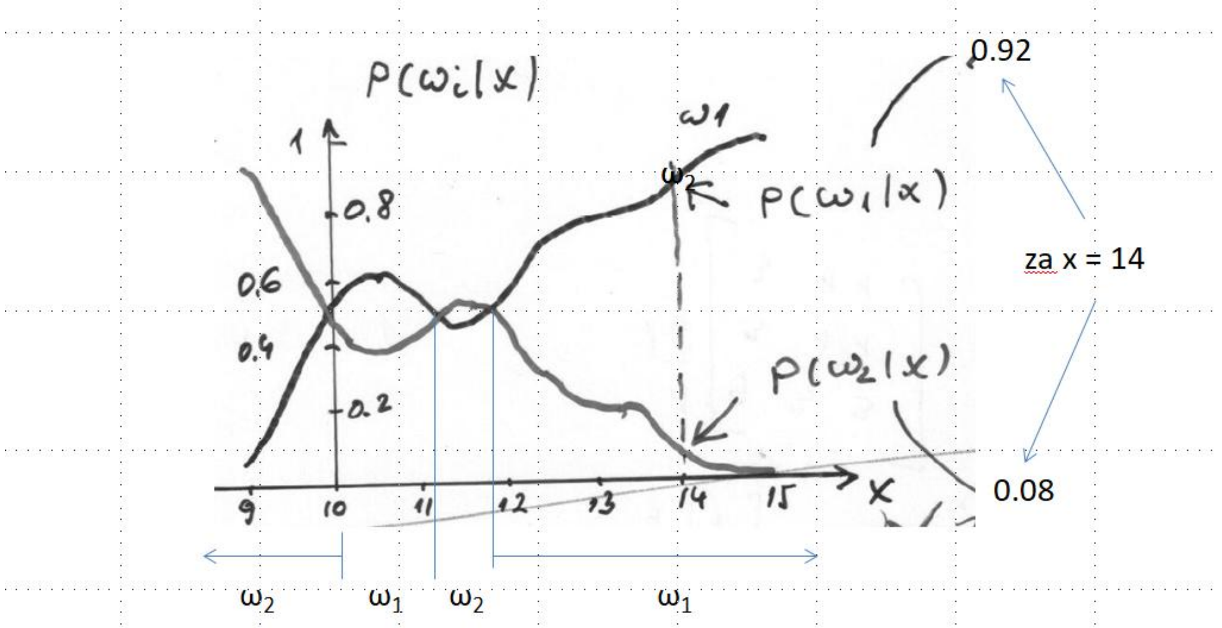
- pretpostavimo da su poznate vjerodostojnosti:
- priorne vjerojatnosti $P(\omega_1) = 2/3$ i $P(\omega_2) = 1/3$

Odredite graf aposteriornih vjerojatnosti i definirajte područja koja odgovaraju razredima ω_1 i ω_2!



Statistički pristup raspoznavanju uzoraka

Skica rješenja:



Statistički pristup raspoznavanju uzoraka

Dva jednostavna primjera:

1. **Primjer**

Nema mjerenja! Broj razreda M = 2; poznate apriorne vjerojatnosti $P(\omega_1) = 0.7$ i $P(\omega_2) = 0.3$

Pozor: nema mjerenja i nemamo vektor značajki!

Kakvo klasifikacijsko pravilo trebamo primijeniti?

- ako je $P(\omega_1) > P(\omega_2)$ onda x razvrstavamo u ω_1

Označimo vjerojatnost klasifikacijske pogreške sa P(error);

U našem slučaju klasifikacijsko pravilo: Uvijek izaberi ω_1 jer je $P(\omega_1) > P(\omega_2)$!

Taj pristup vodi do pogreške:

$$P(\text{error}) = P(\text{izaberi } \omega_1 | \omega_2)\, P(\omega_2)$$

$$= 1 \cdot 0.3 = 0.3$$

Statistički pristup raspoznavanju uzoraka

2. Primjer

Jedno mjerenje - dimenzionalnost vektora značajki n= 1;

Dva razreda M = 2;

Pretpostavimo da su apriorne vjerojatnosti jednaka $P(\omega_1) = P(\omega_2)$;

Pretpostavimo Gaussovu gustoću vjerojatnosti:

$$p(x|\omega_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu_i}{\sigma_i}\right)^2\right\}$$

Poseban slučaj: standardna devijacija $\sigma_1 = \sigma_2$ ali $\mu_1 \neq \mu_2$; $\mu_2 > \mu_1$

Statistički pristup raspoznavanju uzoraka

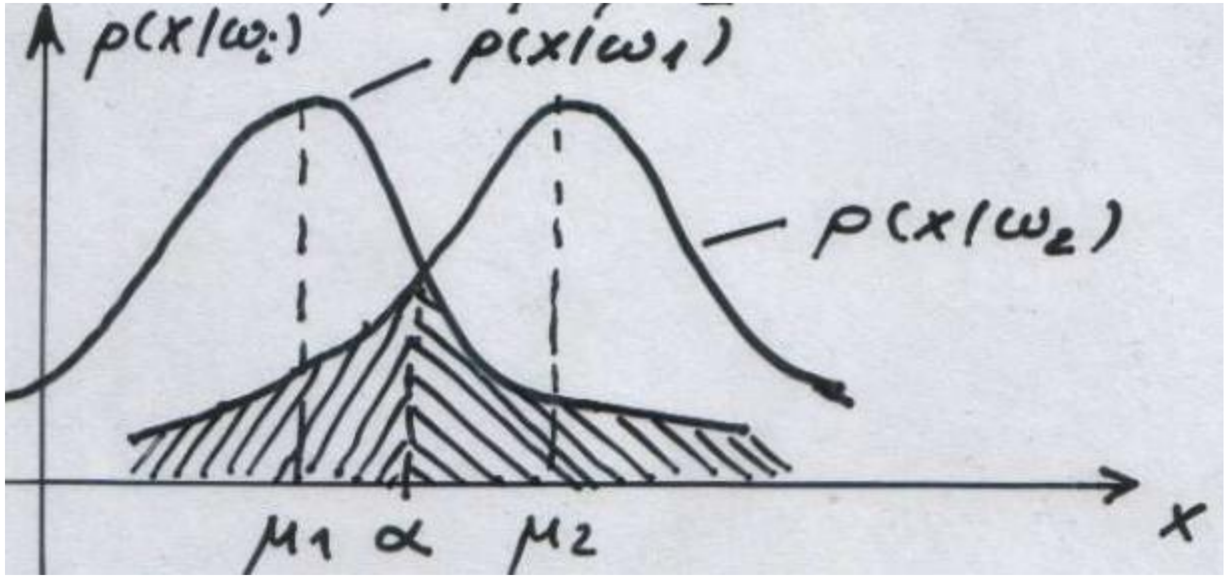
2. Primjer (nastavak)

$P(\omega_1) = P(\omega_2)$

Bayes:

$$P(\omega_j|x) = \frac{p(x|\omega_j)P(\omega_j)}{p(x)}$$

$$p(x) = \sum_{j=1}^{2} p(x|\omega_j)P(\omega_j)$$



Modificirano pravilo:

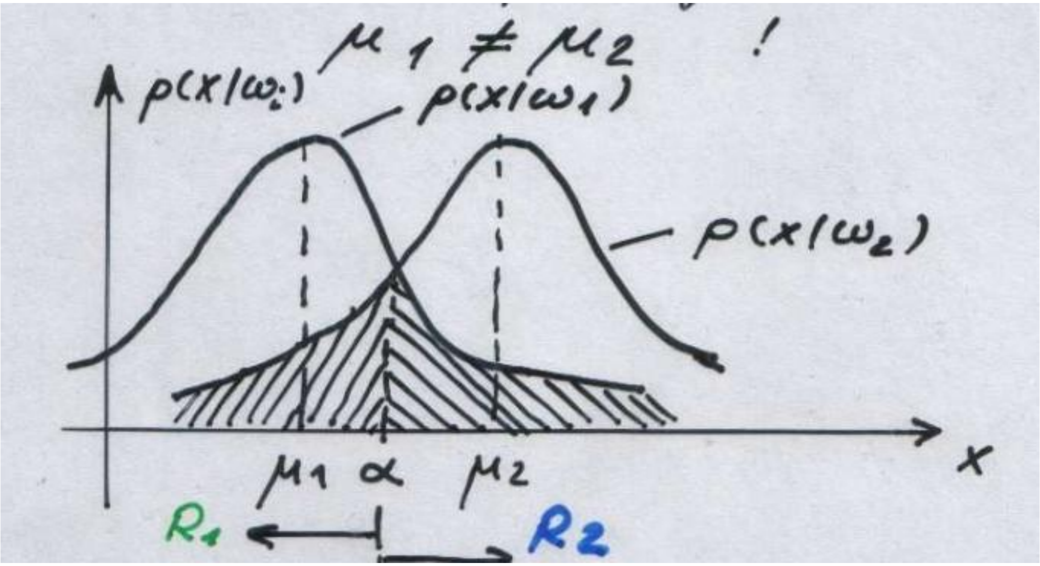
odluči se za ω_1 ako je $p(x|\omega_1)P(\omega_1) > p(x|\omega_2)P(\omega_2)$, inače ω_2

Statistički pristup raspoznavanju uzoraka

2. Primjer (nastavak)

Decizijsko pravilo:
izaberi ω_1 ako je $p(x|\omega_1) > p(x|\omega_2)$ inače ω_2

Prag α određuje klasifikaciju na temelju vrijednosti mjerenja *x*
R_1 – označava područje razreda ω_1
R_2 – označava područje razreda ω_2



Statistički pristup raspoznavanju uzoraka

2. Primjer (nastavak)

- Za izabranu klasifikacijsku strategiju možemo razmotriti pogrešku klasifikacije:

P(error) = P(x je dodjeljen pogrešnom razredu)

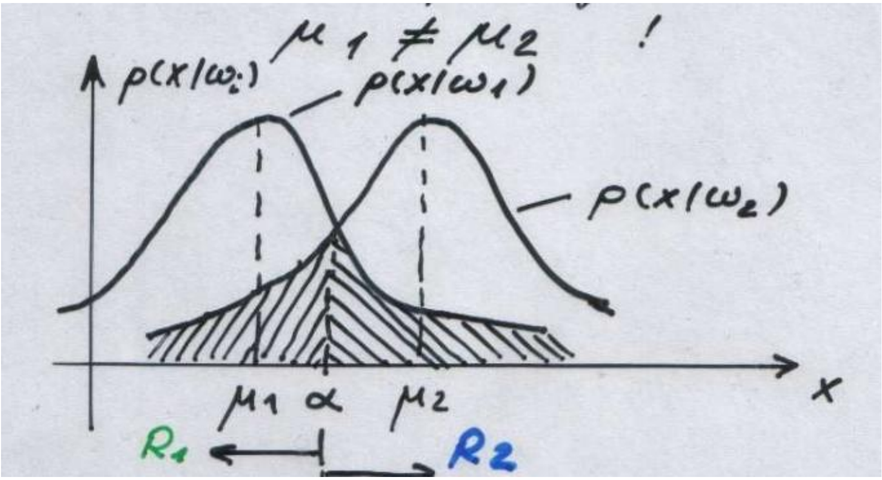
za   M = 2 imamo:

P(error)= P(izaberi ω_1, a stvarno je x iz razreda ω_2) + P(izaberi ω_2, a stvarno je x iz razreda ω_1)

$$P(\text{error}) = P(\text{error}|\omega_1)P(\omega_1) + P(\text{error}|\omega_2)P(\omega_2)$$

$$P(\text{error}) = P(x \in R_2|\omega_1)P(\omega_1) + P(x \in R_1|\omega_2)P(\omega_2)$$

$$= P(x < \alpha|\omega_2)P(\omega_2) + P(x > \alpha|\omega_1)P(\omega_1)$$



Statistički pristup raspoznavanju uzoraka

2. Primjer (nastavak)

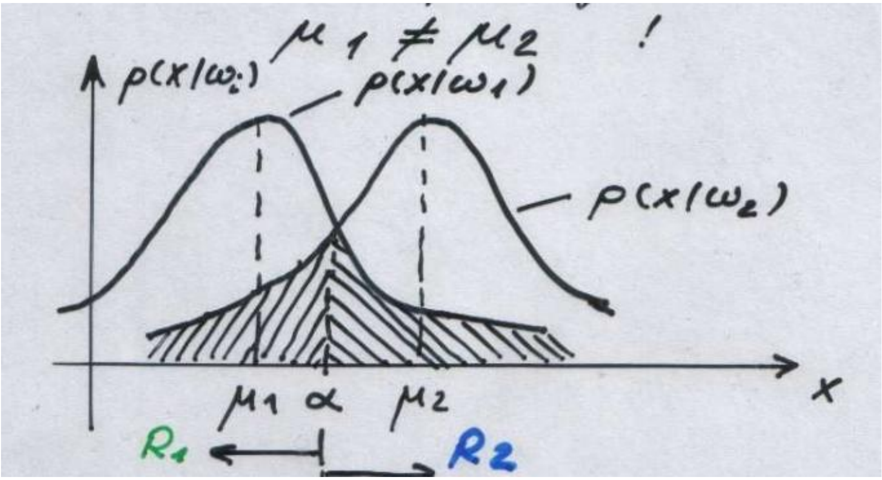
$$P(\text{error}) = P(\text{error}|\omega_1)P(\omega_1) + P(\text{error}|\omega_2)P(\omega_2)$$

$$P(\text{error}) = P(x \in R_2|\omega_1)P(\omega_1) + P(x \in R_1|\omega_2)P(\omega_2)$$

$$= P(x < \alpha|\omega_2)P(\omega_2) + P(x > \alpha|\omega_1)P(\omega_1)$$

$$\int_{-\infty}^{\alpha} p(\xi \,|\, \omega_2)\, d\xi$$

$$\int_{\alpha}^{+\infty} p(\xi \,|\, \omega_1)\, d\xi$$



Integrali određuju šrafiranu površinu!

Statistički pristup raspoznavanju uzoraka

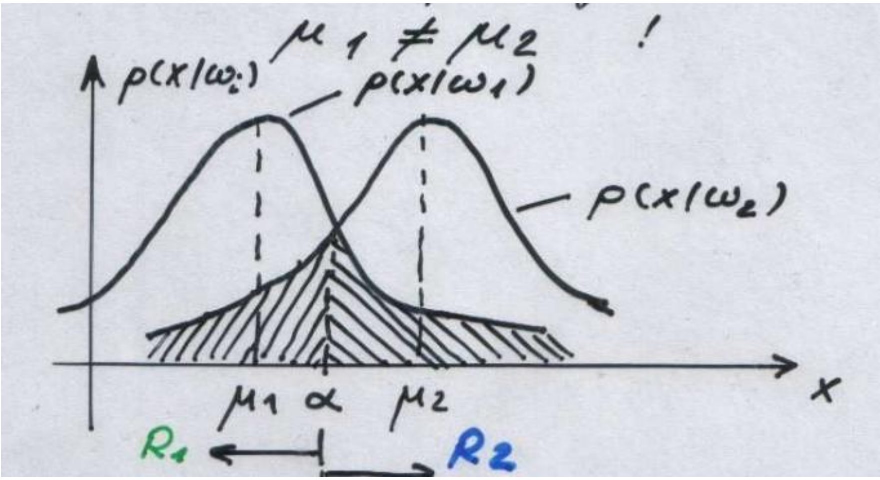
2. Primjer (nastavak)

- Izbor praga α vodi dijeljenju prostora R^1 potprostore (područja) R_1 i R_2

za slučaj $P(\omega_1) = P(\omega_2)$ možemo reči:

"Dodijeli x u R_i pri čemu je x najbliži μ_i–ju"

Kakav je to klasifikator?



Statistički pristup raspoznavanju uzoraka

- Odgovor:

Klasifikator na temelju udaljenosti!

Statistički pristup raspoznavanju uzoraka

Potrebna poopćenja:

1. Upotreba više od jedne značajke,
2. Više od dva stanja prirode (broj razreda $M > 2$),
3. Dopustiti ćemo i akciju umjesto same odluke o stanju prirode,
4. Uvest ćemo funkciju gubitka kao općenitiju mjeru u odnosu na vjerojatnost pogreške

Statistički pristup raspoznavanju uzoraka

1. **Upotreba više od jedne značajke**

- umjesto skalara *x* koristimo vektor značajki $\mathbf{x} \in R^n$; $\mathbf{x}$ je *n*-dimenzionalni vektor iz Euklidskog prostora R^n

Primjer

M = 2;

ω_1 = jasen

ω_2 = jela

vektor značajki: $\mathbf{x} = (x_1, x_2)^T$; n = 2

x_1 – ton drveta; x_2 – svjetlina drveta

$p(x_1, x_2|\omega_i) = p(\mathbf{x}|\omega_i)$; i = 1, 2

$P(\omega_i)$; i= 1, 2

Statistički pristup raspoznavanju uzoraka

- Multidimenzionalna Gaussova distribucija

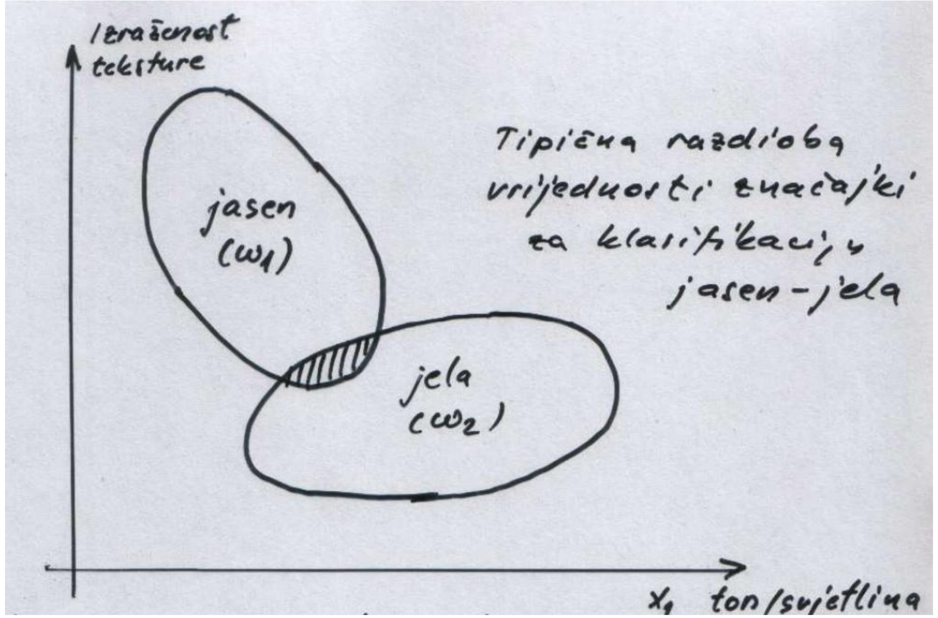
$$p(\mathbf{x}) = (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu})\right]$$

$\mathbf{x}$ - n-dimenzionalni vektor

$\boldsymbol{\mu}$ - n-dimenzionalni vektor srednjih vrijednosti

Σ - kovarijacijska matrica $n \times n$

$|\Sigma|$ - determinanta kovarijacijske matrice



Statistički pristup raspoznavanju uzoraka

Primjer (nastavak)

Pretpostavimo da se četiri puta više proizvede jasenovih dasaka negoli jelovih:

$P(\omega_1) = 0.8; \quad \omega_1 =$ jasen

$P(\omega_2) = 0.2; \quad \omega_2 =$ jela

Treba odrediti strategiju za klasifikaciju dasaka na temelju mjerenja **x** i apriornih vjerojatnosti $P(\omega_1)$ i $P(\omega_2)$!

- Uz pretpostavku da su poznate funkcije uvjetne gustoće vjerojatnosti

$$p(\mathbf{x}|\omega_i); \; i = 1, 2$$

strategiju temeljimo na Bayesovom pravilu

Statistički pristup raspoznavanju uzoraka

Bayesovo pravilo:

$$P(\omega_i \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid \omega_i) P(\omega_i)}{p(\mathbf{x})}$$

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

$$\text{izaberi} \begin{cases} \omega_1 & \text{ako } p(\mathbf{x}|\omega_1)P(\omega_1) > p(\mathbf{x}|\omega_2)P(\omega_2) \\ \omega_2 & \text{ako } p(\mathbf{x}|\omega_2)P(\omega_2) > p(\mathbf{x}|\omega_1)P(\omega_1) \end{cases}$$

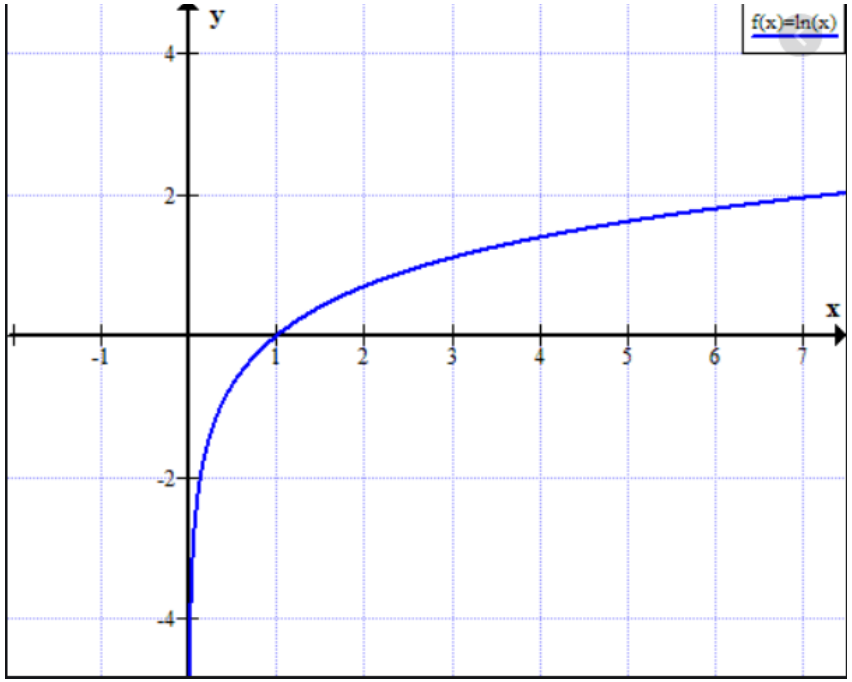
VAŽNO:

- u klasifikacijskom pravilu kombinirane su apriorna vjerojatnost i mjerno zavisna informacija
- Monotono ne padajuća funkcija od $P(\omega_i|\mathbf{x})$ može se rabiti za gornji test!

Statistički pristup raspoznavanju uzoraka

Monotono rastuća (nepadajuća) funkcija od $P(\omega_i|\mathbf{x})$:

Npr. $\log_2 P(\omega_i|\mathbf{x})$ ili $\ln P(\omega_i|\mathbf{x})$



Pravila (općenito):

$\log_b(xy) = \log_b(x) + \log_b(y)$

$\log_b(x/y) = \log_b(x) - \log_b(y)$

$\log_b(x^p) = p \log_b(x)$

$\log_b(\sqrt[p]{x}) = 1/p \log_b(x)$

$\ln(e^x) = x$

Statistički pristup raspoznavanju uzoraka

Primjer:

M = 3 razreda; dimenzionalnost n = 2; Gaussova razdioba

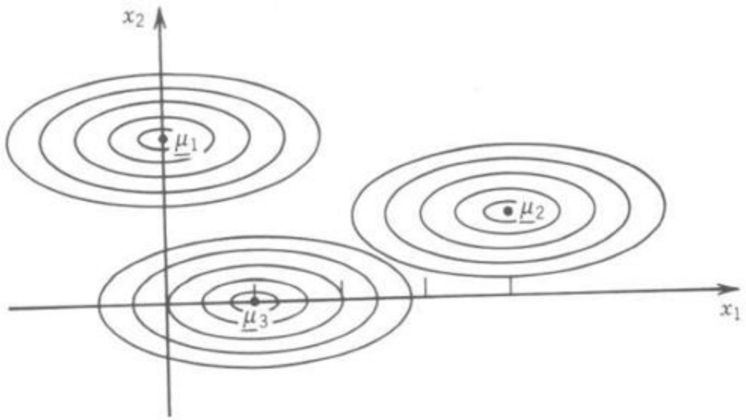
$$p(\mathbf{x}) = (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu})\right]$$

Pretpostavimo da je $P(\omega_1) = P(\omega_2) = P(\omega_3)$ – apriorne vjerojatnosti razreda jednake i da su kovarijacijske matrice jednake:

$$\Sigma = \Sigma_i;\ i = 1, 2, 3 \qquad \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1/2 \end{bmatrix}$$

$$\boldsymbol{\mu}_1 = (0, 2)^T;\ \boldsymbol{\mu}_2 = (4, 1)^T;\ \boldsymbol{\mu}_3 = (1, 0)^T$$

Statistički pristup raspoznavanju uzoraka



$P(\omega_1) = P(\omega_2) = P(\omega_3)$

$\boldsymbol{\mu}_1 = (0, 2)^T$; $\boldsymbol{\mu}_2 = (4, 1)^T$; $\boldsymbol{\mu}_3 = (1, 0)^T$

$$d_i(\mathbf{x}) = \log\{p(\mathbf{x}|\omega_i)\}$$

$$= \log\{(2\pi)^{-n/2} \, |\Sigma|^{-\frac{1}{2}} \exp[-\tfrac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}_i)]\}$$

$$= -\frac{1}{2}\underbrace{(\mathbf{x} - \boldsymbol{\mu}_i)^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}_i)} - \underbrace{\left(\frac{n}{2}\right)\log(2\pi) - \frac{1}{2}\log|\Sigma|}_{\text{ne zavisi od } i}$$

$$mdi^2 = (\mathbf{x} - \boldsymbol{\mu}_i)^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}_i)$$

- (kvadrat) Mahalanobisove udaljenosti

Statistički pristup raspoznavanju uzoraka

$$mdi^2 = (\mathbf{x} - \boldsymbol{\mu}_i)^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}_i)$$

Kada $d_i(\mathbf{x})$ ima najveću vrijedost?

$$d_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) - \underbrace{\left(\frac{n}{2}\right) \log(2\pi) - \frac{1}{2} \log|\Sigma|}_{\text{konstanta}}$$

Onda kada je $mdi^2\ najmanja$!

$$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1/2 \end{bmatrix} \qquad \Sigma^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

$$\text{et } A = \det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

$$A^{-1} = \frac{1}{\det A} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$$

Statistički pristup raspoznavanju uzoraka

Oblik funkcija gustoće vjerojatnosti:

- za ω_1: $$(\mathbf{x}-\boldsymbol{\mu}_i)^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)$$

$$\left[\begin{pmatrix}x_1\\x_2\end{pmatrix}-\begin{pmatrix}0\\2\end{pmatrix}\right]^T\begin{bmatrix}1&0\\0&2\end{bmatrix}\left[\begin{pmatrix}x_1\\x_2\end{pmatrix}-\begin{pmatrix}0\\2\end{pmatrix}\right] = \text{k (neka konstanta)}$$

$$\begin{pmatrix}x_1\\x_2-2\end{pmatrix}^T\begin{bmatrix}1&0\\0&2\end{bmatrix}\begin{pmatrix}x_1\\x_2-2\end{pmatrix} = k$$

$$(x_1, x_2-2)\begin{pmatrix}x_1\\2(x_2-2)\end{pmatrix} = k$$

$$x_1^2+2(x_2-2)^2 = k \quad \text{ili}$$

$$\left(\frac{x_1}{\sqrt{2}}\right)^2+\left(\frac{x_2-2}{1}\right)^2 = \frac{k}{2} \quad \text{jednadžba elipse}$$

Statistički pristup raspoznavanju uzoraka

Jednadžba elipse

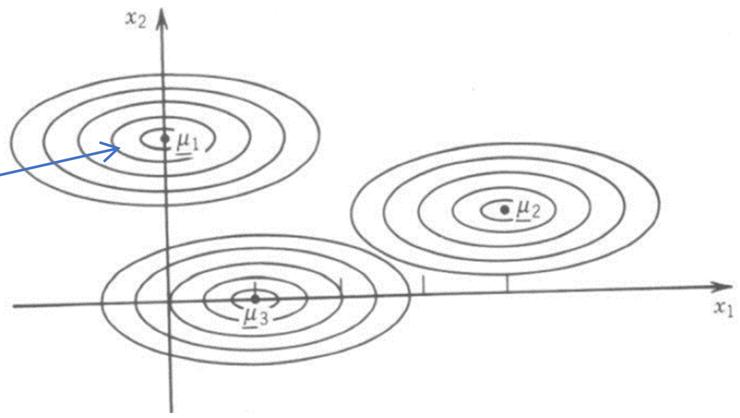
- ako se koordinatne osi poklapaju s osima elipse, tada jednadžba elipse u normalnom obliku glasi: $$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

- jednadžba elipse sa središtem točki S određenoj koordinatama S(p, q) i poluosima a i b

$$\frac{(x-p)^2}{a^2} + \frac{(y-q)^2}{b^2} = 1$$

$$\left(\frac{x_1}{\sqrt{2}}\right)^2 + \left(\frac{x_2-2}{1}\right)^2 = \frac{k}{2}$$

S(p, q) = (0, 2)



Statistički pristup raspoznavanju uzoraka

Oblik funkcija gustoće vjerojatnosti:

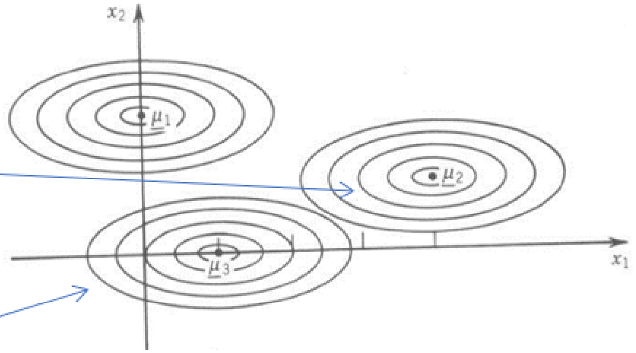
- za ω_2 (izvodimo slično kao za ω_1):

$$(x_1 - 4)^2 + 2(x_2 - 1)^2 = k$$

$$\left(\frac{x_1 - 4}{\sqrt{2}}\right)^2 + \left(\frac{x_2 - 1}{1}\right)^2 = \frac{k}{2}$$

- za ω_3:

$$\left(\frac{x_1 - 1}{\sqrt{2}}\right)^2 + \left(\frac{x_2}{1}\right)^2 = \frac{k}{2}$$



Statistički pristup raspoznavanju uzoraka

Odredimo decizijske funkcije: ; i = 1, 2, 3!

$$d_i(\mathbf{x}) = -(\mathbf{x}-\boldsymbol{\mu}_i)^T\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)$$

$$d_i(\mathbf{x}) = -\|\mathbf{x}-\boldsymbol{\mu}_i\|^2_{\Sigma^{-1}}$$ oznaka za Mahalanobisovu udaljenost

$$d_i(\mathbf{x}) = \underbrace{-\mathbf{x}^T\Sigma^{-1}\mathbf{x}}_{\text{ne zavisi od } i} + 2\boldsymbol{\mu}_i{}^T\Sigma^{-1}\mathbf{x} - \boldsymbol{\mu}_i{}^T\Sigma^{-1}\boldsymbol{\mu}_i$$

Statistički pristup raspoznavanju uzoraka

$$d_i(\mathbf{x}) = 2\boldsymbol{\mu}_i{}^T\Sigma^{-1}\mathbf{x} - \boldsymbol{\mu}_i{}^T\Sigma^{-1}\boldsymbol{\mu}_i \;/2$$

$$d_i(\mathbf{x}) = \boldsymbol{\mu}_i{}^T\Sigma^{-1}\mathbf{x} - \frac{1}{2}\boldsymbol{\mu}_i{}^T\Sigma^{-1}\boldsymbol{\mu}_i$$

$$d_i(\mathbf{x}) = \boldsymbol{\mu}_i{}^T\Sigma^{-1}\mathbf{x} - \frac{1}{2}\boldsymbol{\mu}_i{}^T\Sigma^{-1}\boldsymbol{\mu}_i$$

Σ je simetrična matrica!

$$d_i(\mathbf{x}) = (\Sigma^{-1}\boldsymbol{\mu}_i)^T\mathbf{x} - \frac{1}{2}\boldsymbol{\mu}_i{}^T\Sigma^{-1}\boldsymbol{\mu}_i$$

$$d_i(\mathbf{x}) = \mathbf{w}_i^T\mathbf{x} - \mathrm{w}_{i0}$$

$$\mathbf{w}_i = \Sigma^{-1}\boldsymbol{\mu}_i$$

$$w_{io} = +\frac{1}{2}\|\boldsymbol{\mu}_i\|^2_{\Sigma^{-1}}$$

Statistički pristup raspoznavanju uzoraka

$$d_i(\mathbf{x}) = (\Sigma^{-1}\boldsymbol{\mu}_i)^T\mathbf{x} - \frac{1}{2}\boldsymbol{\mu}_i{}^T\Sigma^{-1}\boldsymbol{\mu}_i$$

$$d_i(\mathbf{x}) = \mathbf{w}_i^T\mathbf{x} - \mathrm{w}_{io}$$

$$\mathbf{w}_i = \Sigma^{-1}\boldsymbol{\mu}_i$$

$$w_{io} = +\frac{1}{2}\|\boldsymbol{\mu}_i\|^2_{\Sigma^{-1}}$$

- Za decizijsku funkciju d_1(**x**) dobivamo:

$$d_1(\mathbf{x}) = \left[\begin{bmatrix}1 & 0\\ 0 & 2\end{bmatrix}\begin{bmatrix}0\\ 2\end{bmatrix}\right]^T \mathbf{x} - \frac{1}{2}\begin{bmatrix}0\\ 2\end{bmatrix}^T\begin{bmatrix}1 & 0\\ 0 & 2\end{bmatrix}\begin{bmatrix}0\\ 2\end{bmatrix}$$

$$d_1(\mathbf{x}) = [0 \;\; 4]\mathbf{x} - \frac{1}{2}[[02]\begin{bmatrix}1 & 0\\ 0 & 2\end{bmatrix}\begin{bmatrix}0\\ 2\end{bmatrix}]$$

$$d_1(\mathbf{x}) = [0 \;\; 4]\,\mathbf{x} - 4$$

$$d_1(\mathbf{x}) = 4\,x_2 - 4$$

Statistički pristup raspoznavanju uzoraka

Na sličan način dobivamo $d_2(\mathbf{x})$ i $d_3(\mathbf{x})$:

$$d_2(\mathbf{x}) = [4 \;\; 2]\,\mathbf{x} - 9 = 4x_1 + 2x_2 - 9$$

$$d_3(\mathbf{x}) = [1 \;\; 0]\,\mathbf{x} - \frac{1}{2} = x_1 - \frac{1}{2}$$

- za vježbu izvesti $d_2(\mathbf{x})$ i $d_3(\mathbf{x})$!

Statistički pristup raspoznavanju uzoraka

- Nađimo (hiper)ravnine , odnosno granice potprostora koji odgovaraju pojedinim razredima

- za M = 3 imamo M(M-1)/2 separatibilnih granica

$$d_i(\mathbf{x}) = d_j(\mathbf{x})$$

$$\mathbf{w}_i^T\mathbf{x} - \mathrm{w}_{0i} = \mathbf{w}_j^T\mathbf{x} - \mathrm{w}_{0j}$$

$$\mathbf{w}_{ij}^T\mathbf{x} - \mathrm{w}_{0ij} = 0$$

$$\mathbf{w}_i = \Sigma^{-1}\boldsymbol{\mu}_i$$

$$\mathbf{w}_{ij} = \mathbf{w}_i - \mathbf{w}_j = \Sigma^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)$$

$$\mathrm{w}_{0ij} = \mathrm{w}_{0i} - \mathrm{w}_{0j}$$

Statistički pristup raspoznavanju uzoraka

- za H_{12} imamo: $$\mathbf{w}_{ij} = \mathbf{w}_i - \mathbf{w}_j = \Sigma^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)$$

$$w_{0ij} = w_{0i} - w_{0j}$$

$$\mathbf{w}_{12} = \mathbf{w}_1 - \mathbf{w}_2 = \begin{bmatrix} 0 \\ 4 \end{bmatrix} - \begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} -4 \\ 2 \end{bmatrix}$$

$$w_{012} = w_{01} - w_{02} = -4 - (-9) = 5$$

$$4x_1 - 2x_2 - 5 = 0$$

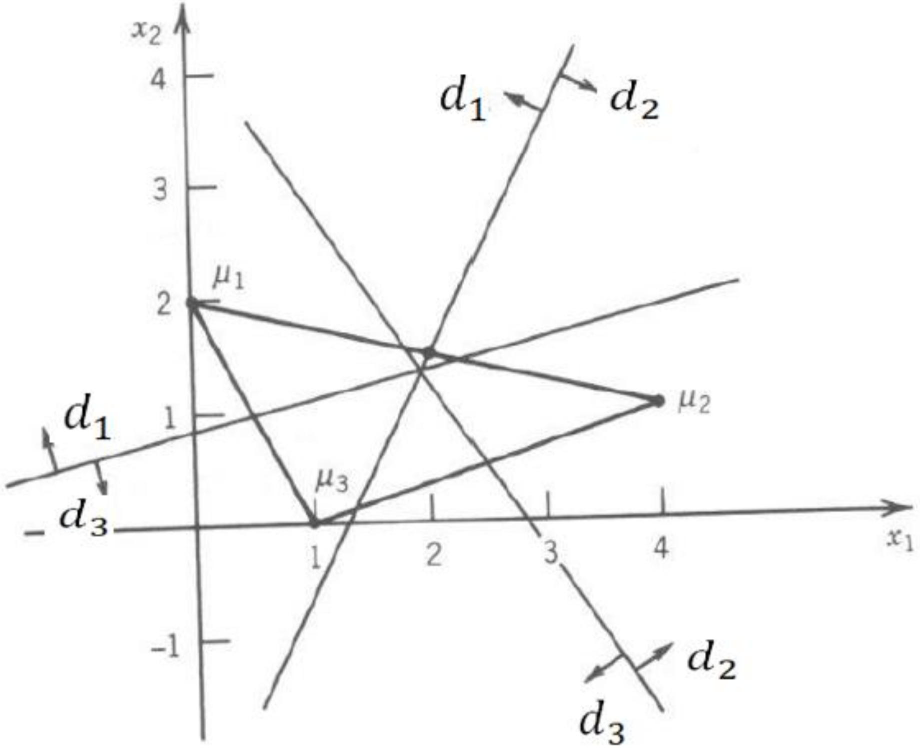
Statistički pristup raspoznavanju uzoraka

za H_{12} imamo: $4x_1 - 2x_2 - 5 = 0$
za H_{23} imamo:

$$6x_1 + 4x_2 - 17 = 0$$

za H_{31} imamo:

$$2x_1 - 8x_2 + 7 = 0$$



Statistički pristup raspoznavanju uzoraka

1) Upotreba više od jedne značajke

- umjesto skalara x koristimo vektor značajki $\mathbf{x} \in R^n$; $\mathbf{x}$ je n-dimenzionalni vektor iz Euklidskog prostora R^n

2) Više od dva stanja prirode (broj razreda M > 2)

- neka je $\{\omega_1, \omega_2, \dots , \omega_M\}$ konačan skup od M stanja (razreda, kategorija)

3) Dopustiti ćemo i akciju umjesto same odluke o stanju prirode

- uvođenje akcije umjesto same odluke o stanju prirode

skup $\{\alpha_1, \alpha_2, \dots , \alpha_a\}$ konačan skup od mogućih *a* akcija

Statistički pristup raspoznavanju uzoraka

3) Dopustiti ćemo i akciju umjesto same odluke o stanju prirode

Primjer:

Testiranje na covid-19

ω_1 = “pozitivan na covid-19 test”

ω_2 =“negativan na covid-19 test”

skup akcija $\{\alpha_1, \alpha_2\}$

akcija α_1 – “uputiti u izolaciju”

akcija α_2 – “nije potrebna izolacija”

Statistički pristup raspoznavanju uzoraka

4) Uvest ćemo funkciju gubitka kao općenitiju mjeru u odnosu na vjerojatnost pogreške

- funkcija gubitka (engl. Loss function)

  $\lambda(\alpha_i|\omega_j)$ opisuje gubitak nastao poduzimanjem akcije α_i kada je stanje prirode ω_j

Primjer:

Testiranje na covid-19

ω_1 = "pozitivan na covid-19 test"

ω_2 ="negativan na covid-19 test"

skup akcija $\{\alpha_1, \alpha_2\}$

akcija α_1 – "uputiti u izolaciju"

akcija α_2 – "nije potrebna izolacija"

gubici $\lambda(\alpha_i|\omega_j)$:

$$\lambda(\alpha_1|\omega_1) = \lambda(\text{uputiti u izolaciju}|\text{pozitivan na covid} - 19 \text{ test})$$

$$\lambda(\alpha_1|\omega_2) = \lambda(\text{uputiti u izolaciju}|\text{negativan na covid} - 19 \text{ test})$$

$$\lambda(\alpha_2|\omega_1) = \lambda(\text{nije potrebna izolacija}|\text{pozitivan na covid} - 19 \text{ test})$$

$$\lambda(\alpha_2|\omega_2) = \lambda(\text{nije potrebna izolacija}|\text{negativan na covid} - 19 \text{ test})$$

Statistički pristup raspoznavanju uzoraka

$$\lambda(\alpha_1|\omega_1) = \lambda(\text{uputiti u izolaciju}|\text{pozitivan na covid} - 19 \text{ test})$$

$$\lambda(\alpha_1|\omega_2) = \lambda(\text{uputiti u izolaciju}|\text{negativan na covid} - 19 \text{ test})$$

$$\lambda(\alpha_2|\omega_1) = \lambda(\text{nije potrebna izolacija}|\text{pozitivan na covid} - 19 \text{ test})$$

$$\lambda(\alpha_2|\omega_2) = \lambda(\text{nije potrebna izolacija}|\text{negativan na covid} - 19 \text{ test})$$

$$\lambda(\alpha_1|\omega_1) < \lambda(\alpha_1|\omega_2)$$

$$\lambda(\alpha_2|\omega_1) > \lambda(\alpha_2|\omega_2)$$

- najveći gubitak:

$$\lambda(\alpha_2|\omega_1) = \lambda(\text{nije potrebna izolacija}|\text{pozitivan na covid} - 19 \text{ test})$$

- najmanji gubitak:

$$\lambda(\alpha_2|\omega_2) = \lambda(\text{nije potrebna izolacija}|\text{negativan na covid} - 19 \text{ test})$$

Statistički pristup raspoznavanju uzoraka

4) Funkcija gubitka

x – *n*-komponentni vektor značajki

$p(\mathbf{x}|\omega_j)$ – funkcija uvjetne gustoće vjerojatnosti

$P(\omega_j)$ – apriorna vjerojatnost da je stanje prirode ω_j

$$\mathrm{P}(\omega_j|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_j)\mathrm{P}(\omega_j)}{p(\mathbf{x})}$$ – aposteriorna vjerojatnost (Bayesova formula)

$$p(\mathbf{x}) = \sum_{j=1}^{M} p(\mathbf{x}|\omega_j)\,\mathrm{P}(\omega_j)$$

Statistički pristup raspoznavanju uzoraka

4) Funkcija gubitka (nastavak)

- Pretpostavimo da smo promotrili **x** i da smo na temelju tog promatranja poduzeli akciju α_i
    - ako je stanje prirode ω_j, po definiciji, pretrpjeli smo gubitak $\lambda(\alpha_i|\omega_j)$
    - budući da je $P(\omega_j|\mathbf{x})$ vjerojatnost da je stanje prirode zaista ω_j (uz realizaciju uzorka **x**), očekujemo gubitak koji nastaje poduzimanjem akcije α_i:

očekivani gubitak
uvjetni rizik

$$R(\alpha_i|\mathbf{x}) = \sum_{j=1}^{M} \lambda(\alpha_i \mid \omega_j) P(\omega_j \mid \mathbf{x})$$

Statistički pristup raspoznavanju uzoraka

$R(\alpha_i|\mathbf{x})$ je uvjetni rizik pridružen akciji α_i a budući da decizijsko pravilo specificira akciju ukupan je rizik:

$$R = \int R(\alpha(\mathbf{x})|\mathbf{x})p(\mathbf{x})d\mathbf{x}$$

$d\mathbf{x}$ – volumski element,

a integral se "proteže" nad cijelim prostorom značajki

- Ako je $\alpha(\mathbf{x})$ izabran tako da je $R(\alpha(\mathbf{x}))$ što je moguće manji za svaki $\mathbf{x}$, onda se i ukupan rizik minimizira

Statistički pristup raspoznavanju uzoraka

4) Funkcija gubitka (nastavak)

Izabiremo strategiju:
Kadgod se susretnemo s određenim **x**, možemo MINIMIZIRATI očekivani gubitak tako da izaberemo akciju koja minimizira uvjetni rizik!
- Bayesova decizijska procedura daje optimalnu performansu

Decizijsko pravilo je sada funkcija $\alpha(\mathbf{x})$ koja nam kaže koju akciju poduzeti za svako promatranje **x**:
za svaki **x** decizijska funkcija $\alpha(\mathbf{x})$ podrazumijeva izbor jedne od akcija $\alpha_1, \alpha_2, \ldots, \alpha_a$
Ukupan rizik *R* je očekivani gubitak pridružen decizijskom pravilu

Statistički pristup raspoznavanju uzoraka

Bayesovo decizijsko pravilo:

Da bismo minimizirali ukupan rizik, izračunajmo uvjetni rizik

$$R(\alpha_i|\mathbf{x}) = \sum_{j=1}^{M} \lambda(\alpha_i \mid \omega_j)P(\omega_j \mid \mathbf{x})$$

za *i* =1, 2,… , a te tada IZABERIMO akciju α_i za koju je $R(\alpha_i|\mathbf{x})$ MINIMUM.

- rezultirajući minimalni ukupni rizik naziva se Bayesov rizik

Statistički pristup raspoznavanju uzoraka

Primjer: Klasifikacija za M = 2

α_1 – akcija koja odgovara odluci da je pravo stanje prirode ω_1

α_2 – akcija koja odgovara odluci da je pravo stanje prirode ω_2

- raspišimo izraze za uvjetni rizik:

$$R(\alpha_i|\mathbf{x}) = \sum_{j=1}^{M} \lambda(\alpha_i \mid \omega_j) P(\omega_j \mid \mathbf{x})$$

$$R(\alpha_1|\mathbf{x}) = \lambda(\alpha_1 \mid \omega_1) P(\omega_1|\mathbf{x}) + \lambda(\alpha_1 \mid \omega_2) P(\omega_2|\mathbf{x})$$

$$R(\alpha_2|\mathbf{x}) = \lambda(\alpha_2 \mid \omega_1) P(\omega_1|\mathbf{x}) + \lambda(\alpha_2 \mid \omega_2) P(\omega_2|\mathbf{x})$$

Statistički pristup raspoznavanju uzoraka

Klasifikacija za M = 2 (nastavak)

Bayesovo decizijsko pravilo

odluči se za ω_1 ako je

$$R(\alpha_1|\mathbf{x}) < R(\alpha_2|\mathbf{x})$$

zapišimo kraće $\lambda(\alpha_i|\omega_j) = \lambda_{ij}$

$$R(\alpha_1|\mathbf{x}) = \lambda(\alpha_1 \mid \omega_1)P(\omega_1|\mathbf{x}) + \lambda(\alpha_1 \mid \omega_2)P(\omega_2|\mathbf{x})$$

$$R(\alpha_2|\mathbf{x}) = \lambda(\alpha_2 \mid \omega_1)P(\omega_1|\mathbf{x}) + \lambda(\alpha_2 \mid \omega_2)P(\omega_2|\mathbf{x})$$

$$\lambda_{11}\, P(\omega_1|\mathbf{x}) + \lambda_{12}\, P(\omega_2|\mathbf{x}) < \lambda_{21}\, P(\omega_1|\mathbf{x}) + \lambda_{22}\, P(\omega_2|\mathbf{x})$$

Statistički pristup raspoznavanju uzoraka

$$\lambda_{11}\, P(\omega_1|\mathbf{x}) + \lambda_{12}\, P(\omega_2|\mathbf{x}) < \lambda_{21}\, P(\omega_1|\mathbf{x}) + \lambda_{22}\, P(\omega_2|\mathbf{x})$$

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

$$-(\lambda_{21} - \lambda_{11})\, P(\omega_1|\mathbf{x}) < -(\lambda_{12} - \lambda_{22})\, P(\omega_2|\mathbf{x}) \;/(-1)$$

$$(\lambda_{21} - \lambda_{11})\, P(\omega_1|\mathbf{x}) > (\lambda_{12} - \lambda_{22})\, P(\omega_2|\mathbf{x})$$

Statistički pristup raspoznavanju uzoraka

Odlučujemo se za ω_1 ako je

$$(\lambda_{21} - \lambda_{11})\, P(\omega_1|\mathbf{x}) > (\lambda_{12} - \lambda_{22})\, P(\omega_2|\mathbf{x})$$

općenito vrijedi da je

$$\lambda_{11} < \lambda_{21} \text{ i } \lambda_{22} < \lambda_{12}$$

(gubitak za ispravnu klasifikaciju manji je negoli za neispravnu (pogrešnu))

$$\lambda_{21} - \lambda_{11} > 0$$

$$\lambda_{12} - \lambda_{22} > 0$$

Statistički pristup raspoznavanju uzoraka

Odlučujemo se za ω_1 ako je

$$(\lambda_{21} - \lambda_{11})\, P(\omega_1|\mathbf{x}) > (\lambda_{12} - \lambda_{22})\, P(\omega_2|\mathbf{x})$$

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

omjer vjerodostojnosti
(engl. Likelihood ratio)

$$\frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} > \frac{(\lambda_{12} - \lambda_{22})}{(\lambda_{21} - \lambda_{11})} \frac{P(\omega_2)}{P(\omega_1)}$$

Statistički pristup raspoznavanju uzoraka

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

$$\frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} > \frac{(\lambda_{12} - \lambda_{22})}{(\lambda_{21} - \lambda_{11})}\frac{P(\omega_2)}{P(\omega_1)}$$

- Odlučujemo se za ω_1 ako je omjer vjerodostojnosti prešao prag koji je nezavisan od **x**!

Prag je:

$$\frac{(\lambda_{12} - \lambda_{22})}{(\lambda_{21} - \lambda_{11})}\frac{P(\omega_2)}{P(\omega_1)}$$

Statistički pristup raspoznavanju uzoraka

Klasifikacija s najmanjom pogreškom

U problemu klasifikacije svakom se stanju prirode (obično) pridružuje jedan od M razreda, i akcija α_i se tumači kao odluka da je pravo stanje prirode ω_i:

- ako je α_i akcija poduzeta i ako je pravo stanje prirode ω_j tada je odluka ispravna ako je $i = j$, a pogrešna ako je $i \neq j$,
- ako se pogreška želi izbjeći prirodno je tražiti decizijsko pravilo koje minimizira vjerojatnost pogreške

Statistički pristup raspoznavanju uzoraka

Funkcija gubitka za ovaj slučaj klasifikacije se obično izabire kao simetrična ili nula-jedan funkcija gubitka:

$$\lambda(\alpha_i|\omega_j) = \begin{cases} 0 \text{ za } i = j \\ 1 \text{ za } i \neq j \end{cases} \quad i, j = 1, 2, \dots, M$$

- funkcija gubitka dodjeljuje gubitak 0 za ispravnu odluku a 1 za bilo koju pogrešnu (sve pogrešne klasifikacije jednako koštaju)

Statistički pristup raspoznavanju uzoraka

Rizik koji odgovara toj funkciji gubitka je:

$$R(\alpha_i|\mathbf{x}) = \sum_{j=1}^{M} \lambda(\alpha_i|\omega_j)P(\omega_j\,|\,\mathbf{x})$$

$$R(\alpha_i|\mathbf{x}) = \lambda(\alpha_i|\omega_1)P(\omega_1|\mathbf{x}) + \lambda(\alpha_i|\omega_2)P(\omega_2|\mathbf{x}) + \ \ldots + \lambda(\alpha_i|\omega_M)P(\omega_M|\mathbf{x})$$

$$R(\alpha_i|\mathbf{x}) = \sum_{i\neq j} P(\omega_j\,|\,\mathbf{x})$$

$$R(\alpha_i|\mathbf{x}) = 1 - \ P(\omega_i\,|\,\mathbf{x})$$

Minimalan rizik je u slučaju najveće aposteriori vjerojatnosti

$$P(\omega_1|\mathbf{x}) + P(\omega_2|\mathbf{x}) + \ \ldots + \ P(\omega_M|\mathbf{x}) = 1$$

uvjetna vjerojatnost da je akcija α_i

Statistički pristup raspoznavanju uzoraka

Primjer:

$$\frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} > \frac{(\lambda_{12}-\lambda_{22})}{(\lambda_{21}-\lambda_{11})}\frac{P(\omega_2)}{P(\omega_1)}$$

"0-1" funkcija gubitka:

$\lambda_{22} = 0;$

$\lambda_{11} = 0;$

$\lambda_{12} = 1;$

$\lambda_{21} = 1;$

$$\frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} > \frac{(\lambda_{12}-\lambda_{22})}{(\lambda_{21}-\lambda_{11})}\frac{P(\omega_2)}{P(\omega_1)} \rightarrow \frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} > \frac{P(\omega_2)}{P(\omega_1)}$$

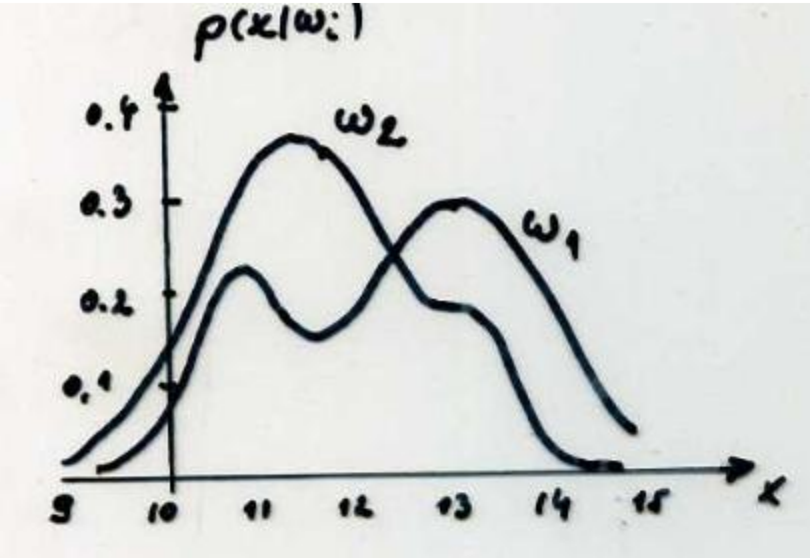
Statistički pristup raspoznavanju uzoraka

Primjer (nastavak):

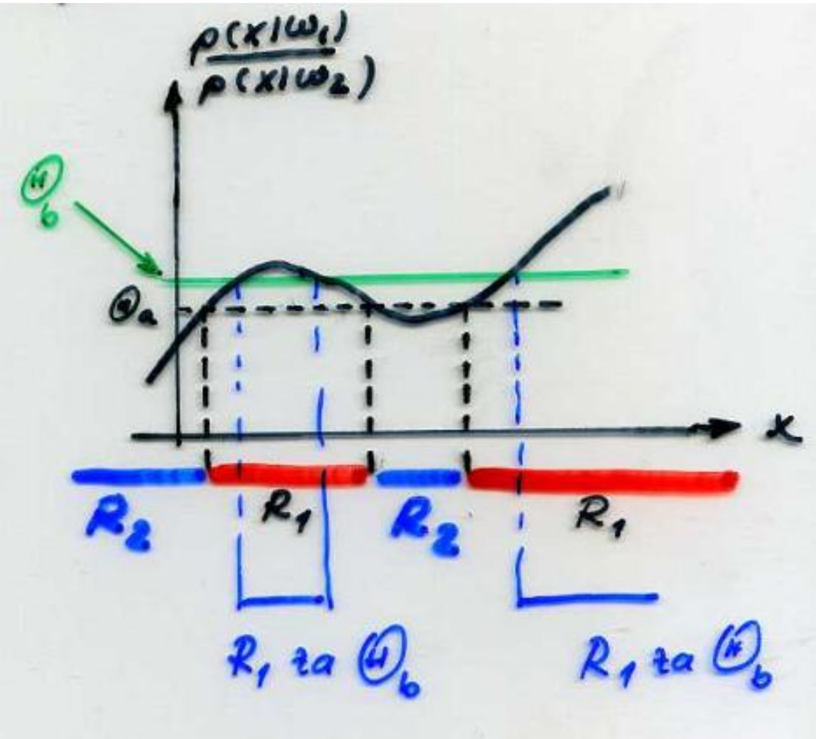
funkcije gustoće vjerojatnosti:

p(**x**|ω_i); *i* = 1, 2

omjer vjerodostojnosti: $\frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)}$

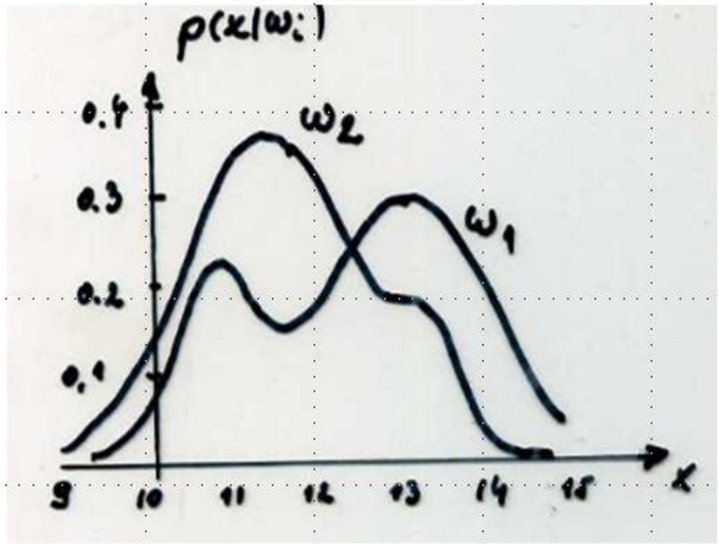


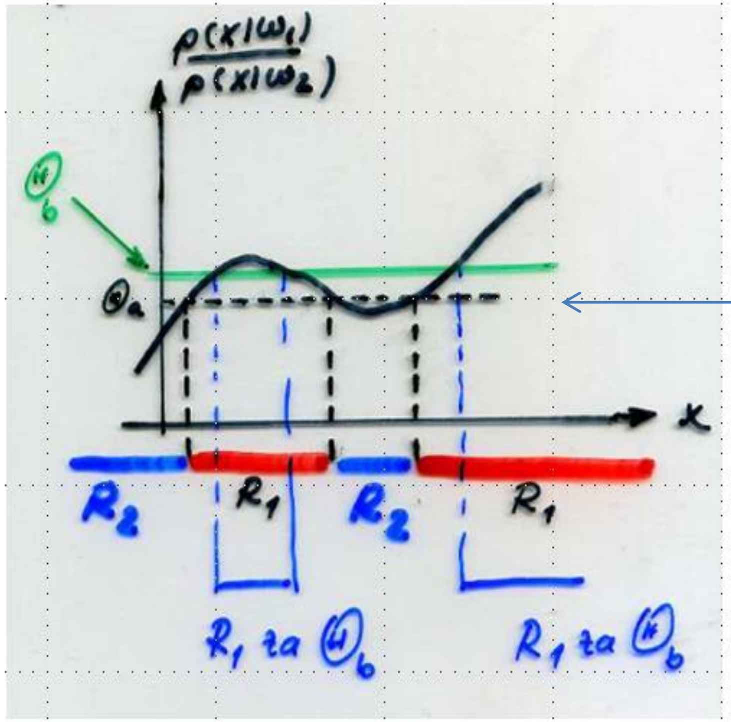
$$\frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} > \frac{P(\omega_2)}{P(\omega_1)}$$



Statistički pristup raspoznavanju uzoraka

Primjer (nastavak):





$$\frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} > \frac{P(\omega_2)}{P(\omega_1)}$$

Prag Θ

npr. ako je $P(\omega_1) = 2/3$ a $P(\omega_2) = 1/3$; $\Theta = \Theta_a = 1/2$

Statistički pristup raspoznavanju uzoraka

Primjer (nastavak):

- ako funkcija gubitka "kažnjava" više pogrešnu klasifikaciju uzorka **x** koji stvarno pripada razredu ω_2 onda je:

$$\lambda_{12} = \lambda(\alpha_1|\omega_2) > \lambda_{21} = \lambda(\alpha_2|\omega_1)$$

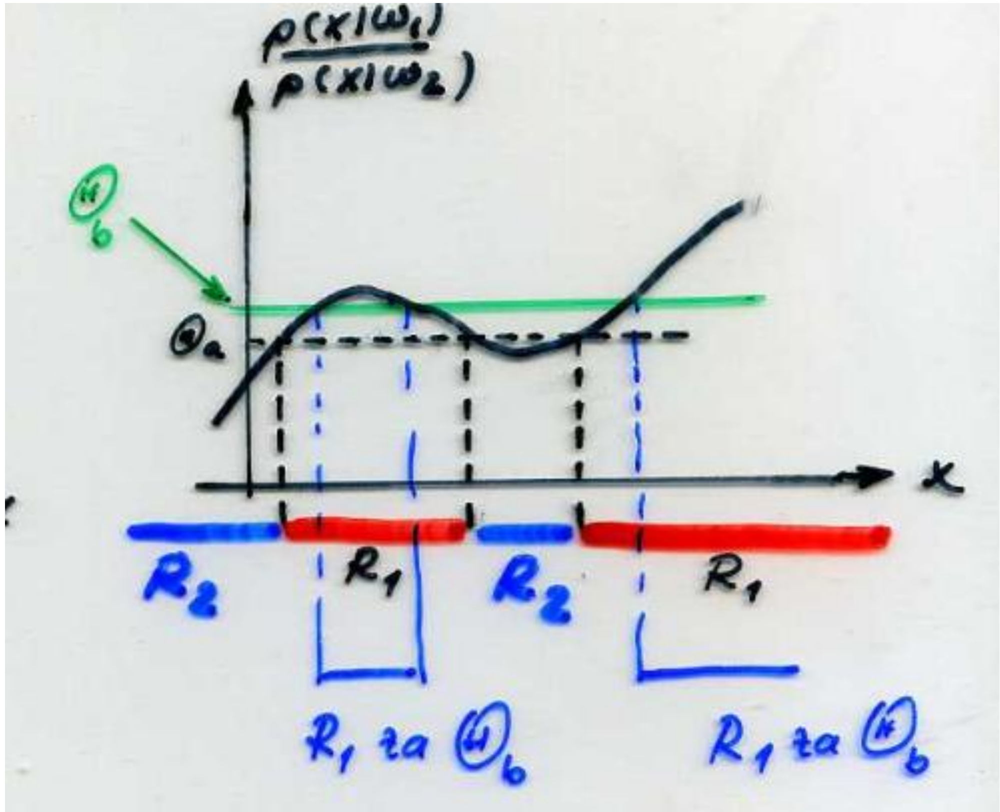
$$\frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} > \frac{(\lambda_{12} - \lambda_{22})}{(\lambda_{21} - \lambda_{11})} \frac{P(\omega_2)}{P(\omega_1)}$$

i pri $\lambda_{11} = \lambda_{22}$

vrijednost praga $\Theta = \Theta_b > \Theta_a$

Kako se to odražava na područja R_1 i R_2 koja odgovaraju razredima $\omega_{1,}$ odnosno ω_2?

Statistički pristup raspoznavanju uzoraka



Vidimo da se područje R_1 smanjuje!!!
(cijena pogrešne klasifikacije uzorka **x,**
koji stvarno pripada razredu ω_2, u razred ω_1
je veća negoli cijena pogrešne klasifikacije
uzorka **x**, koji stvarno pripada razredu ω_1,
u razred ω_2)

Statistički pristup raspoznavanju uzoraka

Primjer:

Testiranje na covid-19

ω_1 = "pozitivan na covid-19 test"

ω_2 ="negativan na covid-19 test"

skup akcija $\{\alpha_1, \alpha_2\}$

akcija α_1 – "uputiti u izolaciju"

akcija α_2 – "nije potrebna izolacija"

$$\lambda(\alpha_1|\omega_1) = \lambda(\text{uputiti u izolaciju}|\text{pozitivan na covid} - 19 \text{ test})$$

$$\lambda(\alpha_1|\omega_2) = \lambda(\text{uputiti u izolaciju}|\text{negativan na covid} - 19 \text{ test})$$

$$\lambda(\alpha_2|\omega_1) = \lambda(\text{nije potrebna izolacija}|\text{pozitivan na covid} - 19 \text{ test})$$

$$\lambda(\alpha_2|\omega_2) = \lambda(\text{nije potrebna izolacija}|\text{negativan na covid} - 19 \text{ test})$$

-neka je $\lambda_{11} = \lambda_{22}$ te $\lambda_{21} > \lambda_{12}$, kako će se to odraziti na područja ω_1 odnosno ω_2?

Statistički pristup raspoznavanju uzoraka

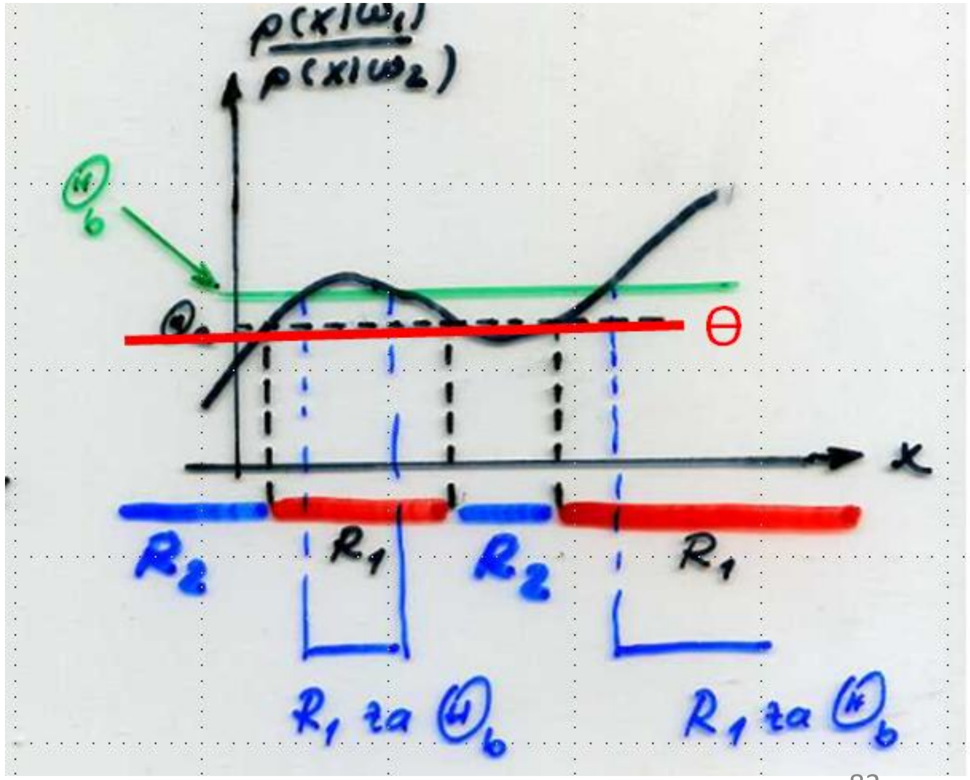
- neka je $\lambda_{11} = \lambda_{22}$ te $\lambda_{21} > \lambda_{12}$,
- kako će se to odraziti na područja ω_1 odnosno ω_2?

$$\frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} > \frac{(\lambda_{12} - \lambda_{22})}{(\lambda_{21} - \lambda_{11})} \frac{P(\omega_2)}{P(\omega_1)}$$

-prag Θ je sada manji od $\Theta_a = \dfrac{P(\omega_2)}{P(\omega_1)}$

-područje za razred ω_1 raste!

razred ω_1 “pozitivan na covid-19 test”



Statistički pristup raspoznavanju uzoraka

Ocjena parametara

x – slučajna varijabla

Za Bayesov klasifikator trebamo poznavati apriornu vjerojatnost $P(\omega_i)$ i gustoću vjerojatnosti $p(\mathbf{x}|\ \omega_i)$!

Tipični slučaj:

- imamo pretpostavku utemeljenu na znanju o problemu
- imamo označene uzorke (uzorke iz skupa za učenje)

Statistički pristup raspoznavanju uzoraka

Pristup:

Uporabom uzoraka za učenje ocijeniti vjerojatnosti $P(\omega_i)$ i $p(\mathbf{x}|\omega_i)$

- apriorna vjerojatnost $P(\omega_i)$ ne predstavlja problem;
- problem je ocjena gustoće vjerojatnosti $p(\mathbf{x}|\omega_i)$!?

(posebno izražen problem kada je dimenzionalnost vektora značajki velika i kada imamo mali broj uzoraka za učenje;

Važno: ako nam znanje o problemu dopušta pretpostavku o razdiobi uzoraka stvar se pojednostavljuje;

Na primjer, ako pretpostavimo da $p(\mathbf{x}|\omega_i)$ ima normalnu (Gaussovu) razdiobu sa srednjom vrijednosti $\boldsymbol{\mu}$ i kovarijacijskom matricom Σ: $N(\boldsymbol{\mu}, \Sigma)$ problem se svodi na ocjenu parametara $\boldsymbol{\mu}$ i Σ.

Statistički pristup raspoznavanju uzoraka

- ocjena parametara klasičan je problem iz statistike
- (metode: parametarske i neparametarske; S. Theeodoridis; K. Koutrumbas, Pattern Recognition, 4th edition,pp.34 – 59)

Bayesov postupak ocjene parametara za Gaussovu razdiobu

$$p(\mathbf{x}) = (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu})\right]$$

$\mathbf{x}_j$ – *n*-dimenzionalni uzorci iz skupa za učenje; j = 1, 2, ..., *N*

$$\boldsymbol{\mu} = \frac{1}{N}\sum_{j=1}^{N} \mathbf{x}_j$$

$$\boldsymbol{\Sigma} = \frac{1}{N}\sum_{j=1}^{N} (\mathbf{x}j - \boldsymbol{\mu})(\mathbf{x}j - \boldsymbol{\mu})^T$$

Statistički pristup raspoznavanju uzoraka

Primjer:

Zadani su uzorci iz skupa za učenje:

$$\omega_1 = \{[0,0,0]^T, [1,0,0]^T, [1,0,1]^T, [1,1,0]^T\}$$

$$\omega_2 = \{[0,0,1]^T, [0,1,0]^T, [0,1,1]^T, [1,1,1]^T\}$$

Pretpostavimo da je $P(\omega_1) = P(\omega_2)$

Izgradite Bayesov klasifikator uz pretpostavku da su uzorci opisani funkcijom *normalne gustoće vjerojatnosti*!

Statistički pristup raspoznavanju uzoraka

Primjer (nastavak):

- ocijenimo srednje vrijednosti uzoraka za pojedine razrede:

$$\boldsymbol{\mu}_1 = \frac{1}{N_1}\sum_{j=1}^{N_1}\mathbf{x}_j$$

$$\boldsymbol{\mu}_2 = \frac{1}{N_2}\sum_{j=1}^{N_2}\mathbf{x}_j$$

- dobivamo:

$$\boldsymbol{\mu}_1 = \frac{1}{4}\left[\begin{bmatrix}0\\0\\0\end{bmatrix} + \begin{bmatrix}1\\0\\0\end{bmatrix} + \begin{bmatrix}1\\0\\1\end{bmatrix} + \begin{bmatrix}1\\1\\0\end{bmatrix}\right] = \frac{1}{4}\begin{bmatrix}3\\1\\1\end{bmatrix}$$

$$\boldsymbol{\mu}_2 = \frac{1}{4}\left[\begin{bmatrix}0\\0\\1\end{bmatrix} + \begin{bmatrix}0\\1\\0\end{bmatrix} + \begin{bmatrix}0\\1\\1\end{bmatrix} + \begin{bmatrix}1\\1\\1\end{bmatrix}\right] = \frac{1}{4}\begin{bmatrix}1\\3\\3\end{bmatrix}$$

Statistički pristup raspoznavanju uzoraka

Primjer (nastavak):

- ocijenimo kovarijacijsku matricu:

$$\boldsymbol{\Sigma}_1 = \frac{1}{N_1}\left(\sum_{j=1}^{N_1}(\mathbf{x}_\mathrm{j} - \boldsymbol{\mu}_1)(\mathbf{x}_\mathrm{j} - \boldsymbol{\mu}_1)^\mathrm{T}\right)$$

$$\boldsymbol{\Sigma}_1 = \frac{1}{N_1}\left\{\left[\begin{bmatrix}0\\0\\0\end{bmatrix} - \frac{1}{4}\begin{bmatrix}3\\1\\1\end{bmatrix}\right]\left[\begin{bmatrix}0\\0\\0\end{bmatrix} - \frac{1}{4}\begin{bmatrix}3\\1\\1\end{bmatrix}\right]^\mathrm{T} + \left[\begin{bmatrix}1\\0\\0\end{bmatrix} - \frac{1}{4}\begin{bmatrix}3\\1\\1\end{bmatrix}\right]\left[\begin{bmatrix}1\\0\\0\end{bmatrix} - \frac{1}{4}\begin{bmatrix}3\\1\\1\end{bmatrix}\right]^\mathrm{T}\right.$$
$$\left. + \left[\begin{bmatrix}1\\0\\1\end{bmatrix} - \frac{1}{4}\begin{bmatrix}3\\1\\1\end{bmatrix}\right]\left[\begin{bmatrix}1\\0\\1\end{bmatrix} - \frac{1}{4}\begin{bmatrix}3\\1\\1\end{bmatrix}\right]^\mathrm{T} + \left[\begin{bmatrix}1\\1\\0\end{bmatrix} - \frac{1}{4}\begin{bmatrix}3\\1\\1\end{bmatrix}\right]\left[\begin{bmatrix}1\\1\\0\end{bmatrix} - \frac{1}{4}\begin{bmatrix}3\\1\\1\end{bmatrix}\right]^\mathrm{T}\right\}$$

Statistički pristup raspoznavanju uzoraka

Primjer (nastavak):

- dobivamo:

$$\boldsymbol{\Sigma}_1 = \frac{1}{16}\begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & -1 \\ 1 & -1 & 3 \end{bmatrix}$$

- za vježbu ocijenite kovarijacijsku matricu $\boldsymbol{\Sigma}_2$
- odredite decizijske funkcije:

$$d_i(\mathbf{x}) = \log\{\mathrm{p}(\mathbf{x}|\omega_i)\};\ i = \mathbf{1},\ 2$$

(Gaussova razdioba, apriorne vjerojatnosti za razrede jednake)

Statistički pristup raspoznavanju uzoraka

Primjer (nastavak):

- rješenje:

$$d_1(\mathbf{x}) = 4x_1 - \frac{3}{2}$$

$$d_2(\mathbf{x}) = -4x_1 + 8x_2 + 8x_3 - \frac{11}{2}$$

- provjerite rješenje!
- odredite granicu koja dijeli prostor uzoraka na dva razreda!
- rješenje:

$$2x_1 - 2x_2 - 2x_3 + 1 = 0$$

Statistički pristup raspoznavanju uzoraka

Primjer:

- za M = 2 razreda i n = 2 (dvodimenzionalni prostor značajki)
- normalna (Gaussova) razdioba

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 2 \\ 0 \end{bmatrix} \qquad \boldsymbol{\Sigma}_1 = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{3} \end{bmatrix}$$

$$\boldsymbol{\mu}_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \boldsymbol{\Sigma}_2 = \begin{bmatrix} \frac{1}{3} & 0 \\ 0 & 1 \end{bmatrix}$$

$$\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$$

$$\mathrm{P}(\omega_1) = \mathrm{P}(\omega_2)$$

- Odredite $d_i(\mathbf{x}) = \log\{\mathrm{p}(\mathbf{x}|\omega_i)\};\ i = 1, 2$

(decizijske funkcije su kvadratne!)

Statistički pristup raspoznavanju uzoraka

Pojednostavljeni postupak ocjene funkcije gustoće vjerojatnosti

(engl. Naive-Bayes Classifier)

- ocijena funkcija gustoće vjerojatnosti (pdf),

$$p(\mathbf{x}|\omega_i), i = 1,2, \ldots, M$$

koje su nam potrebne za Bayesovo klasifikacijsko pravilo, temelji na raspoloživom skupu uzoraka za učenje.

- za dobre ocjene pdfs broj uzoraka *N* u skupu za učenje mora biti dovoljno velik – raste eksponencijalno s *n,* gdje je *n* dimenzionalnost prostora značajki
- za velike vrijednost *n* točna procjena multidimenzionalnih pdfs “is a bit of an “illusion”…”

Statistički pristup raspoznavanju uzoraka

Pojednostavljeni postupak ocjene funkcije gustoće vjerojatnosti

(engl. Naive-Bayes Classifier)- nastavak

- Jedan ustupak u vezi stupnja točnosti ocjene pdf:

pretpostavlja se da su pojedine značajke $x_j, \quad j = 1, 2, \dots, n$ statistički nezavisne (naivna pretpostavka) te dobivamo

$$p(\mathbf{x}|\omega_i) = \prod_{j=1}^{n} p\,(x_j \mid \omega_i), \quad i = 1, 2, \dots, M$$

Naivni Bayesov klasifikator:

$$P(\omega_i|\mathbf{x}) = \mathrm{P}(\omega_i) \prod_{j=1}^{n} p\,(x_j \mid \omega_i), \quad i = 1, 2, \dots, M$$

$\mathbf{x} \in \omega_i$ ako $\quad P(\omega_i|\mathbf{x}) > P(\omega_k|\mathbf{x}), \quad$ za sve $\quad k \neq i$