OPĆENITO

- generički izraz za objekte - uzorak
- RU redukcija -> iz mnogo u 1 (1 od razreda w_i)
- RU ima svojstvo generalizacije

3 vrste raspoznavanja (S. Watanabe)

supervised:
1) Raspoznavanje temeljeno na primjerima
2) Raspoznavanje temeljeno na strukturi (složenih) uzoraka

3) Grupiranje (učenje bez učitelja

1) na temelju nekoliko uzoraka moći za novi odlučiti dal pripada razredu w_i

I. faza učenja - traži se oblik decizijske fje koja će uz najmanju pogrešku (ili bez nje) klasificirati uzorke iz skupa za učenje

II faza odlučivanja - sustav je naučio decizijsku fju i primjenjuje ju na nepoznate nove uzorke

2) - uzorak opisan primitivima i odnosom među njima, gramatika (V_N, V_T, P, S)
- jezik => za sintaksni pristup
- svakom razredu pridružen je jezik i stroj koji prepoznaje njegovu gramatiku

3) neograničeni uzorci, sustav ih „prirodno grupira"
eksperimentira se s različitim postupcima grupiranja
na temelju rezultata korisnik potvrđuje ili odbacuje hipoteze

učenje s kritikom

- na ulazu je uzorak za koji se odredi (tentativni) razred
- ako je u krivom razredu => korekcija decizijske fje proporcionalno „veličini" uzorka
- sve skupa povratni signal je samo: T ili N (nema zašto, kako...)

Prikaz RU - redukcija informacije

$\approx R^{\infty} \longrightarrow R^{n} \longrightarrow R^{1}$

prostor mjerenja — realizirani prostor uzoraka — pripadnost razredu, prostor opisa

- osnovne faze u oblikovanju sustava:

1) senzor
2) generiranje značajki
3) izbor značajki
4) oblikovanje klasifikatora
5) evaluacija sustava

INTERSET značajke su bitne $\vec{x}$

- decizijske granice - decizijskim funkcijama: $d_1(x) \dots d_M(x)$

  └ pravilo razvrstavanja: Ako $d_i(\vec{x}) > d_j(\vec{x}) \Rightarrow \vec{x}$ pripada ω_i

∞ dim. → senzor → (r) → pretprocesiranje → izlučivanje značajki → ($n << r$) → pravilo odlučivanja → R^1

prostor značajki

linearna decizijska f-ja $d(\vec{x}) = w_1 x_1 + w_2 x_2 + \dots + w_n x_n + w_{n+1} = \vec{w}_0^T \vec{x} + w_{n+1}$

težinski koef. — pomaknuće

Decizijsko pravilo $M=2$

$$\boxed{\begin{array}{lll} \text{ako } d(\vec{x}) > 0 & \text{onda} & \vec{x} \in \omega_1 \\ d(\vec{x}) < 0 & \text{onda} & \vec{x} \in \omega_2 \\ \text{ako } d(\vec{x}) = 0 & \text{nije definirano} & \end{array}}$$

neki k: hiperravnina: $\vec{u}^T(\vec{x} - \vec{p}) = 0$

$\vec{w}_0^T \vec{x} + w_{n+1} = 0 \quad / : \|\vec{w}_0\|$

$\frac{\vec{w}_0^T \vec{x}}{\|\vec{w}_0\|} = \frac{-w_{n+1}}{\|\vec{w}_0\|} \qquad \vec{u}^T \vec{x} = p \vec{u}^T$

$\vec{u} = \frac{\vec{w}_0}{\|\vec{w}_0\|}$ → orjentacija hiper-ravnine

ako je neka komponenta $= 0$

⇒ paralelna s tom osi

⇒ ako $w_{n+1} = 0$ ⇒ prolazi kroz ishodište

$D_u = \frac{|w_{n+1}|}{\|\vec{w}_0\|}$ - udaljenost hiperravnine od ishodišta

Više razreda: M>2

→ zapravo m problema klasifikacije u 2 razreda

1. slučaj: granica između w_i i ostalih razreda:

$$d_i(\vec{x}) = \vec{w}_i^T \vec{x} + w_{in+1}$$

2. slučaj: $\frac{c(c-1)}{2}$ f-ja, svaka odjeljuje po 2 razreda

$$d_{ij}(\vec{x}) = w_{ij1} x_1 + w_{ij2} x_2 + \dots + w_{ijn+1} = 0$$

3. slučaj: $d_{ij}(\vec{x}) = d_i(\vec{x}) - d_j(\vec{x}) = (\vec{w}_i - \vec{w}_j)^T \vec{x} + (w_{in+1} - w_{jn+1}) = 0$

ako je $d_i(\vec{x}) > d_j(\vec{x})$ onda $\vec{x} \in w_i$

Određivanje $d(\vec{x})$

$$\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \\ 1 \end{bmatrix} \Rightarrow d(\vec{x}) = \vec{w}^T \vec{x}$$

Uzorak $\vec{x}$ je pravilno razvrstan ako: $\forall \vec{x} \in w_1 : \vec{w}^T \vec{x} > 0$
$\forall \vec{x} \in w_2 : \vec{w}^T \vec{x} < 0$

⇒ jedinstven uvjet $\vec{w}^T \vec{x} > 0$ ako $\forall \vec{x} \in w_2$ $(\cdot -1)$

Tražimo $\vec{w}$ t.d. je $\vec{w}^T \vec{x} > 0$ $\forall \vec{x}$ iz skupa uzoraka, $\vec{x} \in w_2 \Rightarrow \vec{x} \cdot (-1)$

Ako radimo matricu $[X]$ gdje upisujemo retke i $[X]\vec{w} > [0]$

$\vec{w}$ → razdvojni vektor

4) GRADIJENTNI POSTUPCI ODREĐIVANJA $\vec{w}$

$$\text{grad } f(\vec{y}) = \frac{df(\vec{y})}{d\vec{y}} = \begin{bmatrix} \frac{df}{dy_1} \\ \frac{df}{dy_2} \\ \vdots \\ \frac{df}{dy_n} \end{bmatrix}$$

vektor → $\boxed{\text{gradijent}}$ → vektor

→ svaka komponenta - veličina promjene f-je u tom smjeru

→ povećanje argumenta u smjeru negativnog gradijenta dovodi do minimuma funkcije

FUNKCIJA $\boxed{J(\vec{w}, \vec{x}) = |\vec{w}^T \vec{x}| - \vec{w}^T \vec{x}}$

Netrivijalan slučaj: $\vec{w}^T \vec{x} > 0$

$\Rightarrow$ povećavamo argument $\vec{w}$ u smjeru negativnog gradijenta sve dok ne dođemo do minimuma f-je!

$\vec{w}(k)$ - k-ti korak

$$\vec{w}(k+1) = \vec{w}(k) - c \left\{ \frac{\partial J(\vec{w}, \vec{x})}{\partial \vec{w}} \right\}_{\vec{w} = \vec{w}(k)} \qquad c \neq 0$$

PERCEPTRON

- klasificira uzorke u 1 od 2 razreda

$$R = \sum_{i=1}^{n+1} w_i x_i = \vec{w}^T \cdot \vec{x} = d(\vec{x})$$

Pravilo: ako $R > 0$ ide u ω_1
$R < 0$ ide u ω_2

- razredi moraju biti linearno separabilni

- modifikacija za $M > 2$

$$\vec{x} \in \omega_i \iff R_i > R_j \quad \forall i, j$$

- modifikacija za nelinearne $d(\vec{x})$
  - $\Rightarrow$ umetanje nelinearnog pretprocesora između A i R
  - $\Rightarrow$ višeslojni perceptron MLP

Algo perceptrona NE mijenja (-1) za $\vec{x} \in \omega_2$

ALI za uzorke iz ω_1 $+$
ω_2 $-$

a) postupak učenja za kriterijsku funkciju $\frac{1}{2}(|\vec{w}^T\vec{x}| - \vec{w}^T\vec{x})$

$$\vec{w}(k+1) = \vec{w}(k) + c\begin{cases} 0 & \vec{w}^T(k)\vec{x}(k) > 0 \\ \vec{x}(k) & \vec{w}^T(k)\vec{x}(k) \le 0 \end{cases}$$

↳ POSTUPAK PERCEPTRONA SA STALNIM PRIRASTOM

b) $$J(\vec{w},\vec{x}) = \frac{1}{4\vec{x}^T\vec{x}}\left(|\vec{w}^T\vec{x}|^2 - |\vec{w}^T\vec{x}|\cdot\vec{w}^T\cdot\vec{x}\right)$$

$$\frac{\partial J}{\partial \vec{w}} = \dots \Rightarrow \vec{w}(k+1) = \vec{w}(k) + \lambda\cdot\frac{|\vec{w}^T(k)\cdot\vec{x}(k)|}{\vec{w}^T(k)\cdot\vec{x}(k)}\cdot\begin{cases} 0, & \vec{w}^T\cdot\vec{x} > 0 \\ \vec{x}(k), & \vec{w}^T\cdot\vec{x} \le 0 \end{cases}$$

Poopćeni algos perceptrona za $M > 2$ (3. slučaj)

Poopćeni oblik lin. $d(\vec{x})$

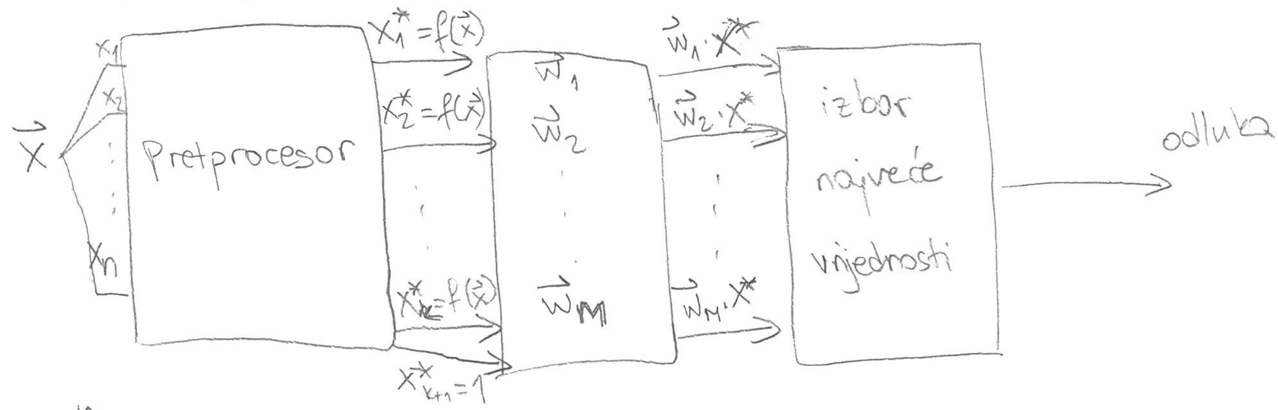
$$\vec{x} \to f(\vec{x})$$

$$d(\vec{x}) = w_1 f_1(\vec{x}) + w_2 f_2(\vec{x}) + \dots + w_{k+1}$$

$$= \sum_{i=1}^{k+1} w_i f_i(\vec{x}) \qquad f_{k+1} = 1$$

$$\vec{x}^* = \begin{bmatrix} f_1(\vec{x}) \\ f_2(\vec{x}) \\ \vdots \\ f_k(\vec{x}) \\ 1 \end{bmatrix} \qquad \boxed{d(\vec{x}) = \vec{w}^T\vec{x}^*}$$

r-stupanj f-je, M-broj razreda



n

$k >> n$

SUSTAV ZA RASPOZNAVANJE

Potencijalne fje

↳ učenje $d(\vec{x})$ bez pretpostavke o konačnom obliku fje

↳ pogle algos

poopćenje na $M>2$ razreda

↳ ona koja je iz tog razreda mora imati najveću vrijednost

postupak: $d_0^1(\vec{x}), d_0^2(\vec{x}), \dots, d_0^M(\vec{x}) = 0$

1 korak - 1 uzorak

$$d_k^{(i)}(\vec{x}) = d_{k-1}^{(i)} + K(\vec{x}, \vec{x}_k)$$

$$d_k^{(l)}(\vec{x}) = d_{k-1}^{(l)} - K(\vec{x}, \vec{x}_k)$$

$$d_k^{(j)}(\vec{x}) = d_{k-1}^{(j)}$$

Bayes

Ako samo na temelju „a priori" vjerojatnosti:

Uvijek odluka za w_1 ako $P(w_1) > P(w_2)$

inače w_2. To je za bez mjerenja.

Svojstva funkcije gustoće razdiobe vjerojatnosti $\Big\} \; p(x|w_j)$

za kontinuiranu varijablu x:

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

$$\left.\begin{array}{l} p(x|w_j) - \text{vjerodostojnost, likelihood of } w_j \\ P(w_j) \end{array}\right\} \text{znamo}$$

$\Rightarrow$ Bayesovo pravilo: $$P(w_j|x) = \frac{p(x|w_j) \cdot P(w_j)}{p(x)}$$

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

$\downarrow$

pokazuje kako x mijenja á priori vjerojatnost $P(w_j)$ u á posteriori $p(w_j|x)$

Vjerojatnost pogreške:

$$P(error|x) = \begin{cases} P(w_1|x), \text{ ako uzmemo } w_2 \\ P(w_2|x), \text{ ako uzmemo } w_1 \end{cases}$$

Odluka: (minimizacija pogreška)

$$\begin{cases} w_1 \text{ ako je } P(w_1|x) > P(w_2|x) \\ w_2 \text{ ako je } P(w_2|x) > P(w_1|x) \end{cases}$$

minimizira jer $P(error) = \int_{-\infty}^{\infty} P(error, x)dx = \int_{-\infty}^{\infty} P(error|x) \cdot p(x)dx$,

a on je uvijek najmanji, pa je i integral najmanji

Modificirano pravilo: uzmi w_1 ako $p(x|w_1) \cdot P(w_1) > p(x|w_2) \cdot P(w_2)$
inače w_2

1) >1 # značajki $\Rightarrow x \Rightarrow \vec{x} \in \mathbb{R}^d$
2) >2 razreda (c)
3) akcija umjesto samo odluka o stanju prirode - skup $\{\alpha_1, \dots, \alpha_k\}$ akcija
4) f-ja gubitka $\lambda(\alpha_i | w_j)$

Očekivani gubitak - rizik - poduzimanjem akcije α_i

$$R(\alpha_i|\vec{x}) = \sum_{j=1}^{c} \lambda(\alpha_i|w_j) P(w_j|\vec{x})$$

$\forall \vec{x}$ decizijska f-ja daje jednu akciju. Ukupan rizik:

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

$R(\alpha_i|\vec{x})$ - uvjetni rizik akcije α_i
d-volumenski el., $\int$ - po cijelom prostoru značajki

Ako $\alpha(\vec{x})$ takav da $R(\alpha_i(\vec{x}))$ što je moguće manji $\forall \vec{x} \Rightarrow$ ukupan rizik minimalan

Bayesovo decizijsko pravilo:

$$R(\alpha_i|\vec{x}) = \sum_{j=1}^{c} \lambda(\alpha_i|w_j) \cdot P(w_j|\vec{x}) \quad \text{za } i = 1, 2, \dots, a$$

i izaberemo α_i t.d. je $R(\alpha_i|\vec{x})$ MINIMALAN.

Rezultirajući minimalni ukupni rizik $\rightarrow$ Bayesov rizik

omjer vjerodostojnosti za 2 razreda: $\frac{p(\vec{x}|w_1)}{p(\vec{x}|w_2)} > \frac{(\lambda_{12} - \lambda_{22})}{(\lambda_{21} - \lambda_{11})} \cdot \frac{P(w_2)}{P(w_1)}$

Odlučujemo se za w_1 ako je omjer vjerodostojnosti prešao prag nezavistan od $\vec{x}$!

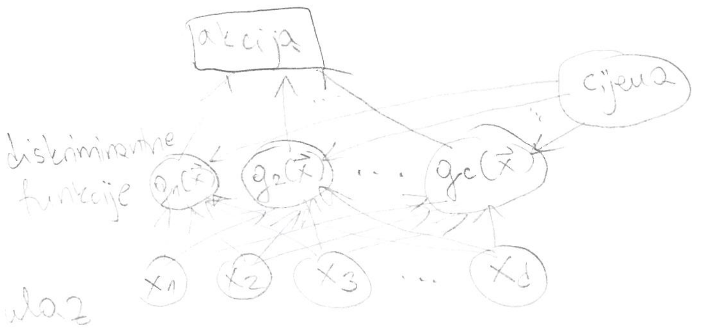
$\lambda_{ij} = \lambda(\alpha_i|w_j)$

$$\lambda(\alpha_i|w_j) = \begin{cases} 0 & i=j \\ 1 & i \neq j \end{cases} \quad i,j = 1, \dots, c$$

$$R(\alpha_i|\vec{x}) = \sum_{j=1}^{c} \lambda(\alpha_i|w_j) \cdot P(w_j|\vec{x}) = \sum_{j \neq i} P(w_j|\vec{x}) = 1 - P(w_i|\vec{x})$$

Pravilo:
w_i ako $P(w_i|\vec{x}) > P(w_j|\vec{x}) \quad \forall j \neq i$

Primjer klasifikacije za $M = C$ $(C > 2)$ razreda



Bayesov klasifikator

- opći slučaj s rizikom: $g_i(\vec{x}) = -R(\alpha_i | \vec{x})$
- slučaj minimalne pogreške: $g_i(\vec{x}) = P(w_i | \vec{x})$

nije jedinstveno. može $\times$ const i + const, te $f \circ g$ s monotono rastućom f-jom

$$= g_i(\vec{x}) = P(w_i|\vec{x}) = \frac{p(\vec{x}|w_i) \cdot P(w_i)}{p(\vec{x})} = p(\vec{x}|w_i) \cdot P(w_i)$$

$$g_i(\vec{x}) = \ln p(\vec{x}|w_i) + \ln P(w_i)$$

I dalje pravilo: $g_i(\vec{x}) > g_j(\vec{x}) \quad \forall i \neq j \Rightarrow w_i$

Decizijska ploha: $g_{ij}(\vec{x}) = g_i(\vec{x}) - g_j(\vec{x}) = 0$, $i, j = 1, 2, \ldots, M$, $i \neq j$

Bayesova klasifikacija za normalne distribucije (Gauss)

$$p(\vec{x}|w_i) = \frac{1}{(2\pi)^{\ell/2} |\Sigma_i|^{1/2}} \exp\left(-\frac{1}{2}(\vec{x} - \vec{\mu}_i)^T \Sigma_i^{-1} (\vec{x} - \mu_i)\right)$$

ℓ - dimenzija značajki
$\vec{\mu}_i = E[\vec{x}] \rightarrow$ srednja vrijednost razreda $\vec{w}_i$
$\Sigma_i \rightarrow \ell \times \ell$ kovarijantna matrica: $\Sigma_i = E[(\vec{x} - \vec{\mu}_i)(\vec{x} - \mu_i)^T]$
$|\Sigma_i|$ determinanta
E - očekivanje sluč. var.

$d(\vec{x}) = p(w_i|\vec{x}) \cdot p(\vec{x})$

$\ln(d(\vec{x})) = \ln p(w_i) - \frac{1}{2}\ln|C_i| - \frac{1}{2}(\vec{x} - \vec{m}_i)^T C_i^{-1} (\vec{x} - \vec{m}_i)$

a) $C_i = C \Rightarrow d_i(\vec{x}) = \ln p(w_i) + \vec{x}^T C^{-1} \vec{m}_i - \frac{1}{2} \vec{m}_i^T C^{-1} \vec{m}_i$

b) $C_i = \sigma^2 I \Rightarrow d_i(\vec{x}) = \ln p(w_i) - \frac{1}{2\sigma^2} \|\vec{x} - \vec{m}_i\|^2$

c) b) + $p(w_i) = \frac{1}{M} \Rightarrow d_i(\vec{x}) = \vec{x}^T \vec{m}_i - \frac{1}{2} \vec{m}_i^T \vec{m}_i$

Gaussova distribucija $\mathcal{N}(\vec{\mu}, \Sigma)$

$$g_i(\vec{x}) = \underbrace{f}_{\ln(\cdot)}(P(w_i|\vec{x}))$$

$$g_i(\vec{x}) = \ln p(\vec{x}|w_i) + \ln P(w_i)$$

$$g_i(\vec{x}) = -\frac{1}{2}\vec{x}^T \Sigma_i^{-1} \vec{x} + \frac{1}{2}\vec{x}^T \Sigma_i^{-1} \vec{\mu}_i - \frac{1}{2}\mu_i^T \Sigma_i^{-1} \vec{\mu}_i + \frac{1}{2}\vec{\mu}_i^T \Sigma_i^{-1} \vec{x} + \ln P(w_i) + C_i$$

$$c_i = -(\ell/2)\ln 2\pi - (1/2)\ln|\Sigma_i|$$

$g_i(\vec{x})$ - nelinearna kvadratna forma. Za $\ell = 2$ dec. krivulje: elipsoidi, parabole, hiperbole

FIIT3

Ocjena parametara (Bay3)

trebamo poznavati: $P(w_i)$ i $p(\vec{x}|w_i)$

$\downarrow$ ako je Gaussova $N(\vec{\mu}_i, \Sigma_i)$

$\rightarrow$ problem se svodi na ocjenu parametara $\vec{\mu}_i$ i Σ_i

Gauss:
$$p(\vec{x}) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp -\frac{1}{2}\left[(\vec{x}-\vec{\mu}_i)^T \Sigma^{-1} (\vec{x}-\vec{\mu}_i)\right]$$

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \dots & \Sigma_{1d} \\ \vdots & & & \\ \Sigma_{d1} & \Sigma_{d2} & \dots & \Sigma_{dd} \end{bmatrix}$$

Skup za učenje: $\vec{\mu}_1 \doteq \frac{1}{N_1}\sum_{j=1}^{N_1} \vec{x}_j$

$\Sigma_1 \doteq \frac{1}{N_1}\left(\sum_{j=1}^{N_1} (\vec{x}_j - \vec{\mu}_1)(\vec{x}_j - \vec{\mu}_1)^T\right)$

za sve razrede

Ako: $P(w_1) = P(w_2)$ \$\$ $\Sigma_1 = \Sigma_2$ \$\$ $\vec{\mu}_1 \neq \vec{\mu}_2$

$\Rightarrow$ razdioba je Gaussova

$$g_i(\vec{x}) = \log\{p(\vec{x}|w_i)\}$$

$$p_i(\vec{x}) = (2\pi)^{-\frac{d}{2}} |\Sigma_i|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\vec{x}-\vec{\mu}_i)^T \Sigma_i^{-1} (\vec{x}-\vec{\mu}_i)\right]$$

$$g_i(\vec{x}) = \log\{p(\vec{x}|w_i)\}$$

$$g_i(\vec{x}) = -\frac{1}{2}(\vec{x}-\mu_i)\Sigma^{-1}(\vec{x}-\vec{\mu}_i) - \frac{d}{2}\log(2\pi) - \frac{1}{2}|\Sigma_i|$$

$\Sigma_1 = \Sigma_2$

$$d_1^2 = (\vec{x}-\mu_i)^T \Sigma^{-1} (\vec{x}-\vec{\mu}_i)$$

$$g_i(\vec{x}) = -(\vec{x}-\vec{\mu}_i)^T \Sigma^{-1} (\vec{x}-\mu_i)$$

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

$$g_i(\vec{x}) = (\Sigma^{-1}\vec{\mu}_i)^T \vec{x} - \frac{1}{2}\vec{\mu}_i^T \Sigma^{-1} \vec{\mu}_i$$

$$g_i(\vec{x}) = \vec{w}_i^T \vec{x} - w_{i0}$$

$$w_i = \Sigma^{-1} \mu_i$$

$$w_{i0} = \frac{1}{2}\mu_i^T \Sigma^{-1} \vec{\mu}_i$$

Granica: $g_1(\vec{x}) - g_2(\vec{x}) = 0$

Klasifikacija pomoću najmanje udaljenosti

- mjera udaljenosti mora zadovoljavati:

a) $D(\vec{x}_k, \vec{x}_l) = 0$ za $\vec{x}_k = \vec{x}_l$

$D(\vec{x}_k, \vec{x}_l) \neq 0$; $D(\vec{x}_k, \vec{x}_l) > 0$ za $\vec{x}_k \neq \vec{x}_l$

b) $D(\vec{x}_k, \vec{x}_l) = D(\vec{x}_l, \vec{x}_k)$

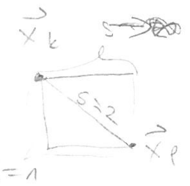
c) $D(\vec{x}_k, \vec{x}_l) \leq D(\vec{x}_k, \vec{x}_j) + D(\vec{x}_j, \vec{x}_l)$

Udaljenost Minkovskog:

$$D(\vec{x}_k, \vec{x}_l) = \left(\sum_{j=1}^{n} |x_{kj} - x_{lj}|^s \right)^{\frac{1}{s}}$$

$s = 2$ Euklidska udaljenost

$s = 1$ Manhattan $\quad D(\vec{x}_k, \vec{x}_l) = \sum_{j=1}^{n} |x_{kj} - x_{lj}|$



$s \to \infty$ Čebišljeva $\quad D(\vec{x}_k, \vec{x}_l) = \max\{|x_{kj} - x_{lj}|\}$

Težinska Minkovskog: $D(\vec{x}_k, \vec{x}_l) = \left(\sum_{j=1}^{n} w_j |x_{kj} - x_{lj}|^s \right)^{\frac{1}{s}}$, $w_j \geq 1$

Mahalanobisova:

$$D(\vec{x}_k, \vec{x}_l) = (\vec{x}_k - \vec{x}_l)^T C^{-1} (\vec{x}_k - \vec{x}_l)$$

C – kovarijacijska matrica iz skupa za učenje

$$C(\vec{x}_k, \vec{x}_l) = \frac{\sum_{j=1}^{n} (x_{kj} - m)(x_{lj} - m)}{\sqrt{\sum_{j=1}^{n} (x_{kj} - m)^2 \sum_{j=1}^{n} (x_{lj} - m)^2}}$$

m - srednja vrijednost od x_j

sličnost $= D^{-1}(\vec{x}_k, \vec{x}_l)$

Hammingova udaljenost $D(\vec{x}_k, \vec{x}_l) = \sum_{j=1}^{n} (w_j) d_j \quad d_j = \begin{cases} 1, & x_{kj} \neq x_{lj} \\ 0, & x_{kj} = x_{lj} \end{cases}$

A) Klasifikacija na temelju jednog prototipa

M razreda $\Rightarrow$ prototipovi $\vec{z}_1, \vec{z}_2, \ldots, \vec{z}_M$

$' = T$

$$D_i = \|\vec{x} - \vec{z}_i\| = \sqrt{(\vec{x} - \vec{z}_i)'(\vec{x} - \vec{z}_i)}$$

Pravilo $\quad \vec{x} \in \omega_i$ ako $D_i < D_j \quad \forall j \neq i$

$$D_i^2 = \|\vec{x} - \vec{z}_i\|^2 = (\vec{x} - \vec{z}_i)'(\vec{x} - \vec{z}_i)$$
$$= \vec{x}'\vec{x} - 2\vec{x}'\vec{z}_i + \vec{z}_i'\vec{z}_i$$
$$= \vec{x}'\vec{x} - 2\left(\vec{x}'\vec{z}_i + \frac{1}{2}\vec{z}_i'\vec{z}_i\right)$$

$\min D_i^2 \rightarrow \min D_i$ jer su sve udaljenosti pozitivne

$\vec{x}'\vec{x}$ - ne zavisi od i

maksimum $= \vec{x}'\vec{z}_i - \frac{1}{2}\vec{z}_i'\vec{z}_i$

a to je kada je D_i minimalan

Decizijska f-ja: $d_i(\vec{x}) = \vec{x}'\vec{z}_i - \frac{1}{2}\vec{z}_i'\vec{z}_i \quad , i = 1, 2, \ldots, M$

$$\vec{x} \in \omega_i \text{ ako je } d_i(\vec{x}) > d_j(\vec{x}) \quad \forall j \neq i$$

$d_i(\vec{x})$ - linearna dec. f-ja.

$z_{ij} \quad j = 1, 2, \ldots, n$ - komponente $\vec{z}_i$

$w_{ij} = z_{ij} \quad j = 1, 2, \ldots, n$

$w_{i,n+1} = -\frac{1}{2}\vec{z}_i'\vec{z}_i \qquad \vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \\ 1 \end{pmatrix}$

$\boxed{d_i(\vec{x}) = \vec{w}_i'\vec{x}} \quad i = 1, 2, \ldots, M$

→ korelacijska metoda

$\vec{w}_i = (w_{i1}, w_{i2}, \ldots, w_{i,n+1})'$

B Klasifikacija na temelju više prototipova za pojedini razred

- razred w_i grupira se oko N_i prototipova $\vec{z}_i^1, \vec{z}_i^2, \ldots, \vec{z}_i^{N_i}$

$D_i = \min_l \|\vec{x} - \vec{z}_i^l\| \quad , l = 1, \ldots, N_i$

Izračunati D_i za $i = 1, 2, \ldots, M$

"Pravilo": $\vec{x} \in \omega_1$ ako $D_i < D_j \quad \forall j \neq i$

Dec. f-ja: $d_i(\vec{x}) = \max_l \{(\vec{x}'\vec{z}_i^l) - \frac{1}{2}(\vec{z}_i^l)'\vec{z}_i^l\} \quad l = 1, 2, \ldots, N_i\}$

Pravilo: $\vec{x} \in \omega_i$ ako je $d_i(\vec{x}) > d_j(\vec{x}) \quad \forall j \neq i$

Klas. pravilo najbližeg susjeda: 1-NN

$D(\vec{s}_i, \vec{x}) = \min_l \{D(\vec{s}_l, \vec{x})\}, \quad l = 1, 2, \ldots, N$

$\vec{s}_i \in \{\vec{s}_1, \vec{s}_2, \ldots, \vec{s}_N\}$, D - mjera udaljenosti def. u prostoru značajki

Nepoznati uzorak $\vec{x}$: $\vec{x} \in \omega_k$ ako je $\vec{s}$ najbliži susjed iz razreda ω_k

q-NN pravilo - koristi q najbližih susjeda, $q>1$

└ na temelju N uzoraka nađi q NN uzorku $\vec{x}$ ($q>1$ da nije višekratnik razreda M)

$\sum_i q_i = q$, $q_i \to$ broj vektora koji pripadaju razredu w_i, $i=1,2,\dots,M$

└ razvrstaj $\vec{x}$ u razred w_k za koji je q_k maksimalan

Ocjena pogreške NN, $N \to \infty$ P_{NN}: $P_B \le P_{NN} \le P_B\left(2 - \frac{M P_B}{M-1}\right) \le 2P_B$

opt. Bay. pogr.

$M=2$: $P_B \le P_{qNN} \doteq P_B + \sqrt{\frac{2P_{NN}}{q}}$ ako $q \to \infty \to$ teži optimalnom

NN i q-NN → za velik skup uzoraka za učenje → vrlo djelotvorno

└ problem - složenost $O(\ell N)$ + ako je velika d, $d >> 1$

Inačice NN

(q, ℓ)-NN pravilo → od q najbl. sus. barem ℓ iz w_i $\ell = \frac{2}{3}q$ npr.

→ u suprotnom se ne klasificira ili nekako drugačije

(q, ℓᵢ)-NN → ℓ_i, $\ell_1 = 2/3\, q$, $\ell_2 = \frac{3}{4}q$, ... $\ell_M = \frac{5}{6}q$

→ ℓ_i na temelju apriorne vjer. razreda uzoraka koje su različite!

Metode - s obzirom na način zapisa: u mem. svi uzorci

u mem. samo predstavnici razreda

- na broj (najbl.) susjeda: 1, q ($q>1$), (q,ℓ) i (q,ℓ_i)-NN

Normalizacija podataka

├ različiti rasponi (redovi veličina), a ne mora značenjem biti tako

└ normalizacija → da budu u sličnim opsezima

N-podataka k-te značajke: $\bar{x}_k = \frac{1}{N}\sum_{i=1}^{N} x_{ik}$, $k=1,2,\dots,\ell$

$$\sigma_k^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_{ik} - \bar{x}_k)^2$$

Normalizirana vrijednost: $\hat{x}_{ik} = \frac{x_{ik} - \bar{x}_k}{\sigma_k}$ ← linearna metoda

- imaju $\bar{x}=0$ i $\sigma_k = 1$

Nelinearne metode - ako se podaci ne gomilaju oko srednje vrijednosti

⇒ „softmax" skaliranje: $y = \frac{x_{ik} - \bar{x}_k}{r \cdot \sigma_k}$, $\hat{x}_{ik} = \frac{1}{1+\exp(-y)}$

r → user defined

Eval

1) Imamo dovoljno uzoraka

→ Holdout metoda

S_u – učenje $S = S_u \cup S_i$

S_i – ispitivanje $\emptyset = S_u \cap S_i$

→ inačica – random subsampling: $\text{prosj. točnost}_k = \frac{1}{k} \sum_{i=1}^{k} \text{točnost}_i$

2) Leave-one-out → ponavlja se N puta i broj greške

→ ali velika računska slož.

3) Metoda ponovne zamjene – isti skup prvo za učenje, pa ispit.

– optimistična procjena

k-fold cross-validation → k puta (leave-1-out → k = N)

Uzorci: – dovoljnost informacije

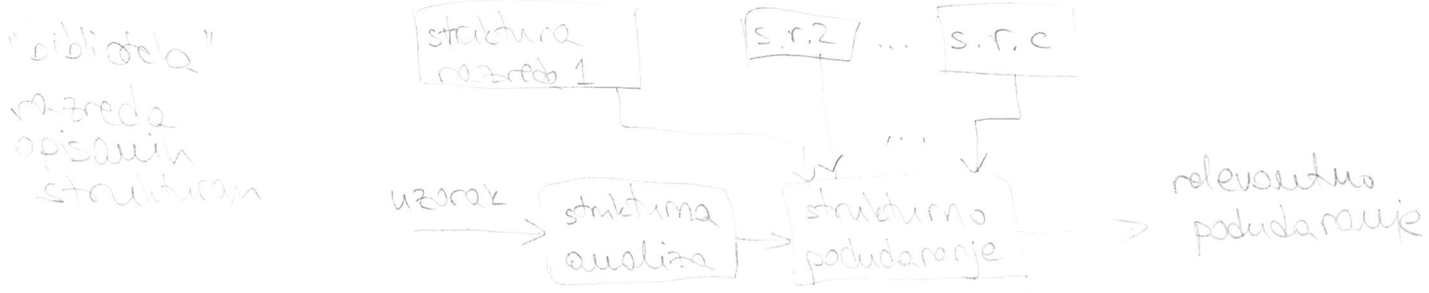
– postojanost značajki

– geometrijska postojanost

Preporuka za $N \doteq (3-5) * d * M$

Jezični pristup

Shematski prikaz sustava za sintaktičko raspoznavanje:



Gram. tipa 0 → slobodna gram.

1 → kontekstno zavisna

2 → kontekstno neovisna → Chomsky

3 → regularna gram.

$L(T_3) \subset L(T_2) \subset L(T_1) \subset L(T_0)$

Problem sintaktičkog raspoznavanja uzoraka

– niz sastavljen od terminala

– pretp. da $\exists$ gramatika G sa svojstvom da se jezik koji generira sastoji od uzoraka koji isključivo pripadaju samo jednom razredu, npr. w_1

Zadani uzorak može biti razvrstan u w_1 ako je on rečenica iz $L(G)$. U drugom slučaju je uzorak iz w_2.

2118

Grupiranje

osnovni koraci u grupiranju:

1. izbor značajki
2. izbor mjere sličnosti/razl.
3. Kriterij grupiranja
4. Algo grupiranja
5. Validacija rez.
6. Interpolacija rez.

Značajke: 1. nominalne (0/1)
2. uređene 1-5
3. intervalno skalirane temp.
4. skalirane omjerom 120kg/60kg

X - skup podataka, $X = \{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N\}$

m - grupiranje X-a : $C_1, \ldots, C_m$ cluster

3 uvjeta: 1. $C_i \neq \emptyset$, $i = 1, 2, \ldots, m$

2. $\bigcup_{i=1}^{m} C_i = X$

3. $C_i \cap C_j = \emptyset$, $i \neq j$, $i, j = 1, 2, \ldots, m$

- izrazito grupiranje

Neizrazito grupiranje: $\mu_j : X \to [0, 1]$, $j = 1, 2, \ldots, m$

$$\sum_{j=1}^{m} u_j(\vec{x}_i) = 1 \quad i = 1, 2, \ldots, N$$

$$0 < \sum_{i=1}^{N} u_j(\vec{x}_i) < N \quad j = 1, 2, \ldots, m$$

membership functions

Mjere bliskosti

- mjera različitosti DM → $d : X \times X \to R$

uvjeti: 1. $\exists d_0 \in R : -\infty < d_0 \leq d(\vec{x}, \vec{y}) < +\infty$

$\forall \vec{x}, \vec{y} \in X$

$d(\vec{x}, \vec{x}) = d_0 \quad \forall \vec{x} \in X$

2. $d(\vec{x}, \vec{y}) = d(\vec{y}, \vec{x}) \quad \forall \vec{x}, \vec{y} \in X$

3. $d(\vec{x}, \vec{y}) = d_0$ iff $\vec{x} = \vec{y}$

dodatni 4. $d(\vec{x}, \vec{z}) \leq d(\vec{x}, \vec{y}) + d(\vec{y}, \vec{z})$, $\forall \vec{x}, \vec{y}, \vec{z} \in X$

- similarity sličnosti obratno: SM → $s : X \times X \to R$

1. $\exists s_0 \in R : -\infty < s(\vec{x}, \vec{y}) \leq s(x)_0 < +\infty \quad \forall \vec{x}, \vec{y} \in X$

2. isto, 3. isto, 4. $s(\vec{x}, \vec{y}) \cdot s(\vec{y}, \vec{z}) \leq [s(\vec{x}, \vec{y}) + s(\vec{y}, \vec{z})] \cdot s(\vec{x}, \vec{z})$ $\forall \vec{x}, \ldots$ 2110

Kriterij grupiranja

Postupci: i) Heuristički - vođeni intuicijom i iskustvom

ii) temeljeni na minimizaciji (ili maksimizaciji) neke kriterijske f-je ili indeksa

najčešći kriterij $$J = \sum_{j=1}^{N_c} \sum_{\vec{x} \in S_j} \|\vec{x} - \vec{m}_j\|^2$$

N_c - broj grupa

S_j - skup uzoraka iz grupe j

$$\vec{m}_j = \frac{1}{N_j} \sum_{\vec{x} \in S_j} \vec{x} \quad , N_j \to \# S_j$$

iii) kombinacija i) i ii)