

UNIVERZA V LJUBLJANI
Fakulteta za elektrotehniko

Simon Dobrišek

ZBIRKA VAJ IN PRIMEROV REŠENIH NALOG
PRI PREDMETU

RAZPOZNAVANJE VZORCEV

UNIVERZITETNI ŠTUDIJSKI PROGRAM II. STOPNJE
ELEKTROTEHNIKA - AVTOMATIKA IN IFORMATIKA

PRVA IZDAJA

Ljubljana, 2013

Predgovor

Skripta z vajami in primeri rešenih nalog so dopolnilno študijsko gradivo, ki je na razpolago študentom predmeta Razpoznavanje vzorcev na podiplomskem drugostopenjskem študijskem programu Elektrotehnike na Fakulteti za elektrotehniko Univerze v Ljubljani na študijski smeri Avtomatika in informatika.

Gradivo vsebuje povzetke teoretičnega ozadja ter pojasnila in rešitve izbranih primerov računskih nalog, ki pridejo v poštev pri pisnem izpitu pri tem predmetu. Vsebina predmeta je sicer v celoti zajeta v učbeniku: "**N. Pavešič**, *Razpoznavanje vzorcev: uvod v analizo in razumevanje vidnih in slušnih signalov*, 3. popravljena in dopolnjena izd., 2 zv., Založba FE in FRI, 2012". Vsebina je zajeta po naslednji straneh te knjige:

- Kaj je razpoznavanje vzorcev? (str. 1-23)
- Razčlenjevanje vzorcev. (str. 71-86, 90-94, 98-119)
- Zapis vzorcev z množico značilnk. (str. 131-165, 170-220)
- Analiza področja uporabe s postopki iskanja rojev. (str. 225-235, 237-238, 239-244, 245-249, 257-263, 283)
- Razvrščanje vzorcev s prileganjem. (str. 287-290, 297-304, 313-324, 467-473) .
- Razvrščanje vzorcev z odločanjem. (str. 330-363, 368-386)
- Razvrščanje vzorcev z večplastnim perceptronom. (str. 396, 399-408)
- Razvrščanje vzorcev z odločitvenimi drevesi. (str. 425-435)
- Preizkušanje razpoznavalnika vzorcev. (str. 23-27)

Primeri, ki niso zajeti v teh skriptah in so bili rešeni na vajah ter tudi lahko pridejo v poštev pri pisnem izpitu, izhajajo iz vsebine in zgledov na naslednjih straneh omenjenega učbenika:

- poglavje 7.3.4 (str. 338-340), zgled 42 (str. 348);
- poglavje 7.3.9.2 (str. 371-375), zgled 49 (str. 376-378);
- poglavje 8.2.2 (str. 432-433); zgled 58 (str. 433-435).

Kazalo

1	Upragovljanje slik z maksimizacijo informacije	5
1.1	Matematični zapis digitalne sive slike	5
1.2	Razčlenjevanje slik z upragovljanjem	6
1.3	Definicija upragovljanja slike	6
1.4	Upragovljanje z maksimizacijo informacije	7
1.5	Vprašanja in naloge	8
1.5.1	Vprašanje	8
1.5.2	Vprašanje	8
1.5.3	Naloga	8
1.5.4	Naloga	8
2	Približni zapis vzorcev z začetnimi členi ortogonalnih transformirank	10
2.1	Transformacija Karhunenena in Loeveja	11
2.2	Vprašanja in naloge	12
2.2.1	Naloga	12
3	Optimalni postopki luščenja značilk	18
3.1	Določanje značilk z izbiranjem	18
3.1.1	Zaporedno iskanje »nazaj«	21
3.1.2	Zaporedno iskanje »naprej«	22
3.2	Vprašanja in naloge	22
3.2.1	Vprašanje	22
3.2.2	Naloga	23
4	Analiza področja uporabe v vzorčnem prostoru s postopki iskanja rojev..	25
4.1	Postopki iskanja rojev v množici vzorcev	26
4.2	Hierarhični postopek iskanja rojev	26
4.2.1	Naloga	29
5	Razvrščanje vzorcev s prileganjem	33
5.1	Pravilo razvrščanja "najbližji sosed" (1-NN)	33
5.1.1	Naloga	34
5.2	Učenje	35

5.3	Predstavitev razreda z značilnim vzorcem	36
5.3.1	Naloga.....	36
5.4	Urejanje učne množice.....	37
5.5	Zgoščevanje učne množice.....	38
5.5.1	Naloga.....	39
5.6	Računanje podobnosti dveh nizov znakov.....	42
5.7	Levenshteinova razdalja med dvema nizoma znakov.....	44
5.7.1	Naloga.....	46
6	Razvrščanje vzorcev s odločanjem	49
6.1	Trije načrti razvrščevalnikov.....	50
6.2	Linearne odločitvene funkcije	51
6.2.1	Naloga.....	52
6.3	Tolmačenje pravila razvrščanja 1 – NN z linearnimi odločitvenimi funkcijami.....	54
6.3.1	Naloga.....	56

1 Upragovljanje slik z maksimizacijo informacije

1.1 Matematični zapis digitalne sive slike

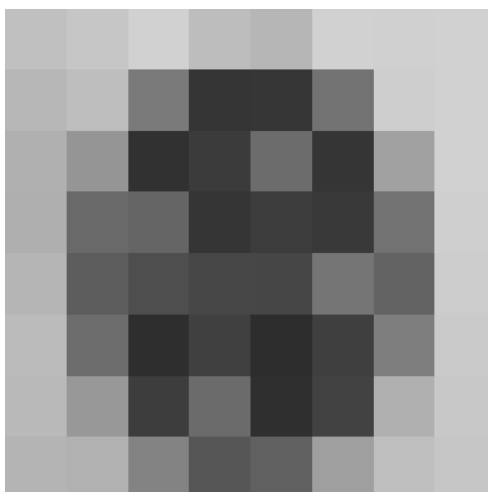
Digitalno sivo sliko $f(x, y)$ matematično zapišemo s preslikavo, ki vsaki točki slike (slikovnemu elementu) priredi neko svetilnost (sivi nivo):

$$f: I_{N_x} \times I_{N_y} \rightarrow G_L ,$$

kjer sta $I_{N_x} = \{1, \dots, N_x\}$ in $I_{N_y} = \{1, \dots, N_y\}$ podmnožici celih ne-negativnih števil ter $G_L = \{0, \dots, L - 1\}$ množica diskretnih vrednosti svetilnosti (sivih nivojev). Pri t.i. osem-bitni sivi sliki, ker število sivih nivojev L enako 256, kjer vrednost svetilnosti 0 predstavlja povsem črno in vrednost svetilnosti 255 povsem belo barvo, torej, $G_L = \{0, \dots, 255\}$.

Primer matematičnega zapisa in grafične ponazoritve sive digitalne slike je podan spodaj

$$f(x, y) = \left\{ \begin{array}{cccccccc} 192, & 198, & 209, & 189, & 182, & 209, & 208, & 209 \\ 183, & 190, & 122, & 53, & 54, & 114, & 206, & 209 \\ 176, & 149, & 49, & 59, & 108, & 53, & 161, & 209 \\ 175, & 105, & 101, & 53, & 60, & 57, & 114, & 207 \\ 181, & 93, & 78, & 71, & 70, & 117, & 99, & 204 \\ 187, & 109, & 46, & 64, & 45, & 63, & 126, & 202 \\ 185, & 152, & 61, & 107, & 47, & 66, & 176, & 200 \\ 180, & 177, & 131, & 86, & 96, & 159, & 191, & 198 \end{array} \right\}$$



1.2 Razčlenjevanje slik z upragovljanjem

Z upragovljanjem razčlenjujemo slike razmeroma enostavnih prizorov. Na primer, v industrijskih aplikacijah pogosto sestavlja prizor en sam predmet (izdelek), ki leži na določeni enolični podlagi. Takšne prizore lahko zelo hitro razčlenimo (v tem primeru ločimo predmet od podlage) na podlagi porazdelitve frekvenc sivih nivojev slike, ker je v tem primeru histogram porazdelitve bi-modalen z dvema izrazitima vrhovoma. En vrh histograma bo ustrezal svetilnosti podlage, drugi pa svetilnosti predmeta. Iz takega histograma lahko določimo mejo, to je prag med sivimi nivoji podlage in sivimi nivoji predmeta. S pragom lahko razvrstimo vse slikovne elemente v dva razreda: v razred $\omega_1 = \text{"podlaga"}$ in v razred $\omega_2 = \text{"predmet"}$ (ali obratno, če je predmet svetlejši od podlage). Primera takšnih prizorov sta podana spodaj.



1.3 Definicija upragovljanja slike

Pri upragovljanju sliko $f(x, y)$ prevedemo v sliko $f_t(x, y)$, ki je dana s preslikavo

$$f_t: I_{N_x} \times I_{N_y} \rightarrow \Omega ,$$

ki vsaki točki slike priredi eno od (navadno dveh) oznak področja na sliki $\Omega = \{\omega_1, \omega_2\}$ (npr. $\omega_1 = \text{"podlaga"}$ in $\omega_2 = \text{"predmet"}$ ali preprosto $\omega_1 = \text{"0"}$ in $\omega_2 = \text{"1"}$).

Preslikavo f_t tako definiramo z izrazom

$$f_t(x, y) = \begin{cases} \omega_1 & f(x, y) > t \\ \omega_2 & f(x, y) \leq t \end{cases} ,$$

kjer je t prag svetilnosti, ki ločuje področja na sliki.

Prag svetilnosti določamo iz porazdelitve relativnih frekvenc sivih nivojev $\mathbf{P} = \{P_0, \dots, P_i, \dots, P_{L-1}\}$, kjer je

$$P_i = \frac{n_i}{n} .$$

Število n_i v gornjem izrazu je število slikovnih elementov s sivim nivojem $i \in G_L$ in n število vseh slikovnih elementov slike ($n = N_x \cdot N_y$).

1.4 Upragovljanje z maksimizacijo informacije

Če sliko upragovljamo s pragom t , delimo množico sivih nivojev slike G_L na dve tuji množici $G_0 \cap G_1 = \emptyset$, kjer sta $G_0 = \{0, \dots, t\}$ in $G_1 = \{t + 1, \dots, L - 1\}$. Množicama G_0 in G_1 pripadata porazdelitvi relativnih frekvenc sivih nivojev

$$\mathbf{P}_0 = \mathbf{P}_0(t) = \left\{ \frac{P_0}{P^*(t)}, \dots, \frac{P_t}{P^*(t)} \right\}$$

in

$$\mathbf{P}_1 = \mathbf{P}_1(t) = \left\{ \frac{P_{t+1}}{1 - P^*(t)}, \dots, \frac{P_{L-1}}{1 - P^*(t)} \right\} ,$$

kjer je

$$P^*(t) = \sum_{i=0}^t P_i .$$

Para (G_0, \mathbf{P}_0) in (G_1, \mathbf{P}_1) lahko obravnavamo kot verjetnostni shemi dveh naključnih spremenljivk z informacijama

$$H_0(t) = - \sum_{i=0}^t \frac{P_i}{P^*(t)} \log\left(\frac{P_i}{P^*(t)}\right)$$

in

$$H_1(t) = - \sum_{i=t+1}^{L-1} \frac{P_i}{1 - P^*(t)} \log\left(\frac{P_i}{1 - P^*(t)}\right) .$$

Optimalni prag svetilnosti t^* , lahko določimo kot prag, ki maksimizira celotno informacijo upragovljene slike, to je:

$$t^* = \arg \max_{t=0, \dots, L-2} \{H_0(t) + H_1(t)\}$$

Optimalni prag določimo z računanjem informacije za vse možne vrednosti praga in ugotavljanjem, pri kateri vrednosti praga je vrednost informacije največja.

1.5 Vprašanja in naloge

1.5.1 Vprašanje

Podajte primer porazdelitve relativnih frekvenc sivih nivojev osem-bitne sive slike tipičnega industrijskega prizora, kjer je svetlejši izdelek na temnejšem tekočem traku.

1.5.2 Vprašanje

Katera najvišja in katera najnižja vrednost lahko nastopa v porazdelitvi relativnih frekvenc sivih nivojev (obrazložite).

1.5.3 Naloga

Vzemimo, da smo pri upagovljanju sivih slik z maksimizacijo informacije izračunali pri možni vrednosti praga $t = 5$ vrednosti funkcij $H_0(5) = 2,3$ in $H_1(5) = 1,7$, pri drugi možni vrednosti praga $t = 10$ pa vrednosti $H_0(10) = 2,2$ in $H_1(10) = 1,9$. Pri kateri od obeh obravnavanih možnih vrednosti praga bi torej bila informacija v upragovljeni sliki višja.

1.5.4 Naloga

Dana je štiri-bitna siva slika s podaj podano množico diskretnih vrednosti svetilnosti (sivih nivojev) in porazdelitvijo njihovih relativnih frekvenc

$$G_8 = \{0, 1, 2, 3, 4, 5, 6, 7\}$$

$$P = \left\{0, \frac{1}{10}, \frac{2}{10}, \frac{1}{10}, \frac{2}{10}, \frac{3}{10}, \frac{1}{10}, 0\right\}$$

Predpostavimo, da smo izbrali postopek upagovljanja slike z maksimizacijo informacije. Izračunajte vrednost informacije slike pri pragu svetilnosti $t = 3$ in $t = 4$ in ugotovite, pri kateri od obeh vrednosti praga je vrednost informacije višja.

Rešitev

Najprej izračunajmo vrednost informacije pri pragu $t = 3$.

$$P^*(3) = 0 + \frac{1}{10} + \frac{2}{10} + \frac{1}{10} = \frac{4}{10} \quad 1 - P^*(3) = 1 - \frac{4}{10} = \frac{6}{10}$$

$$G_0 = \{0, 1, 2, 1\} \quad G_1 = \{2, 3, 1, 0\}$$

$$P_0 = \left\{0, \frac{1}{4}, \frac{2}{4}, \frac{1}{4}\right\} = \left\{0, \frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right\} \quad P_1 = \left\{\frac{2}{6}, \frac{3}{6}, \frac{1}{6}, 0\right\} = \left\{\frac{1}{3}, \frac{1}{2}, \frac{1}{6}, 0\right\}$$

$$H_0(3) = H\left(0, \frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right) = -\frac{1}{4}\log_2\left(\frac{1}{4}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right) = 1,5$$

$$H_1(3) = H\left(0, \frac{1}{3}, \frac{1}{2}, \frac{1}{6}\right) = -\frac{1}{3}\log_2\left(\frac{1}{3}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{6}\log_2\left(\frac{1}{6}\right) \approx 1,459$$

$$H_0(3) + H_1(3) \approx 2,959$$

Nato izračunajmo še vrednost informacije pri pragu $t = 4$.

$$P^*(4) = 0 + \frac{1}{10} + \frac{2}{10} + \frac{1}{10} + \frac{2}{10} = \frac{6}{10} \quad 1 - P^*(4) = 1 - \frac{6}{10} = \frac{4}{10}$$

$$G_0 = \{0, 1, 2, 1, 2\} \quad G_1 = \{3, 1, 0\}$$

$$P_0 = \left\{0, \frac{1}{6}, \frac{2}{6}, \frac{1}{6}, \frac{2}{6}\right\} = \left\{0, \frac{1}{6}, \frac{1}{3}, \frac{1}{6}, \frac{1}{3}\right\} \quad P_1 = \left\{\frac{3}{4}, \frac{1}{4}, 0\right\}$$

$$H_0(4) = H\left(0, \frac{1}{6}, \frac{1}{3}, \frac{1}{6}, \frac{1}{3}\right) =$$

$$= -\frac{1}{6}\log_2\left(\frac{1}{6}\right) - \frac{1}{3}\log_2\left(\frac{1}{3}\right) - \frac{1}{6}\log_2\left(\frac{1}{6}\right) - \frac{1}{3}\log_2\left(\frac{1}{3}\right) \approx 1,918$$

$$H_1(4) = H\left(\frac{3}{4}, \frac{1}{4}, 0\right) = -\frac{3}{4}\log_2\left(\frac{3}{4}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right) \approx 0,811$$

$$H_0(4) + H_1(4) \approx 2,729$$

Vrednost informacije je torej očitno višja pri vrednosti praga svetilnosti $t = 3$ in nižja pri vrednosti praga svetilnosti $t = 4$.

2 Približni zapis vzorcev z začetnimi členi ortogonalnih transformirank

Postopki luščenja značilnik temeljijo na predpostavki, da lahko vsak vzorec opišemo z vrednostmi r spremenljivk, ki jih obravnavamo kot krajevni vektor $\mathbf{y} = (y_1, \dots, y_r)^T$. Spremenljivkam, ki črpajo iz zaloge realnih števil, pravimo *meritve*. Pri postopkih luščenja značilnik, ki temelje na uporabi ortogonalnih linearnih transformacij, dosežemo zgostitev informacije tako, da zapišemo vzorec \mathbf{y} z $n \leq r$ začetnih členov izbrane ortogonalne razvrstitve, ker le-ti vsebujejo največ, ali celo celotno, informacijo o vzorcu.

Linearno transformacijo, ki preslikuje vektorje \mathbf{y} v vektorje \mathbf{c} , definiramo z matrično enačbo:

$$\mathbf{c} = \mathbf{A}\mathbf{y} ,$$

kjer \mathbf{A} označuje transformacijsko matriko, ki jo po vrsticah sestavljajo linearno neodvisni r -razsežni vektorji $\mathbf{a}_1^T, \dots, \mathbf{a}_n^T$, torej

$$\mathbf{A}_{n \times r} = \begin{bmatrix} a_{11} & \cdots & a_{1r} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nr} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1^T \\ \vdots \\ \mathbf{a}_n^T \end{bmatrix}$$

Vektor \mathbf{c} je vektor značilnik, ki so določeni kot koeficienti ortogonalne razvrstitve $\mathbf{c} = (c_1, \dots, c_n)^T$. Podana linearna transformacija je ortogonalna oziroma ortonormalna, če velja $\mathbf{A}\mathbf{A}^T = \mathbf{I}$.

Podano linearno transformacijo lahko zapišemo tudi kot sistem linearnih enačb:

$$c_i = \mathbf{a}_i^T \mathbf{y} = \sum_{j=1}^r a_{ij} y_j; \quad i = 1, \dots, n .$$

V primeru, ko je $n < r$, pri linearni transformaciji izvorni vzorec \mathbf{y} izrazimo s približkom $\hat{\mathbf{y}}$, ki je določen s koeficienti ortogonalne razvrstitve

$$\hat{\mathbf{y}} = \mathbf{A}^T \mathbf{c} .$$

Napako, ki jo pri tej izražavi naredimo, pa določimo kot srednjo kvadratno napako

$$\|\Delta_y\|^2 = \|\mathbf{y} - \hat{\mathbf{y}}\|^2 = \|\mathbf{y} - \mathbf{A}^T \mathbf{c}\|^2 = \|\mathbf{y}\|^2 - \|\mathbf{c}\|^2 .$$

V primeru, ko je $n = r$, pri izražavi ne naredimo napake in je $\mathbf{y} = \hat{\mathbf{y}}$.

Na tem področju se uporablja postopke približnega zapisa vzorcev, ki temeljijo na ortogonalnih razvrstitvah Karhunena in Loeveja, Fourierja, Walsha in Hadamarda ter na valčni Haarovi transformaciji idr.

2.1 Transformacija Karhunena in Loeveja

Transformacijo Karhunena in Loeveja lahko definiramo kot postopek minimizacije srednje kvadratne napake približnega zapisa vzorcev. Izpeljava tega postopka pokaže, da pri izbranem številu koeficientov ortogonalne razvrstitve $n < r$ in dani množici vzorcev $\mathcal{S}_N = \{\mathbf{y}_j\}$ pri $j = 1, \dots, N$ dobimo najmanjšo srednjo kvadratno napako $\bar{\Delta}_y^2 = \mathbf{E}\{\|\Delta_y\|^2\}$, če transformacijsko matriko \mathbf{A} določimo iz prvih $n \leq r$ normiranih lastnih vektorjev $\mathbf{e}_1^T, \dots, \mathbf{e}_n^T$ kovariančne matrike

$$\begin{aligned} \mathbf{K} &= \mathbf{E}\{(\mathbf{y} - \mathbf{E}(\mathbf{y}))(\mathbf{y} - \mathbf{E}(\mathbf{y}))^T\} = \mathbf{E}\{\mathbf{y}\mathbf{y}^T\} - \mathbf{E}(\mathbf{y})\mathbf{E}(\mathbf{y})^T \\ &= \mathbf{E}\{\mathbf{y}\mathbf{y}^T\} - \mathbf{m}\mathbf{m}^T, \end{aligned}$$

ki so urejeni po velikosti pripadajočih lastnih vrednosti tako, da velja

$$\lambda_1 \geq \lambda_2 \geq \dots \lambda_n \geq \dots \lambda_r ,$$

pri čemer lastne vektorje \mathbf{e}_j in lastne vrednosti λ_j pri $j = 1, \dots, n$ določa enačba

$$\mathbf{K}\mathbf{e}_j = \lambda_j \mathbf{e}_j ,$$

oziroma enačba

$$(\mathbf{K} - \lambda_j \mathbf{I})\mathbf{e}_j = \mathbf{0} .$$

Transformacijsko matriko \mathbf{A} tako določajo lastni vektorji $\mathbf{e}_1^T, \dots, \mathbf{e}_n^T$ po izrazu

$$\mathbf{A}_{n \times r} = \begin{bmatrix} e_{11} & \dots & e_{1r} \\ \vdots & \ddots & \vdots \\ e_{n1} & \dots & e_{nr} \end{bmatrix} = \begin{bmatrix} \mathbf{e}_1^T \\ \vdots \\ \mathbf{e}_n^T \end{bmatrix}$$

Vektorje značilik $= (c_1, \dots, c_n)^T$ pridobimo s transformacijo

$$\mathbf{c} = \mathbf{A}(\mathbf{y} - E(\mathbf{y})) = \mathbf{A}(\mathbf{y} - \mathbf{m}),$$

kjer \mathbf{m} označuje povprečni vektor vseh vzorcev \mathbf{y} . Središčenje moramo upoštevati tudi pri povratni izražavi približka

$$\hat{\mathbf{y}} = \mathbf{A}^T \mathbf{c} + \mathbf{m}.$$

V primeru, ko je $n < r$, pri izražavi približka naredimo določeno napako in ob upoštevanju ortonormalnosti lastnih vektorjev lahko izpeljemo srednjo kvadratno napako vseh izraženih vektorjev

$$\bar{\Delta}_{\mathbf{y}}^2 = \sum_{j=n+1}^r \lambda_j$$

Srednja kvadratna napaka vseh izraženih vzorcev je torej kar seštevek lastnih vrednosti lastnih vektorjev, ki jih ne upoštevamo pri linearni transformaciji. Iz tega je očitno, zakaj za tvorjenje transformacijske matrike izbiramo lastne vektorje, ki ustrezajo višjim lastnim vrednostim.

2.2 Vprašanja in naloge

2.2.1 Naloga

Izračunajte transformacijo Karhunen in Loeveja za vzorce iz učne množice $\mathcal{U}_2 = \{S_{10}, \Omega\} = \{U_1, U_2\}$, kjer so

$$S_{10} = \{(1,3)^T, (2,2)^T, (3,3)^T, (2,4)^T, (4,1)^T, (5,2)^T, (5,4)^T, (6,1)^T\},$$

$$\Omega = \{\omega_1, \omega_2\},$$

$$U_1 = \{(1,3)^T, (2,2)^T, (3,3)^T, (2,4)^T\}$$

$$U_2 = \{(4,1)^T, (5,2)^T, (5,4)^T, (6,1)^T\}.$$

Rešitev

Če upoštevamo, da je ocena povprečnega vektorja vseh vzorcev

$$\hat{\mathbf{m}} = E(\mathbf{y}) = \frac{1}{N} \sum_{\mathbf{y} \in S_N} \mathbf{y} = \frac{1}{8} \sum_{j=1}^8 \mathbf{y}_j = \frac{1}{8} (28, 20)^T = (3.5, 2.5)^T$$

in, da je ocena kovariančne matrike

$$\begin{aligned}\hat{\mathbf{K}} &= E\{\mathbf{y}\mathbf{y}^T\} - \hat{\mathbf{m}}\hat{\mathbf{m}}^T = \left(\frac{1}{N} \sum_{\mathbf{y} \in S_N} \mathbf{y}\mathbf{y}^T \right) - \hat{\mathbf{m}}\hat{\mathbf{m}}^T = \left(\frac{1}{8} \sum_{j=1}^8 \mathbf{y}_j \mathbf{y}_j^T \right) - \hat{\mathbf{m}}\hat{\mathbf{m}}^T = \\ &= \frac{1}{8} \begin{bmatrix} 1 \cdot 1 + 2 \cdot 2 + \dots & 1 \cdot 3 + 2 \cdot 2 + \dots \\ 3 \cdot 1 + 2 \cdot 2 + \dots & 3 \cdot 3 + 2 \cdot 2 + \dots \end{bmatrix} - \begin{bmatrix} 12.25 & 8.75 \\ 8.75 & 6.25 \end{bmatrix} \\ &= \frac{1}{8} \begin{bmatrix} 120 & 64 \\ 64 & 60 \end{bmatrix} - \begin{bmatrix} 12.25 & 8.75 \\ 8.75 & 6.25 \end{bmatrix} \\ &= \begin{bmatrix} 15 & 8 \\ 8 & 7.5 \end{bmatrix} - \begin{bmatrix} 12.25 & 8.75 \\ 8.75 & 6.25 \end{bmatrix} = \begin{bmatrix} 2.75 & -0.75 \\ -0.75 & 1.25 \end{bmatrix},\end{aligned}$$

dobimo lastni vrednosti λ_1 in λ_2 matrike $\hat{\mathbf{K}}$ z izračunom enačbe

$$|\hat{\mathbf{K}} - \lambda \mathbf{I}| = \det \left(\begin{bmatrix} 2.75 - \lambda & -0.75 \\ -0.75 & 1.25 - \lambda \end{bmatrix} \right) = 0,$$

Ki nas pripelje do kvadratne enačbe

$$(2.75 - \lambda)(1.25 - \lambda) - 0.75^2 = \lambda^2 - 4\lambda + 2.875 = 0.$$

Rešitvi kvadratne enačbe in s tem lastni vrednosti sta

$$\lambda_{1,2} = \frac{4 \pm \sqrt{4^2 - 4 \cdot 2.875}}{2} \Rightarrow \lambda_1 \approx 3.0607, \lambda_2 \approx 0.9393$$

Lastna vektorja \mathbf{e}_j pri $j = 1, 2$ izračunamo iz enačbe:

$$(\hat{\mathbf{K}} - \lambda_j \mathbf{I})\mathbf{e}_j = \mathbf{0}.$$

Najprej izračunamo prvi lastni vektor $\mathbf{e}_1 = (e_{11}, e_{12})^T$

$$\begin{bmatrix} 2.75 - \lambda_1 & -0.75 \\ -0.75 & 1.25 - \lambda_1 \end{bmatrix} \begin{bmatrix} e_{11} \\ e_{12} \end{bmatrix} \approx \begin{bmatrix} -0.3107 & -0.75 \\ -0.75 & -1.8107 \end{bmatrix} \begin{bmatrix} e_{11} \\ e_{12} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow$$

$$-0.3107 e_{11} - 0.75 e_{12} = 0$$

$$-0.75 e_{11} - 1.8107 e_{12} = 0$$

Iz obeh enačb lahko izpeljemo zvezo $e_{11} \approx -2.4142 e_{12}$, kar pomeni, da so vsi vektorji $\hat{\mathbf{e}}_1 = (e_{11}, e_{12})^T$, katerih komponenti sta različni od nič in v podani zvezi, lastni vektorji matrike \mathbf{K} , ki pripadajo prvi lastni vrednosti λ_1 . Če

izberemo vrednost $e_{12} = 1$, je takšen vektor tudi vektor $\hat{\mathbf{e}}_1 = (-2.4142, 1)^T$. Glede na to, da za določitev transformacijske matrike potrebujemo normirani vektor, dobljeni vektor še normirajmo

$$\mathbf{e}_1 = \frac{\hat{\mathbf{e}}_1}{\|\hat{\mathbf{e}}_1\|} \approx \frac{\hat{\mathbf{e}}_1}{2.6131} \approx (-0.9239, 0.3827)^T$$

Na podoben način lahko določimo še drugi lastni vektor $\mathbf{e}_2 = (e_{21}, e_{22})^T$

$$\begin{bmatrix} 2.75 - \lambda_2 & -0.75 \\ -0.75 & 1.25 - \lambda_2 \end{bmatrix} \begin{bmatrix} e_{21} \\ e_{22} \end{bmatrix} \approx \begin{bmatrix} 1.8107 & -0.75 \\ -0.75 & 0.3107 \end{bmatrix} \begin{bmatrix} e_{21} \\ e_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow$$

$$1.8107 e_{21} - 0.75 e_{22} = 0$$

$$-0.75 e_{21} - 0.3107 e_{22} = 0$$

Iz obeh enačb lahko izpeljemo zvezo $e_{21} \approx 0.4142 e_{22}$. Če izberemo vrednost $e_{12} = 1$, je iskani lastni vektor tudi vektor $\hat{\mathbf{e}}_2 = (0.4142, 1)^T$. Glede na to, da za določitev transformacijske matrike potrebujemo normirani vektor, tudi ta dobljeni vektor normirajmo

$$\mathbf{e}_2 = \frac{\hat{\mathbf{e}}_2}{\|\hat{\mathbf{e}}_2\|} \approx \frac{\hat{\mathbf{e}}_2}{1.0824} \approx (0.3827, 0.9239)^T$$

V primeru, ko transformacijsko določimo iz obeh lastnih vektorjev, torej

$$\mathbf{A} = \begin{bmatrix} \mathbf{e}_1^T \\ \mathbf{e}_2^T \end{bmatrix} = \begin{bmatrix} -0.9239 & 0.3827 \\ 0.3827 & 0.9239 \end{bmatrix},$$

dobimo s transformacijo

$$\mathbf{c} = \mathbf{A}(\mathbf{y} - E(\mathbf{y})) = \mathbf{A}(\mathbf{y} - \hat{\mathbf{m}})$$

vektorje značilik

$$U'_1 \approx \{(2.50, -0.49)^T, (1.19, -1.04)^T, (0.65, 0.27)^T, (1.96, 0.81)^T\}$$

$$U'_2 \approx \{(-1.04, -1.19)^T, (-1.58, 0.11)^T, (-0.81, 1.96)^T, (-2.88, -0.43)^T\} .$$

Povprečni vektor teh vektorjev značilik je $\mathbf{0}$, kovariančna matrika pa je diagonalna matrika, ki po diagonali vsebuje lastne vrednosti, torej

$$\hat{\mathbf{K}} = \begin{bmatrix} 3.0607 & 0 \\ 0 & 0.9393 \end{bmatrix}.$$

V primeru, ko želimo izluščiti eno samo značilko, izberemo za določitev linearne transformacije prvi lastni vektor \mathbf{e}_1 , ker mu pripada višja lastna vrednost λ_1 . Vrednosti značilk pridobimo s transformacijo

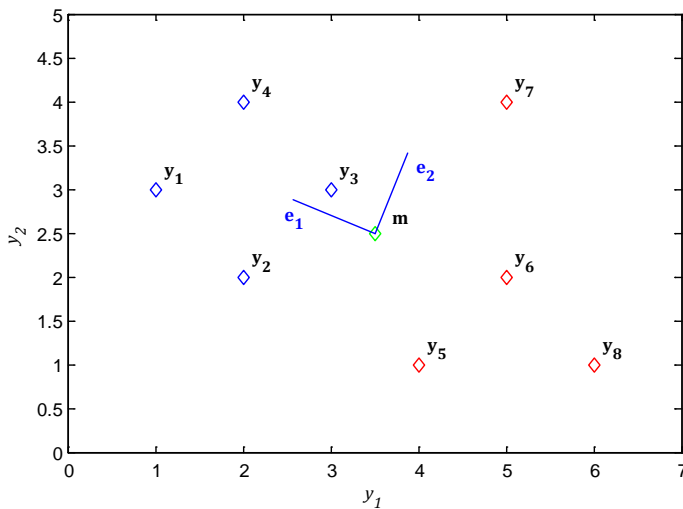
$$\mathbf{c} = \mathbf{e}_1^T (\mathbf{y} - E(\mathbf{y})) = \mathbf{e}_1^T (\mathbf{y} - \hat{\mathbf{m}}),$$

torej

$$U'_1 \approx \{2.50, 1.19, 0.65, 1.96\}$$

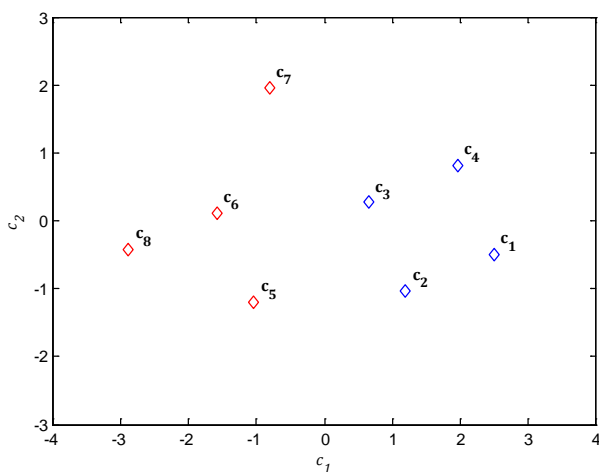
$$U'_2 \approx \{-1.04, -1.58, -0.81, -2.88\}.$$

Rezultat izračunov lahko ponazorimo v \mathbb{R}^2 . Najprej ponazorimo vzorce in lastna vektorja v prostoru izvirnih krajevnih vektorjev.

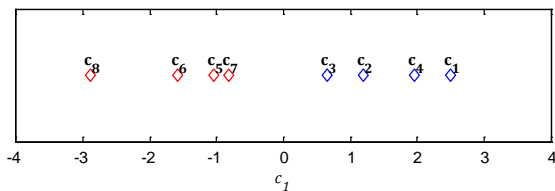


Nato ponazorimo vzorce preslikane v prostor obeh lastnih vektorjev.

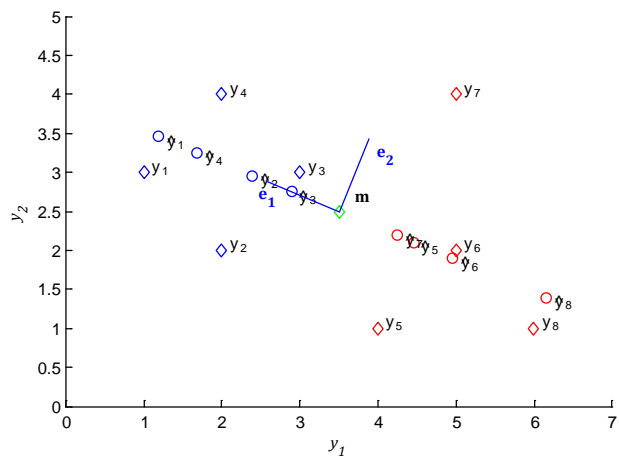
Približni zapis vzorcev z začetnimi členi ortogonalnih transformirank



V primeru, ko za preslikavo izberemo le prvi lastni vektor z višjo lastno vrednostjo, torej e_1 , dobimo preslikavo v eno samo značilko, ki jo ponazarja spodnja slika.



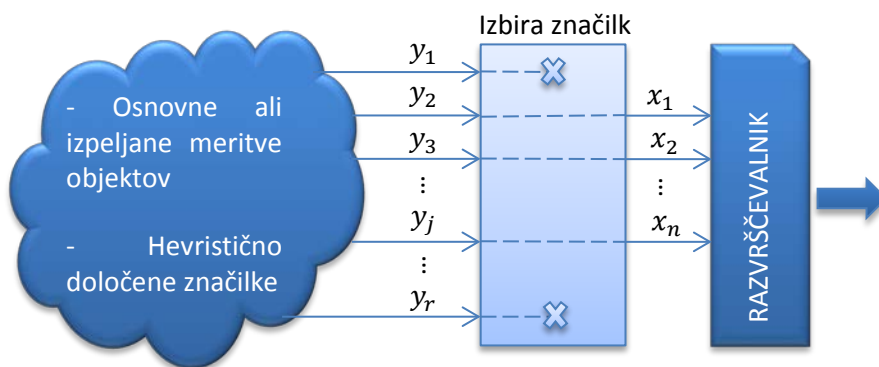
Po povratni preslikavi izraženih značilk v izvorni prostor vzorcev dobimo približke, na osi, ki jo določa prvi lastni vektor. To ponazarja naslednja slika



3 Optimalni postopki luščenja značil

3.1 Določanje značil z izbiranjem

Postopki luščenja značil temeljijo na predpostavki, da lahko vsak vzorec opišemo z vrednostmi r spremenljivk, ki jih obravnavamo kot krajevni vektor $\mathbf{y} = (y_1, \dots, y_r)^T$. Spremenljivkam, ki črpajo iz zaloge realnih števil, pravimo *meritve*. Določanje značil z izbiranjem temelji na iskanju tistih spremenljivk vzorcev, ki ne prispevajo k natančnosti razpoznavanja, oz. ne prispevajo k boljši ločljivosti razredov vzorcev v prostoru značil. Takšne spremenljivke nato izločimo iz opisa objekta razpoznavanja.



Izbiranje značil formalno definiramo kot problem pri katerem v množici r spremenljivk (meritev) \mathcal{Y} iščemo tako podmnožico n (pri $n < r$) spremenljivk, ki je najboljša glede na izbrano kriterijsko funkcijo J , oziroma

$$J(\mathcal{X}) = \max_{P \in \mathcal{P}} \{J(P)\} ,$$

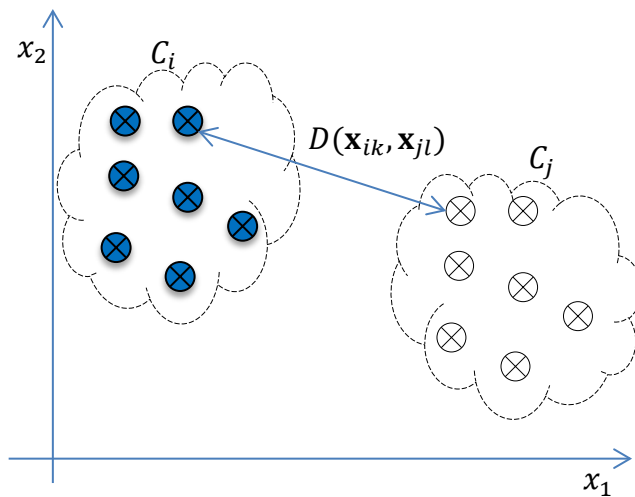
kjer P poljubna podmnožica \mathcal{Y} in \mathcal{P} množica vseh možnih podmnožic \mathcal{Y} . Množica \mathcal{X} je najboljša podmnožica, ki daje najvišjo/najnižjo vrednost izbrane kriterijske funkcije. Izbranim spremenljivkam v \mathcal{X} pravimo značilke vzorca, ki objekt razpoznavanja opisujejo s krajevnim vektorjem $\mathbf{x} = (x_1, \dots, x_n)^T$. Z izbiro značil tako izvedemo preslikavo

$$\mathbf{y} = (y_1, \dots, y_r)^T \rightarrow \mathbf{x} = (x_1, \dots, x_n)^T ; n < r$$

Za kriterijsko funkcijo navadno izberemo eno od mer ločljivosti med razredi vzorcev. Med tovrstnimi merami je računsko manj zahtevna povprečna razdalja med vzorci, ki jo določa naslednji izraz

$$J_D = \frac{1}{2} \sum_{i=0}^M P(C_i) \sum_{\substack{j=1 \\ j \neq i}}^M P(C_j) \frac{1}{N_i N_j} \sum_{k=1}^{N_i} \sum_{l=1}^{N_j} D(\mathbf{x}_{ik}, \mathbf{x}_{jl}) ,$$

kjer $P(C_i)$ označuje a priori verjetnost razreda vzorcev C_i in N_i število vzorcev iz razreda C_i pri $i = 1, \dots, M$. Funkcija $D(\mathbf{x}_{ik}, \mathbf{x}_{jl})$ označuje osnovno mero razdalje med k -tim vzorcem v i -tem razredu in l -tim vzorcem v j -tem razredu. Čim večja bo povprečna razdalja med vzorci različnih razredov, tem večja naj bi bila ločljivost med razredi vzorcev. Mero razdalje med vzorci različnih razredov ponazarja spodnja slika.



A priori verjetnosti razredov navadno ocenimo s posebno statistično analizo za dano področje uporabe ali kar iz števila podanih vzorcev. Če je dano N vseh vzorcev in je od tega N_i vzorcev iz razreda C_i , potem je ocena za $P(C_i)$ kar $P(C_i) = N_i/N$. Če ocena a priori verjetnosti ni zanesljiva, potem navadno predpostavljamo, da so vsi razredi enako-verjetni, torej $P(C_i) = 1/M$.

Mere razdalje med vzorci različnih razredov, ki jih najbolj pogosto uporabljamo, so matematične razdalje v vektorskem prostoru, ki jim pravimo razdalje Minkovskega.

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

kjer je sta $\mathbf{x}_k = (x_{k,1}, \dots, x_{k,n})^T$ in $\mathbf{x}_l = (x_{l,1}, \dots, x_{l,n})^T$ poljubna dva vektorja značilik.

Za mero razdalje med vzorci tako najbolj pogosto uporabljamo razdalje Minkovskega pri vrednostih $s = 1$, $s = 2$ in $s \rightarrow \infty$.

Razdalji pri $s = 1$ pravimo razdalja "City Block" ali "Manhattan". Določa jo izraz

$$D_{CB}(\mathbf{x}_k, \mathbf{x}_l) = \sum_{j=1}^n |x_{k,j} - x_{l,j}|.$$

Razdalji pri $s = 2$ pravimo Evklidova razdalja

$$D_E(\mathbf{x}_k, \mathbf{x}_l) = \sqrt{\sum_{j=1}^n (x_{k,j} - x_{l,j})^2} = \sqrt{(\mathbf{x}_k - \mathbf{x}_l)^T (\mathbf{x}_k - \mathbf{x}_l)} = \|\mathbf{x}_k - \mathbf{x}_l\|.$$

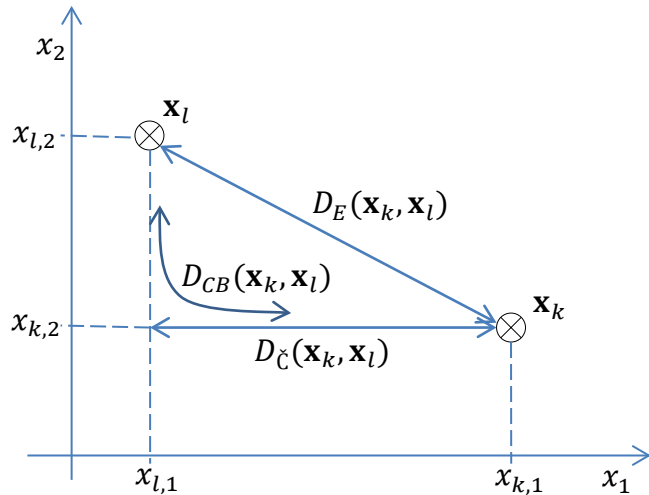
Razdalji pri $s \rightarrow \infty$ pa Chebysheva (Čebiševljeva) razdalja

$$D_{\check{c}}(\mathbf{x}_k, \mathbf{x}_l) = \max_{j=1, \dots, n} \{|x_{k,j} - x_{l,j}|\}.$$

Za omenjen razdalje velja za vsak $\mathbf{x}_k, \mathbf{x}_l \in \mathbb{R}^n$

$$D_{\check{c}}(\mathbf{x}_k, \mathbf{x}_l) \leq D_E(\mathbf{x}_k, \mathbf{x}_l) \leq D_{CB}(\mathbf{x}_k, \mathbf{x}_l)$$

Odnos med razdaljami ponazarja spodnji primer



Pri izbiri značilk se najprej odločimo za ustrezno mero ločljivosti razredov vzorcev, nato še za ustrezen postopek iskanja najboljše podmnožice n meritev. Med postopki iskanja sta najbolj preprosta postopka zaporednega iskanja.

3.1.1 Zaporedno iskanje »nazaj«

Pri tem postopku najprej predpostavimo, da so vse dane meritve značilne, nato pa v vsaki ponovitvi postopka izločimo tisto značilko (meritev), brez katere je kriterijska funkcija največja.

Postopek:

1. korak: Množica značilk \mathcal{X} je enaka množici danih meritev \mathcal{Y} , torej $\mathcal{X} = \mathcal{Y}$. Izračunamo vrednost kriterijske funkcije $J(\mathcal{X})$.
2. korak: V k -ti ponovitvi postopka imamo v podmnožici $\mathcal{X}(k)$ še $(r - k)$ značilk. Le-te uredimo tako, da je:

$$J(\mathcal{X}(k) - \{x_1\}) \geq J(\mathcal{X}(k) - \{x_2\}) \geq \dots \geq J(\mathcal{X}(k) - \{x_{r-k}\}).$$

3. korak: Značilko x_1 , brez katere podmnožica $\mathcal{X}(k)$ daje največjo vrednost kriterijske funkcije, odvezamo iz $\mathcal{X}(k)$ ter povečamo k za ena. Torej:

$$\mathcal{X}(k+1) = \mathcal{X}(k) - \{x_1\} \quad \text{in} \quad k = k + 1$$

4. korak: Če je $k = r - n$ končamo, sicer ponovno preidemo na korak 2.

3.1.2 Zaporedno iskanje »naprej«

To je preprost postopek iskanja, kjer na vsakem koraku postopka dodamo trenutni podmnožici značilk tisto meritev, ki najbolj poveča vrednost kriterijske funkcije.

Postopek:

1. korak: Množica značilk \mathcal{X} je prazna, množici \mathcal{Y} pa je vseh r spremenljivk (meritev).
2. korak: V k -ti ponovitvi postopka imamo v podmnožici $\mathcal{X}(k)$ še k značilk, v množici $\{\mathcal{Y} - \mathcal{X}(k)\}$ pa $r - k$ spremenljivk (meritev). Le-te uredimo tako, da je:

$$J(\mathcal{X}(k) + \{x_1\}) \geq J(\mathcal{X}(k) + \{x_2\}) \geq \dots \geq J(\mathcal{X}(k) + \{x_{r-k}\}).$$

3. korak: Značilko x_1 , ki skupaj s podmnožico $\mathcal{X}(k)$ daje največjo vrednost kriterijske funkcije, odvezamo iz $\{\mathcal{Y} - \mathcal{X}(k)\}$ in dodamo v podmnožico $\mathcal{X}(k)$ ter povečamo k za ena. Torej:

$$\mathcal{X}(k+1) = \mathcal{X}(k) + \{x_1\} \quad \text{in} \quad k = k + 1$$

4. korak: Če je $k = r - n$ končamo, sicer ponovno preidemo na korak 2.

3.2 Vprašanja in naloge

3.2.1 Vprašanje

V čem se bistveno razlikujeta postopka luščenja značilk z zaporednim iskanjem »naprej« in zaporednim iskanjem »nazaj«.

3.2.2 Naloga

Objekte razpoznavanja opisujejo vrednosti treh osnovnih spremenljivk (meritev). Podani so štirje vzorci, ki so razvrščeni v dva razreda

$$\mathcal{U}_2 = \{U_1, U_2\}$$

$$U_1 = \{(2,2,3)^T, (3,3,3)^T\} \quad U_2 = \{(0,3,1)^T, (1,2,0)^T\}$$

Vzemimo, da želimo z luščenjem značilik iz podanih treh ($r = 3$) osnovnih meritev izbrati dve ($n = 2$) najbolj značilni in to s postopkom izbire značilik z zaporednim iskanjem »naprej«, pri čemer uporabljamo kriterijsko funkcijo, ki je določena kot povprečna razdalja med vzorci različnih razredov, za mero razdalje med posameznimi vzorci pa uporabljamo razdaljo »City Block«. A priori verjetnosti obeh razredov ocenimo kar iz števila in porazdelitve podanih vzorcev med razredoma, torej

$$M = 2, \quad N_1 = 2, \quad N_2 = 2, \quad N = 4$$

$$P(C_1) = \frac{N_1}{N} = \frac{2}{4} = \frac{1}{2} \quad P(C_2) = \frac{N_2}{N} = \frac{2}{4} = \frac{1}{2}$$

V prvem koraku je podmnožica značilik prazna, torej $\mathcal{X}(1) = \{\}$. Nato izračunamo vrednosti kriterijske funkcije za vsako spremenljivko iz \mathcal{U} , ki bi jo lahko preselili v $\mathcal{X}(1)$, torej

$$\begin{aligned} J(\{y_1\}) &= \frac{1}{2} \sum_{i=0}^2 \frac{1}{2} \sum_{\substack{j=1 \\ j \neq i}}^2 \frac{1}{2} \frac{1}{2 \cdot 2} \sum_{k=1}^2 \sum_{l=1}^2 D(\mathbf{x}_{ik}, \mathbf{x}_{jl}) \\ &= \frac{1}{32} (2 + 1 + 3 + 2 + 2 + 3 + 1 + 2) = \frac{16}{32} = 0,5 \end{aligned}$$

$$J(\{y_2\}) = \frac{1}{32} (1 + 0 + 0 + 1 + 1 + 0 + 0 + 1) = \frac{4}{32} = 0,125$$

$$J(\{y_3\}) = \frac{1}{32} (2 + 3 + 2 + 3 + 3 + 2 + 3 + 2) = \frac{20}{32} = 0,625$$

Vidimo, da je

$$J(\{y_3\}) \geq J(\{y_1\}) \geq J(\{y_2\})$$

torej, preselimo tretjo spremenljivko y_3 iz \mathcal{Y} v $\mathcal{X}(2)$, oziroma, $x_1 = y_3$ in $\mathcal{X}(2) = \{x_1\} = \{y_3\}$.

Ker je moramo izbrati dve spremenljivki se vrnemo v drugi korak postopka, pri čemer upoštevamo, da je $\{\mathcal{Y} - \mathcal{X}(2)\} = \{y_1, y_2\}$. Zdaj izračunajmo vrednosti kriterijske funkcije za vsako spremenljivko iz $\{\mathcal{Y} - \mathcal{X}(2)\}$, ki bi jo lahko preselili v $\mathcal{X}(2)$, torej

$$J(\{y_3, y_1\}) = \frac{1}{32}(4 + 4 + 5 + 5 + 4 + 5 + 4 + 5) = \frac{36}{32} = 1,125$$

$$J(\{y_3, y_2\}) = \frac{1}{32}(3 + 3 + 2 + 4 + 3 + 2 + 3 + 4) = \frac{24}{32} = 0,75$$

Vidimo, da je

$$J(\{y_3, y_1\}) \geq J(\{y_3, y_2\})$$

torej, preselimo prvo spremenljivko y_1 iz $\{\mathcal{Y} - \mathcal{X}(2)\}$ v $\mathcal{X}(3)$, oziroma, $x_2 = y_1$ in $\mathcal{X}(3) = \{x_1, x_2\} = \{y_3, y_1\}$.

S tem smo dve spremenljivki $\{y_3, y_1\}$ izbrali kot v dve najbolj značilni, oziroma kot značilki $\{x_1, x_2\}$. S tem smo izvedeli preslikavo $y_3 \rightarrow x_1$ in $y_1 \rightarrow x_2$.

4 Analiza področja uporabe v vzorčnem prostoru s postopki iskanja rojev

Razpoznavanje objektov je možno šele potem, ko področje uporabe ustrezno spoznamo. Spoznavanje področja uporabe iz vzorcev objektov razpoznavanja temelji na nenadzorovanem razvrščanju (rojenju, samoorganizaciji) vzorcev iz končne množice vzorcev objektov razpoznavanja \mathcal{S}_N ter na označevanju rojev vzorcev z oznakami razredov vzorcev. Zato si spoznavanje področja uporabe lahko predstavljamo kot preslikavo znotraj prostora vzorcev, ki preslika množico vzorcev objektov razpoznavanja \mathcal{S}_N v učno množico vzorcev \mathcal{U}_M . Ta preslikava ni v celoti samodejna, ker označevanje vzorcev iz množice \mathcal{S}_N opravi lahko le človek — "učitelj" oziroma strokovnjak za določeno področje uporabe.

Razbitje množice vzorcev \mathcal{S}_N v $N_C \leq N$ nepraznih podmnožic je množica

$$\mathcal{D}_{N_C} = \{S_1, \dots, S_j, \dots, S_{N_C}\},$$

kjer je unija vseh nepraznih podmnožic S_j enaka množici \mathcal{S}_N in presek poljubnega para podmnožic $S_j \cap S_k$ prazna množica, torej

$$\bigcup_{j=1}^{N_C} S_j = \mathcal{S}_N, \quad S_j \cap S_k = \emptyset; \forall j, k = 1, \dots, N_C \wedge j \neq k.$$

Na splošno se množice \mathcal{S}_N v N_C podmnožice določi s preslikavo $\mathcal{R}: \mathbf{x}_i \xrightarrow{\mu_{ji}} S_j$, ki vsak vzorec \mathbf{x}_i (bolj splošno $f_i(\mathbf{x})$) preslika (razvrsti) v podmnožice z neko stopnjo pripadnosti $\mu_{ji} \in [0,1]$, kjer velja, da vsaki podmnožici pripada vsaj en vzorec z stopnjo večjo od nič in je seštevek vseh pripadnosti za dani vzorec enak ena, torej

$$\sum_{i=1}^N \mu_{ji} > 0, \quad \sum_{j=1}^{N_C} \mu_{ji} = 1.$$

V primeru, ko stopnje pripadnosti lahko zavzemajo poljubno vrednost iz intervala $[0,1]$ takšnemu razbitju pravimo neizrazito (angl. fuzzy). V primeru, ko pa pripadnosti lahko zavzemajo lahko zavzamejo le vrednost 0 ali 1, pa takšnemu razbitju pravimo izrazito (angl. crisp).

Za razbitje potrebujemo neko mero podobnosti med vzorci. Predlagana je cela paleta takšnih mer, ki so odvisne od zapisa in »narave« vzorcev. Namesto mer

podobnosti pogosto uporabljamo tudi mere različnosti, ki jih pod določenimi pogoji lahko obravnavamo kot metrike (funkcije razdalje). V primeru, ko so vzorci predstavljeni z množico vrednosti značilik (vektorji značilik), za mere različnosti najbolj pogosto uporabljajo kar razdalje Minkovskega, ki smo jih spoznali pri postopkih luščenja značilik.

4.1 Postopki iskanja rojev v množici vzorcev

S postopki iskanja rojev iščemo najboljše razbitje \mathcal{D}_{N_C} dane množice vzorcev \mathcal{S}_N glede na nek izbran kriterij, na osnovi katerega poskušamo določiti najboljše razbitje. Množico \mathcal{S}_N lahko delimo v N_C nepraznih podmnožic na

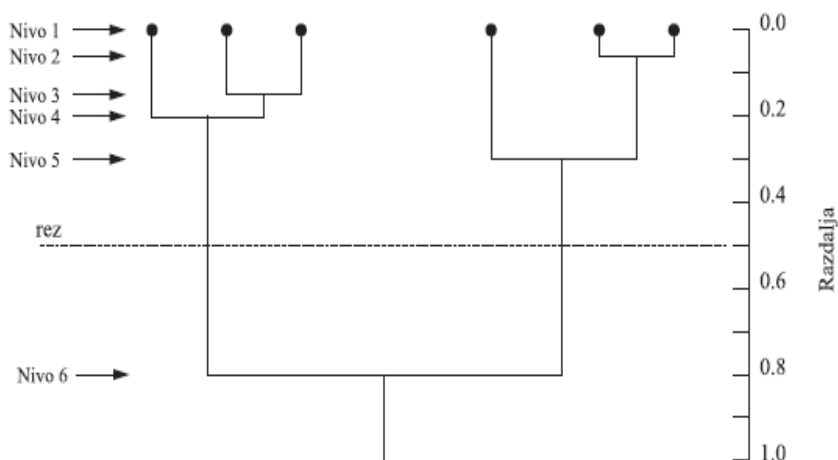
$$\frac{1}{N_C} \sum_{j=1}^{N_C} \binom{N_C}{j} (-1)^{N_C-j} j^{N_C}$$

načinov. Na primer, 100 vzorcev iz množice \mathcal{S}_{100} lahko razdelimo v 3 roje na približno 10^{47} načinov. V tako obsežni množici razbitij je skorajda nemogoče določiti najboljše razbitje, zato je predlaganih vrsta različni postopkov, ki takšno razbitje izvedejo bolj ali manj optimalno. Postopke iskanja rojev delimo na postopke, ki temelje na *teoriji grafov*, na *hierarhične postopke*, na *delitvene postopke* ter na postopke iskanja rojev z *nevronskimi omrežji*.

4.2 Hierarhični postopek iskanja rojev

S hierarhičnimi postopki iskanja rojev ne dobimo le enega razbitja \mathcal{D}_{N_C} množice \mathcal{S}_N , temveč zaporedje razbitij za vsak $N_C = N, N - 1, \dots, 1$. Prvi člen zaporedja, ki je najnižji v hierarhiji, predstavlja razbitje množice \mathcal{S}_N v $N_C = N$ rojev vzorcev. V vsakem roju je torej natanko en vzorec iz množice \mathcal{S}_N . Zadnji oziroma N -ti člen zaporedja je najvišji v hierarhiji rojenj. Predstavlja rojenje vseh vzorcev iz \mathcal{S}_N v en sam roj. Na vsakem nivoju iskanja rojev, od najnižjega do najvišjega, združimo dva najbolj podobna roja v en sam večji roj.

Zaporedje razbitij: $\mathcal{D}_N, \mathcal{D}_{N-1}, \dots, \mathcal{D}_1$ dane množice vzorcev \mathcal{S}_N lahko ponazorimo z drevesom, ki ga imenujemo *dendrogram*. Primer dendrograma je ponazorjen na spodnji sliki



Hierarhično rojenje vzorcev lahko strnemo v naslednje zaporedje korakov:

1. korak: Izračunaj trikotniško matriko razdalj $D(S_k, S_l)$ med roji prvega nivoja iskanja rojev (to je dejansko matrika razdalj med danimi vzorci).
2. korak: Če je razdalja med rojema S_i in S_j najmanjša (največja podobnost), združimo roja v en roj. Dobimo roj $S_i + S_j$. Razdalje med novim rojem in vsemi ostalimi roji $D(S_i + S_j, S_k)$ nato izračunamo po obrazcu:

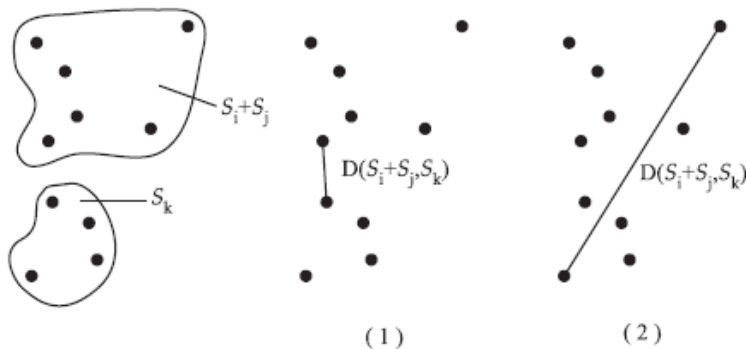
$$D(S_i + S_j, S_k) = a_i D(S_i, S_k) + a_j D(S_j, S_k) + b D(S_i, S_j) + c |D(S_i, S_k) - D(S_j, S_k)|,$$

kjer so a_i , a_j , b in c koeficienti, ki jih izbiramo na več načinov, odvisno od tega, kako merimo razdaljo med roji vzorcev. Med različnimi načini se posvetimo le dvema:

1. $a_i = a_j = 1/2$, $b = 0$ in $c = 1/2$. V tem primeru je razdalja $D(S_i + S_j, S_k)$ enaka razdalji med najbližjima vzorcema iz rojev $S_i + S_j$ in S_k .
2. $a_i = a_j = 1/2$, $b = 0$ in $c = -1/2$. V tem primeru je razdalja $D(S_i + S_j, S_k)$ enaka razdalji med najbolj

oddaljenima vzorcema iz rojev $S_i + S_j$ in S_k .

3. korak Če ima nova matrika razdalj več kot en stolpec, ponovimo korak 2. Če ne, postopek končamo.



Na gornji sliki sta ponazorjena oba obravnavana načina merjenja razdalje med roji vzorcev.

Pri različnih osnovnih merjenjih podobnosti med samimi vzorci in merjenji podobnosti med roji dobimo kot rezultat različno hierarhijo rojenja. Vprašanje je, katera hierarhija je boljša ali slabša. Predlaganih je nekaj mer prileganja hierarhije rojenja danim vzorcem, med njimi sta meri Q in $CPCC$, ki temeljita na t.i. kofenetični matriki \mathbf{D}_C , ki je določena kot matrika razsežnosti $N \times N$, kjer je posamezni element $D_C(i, j)$ razdalja rojev (višina dendrograma), na kateri se vzorca \mathbf{x}_i in \mathbf{x}_j prvič pojavita v istem roju. Zaradi zahtevnosti numeričnega izračuna obe meri navadno izračunavamo s pomočjo računalniškega programa.

Iz dendrograma lahko razberemo tudi odgovor na vprašanje: koliko je rojev v dani množici vzorcev? V ta namen prekinemo proces združevanja rojev na tistem nivoju podobnosti rojev, pri katerem je podobnost vzorcev zunaj rojev "precej manjša" od podobnosti vzorcev znotraj rojev, kar pomeni, da presekamo dendrogram na mestu, kjer ima najdaljše navpične črte. Najdaljše navpične črte v dendrogramu pomenijo namreč največjo razliko med podobnostjo, pri kateri se roji združijo v nov roj, in podobnostjo, pri kateri so nastali. Rez najdaljših navpičnih črt v dendrogramu odkrije "naravne" roje vzorcev oziroma naravno razbitje dane množice vzorcev v roje. Rez torej lahko določimo tudi z vizualnim pregledom podanega dendrograma.

Število rojev v dani množici vzorcev lahko računsko določimo tudi iz ocen dvovrstnega korelacijskega koeficienta PB . Rojenje, pri katerem je vrednost dvovrstnega korelacijskega koeficienta največja, štejemo za razbitje na "naravne" roje vzorcev. Zaradi zahtevnosti numeričnega izračuna tudi ta koeficient navadno izračunavamo s pomočjo računalniškega programa.

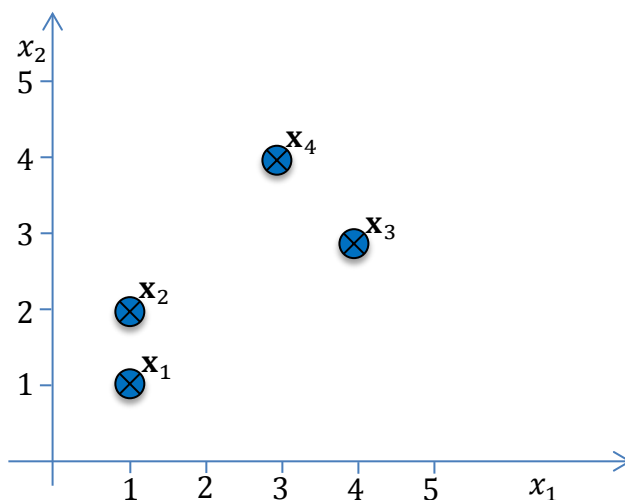
4.2.1 Naloga

Izvedite hierarhični postopek rojenja vzorcev za podano množico vzorcev

$$S_4 = \{(1,1)^T, (1,2)^T, (4,3)^T, (3,4)^T\},$$

pri čemer razdaljo med roji vzorcev računajte kot razdaljo med njunima najbližjima vzorcema. Za osnovno mero podobnosti med vzorci izberite Evklidovo razdaljo. Določite še kofenetično matriko dobljene hierarhije rojenja.

Podano množico vzorcev lahko ponazorimo v \mathbb{R}^2 , kar je razvidno iz spodnje



slike.

Razdalje med roji vzorcev lahko nato ocenjujemo kar vizualno.

Najprej določimo trikotno matriko vseh razdalj med vzorci. Denimo, razdalja med $D(S_1, S_4)$ je tako določena kar kot Evklidova razdalja $D(\mathbf{x}_1, \mathbf{x}_4) = \sqrt{(1-3)^2 + (1-4)^2} = \sqrt{4+9} = \sqrt{13} \approx 3.16$. Podobno izračunamo še razdalje za preostale kombinacije vzorcev.

Analiza področja uporabe v vzorčnem prostoru s postopki iskanja rojev

$S_1 = \{\mathbf{x}_1\}$	$S_2 = \{\mathbf{x}_2\}$	$S_3 = \{\mathbf{x}_3\}$	$S_4 = \{\mathbf{x}_4\}$	$D(S_k, S_l)$	
0	<u>1.0</u>	3.61	3.61		$S_1 = \{\mathbf{x}_1\}$
	0	3.16	2.83		$S_2 = \{\mathbf{x}_2\}$
		0	1.41		$S_3 = \{\mathbf{x}_3\}$
			0		$S_4 = \{\mathbf{x}_4\}$

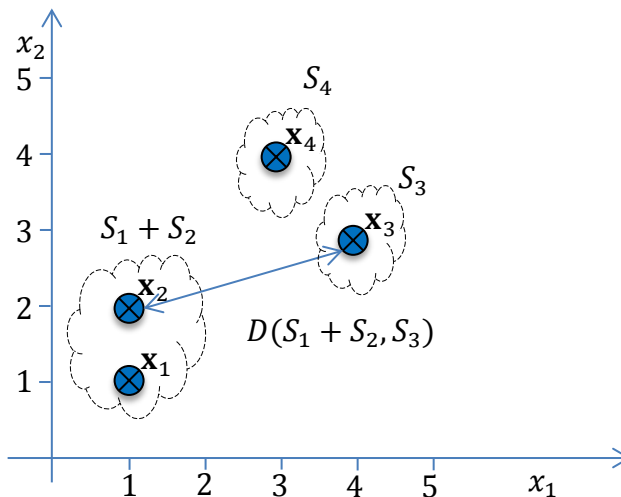
Vidimo, da ima razdalja $D(S_i, S_j) = D(S_1, S_2) = 1.0$ najmanjšo vrednost, zato združimo ta dva roja v enega in izračunamo razdalje preostalih rojev S_k (pri $k \neq i, j$) do združenega roja $S_i + S_j$ po obrazcu

$$D(S_i + S_j, S_k) = \frac{1}{2}D(S_i, S_k) + \frac{1}{2}D(S_j, S_k) - \frac{1}{2}|D(S_i, S_k) - D(S_j, S_k)|$$

Po izračunih razdalj $D(S_1 + S_2, S_3)$ in $D(S_1 + S_2, S_4)$ dobimo zmanjšano matriko

$S_1 = \{\mathbf{x}_1 + \mathbf{x}_2\}$	$S_2 = \{\mathbf{x}_3\}$	$S_3 = \{\mathbf{x}_4\}$	$D(S_i, S_j)$
0	3.16	2.83	$S_1 = \{\mathbf{x}_1 + \mathbf{x}_2\}$
	0	<u>1.41</u>	$S_2 = \{\mathbf{x}_3\}$
		0	$S_3 = \{\mathbf{x}_4\}$

Izračun razdalje $D(S_1 + S_2, S_3)$, ki je v bistvu kar enaka razdalji $D(\mathbf{x}_2, \mathbf{x}_3)$

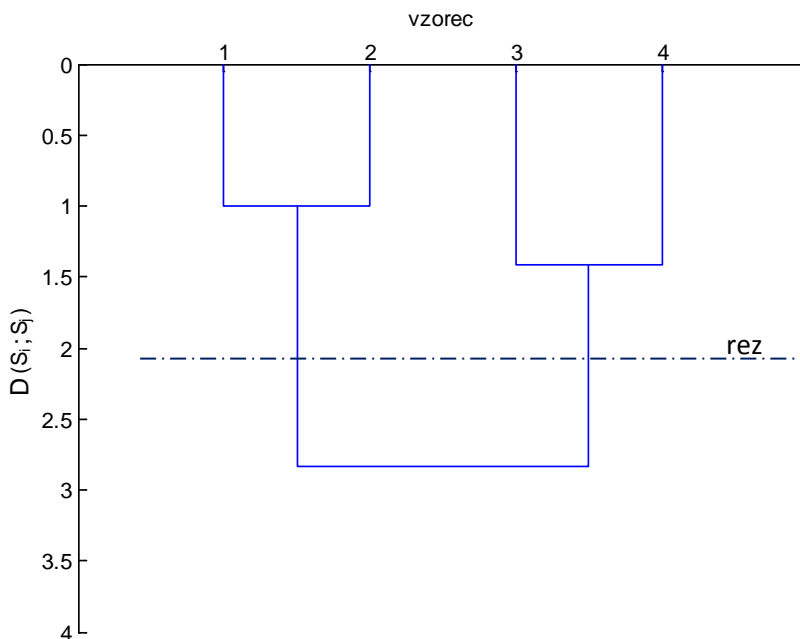


ponazarja spodnja slika

V zmanjšani matriki nato ponovno poiščemo minimum, ki je pri $D(S_2, S_3) = 1.41$, zato tokrat združimo roja S_2, S_3 in izvedemo izračun razdalje $D(S_3 + S_4, S_1) = D(S_1, S_3 + S_4)$. Dobimo matriko

$$\begin{array}{ccc}
 S_1 = \{\mathbf{x}_1 + \mathbf{x}_2\} & S_2 = \{\mathbf{x}_3 + \mathbf{x}_4\} & D(S_i, S_j) \\
 0 & 2.83 & S_1 = \{\mathbf{x}_1 + \mathbf{x}_2\} \\
 & 0 & S_2 = \{\mathbf{x}_3 + \mathbf{x}_4\}
 \end{array}$$

Izvedemo še zadnjo združitev, s katerim so vsi vzorci v zadnjem enem roju. Na osnovi podanih izračunov skiciramo spodnji dendrogram rojenja



Iz dendrograma je razvidno, da je naravno število rojev dve, ker so najdaljše črte pri zadnji združitvi dveh rojev v enega samega. Kofenetična matrika dobljene hierarhije pa bi bila

$$\mathbf{D}_c = \begin{bmatrix} 0 & 1.0 & 2.83 & 2.83 \\ 1.0 & 0 & 2.83 & 2.83 \\ 2.83 & 2.83 & 0 & 1.41 \\ 2.83 & 2.83 & 1.41 & 0 \end{bmatrix}$$

Začetna matrika razdalj med vsemi možnimi pari vzorcev, ki smo jo izračunali na začetku postopka, označimo tudi kot

$$\mathbf{D} = \begin{bmatrix} 0 & 1.0 & 3.61 & 3.61 \\ 1.0 & 0 & 3.16 & 2.83 \\ 3.61 & 3.16 & 0 & 1.41 \\ 3.61 & 2.83 & 1.41 & 0 \end{bmatrix}$$

Iz kofenetične matrike \mathbf{D}_C in začetne matrike razdalj \mathbf{D} med vsemi vzorci bi lahko po potrebi izračunali meri Q in $CPCC$. Vrednost obeh mer nam bi podala informacijo o ujemanju pridobljene hierarhije dani učni množici (glede na uporabljen način osnovnega merjenja različnosti med vzorci in med roji).

Z ugotavljanjem števila in porazdelitve vzorcev po posameznih rojih pa bi lahko za vsako razbitje na roje izračunali tudi dvovrstni korelacijski koeficient PB . Največja vrednost tega koeficienta bi v našem primeru tako morala biti pri dveh rojih.

5 Razvrščanje vzorcev s prileganjem

Postopki razvrščanja, ki jih uvrščamo v to skupino postopkov in jih obravnavajo naloge v nadaljevanju, temelje na merjenju podobnosti med vzorcem (v celoti), ki ga razvrščamo, in vzorci iz učne množice vzorcev ter na razvrstitvi vzorca v razred vzorcev C_j , če so mu najbolj podobni učni vzorci iz razreda $U_j \subset \mathcal{U}_M$.

Razviti so bili številni postopki razvrščanja vzorcev s prileganjem. Delimo jih glede na to, katere vzorce razvrščevalnik hrani v pomnilniku (celotno učno množico ali samo značilne predstavnike), in glede na to, koliko najbolj podobnih vzorcev upoštevamo pri razvrščanju vzorcev.

5.1 Pravilo razvrščanja “najbližji sosed” (1-NN)

Dana je učna množica vzorcev $\mathcal{U}_M = \{U_1, \dots, U_M\}$, kjer je

$$U_j = \{(\mathbf{x}_{j1}, \omega_j), (\mathbf{x}_{j2}, \omega_j), \dots, (\mathbf{x}_{jN_j}, \omega_j)\}.$$

Vzorec, ki ga razvrščamo, \mathbf{x} , razvrstimo v razred C_j , če mu je med vsemi vzorci v učni množici najbolj podoben vzorec $\mathbf{x}_{jk} \in U_j$, to je, če velja

$$P(\mathbf{x}, (\mathbf{x}_{jk}, \omega_j)) = \max_{\mathbf{x}_i \in \mathcal{U}_M} \{P(\mathbf{x}, \mathbf{x}_i)\},$$

oziroma

$$D(\mathbf{x}, (\mathbf{x}_{jk}, \omega_j)) = \min_{\mathbf{x}_i \in \mathcal{U}_M} \{D(\mathbf{x}, \mathbf{x}_i)\},$$

kjer sta P in D meri podobnosti oziroma različnosti med vzorci. Z drugimi besedami, vzorec, ki ga razvrščamo, razvrstimo v tisti razred, v katerega je razvrščen vzorec učne množice, ki mu je najbolj podoben oziroma najmanj različen. Za mere različnosti med vzorci najbolj pogosto uporabljamo kar razdalje Minkovskega.

Pravilo razvrščanja, ki upošteva samo najbolj podoben vzorec iz učne množice, ne izkorišča dovolj informacije, ki jo vsebuje učna množica vzorcev. Zato je logična razširitev pravila 1-NN na pravilo, ki upošteva oznake k najbolj podobnih vzorcev. Pri tem je k liho (ali še bolje pra) število, večje od 1.

Pravila razvrščanja k -najbližjih sosedov temelje na določitvi k vzorcev iz učne množice vzorcev \mathcal{U}_M , ki so najbolj podobni (najmanj različni) danemu vzorcu \mathbf{x} .

Le-tega uvrstimo v razred C_j , če ima večina vzorcev med k najbolj podobnimi vzorci oznako ω_j . Za opisano pravilo uporabljamo kratico k -NN

Pravilo razvrščanja k -najbližjih sosedov lahko razširimo na

- pravilo (k, l) -NN, če postavimo, da mora doseči v množici k najbolj podobnih vzorcev večina oznak vnaprej predpisano število l . V nasprotnem ga pustimo nerazvrščenega ali pa ga poskusimo razvrstiti s pomočjo kakšnega drugega pravila razvrščanja.
- pravilo (k_j, l) -NN, kjer dopustimo, da je potrebna večina za razvrščanje vzorcev določena za vsak razred posebej. To različico uporabimo takrat, ko so a priori verjetnosti razredov medsebojno različne.
- pravilo Fk -NN, ki priredi vsakemu vzorcu k ga razvrščamo, vektor stopnje pripadnosti vzorca posameznemu razredu vzorcev. Nato vzorec razvrstimo v razred z največjo stopnjo pripadnosti.

5.1.1 Naloga

Dana je učna množic vzorcev dveh razredov $\mathcal{U}_2 = \{U_1, U_2\}$, kjer je

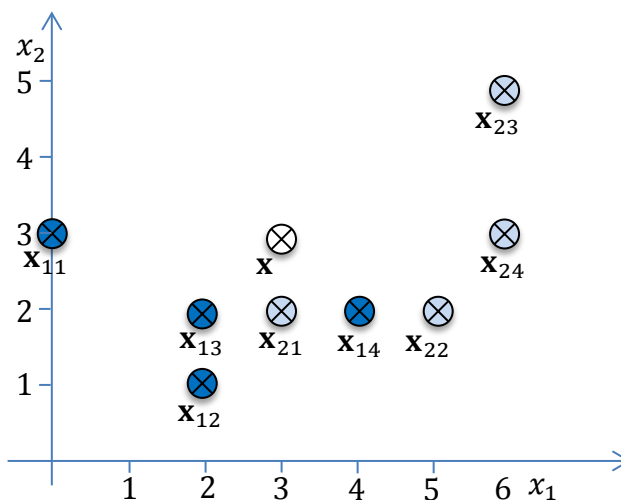
$$U_1 = \{((0,3)^T, \omega_1), ((2,1)^T, \omega_1), ((2,2)^T, \omega_1), ((4,2)^T, \omega_1)\},$$

$$U_2 = \{((3,2)^T, \omega_2), ((5,2)^T, \omega_2), ((6,5)^T, \omega_2), ((6,3)^T, \omega_2)\}.$$

Razvrstite vzorec $\mathbf{x} = (3,3)^T$ s prileganjem najprej s pravilom 1-NN in nato še s pravilom (k, l) -NN, pri $k = 3$ in $l = 2$, pri čemer za mero različnosti med vzorci uporabite Evklidovo razdaljo D_E .

Rešitev

Za lažji izračun najprej ponazorimo vzorce kot točke v dvorazsežnem prostoru značilk.



Evklidove razdalje danega vzorca \mathbf{x} do vsakega od vzorcev v podani učni množici so

$$D_E(\mathbf{x}, \mathbf{x}_{11}) = \sqrt{(3-0)^2 + (3-3)^2} = \sqrt{3^2 + 0^2} = 3$$

$$D_E(\mathbf{x}, \mathbf{x}_{12}) = \sqrt{(3-2)^2 + (3-1)^2} = \sqrt{1^2 + 2^2} = \sqrt{5}$$

$$D_E(\mathbf{x}, \mathbf{x}_{13}) = \sqrt{(3-2)^2 + (3-2)^2} = \sqrt{1^2 + 1^2} = \sqrt{2}$$

$$D_E(\mathbf{x}, \mathbf{x}_{14}) = \sqrt{(3-4)^2 + (3-2)^2} = \sqrt{1^2 + 1^2} = \sqrt{2}$$

$$D_E(\mathbf{x}, \mathbf{x}_{21}) = \sqrt{(3-3)^2 + (3-2)^2} = \sqrt{0^2 + 1^2} = 1$$

$$D_E(\mathbf{x}, \mathbf{x}_{22}) = \sqrt{(3-5)^2 + (3-2)^2} = \sqrt{2^2 + 1^2} = \sqrt{5}$$

$$D_E(\mathbf{x}, \mathbf{x}_{23}) = \sqrt{(3-6)^2 + (3-5)^2} = \sqrt{3^2 + 2^2} = \sqrt{13}$$

$$D_E(\mathbf{x}, \mathbf{x}_{24}) = \sqrt{(3-6)^2 + (3-3)^2} = \sqrt{3^2 + 0^2} = 3$$

Iz gornjih izračunov vidimo, da je vzorcu \mathbf{x} najmanj različen vzorec \mathbf{x}_{21} in s pravilom 1-NN torej vzorec \mathbf{x} razvrstimo v razred C_2 .

Pri razvrščanju s pravilom (k, l) -NN, pri $k = 3$ in $l = 2$ pa ugotovimo, da med tremi vzorci \mathbf{x}_{21} , \mathbf{x}_{13} , in \mathbf{x}_{14} , ki so danemu vzorcu \mathbf{x} najmanj različni, vzorca \mathbf{x}_{13} in \mathbf{x}_{14} iz razreda C_1 predstavljata večino, torej s tem pravilom vzorec razvrstimo razred C_1 , kar je drugače kot pri pravilu 1-NN.

5.2 Učenje

Če je moč učne množice vzorcev velika, lahko s pravili razvrščanja vzorcev s prileganjem razvrščamo vzorce v razrede zelo zanesljivo. Istočasno pa velika moč učne množice vzorcev zelo upočasnjuje razvrščanje, ker pravila razvrščanja s prileganjem zahtevajo določitev podobnosti med vsakim še nerazvrščenim vzorcem in vsemi vzorci v učni množici.

Da bi zmanjšali računsko zahtevnost razvrščanja s pravili najbližjega soseda, poskušamo zmanjšati število vzorcev v \mathcal{U}_M s preslikavo v množico \mathcal{U}'_M , ki ima manjše število vzorcev. Preslikavi množice \mathcal{U}_M v množico \mathcal{U}'_M pravimo *učenje*.

5.3 Predstavitev razreda z značilnim vzorcem

Razvrščanje vzorcev po postopku 1-NN najbolj pospešimo, če zgotavimo informacijo, ki jo nosijo učni vzorci nekega razreda, v en sam vzorec, ki ga proglašimo za značilnega predstavnika razreda vzorcev. Le-ta je za zvezne značilke navadno vektor \mathbf{m}_j , določen kot vektor povprečnih vrednosti značilk vzorcev \mathbf{x}_{ji} iz razreda C_j , torej

$$\mathbf{m}_j = \frac{1}{N_j} \sum_i^{N_j} \mathbf{x}_{ji} \quad \text{za } \mathbf{x}_{ji} \in C_j,$$

kjer je N_j število učnih vzorcev iz razreda vzorcev C_j . V tem primeru imamo v učni množici \mathcal{U}'_M natanko M vzorcev.

5.3.1 Naloga

Preslikajte učno množico vzorcev \mathcal{U}_2 iz zadnje naloge v \mathcal{U}'_2 tako, da vsak razred vzorcev predstavite z značilnim predstavnikom, določen kot vektor povprečnih vrednosti značilk vzorcev svojega razreda. S prileganjem s pravilom 1-NN nato razvrstite vzorec $\mathbf{x} = (3,3)^T$.

Rešitev

Najprej izračunamo vektorja povprečnih vrednosti značilk vzorcev obeh razredov vzorcev.

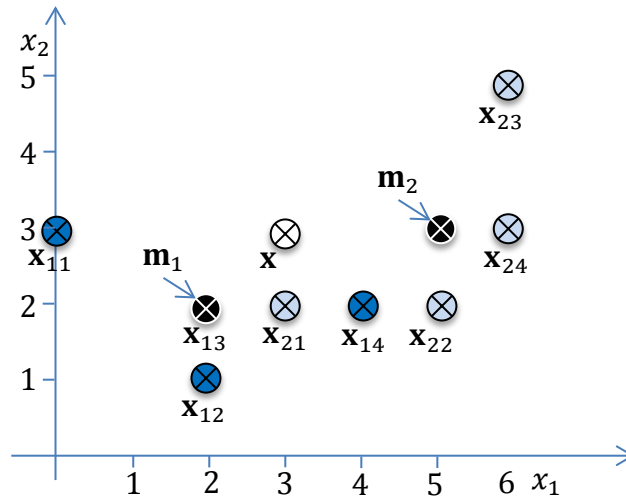
$$\begin{aligned} \mathbf{m}_1 &= \frac{1}{4} \sum_i^4 \mathbf{x}_{1i} = \frac{1}{4} [(0,3)^T + (2,1)^T + (2,2)^T + (4,2)^T] = \\ &= \frac{1}{4} (8,8)^T = (2, 2)^T \end{aligned}$$

$$\begin{aligned} \mathbf{m}_2 &= \frac{1}{4} \sum_i^4 \mathbf{x}_{2i} = \frac{1}{4} [(3,2)^T + (5,2)^T + (6,5)^T + (6,3)^T] = \\ &= \frac{1}{4} (20,12)^T = (5, 3)^T \end{aligned}$$

Po preslikavi tako dobimo $\mathcal{U}'_2 = \{U'_1, U'_2\}$, kjer je

$$U'_1 = \{((2,2)^T, \omega_1)\},$$

$$U'_2 = \{((5,2)^T, \omega_2)\}.$$



Povprečna vektorja sta ponazorjena na gornji sliki

Evklidove razdalje danega vzorca \mathbf{x} do vsakega od vzorcev v učni množici U'_2 so

$$D_E(\mathbf{x}, \mathbf{m}_1) = \sqrt{(3-2)^2 + (3-2)^2} = \sqrt{1^2 + 1^2} = \sqrt{2}$$

$$D_E(\mathbf{x}, \mathbf{m}_2) = \sqrt{(3-5)^2 + (3-3)^2} = \sqrt{2^2 + 0^2} = 2$$

Iz gornjih izračunov vidimo, da je vzorcu \mathbf{x} najmanj različen vzorec \mathbf{m}_1 in s pravilom 1-NN torej vzorec \mathbf{x} razvrstimo v razred C_1 .

5.4 Urejanje učne množice

Učno množico urejemo tako, da iz nje izločimo vse tiste vzorce iz enega razreda, ki se "mešajo" v prostoru značilik z vzorci iz drugega razreda. To so tisti vzorci učne množice, ki jih z Bayesovim odločitvenim pravilom (obravnavano v nadaljevanju) napačno razvrstimo.

Postopek

1. korak: Naključno razdeli dano učno množico vzorcev \mathcal{U}_M v K podmnožic $L_0, L_1, \dots, L_j, \dots, L_{K-1}$; $K \geq 3$.
2. korak: Razvrsti vzorce iz L_j s pravilom 1-NN tako, da uporabiš vzorce iz $L_{(j+1) \bmod (K+1)}$ kot učno množico; $j = 0, 1, 2, \dots, K - 1$.
3. korak: Izloči iz učne množice UM vse tiste vzorce, ki so bili v koraku 2 napačno razvrščeni.
4. korak: Če v zadnji ponovitvi postopka nisi izločil nobenega vzorca iz učne množice, končaj postopek, sicer pa se vrni na 1. korak postopka.

Konec postopka

5.5 Zgoščevanje učne množice

Z urejevanjem učne množice vzorcev \mathcal{U}_M dobimo v prostoru vzorcev strnjene roje vzorcev, ki ustrezajo razredom vzorcev v učni množici. Na pravilo razvrščanja 1-NN vplivajo le vzorci na robu rojev, zato lahko, ne da bi vplivali na zanesljivost razpoznavanja, izločimo iz učne množice vzorcev vse vzorce iz notranjosti rojev. S postopki zgoščanja lahko dobimo tako podmnožico učne množice UM, ki z manj učnih vzorcev zagotavlja približno enako zanesljivost razvrščanja.

Postopek

1. korak: Vpiši v seznam STORE prvi vzorec učne množice \mathcal{U}_M , vse ostale vzorce pa v seznam GRABBAG.
2. korak: S pravilom 1-NN in učno množico vzorcev iz seznama STORE razvrsti vzorce iz seznama GRABBAG. Vse pravilno razvrščene vzorce pusti v seznamu GRABBAG, napačno razvrščene pa vpiši v seznam STORE.
3. korak: Če v koraku 2 postopka noben vzorec ni vpisan v seznam STORE ali če je seznam GRABBAG prazen, postopek končaj. Sicer pojdi na korak 2.

Konec postopka

Podmnožica učnih vzorcev \mathcal{U}'_M je zapisana v seznamu STORE. Postopek ne zagotavlja minimalnosti podmnožic učnih vzorcev.

5.5.1 Naloga

Dana je učna množica vzorcev dveh razredov $\mathcal{U}_2 = \{U_1, U_2\}$ iz zadnje naloge, kjer množici U_1 in U_2 zapišemo krajše (brez oznak razredov)

$$U_1 = \{\mathbf{x}_{11}, \mathbf{x}_{12}, \mathbf{x}_{13}, \mathbf{x}_{14}\} = \{(0,3)^T, (2,2)^T, (2,1)^T, (4,2)^T\},$$

$$U_2 = \{\mathbf{x}_{21}, \mathbf{x}_{22}, \mathbf{x}_{23}, \mathbf{x}_{24}\} = \{(3,2)^T, (5,3)^T, (6,5)^T, (6,3)^T\},$$

Denimo, da smo za razvrščanje vzorcev izbrali pravilo 1-NN in Evklidovo razdaljo za mero različnosti med vzorci. Vzorce grafično ponazorite v \mathbb{R}^2 . Evklidove razdalje med vzorci potem ocenjujete kar grafično.

Grafično ponazorite izvedbo postopka urejanja učne množice, pri čemer predpostavite, da smo v prvem koraku postopka vse vzorce naključno razdelili v $K = 2$ podmnožici tako, da sta:

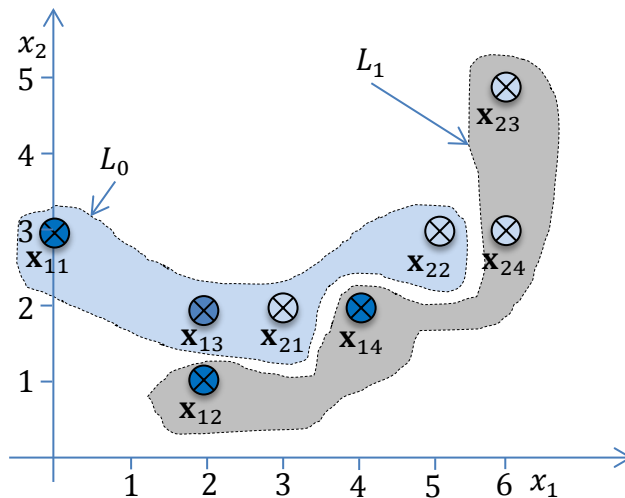
$$L_0 = \{\mathbf{x}_{11}, \mathbf{x}_{13}, \mathbf{x}_{21}, \mathbf{x}_{22}\},$$

$$L_1 = \{\mathbf{x}_{12}, \mathbf{x}_{14}, \mathbf{x}_{23}, \mathbf{x}_{24}\},$$

Dobljeno urejeno učno množico nato še zgostite s postopkom zgoščanja učne množice, pri čemer v prvem koraku postopka v seznam STORE preselite prvi vzorec iz urejene učne množice.

Rešitev

Najprej grafično ponazorimo podane vzorce v \mathcal{U}_2 v ravnini \mathbb{R}^2 .



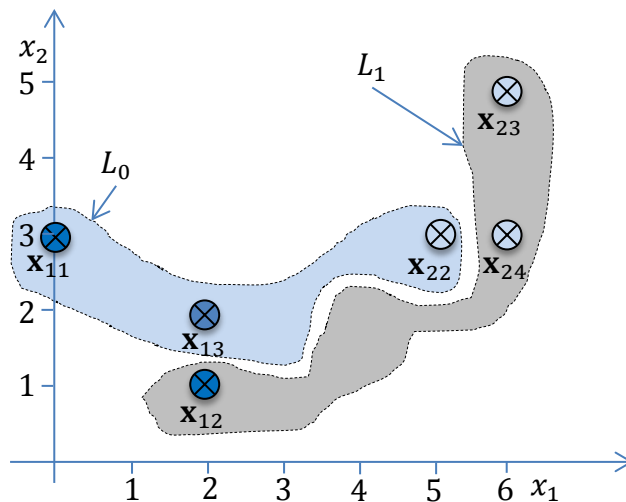
Pri razvrščanju vzorcev iz množice L_0 s pravilom 1-NN tako, da uporabimo vzorce iz L_1 kot učno množico (pri tem si pomagamo z gornjo grafično ponazoritvijo), bi se iz množice L_0 napačno razvrstil vzorec \mathbf{x}_{21} , ki je najmanj različen (najmanjša Evklidova razdalja) od vzorca \mathbf{x}_{41} iz L_1 , ki ni iz istega razreda, zato vzorec \mathbf{x}_{21} izločimo iz množice L_0 in s tem iz učne množice. Preostali vzorci se pravilno razvrstijo, zato jih ne odstranimo iz učne množice.

Podobno bi se pri razvrščanju vzorcev iz množice L_1 tako, da uporabimo vzorce iz L_0 kot učno množico, bi se napačno razvrstil vzorec \mathbf{x}_{14} , ki je najmanj različen (najmanjša Evklidova razdalja) od vzorca \mathbf{x}_{22} iz L_0 , ki ni iz istega razreda, zato tudi vzorec \mathbf{x}_{14} izločimo iz množice L_1 in s tem iz učne množice. Stanje po opisanih dveh korakih in izločenih dveh vzorcih ponazarja gornja slika.

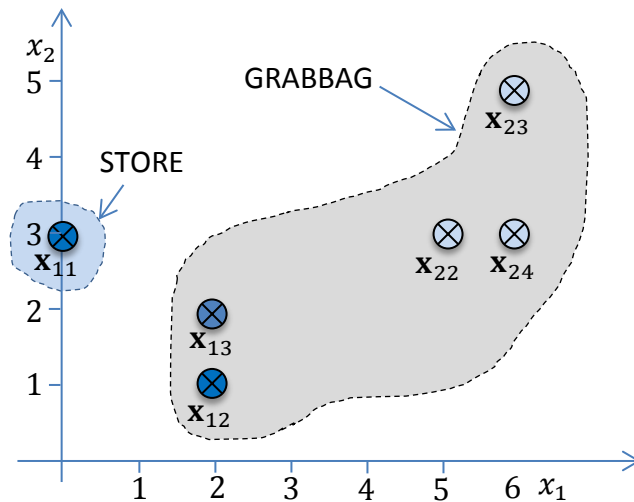
Pri ponovitvi obeh korakov ugotovimo, da se noben od obeh vzorcev iz L_0 ne razvrsti napačno in enako velja tudi za vzorca iz L_1 , zato iz obeh množic ne bi izločili nobenega vzorca. To pomeni, da se postopek urejanja zaključi in izhodiščna učna množica $\mathcal{U}_2 = \{U_1, U_2\}$ se preslika v urejeno učno množico $\mathcal{U}'_2 = \{U'_1, U'_2\}$, kjer je

$$U'_1 = \{\mathbf{x}_{11}, \mathbf{x}_{12}, \mathbf{x}_{13}\} = \{(0,3)^T, (2,1)^T, (2,2)^T\},$$

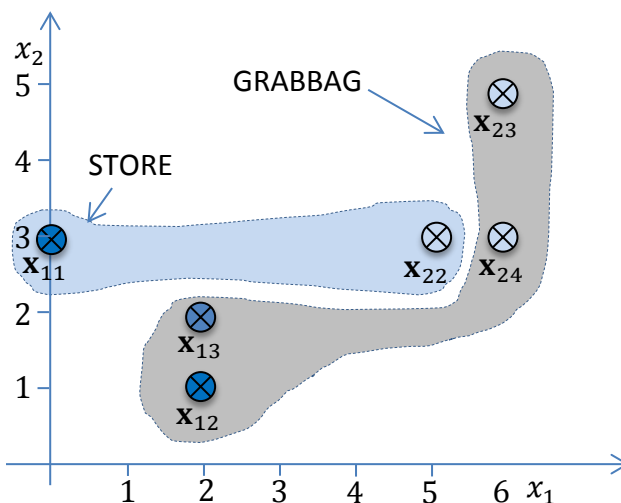
$$U'_2 = \{\mathbf{x}_{22}, \mathbf{x}_{23}, \mathbf{x}_{24}\} = \{(5,3)^T, (6,5)^T, (6,3)^T\}.$$



Po urejanju izvedemo še postopke zgoščanje, pri čemer v prvem koraku v seznam STORE vpišemo vzorec x_{11} , preostale vzorce pa v seznam GRABBAG. Stanje obeh seznamov grafično ponazarja spodnja slika.



Nato s pravilom 1-NN in učno množico vzorcev iz seznama STORE razvrščamo vzorce iz seznama GRABBAG in vse napačno razvrščene sproti vpisujemo v seznam STORE. Iz grafične ponazoritve vidimo, da bi bil prvi napačno razvrščeni vzorec vzorec x_{22} , ki bi ga preselili v seznam STORE. Pri nadaljnjem razvrščanju vzorcev iz seznama GRABBAG pa bi bili vsi vzorci pravilno razvrščeni (tudi pri ponovitvi tega koraka postopka) in nobenega vzorca ne bi več preselili iz seznama GRABBAG v seznam STORE. Na spodnji sliki je grafično ponazorjeno končno stanje obeh seznamov.



Vzorci v seznamu STORE predstavljajo končno zgoščeno učno množico $\mathcal{U}_2'' = \{U_1'', U_2''\}$, kjer je

$$U_1'' = \{\mathbf{x}_{11}\} = \{(0,3)^T\},$$

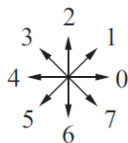
$$U_2'' = \{\mathbf{x}_{22}\} = \{(5,3)^T\}.$$

Od vseh vzorcev izhodiščne učne množice sta torej izbrana le dva značilna predstavnika (za vsak razred po en značilni predstavnik).

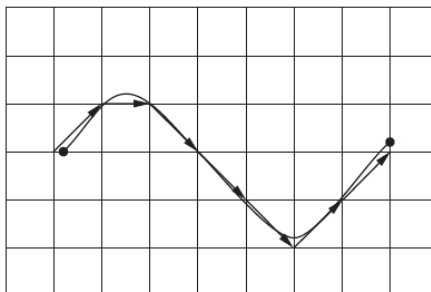
5.6 Računanje podobnosti dveh nizov znakov

V določenih primerih vzorce zapišemo kot urejene sestave znakov, ki označujejo osnovne sestavne dele vzorca. Ti osnovni sestavni deli vzorca so po definiciji takšni, da lahko za njih pravimo, da niso sestavljeni iz še preprostejših delov. Z vidika urejenosti sestave osnovnih delov vzorca je najbolj preprosta ureditev niz, ko osnovni sestavni deli gradijo vzorec tako, da se zaporedno navezujejo drug na drugega.

Primer takšnih vzorcev so nizi oznak smernih vektorjev, ki opisujejo krivuljo ali obris v ravnini. Če so smernim vektorjem priredimo oznake iz množice $\{0,1,\dots,7\}$ tako, kot je ponazorjeno spodaj



potem niz $x = 1077711$ opisuje spodnjo krivuljo v ravnini



V izrazoslovju, ki ga poznamo iz teorije formalnih jezikov, pravimo oznakam osnovnih delov vzorca tudi *znaki* ali *črke*, zaporedjem imen osnovnih delov pa nizi ali *stavki*. Končni neprazni množici znakov, s katerimi smo zapisali vzorec, pravimo *abeceda*.

V nize znakov pogosto pretvarjamo tudi določene vrste signalov, kot je na primer govorni signal. Pri tem posamezne odtipke signalov ali krajše odseke signalov s postopkom vektorske kvantizacije preslikamo v nize znakov iz končne abecede, ki vsebujejo od nekaj deset do nekaj sto znakov.

Pri govornih signalih pogosto uporabljamo kar fonetično abecedo in v znake te abecede preslikujemo odseke govornega signala, ki predstavljajo uresničitve osnovnih govornih enot, ki jim pravimo *fonemi*. Preslikavo ponavadi izvedemo z razpoznavalniki glasov, ki temeljijo na prikritih Markovovih modelih ipd. Namesto fonemov pogosto uporabljamo kar *grafeme* ali *črke*.

Govorni signal danega ločeno izgovorjenega govornega ukaza bi tako, denimo, lahko preslikali v niz grafemov $x = desno$. Zaradi napak, ki jih dela samodejni razpoznavalnik glasov, bi se govorni ukaz lahko preslikal tudi v $x = tesmo$.

Vzorci, ki smo jih zapisali z nizi znakov, lahko razpoznavamo tako, da merimo podobnost med njimi in med značilnimi predstavniki (prototipi) razredov, ki so prav tako zapisani z nizi znakov iz abecede. Vzorec razvrstimo v razred, katerega značilni predstavnik mu je najbolj podoben.

Podobnost dveh nizov znakov najbolj pogosto "merimo" z Levenshteinovo razdaljo. Le-ta je primernejša kot na primer Hammingova razdalja, ker ne zahteva, da sta oba niza enako dolga, hkrati pa ni tako občutljiva na manjša popačenja vzorcev.

5.7 Levenshteinova razdalja med dvema nizoma znakov

Množico vseh nizov, ki jih lahko sestavimo z znaki iz abecede V , vključno z nizom dolžine nič λ , označimo z V^* . Vsak stik nizov iz V^* je tudi iz V^* .

Levenshteinova razdalja med nizoma znakov x in y iz V^* je najmanjše število preslikav znakov, ki jih potrebujemo za prevedbo niza x v niz y . Pri tem so možne naslednje preslikave znakov

1. zamenjava znaka

$$\alpha a \beta \xrightarrow{T_Z} \alpha b \beta \quad \forall a, b \in V, \quad a \neq b, \quad \alpha, \beta \in V^*,$$

2. brisanje znaka

$$\alpha a \beta \xrightarrow{T_B} \alpha \beta \quad \forall a \in V, \quad \alpha, \beta \in V^* \text{ in}$$

3. vrivanje znaka

$$\alpha \beta \xrightarrow{T_V} \alpha a \beta \quad \forall a \in V, \quad \alpha, \beta \in V^*.$$

Matematično Levenshteinovo razdaljo zapišemo kot

$$D_L(x, y) = \min_j \{Z_j + B_j + V_j\}, \quad j = 1, \dots, J,$$

kjer Z_j , B_j in V_j označujejo število zamenjanih, brisanih in vrinjenih znakov, pri prevedbi niza znakov x v niz y ter J število vseh možnih prevedb enega niza v drugega.

Pri uteženi Levenshteinovi razdalji posameznim preslikavam znakov pripisujemo nenegativno ceno. Cena je lahko odvisna od posameznih znakov in določena na naslednji način:

1. cena preslikave znaka a iz niza x v znak b niza y je $Z(a, b)$, $\forall a, b \in V$, pri čemer je cena zamenjave istih znakov $Z(a, a) = 0$, $\forall a \in V$.
2. cena brisanja znaka a iz niza x je $B(a)$, $\forall a \in V$.
3. cena vrivanja znaka b v niz y je $V(b)$, $\forall b \in V$.

Utežena Levenshteinova razdalja med nizoma x in y je potem določena kot najmanjša vsota cen preslikav znakov, ki prevedejo niz x v niz y . Utežena Levenshteinova razdalja je enaka navadni Levenshteinovi razdalji, ko za cene vseh preslikav znakov uporabljamo vrednost 1.

Ker je cena prevedbe niza x v niz y enaka vsoti cen posameznih preslikav znakov iz zaporedja preslikav, ki določa prevedbo, lahko izračunamo najcenejšo prevedbo z dinamičnim programiranjem, to je z zaporednim minimiziranjem delnih vsot.

Vzemimo, da niza x in y sestavlja n oziroma m znakov, torej

$$\begin{aligned}x &= a_1 a_2 \cdots a_n \text{ in} \\y &= b_1 b_2 \cdots b_m.\end{aligned}$$

Vzemimo, da $x(i)$ in $y(i)$ označujeta delna niza nizov x in y , torej

$$\begin{aligned}x(i) &= a_1 a_2 \cdots a_i, \quad i = 1, \dots, n \quad \text{in} \\y(i) &= b_1 b_2 \cdots b_j, \quad j = 1, \dots, m.\end{aligned}$$

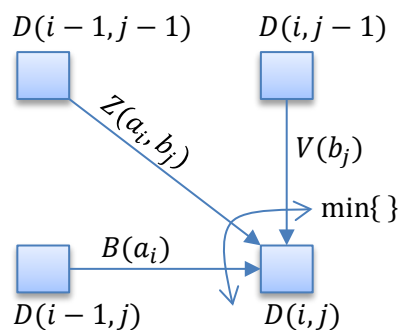
Levenshteinovo razdaljo med nizoma $x(i)$ in $y(i)$ krajše zapišimo

$$D(i, j) = D_L(x(i), y(i)).$$

Razdaljo potem izračunamo kot

$$D(i, j) = \min \begin{cases} D(i-1, j-1) + Z(a_i, b_j) \\ D(i-1, j) + B(a_i) \\ D(i, j-1) + V(b_j) \end{cases}$$

Izračun grafično ponazorimo z naslednjo sliko.



Celoten rekurzivni izračun, ki upošteva vse možne vrednosti indeksov i in j , je potem določne kot

$$D(0,0) = 0$$

$$D(i, 0) = D(i - 1, 0) + B(a_i), \quad i = 1, \dots, n$$

$$D(0, j) = D(0, j - 1) + V(b_j), \quad j = 1, \dots, m$$

$$D(i, j) = \min \left\{ \begin{array}{l} D(i - 1, j - 1) + Z(a_i, b_j) \\ D(i - 1, j) + B(a_i) \\ D(i, j - 1) + V(b_j) \end{array} \right\}, \quad \begin{array}{l} i = 1, \dots, n \wedge \\ j = 1, \dots, m \end{array}$$

Rekurzivni izračun razdalje lahko grafično ponazorimo z matriko razdalj med vsemi delnimi nizi obeh nizov, med katerim računamo razdaljo, torej z matriko z elementi $D(i, j)$ pri $i = 0, \dots, n$ in $j = 0, \dots, m$. Iz takšne ponazoritve lahko razberemo preslikave znakov, ki prevedejo en niz v drugega. Primer takšne ponazoritve je podan v naslednji nalogi.

5.7.1 Naloga

Izračunajte Levenshteinovo razdaljo med nizoma znakov

$$x = \textit{beseda} \quad \text{in} \quad y = \textit{bedra}.$$

Rekurzivni izračun razdalje ponazorite grafično z matriko razdalj med vsemi delnimi nizi obeh nizov in podajte preslikave, ki določajo Levenshteinovo razdaljo in prevedejo en niz znakov v drugega.

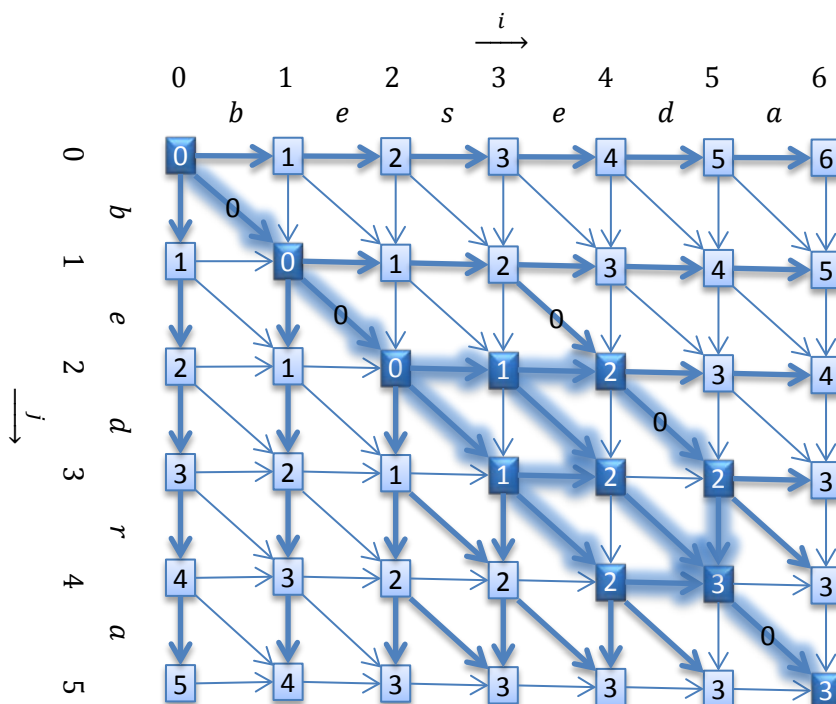
Rešitev

Pri izračunu Levenshteinove razdalja upoštevamo, da so cene vseh preslikav znakov enake ena, samo preslikava enakih znakov je enaka nič, torej, za vsak $\forall a_i, b_j \in V$ so cene za zamenjavo brisanje in vrivanje znakov določene kot

$$Z(a_i, b_j) = \begin{cases} 1 & a_i \neq b_j \\ 0 & a_i = b_j \end{cases} \quad \text{in} \quad B(a_i) = V(b_j) = 1.$$

Celoten rekurzivni izračun ponazorimo z naslednjo sliko. Slika ponazarja matriko vseh vrednosti $D(i, j)$ pri $i = 0, \dots, 6$ in $j = 0, \dots, 5$, pri čemer upoštevamo, da je dolžina niza x enaka 6 in dolžina niza y enaka 5. V sliki so vpisane samo cene preslikav enakih znakov, ki imajo vrednosti 0. Vse preostale cene so enake 1. Zadebeljeno puščice ponazarjajo minimum, ki se upošteva pri rekurzivni določitvi vrednosti razdalje $D(i, j)$ iz razdalj, ki se nanašajo na krajše delne nize. Dodatno zasenčena je t.i. sled preslikav z najmanjšo skupno ceno, ki prevedejo en niz v drugega.

Levenshteinovo razdaljo tako določa zadnja vrednost razdalje pri $D(6,5) = D(n,m) = 3$. To pomeni, da je mogoče niz x prevesti v niz y z najmanj tremi preslikavami znakov. Te preslikave lahko razberemo iz označene sledi, ki jo določimo od zadnje vrednosti razdalja $D(6,5) = D(n,m) = 3$ proti prvi vrednosti razdalje $D(0,0) = 0$, pri čemer se selimo po zadebeljenih puščicah, ki označujejo minime.



Iz označenih sledi razberemo, da je mogoče niz x prevesti v niz y z različnimi zaporedji najmanj treh preslikav znakov. Primer takšne prevedbe, ki jo določajo preslikave na povsem spodnji sledi je tako

$$x = \underline{b}e\underline{s}e\underline{d}a \xrightarrow{T_Z} \underline{b}e\underline{\bar{e}}\underline{d}a \xrightarrow{T_Z} \underline{b}e\underline{\bar{r}}\underline{d}a \xrightarrow{T_B} \underline{b}e\underline{d}r\underline{a} = y$$

Druga povsem zgornja sled pa določa preslikave

$$x = \underline{b}e\underline{s}e\underline{d}a \xrightarrow{T_B} \underline{b}e\underline{e}d\underline{a} \xrightarrow{T_B} \underline{b}e\underline{d}a \xrightarrow{T_V} \underline{b}e\underline{\bar{r}}a = y$$

Na podoben način bi lahko identificirali še tri (skupaj jih je pet) zaporedja treh preslikav, ki niz x prevede v niz y .

Prevajanje enega niza v drugega pogosto ponazorimo tudi s t.i. poravnavo. Prvo zgoraj navedeno prevedbo bi tako ponazorili s poravnavo

$$\begin{array}{c} b e s e d a \\ b e d r _ a \end{array}$$

In podobno lahko ponazorimo poravnavo še drugi zgornji primer

$$\begin{array}{c} b e s e d _ a \\ b e _ _ d r a \end{array}$$

6 Razvrščanje vzorcev s odločanjem

Vzorci, ki jih navadno zapišemo kot vektorje značilik, lahko razvrščamo hitreje kot s prileganjem, če jih razvrščamo s funkcijami, ki delijo n -razsežni prostor značilik vzorcev, \mathbb{R}^n , v neprekrivajoča se področja, kjer vsako področje predstavlja domeno enega razreda vzorcev. V teoriji razpoznavanja vzorcev imenujemo takšne funkcije *odločitvene funkcije*, razvrščanje, ki temelji na uporabi odločitvenih funkcij, pa *razvrščanje z odločanjem*.

Odločitvene funkcije so vse funkcije n -spremenljivk, ki zadoščajo pogoju

$$d_i(\mathbf{x}) > d_j(\mathbf{x}), \quad \forall \mathbf{x} \in C_i, \quad \forall j \neq i$$

Na področju \mathcal{P}_i , ki pripada vzorcem iz razreda C_i , ima torej odločitvena funkcija $d_i(\mathbf{x})$ večjo vrednost kot katerakoli druga odločitvena funkcija $d_j(\mathbf{x})$ iz sistema M -tih odločitvenih funkcij.

Mejo, ki ločuje področji \mathcal{P}_i in \mathcal{P}_j , sestavljajo vse tiste točke prostora značilik \mathbb{R}^n , ki zadoščajo pogoju:

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

Za razmejitev prostora vzorcev iz M razredov objektov razpoznavanja potrebujemo največ $M(M - 1)/2$ mej:

$$d_{ij}(\mathbf{x}) = d_i(\mathbf{x}) - d_j(\mathbf{x}) = 0, \quad i, j = 1, 2, \dots, M; \quad i \neq j.$$

Meje, ki razmejujejo prostor \mathbb{R}^n , imenujemo *ločilne meje*. Vidimo, da lahko z ločilnimi mejami zapišemo pravilo razvrščanja vzorcev:

$$d_{ij}(\mathbf{x}) > 0 \quad \forall \mathbf{x} \in C_i$$

oziroma

$$d_{ij}(\mathbf{x}) < 0 \quad \forall \mathbf{x} \in C_j$$

Odločitvene funkcije oziroma ločilne meje največkrat zapišemo s polinomi, vsotami potencialnih funkcij ali s funkcijami pogojnih gostot verjetnosti razredov

Parametre odločitvenih funkcij oziroma ločilnih mej (npr. koeficiente polinomov) ocenimo iz učne množice vzorcev \mathcal{U}_M . Ker so z določitvijo parametrov odločitvenih funkcij ali ločilnih mej vzpostavljene zveze med oznakami razredov in objekti na danem področju uporabe, in ker se teh zvez

človek uči, pravimo procesu ocenjevanja parametrov odločitvenih funkcij ali ločilnih mej *učenje*.

6.1 Trije načrti razvrščevalnikov

Če znanje o področju uporabe v razvrščevalnik vzorcev zapišemo z uporabo odločitvenih funkcij ali ločilnih mej, lahko upoštevamo naslednje tri postopke načrtovanja razvrščevalnikov vzorcev:

- 1) Iz učne množice \mathcal{U}_M za vsak razred vzorcev določimo natanko eno odločitveno funkcijo $d_i(\mathbf{x})$, $i = 1, \dots, M$. Vzorce nato razvrščamo po pravilu najvišje vrednosti odločitvene funkcije, torej: $\mathbf{x} \in C_i$, če velja $d_i(\mathbf{x}) > d_j(\mathbf{x})$ za $j = 1, \dots, M; j \neq i$, kjer \mathbf{x} označuje vzorec, ki ga razpoznavamo.
- 2) Za vse možne kombinacije parov M razredov (teh je $M(M - 1)/2$) iz učne množice \mathcal{U}_M določimo ločilne meje $d_{ij}(\mathbf{x})$ med vzorci para razredov. Vzorce nato razvrščamo po pravilu pozitivne vrednosti funkcije ločilne meje, torej: $\mathbf{x} \in C_i$, če velja $d_{ij}(\mathbf{x}) > 0$ za $j = 1, \dots, M; j \neq i$.
- 3) Za vsak razred vzorcev določamo ločilno mejo $d_i^*(\mathbf{x})$ med vzorci i -tega razreda in vzorci vseh preostalih razredov. Vzorce nato ponovno razvrščamo po pravilu pozitivne vrednosti funkcije ločilne meje, torej: $\mathbf{x} \in C_i$, če velja $d_i^*(\mathbf{x}) > 0$ pri $i = 1, \dots, M$.

Pri vseh treh načrtih gre torej za preslikavo učne množice vzorcev v odločitvene funkcije oziroma ločilne meje. V prvem primeru je učenje preslikava

$$\mathcal{U}_M \rightarrow \{d_1(\mathbf{x}), \dots, d_M(\mathbf{x})\},$$

v prvem primeru preslikava

$$\mathcal{U}_M \rightarrow \{d_{11}(\mathbf{x}), d_{12}(\mathbf{x}), \dots, d_{M-1,M}(\mathbf{x})\},$$

in v tretjem primeru preslikava

$$\mathcal{U}_M \rightarrow \{d_1^*(\mathbf{x}), \dots, d_M^*(\mathbf{x})\},$$

Pri drugem načrtu in določanju do največ $M(M - 1)/2$ ločilnih mej lahko povečamo učinkovitost razvrščanja z določanjem ločilnih mej najprej med pari večjih podmnožic učne množice, ki lahko vključujejo vzorce več razredov, in nato še med pari vedno manjših podmnožic, vse do parov podmnožic, ki vsebujejo vsaka samo še vzorce posameznih razredov. Pri razvrščanju vzorcev nato ločilne meje uredimo v hierarhično drevesno strukturo, ki ji pravimo

odločitveno drevo. Na ta način je potrebno za razvrščanje vzorcev določiti samo še $M - 1$ ločilnih mej.

6.2 Linearne odločitvene funkcije

Najbolj preprosta oblika odločitvena funkcija je linearna funkcija vzorca, ki jo za n značilk vzorcev zapišemo kot:

$$d(\mathbf{x}) = w_1x_1 + w_2x_2 + \dots + w_nx_n + w_{n+1} = \mathbf{w}_0^T \mathbf{x} + w_{n+1}$$

kjer so:

$$\begin{array}{ll} \mathbf{w}_0 = (w_1, w_2, \dots, w_n)^T & \text{vektor koeficientov,} \\ \mathbf{x} = (x_1, x_2, \dots, x_n)^T & \text{vzorec - vektor vrednosti značilk in} \\ w_{n+1} & \text{prag.} \end{array}$$

Meja, ki ločuje C_i in C_j oziroma \mathcal{P}_i in \mathcal{P}_j v prostoru \mathbb{R}^n , je prav tako linearna

$$d_{ij}(\mathbf{x}) = d_i(\mathbf{x}) - d_j(\mathbf{x}) = 0$$

oziroma

$$(\mathbf{w}_{i0} - \mathbf{w}_{j0})^T \mathbf{x} = -(w_{i,n+1} - w_{j,n+1}) = w_{j,n+1} - w_{i,n+1}$$

Vidimo, da lahko z linearnimi ločilnimi mejami zapišemo pravilo razvrščanja vzorcev:

$$(\mathbf{w}_{i0} - \mathbf{w}_{j0})^T \mathbf{x} > w_{j,n+1} - w_{i,n+1} \quad \forall \mathbf{x} \in C_i$$

oziroma

$$(\mathbf{w}_{i0} - \mathbf{w}_{j0})^T \mathbf{x} < w_{j,n+1} - w_{i,n+1} \quad \forall \mathbf{x} \in C_j$$

Linearne ločilne meje so točke, premice, ravnine oziroma hiper-ravnine, odvisno od dimenzije prostora značilk.

Linearne odločitvene funkcije $d(\mathbf{x}) = w_1x_1 + w_2x_2 + \dots + w_nx_n + w_{n+1}$ navadno zapišemo v splošno sprejeti homogeni obliki:

$$d(\mathbf{x}) = \mathbf{w}^T \mathbf{x}$$

kjer so:

$$\begin{array}{ll} \mathbf{w} = (w_1, w_2, \dots, w_n, w_{n+1})^T & \text{vektor koeficientov in} \\ \mathbf{x} = (x_1, x_2, \dots, x_n, 1)^T & \text{»razširjeni« vzorec.} \end{array}$$

Linearno ločilno mejo potem zapišemo

$$(\mathbf{w}_i - \mathbf{w}_j)^T \mathbf{x} = 0,$$

pravilo razvrščanja z linearnimi ločilnimi mejami pa

$$(\mathbf{w}_i - \mathbf{w}_j)^T \mathbf{x} > 0 \quad \forall \mathbf{x} \in C_i$$

oziroma

$$(\mathbf{w}_i - \mathbf{w}_j)^T \mathbf{x} < 0 \quad \forall \mathbf{x} \in C_j$$

6.2.1 Naloga

Dana je učna množica vzorcev dveh razredov $\mathcal{U}_2 = \{U_1, U_2\}$, kjer je

$$U_1 = \{((1,3)^T, \omega_1), ((2,1)^T, \omega_1), ((2,2)^T, \omega_1), ((3,1)^T, \omega_1)\},$$

$$U_2 = \{((4,2)^T, \omega_2), ((5,2)^T, \omega_2), ((3,3)^T, \omega_2), ((6,3)^T, \omega_2)\}.$$

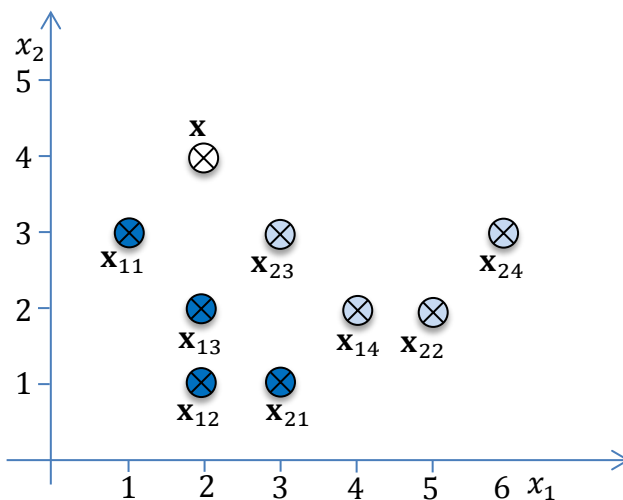
Vzemimo, da sta podana vektorja koeficientov obeh linearnih odločitvenih funkcij za podana dva razreda vzorcev in sicer

$$\mathbf{w}_1 = (-1, -1, 6)^T \quad \text{in} \quad \mathbf{w}_2 = (2, 2, -9)^T.$$

Določite ločilno mejo med razredoma vzorcev in jo skupaj z vzorci skicirajte v \mathbb{R}^2 . Z uporabo ločilne meje razvrstite vzorec $\tilde{\mathbf{x}} = (2, 4)^T$ v enega od obeh razredov.

Rešitev

Najprej ponazorimo vzorce kot točke v dvorazsežnem prostoru značilk.



Linearno ločilno mejo določa izraz $(\mathbf{w}_1 - \mathbf{w}_2)^T \mathbf{x} = \mathbf{w}^T \mathbf{x} = 0$, kjer \mathbf{x} označuje »razširjen« vektor in je vektor razlike med vektorjema koeficientov

$$\mathbf{w} = \mathbf{w}_1 - \mathbf{w}_2 = (-1, -1, 6)^T - (2, 2, 9)^T = (-3, -3, 15)^T$$

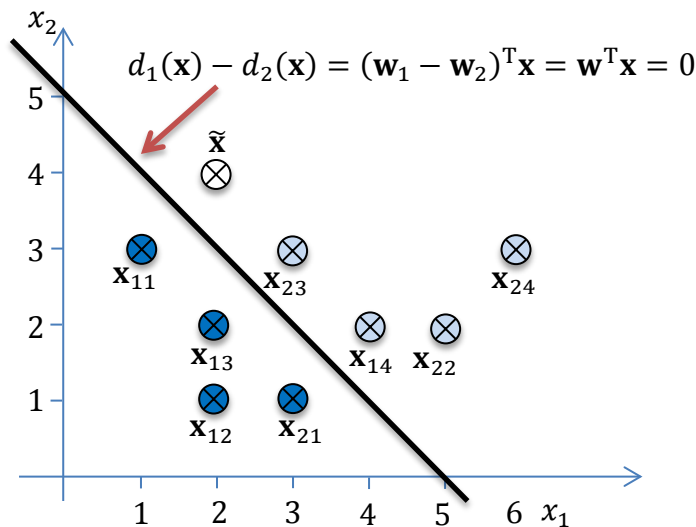
Funkcijo ločilne meja lahko tako zapišemo

$$\mathbf{w}^T \mathbf{x} = [-3, -3, 15] \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix} = -3x_1 - 3x_2 + 15 = 0$$

Za lažje skiciranje lege linearne ločilne meje izpostavimo x_2 in dobimo

$$x_2 = \frac{-3x_1 + 15}{3} = -x_1 + 5$$

Linearno ločilno mejo skicirana skiciramo v \mathbb{R}^2 tako, kot je prikazano na spodnji sliki.



Za razvrstitev s podano ločilno mejo moramo podani vzorec $\tilde{\mathbf{x}} = (2, 4)^T$ razširiti v razširjeni vzorec $\mathbf{x} = (2, 4, 1)^T$. Nato izračunamo vrednost funkcije ločilne meje

$$d_{12}(\mathbf{x}) = \mathbf{w}^T \mathbf{x} = [-3, -3, 15] \begin{bmatrix} 2 \\ 4 \\ 1 \end{bmatrix} = -6 - 12 + 15 = -3 < 0.$$

Ker je vrednost funkcije $d_{12}(\mathbf{x}) < 0$, se v skladu s tem pravilom razvrščanja vzorec razvrsti v razred $\mathbf{x} \in C_2$, kar je razvidno tudi iz slike.

6.3 Tolmačenje pravila razvrščanja 1 – NN z linearnimi odločitvenimi funkcijami

Pri tolmačenju omenjenega pravila navadno predpostavimo, da pri razpoznavanju s prileganjem podobnost med vzorci - vektorji značilnik – merimo z Evklidovo razdaljo, katere kvadrat lahko zapišemo

$$D^2(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^T (\mathbf{x}_i - \mathbf{x}_j)$$

Če je razred vzorcev C_i predstavljen samo z enim značilnim predstavnikom \mathbf{x}_i , potem lahko odločitveno funkcijo $d_i(\mathbf{x})$ izpeljemo iz negativnega kvadrata Evklidove razdalje $-D^2(\mathbf{x}, \mathbf{x}_i)$, to je

$$d_i(\mathbf{x}) = \mathbf{x}^T \mathbf{x}_i - \frac{1}{2} \mathbf{x}_i^T \mathbf{x}_i = \mathbf{w}_i^T \mathbf{x} + w_{n+1}$$

kjer so:

$\mathbf{w}_i = \mathbf{x}_i = (x_{i_1}, x_{i_2}, \dots, x_{i_n})^T$ vektor koeficientov – značilni predstavnik

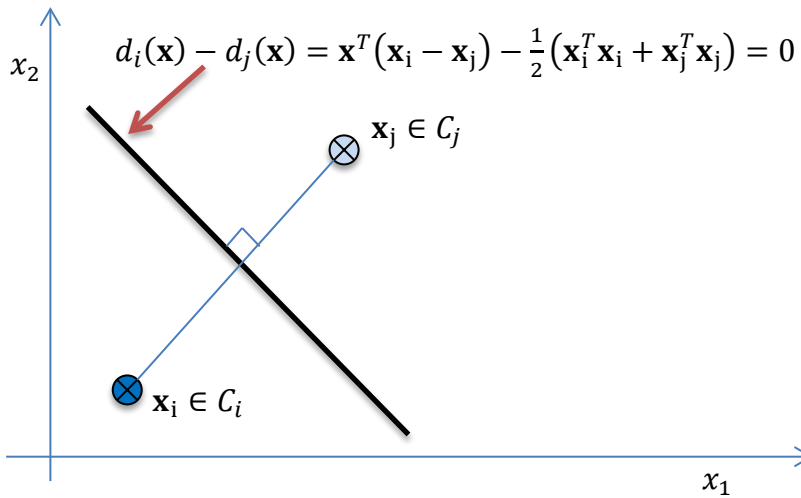
$\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ vzorec - vektor vrednosti značilnik in

$w_{n+1} = -\frac{1}{2} \mathbf{x}_i^T \mathbf{x}_i$ prag.

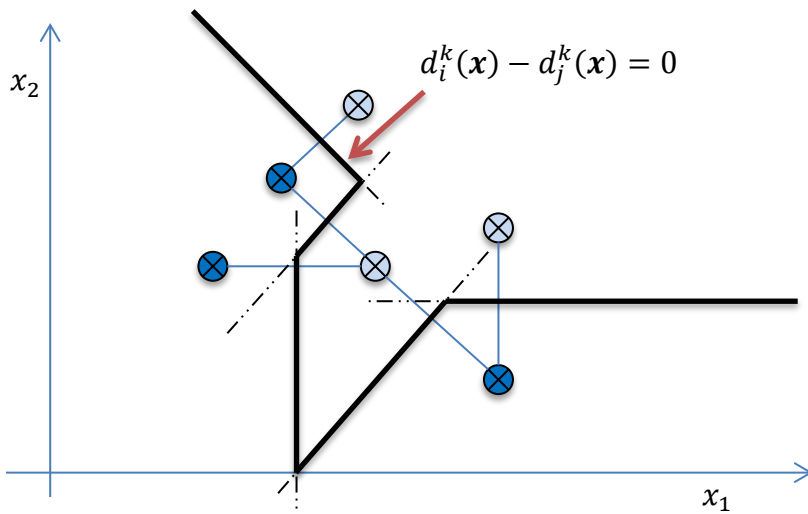
Odločitvene funkcije so linearne funkcije vzorcev \mathbf{x} , zato je linearna tudi ločilna meja med razredoma C_i in C_j prav tako linearna

$$d_{ij}(\mathbf{x}) = d_i(\mathbf{x}) - d_j(\mathbf{x}) = \mathbf{x}^T (\mathbf{x}_i - \mathbf{x}_j) - \frac{1}{2} (\mathbf{x}_i^T \mathbf{x}_i + \mathbf{x}_j^T \mathbf{x}_j) = 0$$

Podana ločilna meja pravokotno razpolavlja spojnico vzorcev (značilnih predstavnikov razredov) \mathbf{x}_i in \mathbf{x}_j , kot je ilustrirano na spodnji sliki.



V primeru, da posamezni razred vsebuje več vzorcev oziroma značilnih predstavnikov, potem je ločilna meja med dvema razredoma vzorcev C_i in C_j odsekoma linearna in razpolavlja spojnice najbližjih vzorcev iz dveh nasprotnih razredov, kot je ilustrirano na spodnji sliki.



6.3.1 Naloga

Dana je učna množica vzorcev dveh razredov $\mathcal{U}_2 = \{U_1, U_2\}$, kjer je

$$U_1 = \{((2,2)^T, \omega_1), ((2,-2)^T, \omega_1), ((-2,-2)^T, \omega_1)\},$$

$$U_2 = \{((-0,2)^T, \omega_2), ((0,0)^T, \omega_2)\}.$$

Najprej grafično skicirajte potek ločilne meje med razredoma vzorcev, če razvrščanje s prileganjem s pravilom 1-NN in uporabo Evklidove razdalje kot osnovne mere različnosti tolmačimo kot razpoznavanje z odločanjem z linearnimi odločitvenimi funkcijami.

Nato predstavite razreda vzorcev z značilnima predstavnikoma \mathbf{x}_1 in \mathbf{x}_2 , ki naj bosta kar vektorja povprečnih vrednosti značilnik, ter določite linearno ločilno mejo, ki jo določata ta dva vektorja.

7 Zaključek

Primeri, ki niso zajeti v teh skriptah in so bili rešeni na vajah ter tudi lahko pridejo v poštev pri pisnem izpitu, izhajajo iz vsebine in zgledov na naslednjih straneh omenjenega učbenika:

- poglavje 7.3.4 (str. 338-340), zgled 42 (str. 348);
- poglavje 7.3.9.2 (str. 371-375), zgled 49 (str. 376-378);
- poglavje 8.2.2 (str. 432-433); zgled 58 (str. 433-435).

Čas pisanja: 60 minut

Naloga 1 (30 %)

Dana je štiri-bitna siva slika s podaj podano množico možnih diskretnih vrednosti svetilnosti (šestnajst možnih sivih nivojev) in porazdelitvijo njihovih relativnih frekvenc

$$G_{16} = \{0, 1, \dots, 15\}$$

$$P = \left\{ 0, 0, \frac{2}{10}, \frac{2}{10}, 0, 0, 0, 0, 0, 0, 0, 0, \frac{3}{10}, \frac{3}{10}, 0, 0 \right\}$$

Predpostavimo, da smo izbrali postopek upravljanja slike z maksimizacijo informacije. Izračunajte in skicirajte vrednost informacije upravljenega slike pri vseh možnih vrednostih praga svetilnosti $t = 0, \dots, 14$. Za poenostavitev izračuna upoštevajte lastnosti entropijske funkcije, ki se nanaša na verjetnosti (relativne frekvence) z vrednostjo nič.

Naloga 2 (35 %)

Dana je urejena učna množica vzorcev treh razredov $\mathcal{U}_3 = \{U_1, U_2, U_3\}$, kjer sta

$$U_1 = \{(2,1)^T, (1,1)^T, (1,2)^T\}, U_2 = \{(3,2)^T, (3,3)^T, (3,4)^T\} \text{ in } U_3 \\ = \{(4,1)^T, (5,1)^T, (5,2)^T\}.$$

Vzemimo, da smo za razvrščanje vzorcev izbrali pravilo 1-NN in Evklidovo razdaljo za osnovno mero različnosti med vzorci. Vzorce grafično ponazorite v ravnini \mathbb{R}^2 . Evklidove razdalje med vzorci potem ocenjujete kar grafično.

- Zgostite podano učno množico, pri čemer v prvem koraku postopka zgoščanja v seznam STORE preselite prvi vzorec po vrsti iz podane urejene učne množice.
- Skicirajte vzorce zgoščene učne množice v \mathbb{R}^2 in linearne ločilne meje med razredi, če pravilo 1-NN tolmačimo z linearnimi odločitvenimi funkcijami.

Naloga 3 (35 %)

Predpostavimo, da so dvo-razsežni vzorci $\mathbf{x} = (x_1, x_2)^T$ po dveh razredih C_1 in C_2 porazdeljeni normalno. Vzemimo, da so parametri normalnih pogojnih porazdelitev za oba razreda, $p(\mathbf{x}|C_1)$ in $p(\mathbf{x}|C_2)$, enaki:

$$\mathbf{m}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}; \mathbf{K}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}, \quad \mathbf{m}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}; \mathbf{K}_2 = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

Vzemimo, da sta apriorni verjetnosti $P(C_1) = P(C_2) = \frac{1}{2}$.

- Določite najboljši odločitveni funkciji $d_1(\mathbf{x})$ in $d_2(\mathbf{x})$ za normalno porazdeljene vzorce iz obeh razredov C_1 in C_2 .
- V ravnini \mathbb{R}^2 približno skicirajte ovojnice enakih vrednosti gostot verjetnosti (kovariančne elipse) obeh funkcij pogojnih gostot verjetnosti $p(\mathbf{x}|C_1)$ in $p(\mathbf{x}|C_2)$.
- V ravnini \mathbb{R}^2 približno skicirajte še ločilno mejo med razredoma vzorcev $d_{12}(\mathbf{x}) = d_1(\mathbf{x}) - d_2(\mathbf{x}) = 0$ in označite področji vzorcev \mathcal{P}_1 in \mathcal{P}_2 .