Raspoznavanje uzoraka

prof. dr.sc. Slobodan Ribarić

18. studenog 2012.

Sadržaj

Poglavlje 1

Uvod

Raspoznavanje uzoraka

Raspoznavanje uzoraka (eng. *Pattern Recognition*, njem. *Mustererkennung*, franc. *le reconnaissance des formes*) je znanstvena disciplina Računarskih znanosti čiji je cilj **klasifikacija** (razvrstavanje) **objekata** u jedan od brojnih **razred** ili **klasa**.

Raspoznavanje uzoraka je relativno mlada znanstvena disciplina nastala 1965. – 1970. godina. Kao takva, sastavni je dio umjetne (strojne) inteligencije (eng. *Machine intelligence*).

Primjeri područja upotrebe

Postoje brojne primjene *Raspoznavanja uzoraka* u stvarnom životu. Neke zanimljivije primjere možemo pronaći u nekoliko različitih domena.

Prilikom raspoznavanja vizualnih objekata možemo izvršavati *klasifikacija znakova* (slovčano-brojčanih, tiskanih, rukom pisanih, OCR sustavi), raspoznavati objekte bitne za *medicinsku dijagnostiku* (X– mamografija, tomografija, građa stanice, klasifikacija kromosoma), detektirati i lokalizirati objekte na slikama, interpretirati 3D scene (robotski ili strojni vid), otkrivati prirodna bogatstava na temelju satelitskih snimaka. Raspoznavanje uzoraka sastavni je dio *računalnog vida* te *biometrijskih sigurnosnih sustava.*

Ako želimo raspoznavati zvučne uzorke, htjet ćemo raspoznavati govor, govornika, jezika te zvuk (pravilan rad stroja, tip vozila, raspoznavanje koraka).

Biomedicina predstavlja veliko područje uporabe za raspoznavanje uzoraka. Pri raspoznavanju biomedicinskih uzoraka raspoznajemo EKG, EEG te dijagnosticiramo različite bolesti.

Raspoznavanjem uzoraka potreba želimo saznati je li potres nastao prirodnim uzrokom ili je potaknut podzemnom atomskom eksplozijom.Također, želimo raspoznati korake te razlikovati ljudske od životinjskih.

Raspoznavanjem ponašanje složenih sustava možemo predviđati vrijeme, raspoznavati smjerove razvoja te razvoj ponude i potražnje na tržištu.

Novo znanje

Najnoviji i najznačajniji rezultati iz područja *raspoznavanja uzoraka* objavljuju se u međunarodnim znanstvenim časopisima. Evo nekoliko najvažnijih:

- IEEE Transactions on Pattern Analysis and Machine Intelligence
- IEEE Transactions on Systems, Man and Cybernetics
- IEEE Transactions on Neural Networks
- IEEE Transactions on Speech and Audio Processing
- IEEE Transactions on Image Processing
- IEEE Transactions on Fuzzy Systems
- Pattern Recognition
- Pattern Recognition Letters
- Computer Vision and Image Understanding
- Computer Speech and Language

1.1 Osnovni pojmovi i terminologija

Okolina

Okolina Θ je skup predmeta, pojava i bića, kraće objekata , koje raspoznajemo:

$$\Theta = \Theta_B \cup V_\Theta$$

gdje je:

$$\Theta_B = \{\sigma_k : k = 1, 2, \ldots\}$$

skup objekata, a

$$V_\Theta = \{v_j : j = 1, 2, \ldots\}$$

skup međusobnih relacija i veza između objekata u prostoru i vremenu.

Pri tome svaki objekt (predmet, fizička pojava, pojam, činjenica ili stanje) možemo opisati odgovarajućim brojem funkcija. Vrijednost tih funkcija daje karakterističnu količinu u prostoru i vremenu (u zavisnosti o vrsti objekata i osjetniku ili mjernom uređaju).

Univerzalni sustav za raspoznavanje

Univerzalni sustav za raspoznavanje koji je u stanju procesirati cijelu okolinu ili čak veći dio okoline **NIJE IZVEDIV (za sada)** - zato pri oblikovanju sustava za raspoznavanje uzoraka se ograničavamo na *područje uporabe*.

Područje uporabe

Područje uporabe sadrži samo one objekte $\sigma_k \in \Theta_B$ i njihove međusobne veze i odnose $v_j \in V_\Theta$ koje raspoznajemo:

$$\mathcal{PU} \in \Theta$$

Skup $\mathcal{PU}$ određen je **zadatkom** sustava za raspoznavanje.

Primjeri:

- *raspoznvanje brojeva u rasponu 0 do 9*
- *raspoznavanje slovčanih znakova*
- *raspoznavanje sastavnih djelova određenog složenog proizvoda*
- *raspoznavanje EKG-a*
- *analiza slika dobivenih iz određenog broja spektralnih kanala*

Uzorak

Uzorak je generički izraz za objekte raspoznavanja (eng. *Pattern*). Uzorak sadrži rezultat percepcije ili mjerenja (mjerna naprava, osjetinik) i predočava stroju podatke o objektu ili objektima i njihovim međusobnim odnosima

$$f_k(x) = \begin{bmatrix} f_{k_1}(x_1, x_2, \ldots, x_q) \\ f_{k_2}(x_1, x_2, \ldots, x_q) \\ \vdots \\ f_{k_p}(x_1, x_2, \ldots, x_q) \end{bmatrix}$$

gdje p i q zavise od sustava osjetnika, odnosno mjernih naprava koje sustav za raspoznavanje uzoraka koristi.

Primjeri:

- *sustav za raspoznavanje EKG signala:*
  f(t); p=q=1, t - vrijeme
- *sustav za raspoznavanje alfanumeričkih znakova:*
  f(x,y); p=1, q=2
- *sustav za raspoznavanje objekata u slikama u boji:*
  $f^R(x,y)$, $f^G(x,y)$, $f^B(x,y)$; *p=3, q=2*

Funkcija koja preslikava objekt raspoznavanja u uzorak mora biti takva da jednoznačno objekte iz razreda Θ_{B_i} preslika u razred uzoraka C_i.

Razred objekata

Razred objekata $\Theta_{B_i} \subset \Theta_B; i = 1, 2, \ldots, M$ je podskup onih objekata iz zadanog područja uporabe, na koje se odnosi *oznaka* (simbol, ime razreda) ω_i iz skupa oznaka razreda objekata:

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_M\}.$$

Primjer: Područje uporabe i raspoznavanje voća mogu činiti razredi označeni oznakam "banana", " limun", "kruška", . . . , itd.

Za zadano područje uporabe broj razreda $M \geq 2$ je određen zadatkom sustava za raspoznavanje (možemo reći da je M apriorno poznat), te vrijedi:

$$\Theta_{B_i} \cap \Theta_{B_j} = \emptyset \ za \ \forall i \neq j$$

$$\bigcup_{i=1}^{M} \Theta_{B_i} = \Theta_B$$

Presjek više razreda objekata mora biti prazan skup. Drugim rječima, jedan objekt može istovremeno pripadati **samo jednom** razredu objekata. Također, a djelomično slijedi iz gornjeg, unija svih razreda objekata mora činiti potpun skup objekata Θ_B.

Razred uzoraka

Razred uzoraka C_i čine slike objekata iz razreda $\Theta_{B_i}; i = 1, 2, \ldots, M$. Za zadano područje uporabe $\mathcal{PU}$ mora vrijediti za svaki par razreda uzoraka (C_i, C_j) da je

$$C_i \cap C_j = \emptyset \ za \ \forall i \neq j$$

te da je svaki uzorak iz razreda C_i sličniji nekom drugom uzorku iz razreda C_i, negoli uzorku iz razreda C_j za $\forall i \neq j$.

Opisivanje područja uporabe

Skup uzoraka koji opisuje područje uporabe čini konačan broj uzoraka iz zadanog područja uporabe. Taj konačan skup S_N koji čine M podskupova uzoraka S_i mora zadovoljavati slijedeće uvjete:

- $S_i \subseteq C_i, \ za \ \forall i = 1, 2, \ldots, M$
- $S_i \neq \emptyset, \ za \ \forall i = 1, 2, \ldots, M$
- $S_i \cap S_j = \emptyset, \ za \ \forall i \neq j$

- $\bigcup_{i=1}^{M} S_i = S_N$

gdje je C_i i-ti razred uzoraka iz zadanog područja uporabe. S N_i označimo kardinalni broj podskupa S_i. Za N - kardinalni broj skupa S_N, vrijedi:

$$N = \sum_{i=1}^{M} N_i.$$

Skup za učenje

Skup uzoraka za učenje U_M je konačan skup uzoraka iz zadanog područja uporabe $\mathcal{PU}$, iz kojeg sustav za raspoznavanje uzoraka može *naučiti* veze između oznake razreda i objekata raspoznavanja:

$$U_M = (S_N, \Omega)$$

gdje je S_N skup uzoraka koji opisuje područje uporabe, a $\Omega = \{\omega_1, \omega_2, \ldots, \omega_M\}$ skup oznaka razreda objekata u zadanom području uporabe.

Skup za učenje sastoji se od *podskupova* koje sačinjavaju parovi $(uzorak, klasa)$, a unutar jednog podskupa U_i svi uzorci pripadaju klasi ω_i:

$$U_M = \{U_1, U_2, \ldots, U_M\}$$

gdje je

$$U_i = \{(f_{i_1}(\vec{x}), \omega_i), (f_{i_2}(\vec{x}), \omega_i), \ldots, (f_{i_N}(\vec{x}), \omega_i)\}.$$

Pri tome moraju biti ispunjeni slijedeći uvjeti:

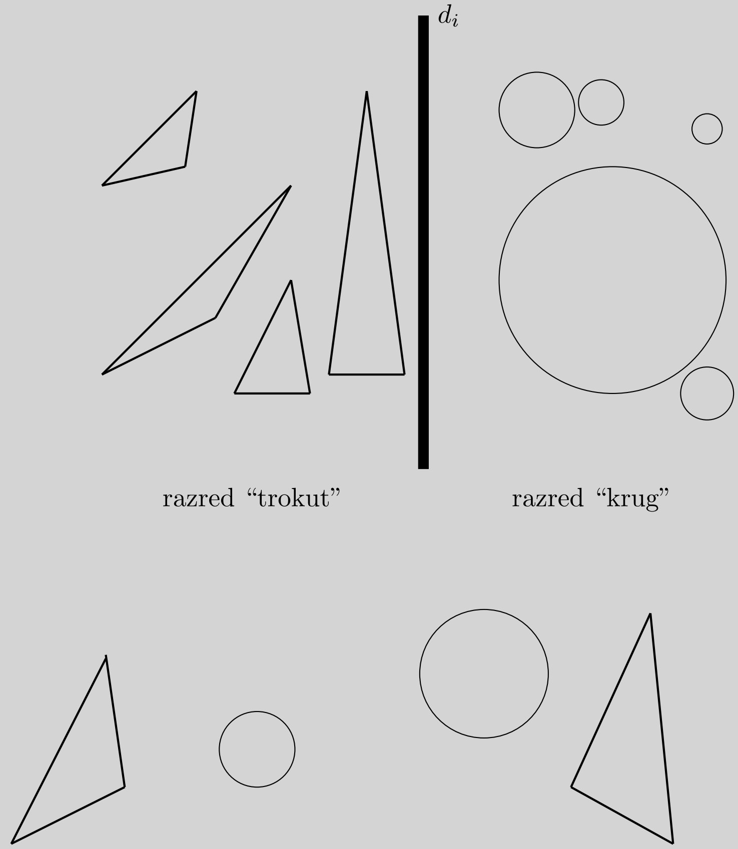
- $U_i \neq \emptyset \ za \ \forall i = 1, 2, \ldots, M$
- uzorci iz U_i moraju biti međusobno slični
- uzorci iz U_i nisu slični uzorcima iz U_j za $\forall i = 1, 2, \ldots, M$
- $U_i \cap U_j = \emptyset \ za \ \forall i \neq j$
- $\bigcup_{i=1}^{M} U_i = U_M$

Uzorke iz S_N u procesu oblikovanja sustava za raspoznavanje uzoraka za zadano područje uporabe, izabire i označava čovjek - stručnjak za zadano područje uporabe (nazovimo ga *učitelj*).

Učenje

Učenje je proces u kojem se upostavljaju veze između uzoraka iz skupa za učenje i oznakama razreda uzoraka. Učenje izvodimo s nepotpunom indukcijom pomoću koje poopćujemo informaciju (koju sadrži relativno mali skup za učenje) na sve uzorke iz danog područja uporabe $\mathcal{PU}$.

Primjer:



Ispitni primjeri geometrijskih likova

Tipovi uzoraka

Uzorke možemo podijeliti u dvije velike skupine bazirane na kompleksnosti informacije koju sadrži. To su *jednostavni* i *složeni* uzorci.

Uzorak je **jednostavan** ako ga raspoznajemo kao cjelinu (korisnik sustava za raspoznavanje uzoraka je zainteresiran samo za oznaku razreda kojem uzorak pripada). Uzorak se smatra **složenim** ako samo ime razreda nije dovoljno korisniku sustava za raspoznavanje uzoraka ili je čak klasifikacija uzorka neizvediva.

Primjer:

- *slika jednog slova ili znaka* $\rightarrow$ *jednostavan uzorak*
- *slika jedne stranice teksta* $\rightarrow$ *složeni uzorak*
- *signal izgovorene riječi* $\rightarrow$ *jednostavan uzorak*
- *signal izgovorene priče* $\rightarrow$ *složeni uzorak*

Raspoznavanje jednostavnih uzoraka

Raspoznavanje jednostavnih uzoraka možemo definirati kao

$$RU \ : \ f_k(\vec{x}) \longmapsto \omega_l, \ \omega_l \in \Omega$$

gdje je Ω skup od M uzoraka razreda iz PU. *Vrlo često se prirodaje $M+1$ oznaka oznaka koja označava razreda odbačenih uzoraka koje sustav ne može raspoznati.*

Raspoznavanje složenih uzoraka

Raspoznavanje složenih uzoraka definiramo kao

$$RU \ : \ f_k(\vec{x}) \longmapsto \lambda_l,$$

gdje je λ_l koristan opis uzorka (eng. *pattern description, pattern interpretation*). Kao rezultat raspoznavanja složenih uzoraka možemo dobiti *popis predmeta ili događaja koji su predmet zanimanja*, *opis promjena utvrđenih iz vremenskih sljedova uzoraka* te *opis sastavaljen iz definicije znakova ili riječi nekog prirodnog ili umjetnog jezika.*

1.2 Šest postulata raspoznavanja uzoraka

Šest postulata

Pri prikupljanju uzoraka za naš sustav moramo obratiti pažnju na nekoliko pravila. Ta pravila možemo formulirati u *šest postulata raspoznavanja uzoraka:*

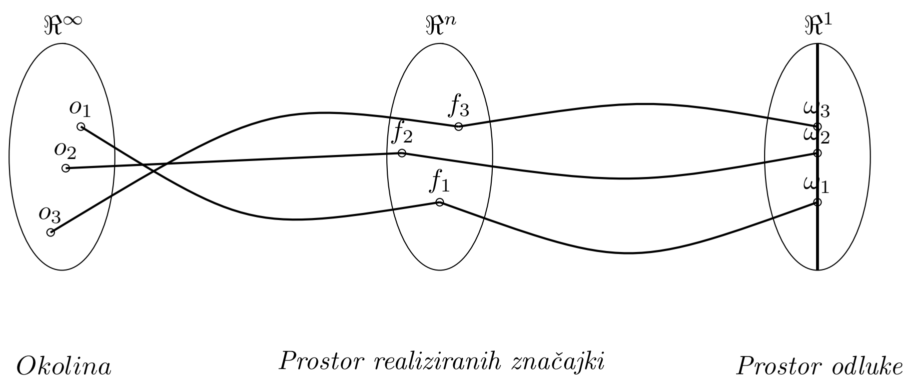
- **Postulat 1:** U cilju prikupljanja informaciju o PU moraju biti raspoloživi reprezentativni uzorci iz M razreda
- **Postulat 2:** Jednostavan uzorak ima značajke koje karakteriziraju njegovu pripadnost razredu
- **Postulat 3:** Značajke uzoraka koji pripadaju jednom razredu uzoraka zauzimaju kompaktno područje u prostoru značajki. Područja okupirana značajkama različitih razreda su odvojena

- **Postulat 4:** Složeni uzorak sastoji se iz jednostavnijih građevnih komponenti ili segmenata objekata koji se nalaze u izvjesnim odnosima. Uzorak se može sastaviti od tih komponenti
- **Postulat 5:** Složeni uzorak koji pripada određenom području uporabe ima određenu strukturu. To implicira da bilo kakvo uređenje jednostavnih građevnih elemenata neće dati uzorak $f_k(\vec{x})$
- **Postulat 6:** Dva uzorka su si slična ako je pogodno definirana mjera udaljenosti u prostoru značajki mala

Poglavlje 2

Model sustava za raspoznavanje uzoraka

2.1 Formalni model



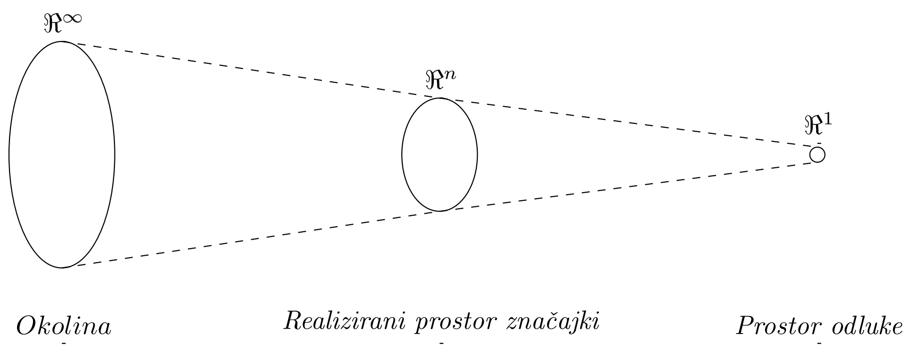
Slika 2.1: Formalni model sustava za raspoznavanje uzoraka

U formalnom modelu sustava za raspoznavanje uzoraka, u okolini Θ pronalazimo objekte i veze među njima: $\sigma_k \in \Theta$, $v_j \in V_\Theta$. Postupkom *mjerenja* iz okoline uzorkujemo objekte i pretvaramo ih u značajke $f_k : \sigma_k \to f_k$. Time dobivamo **prostor realiziranih uzoraka.** Postupkom *klasifikacije* dobivene uzorke smještamo u jednu od mogućih klasa, te dobivamo **prostor razreda uzoraka:** $\Omega = \{\omega_1, \ldots, \omega_M\}$.

Mjerenje

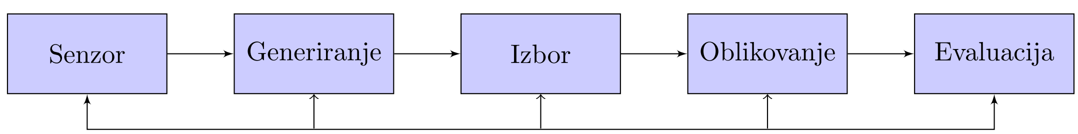
Klasifikacija

Postupak redukcije značajki možemo zorno prikazati *Uhrovim stožcem raspoznavanja* prikazanim na slici 2.2.



Slika 2.2: Uhrov stožac raspoznavanja – redukcija značajki

2.2 Značajke, vektor značajki i klasifikator



Slika 2.3: Osnove faze u postupku oblikovanja sustava RU

Značajke

Informacije o objektima iz fizičkog svijeta formiramo u **značajke**. Značajke promatramo kao slučajne varijable x_i.

Vektor značajki

Značajke jednog objekta formiramo u **vektor značajki**. Vektor značajki promatramo kao slučajni vektor $\vec{x}$:

$$\vec{x} = [x_1, x_2, \ldots, x_n]^T$$

Tipovi značajki

Značajke možemo podijeliti u dvije skupine - *interset* i *intraset*.

Značajke koje predstavljaju razlike između razreda uzoraka nazivaju se **interset** značajkama i to su značajke koje nose informacije potrebne za klasifikaciju objekata.

Intraset značajke zajedničke su **svim** razredima iz $\mathcal{PU}$ (područja uporabe) i <u>ne nose</u> diskriminacijsku informaciju – **takve značajke mogu se zanemariti**.

Izbor značajki

Kada biramo značajke koje će čini *skup za učenje*, biramo *interset* značajke. Pri tome ne smijemo zaboraviti da je u većini slučajeva određivanje potpunog skupa diskriminacijskih značajki **iznimno teško ili čak nemoguće**. Neke diskriminacijske značajke mogu se naći na temelju raspoloživih rezultata mjerenja (senzoriranja).

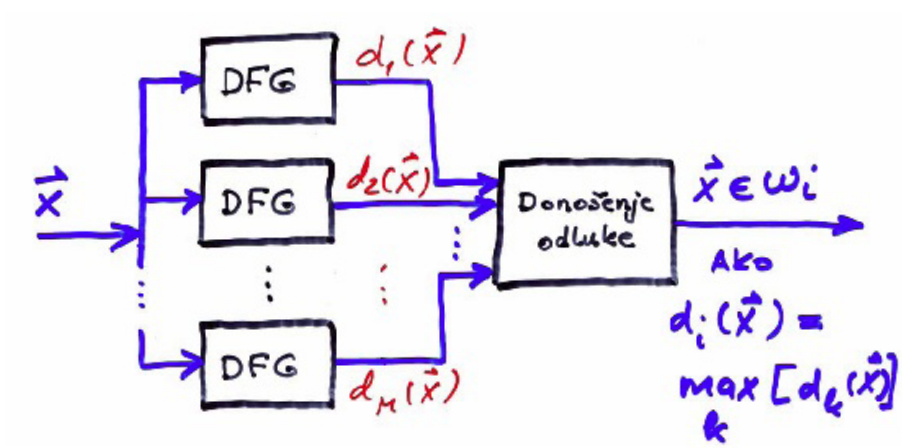
Vektor značajki predočujemo kao točku u n–dimenzijskom prostoru značajki. Obično definiramo i neku vrstu metrike u takvom prostoru značajki. Često ćemo se susresti s velikim brojem značajki te ćemo htjeti smanjiti dimenzionalnost prostora značajki. Redukciju dimenzionalnosti vektora značajki postižemo uporabom različitih transformacija uz minimalni gubitak informacija.

Klasifikaciju samog uzorka temeljimo na decizijskim funkcijama, te se susrećemo s problemom određivanja optimalne decizijske procedure. Problem klasifikacije može se promatrati kao razvrstavanje nepoznatog uzorka u potprostor prostora značajki na temelju decizijskih granica koje definiraju te procedure.

Pravilo razvrstavanja

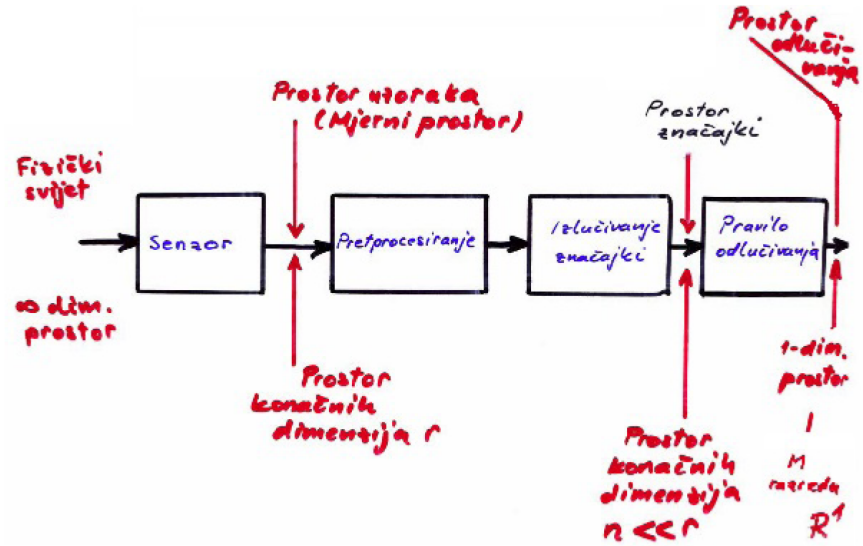
Decizijske granice određene su decizijskim funkcijama: $d_1(\vec{x}), d_2(\vec{x}), \ldots, d_n(\vec{x})$. Važno je zapamtiti da je d_i *funkcija koja ima za argument* ***vektor*** *a vraća* ***skalar***. Sada možemo definirati i **pravilo razvrstavanja:**

Ako $d_i(\vec{x}) > d_j(\vec{x})$ za $i, j = 1, 2, \ldots, M$ te $i \neq j$ tada nepoznati uzorak $\vec{x}$ pripada razredu ω_i



Slika 2.4: Blok dijagram klasifikatora

Vratimo se još jednom osnovnim fazama u postupku sustava za raspoznavanje uzoraka.



Slika 2.5: Osnove faze u postupku oblikovanja sustava RU

Primjetimo slijedeće stvari. Vanjski fizički svijet je tzv. *analogni svijet*, i sadrži praktički ∞ mnogo značajki ($\Re^\infty$). Senzor ili pretvarač pretvara analogni svijet u zapis koji sadrži r vrijednosti ($\Re^r$). Postupkom *pretprocesiranja* izlučujemo šum i poboljšavamo mjerna svojstva podataka. *Izlučivanjem značajki* taj prostor dodatno smanjujemo na prostor s n vrijednosti ($\Re^r$, $n << r$), te dobivamo konačni vektor značajki $\vec{x} = [x_1, x_2, \ldots, x_n]^T$. Donošenjem odluke *pravilom odlučivanja* dolazimo u **prostor odlučivanja** sa samo jednom dimenzijom ($\Re^1$) - *odlukom.*

2.3 Case study: automatska dijagnoza EKG signala na testu opterećenja

Pogledajmo slijedeći primjer iz biomedicinske domene. Promatrat ćemo signal EKG-a pacijenata na testu opterećenja, tj. provodit ćemo automatizaciju *Lesterove dijagnostičke metode.* Definirajmo prvo osnovne pojmove.

Područje uporabe $\mathcal{PU}$ je *automatska dijagnoza signala EKG-a u testu opterećenja.*

Objekti iz skupa objekata $\Theta_B = \{\sigma_k; k = 1, 2, \ldots, n\}$ su σ_i = *signal EKG-a pacijenata.*

Koristit ćemo jednostavan uzorak koji će se sastojati od dvije značajke:

$$f_k(\vec{x}) \longmapsto \vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Moramo još definirati razred objekata, skup oznaka te razred uzoraka. Razred objekata sastoji nam se od tri oznake:

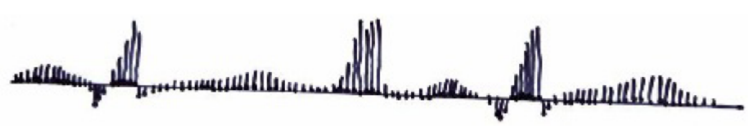
- Normalni EKG
- Granični EKG
- Nenormalni EKG

Skup oznaka sastoji se od oznaka istoimenih objekata iz razred objekata. Razred uzoraka sastoji se od slijedećih klasa:

- C_1 / ω_1 - normalni EKG
- C_2 / ω_2 - granični EKG
- C_3 / ω_3 - nenormalni EKG

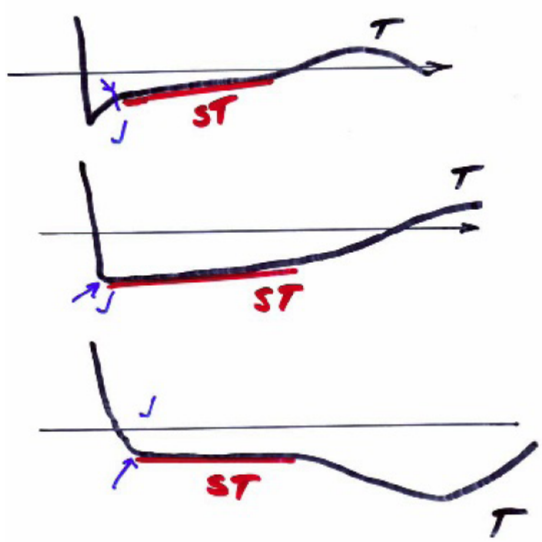
EKG signal iz fizičkog svijeta prikazan je na slici 2.6., a izmjereni (uzorkovani) signal dobiven A/D konverzijom prikazan je na slici 2.7.

Slika 2.6: Signal iz fizičkog svijeta



Slika 2.7: Uzorkovani signal ($f_u = 800Hz$)

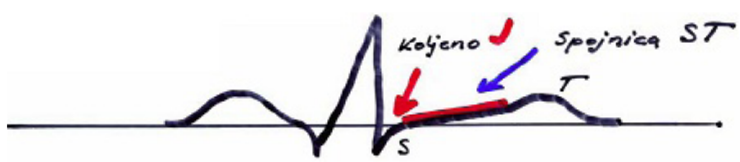
U procesu pretprocesiranja prije svega želimo eliminirati *ekstrasistolu* jer to predstavlje smetnju u radu srca koja nam ne donosi puno informacije, te utvrditi broj otkucaja srca u minuti. Zatim, želimo detektirati stavilne točke, tj.



Slika 2.9: Primjeri kliničkih oblika koljena J i nagiba segmenata ST

minimalnu i maksimalnu derivaciju, te filtrirati usrednjavanjem pri čemu eliminiramo šum za faktor $\sqrt{n}$, gdje je n broj analiziranih perioda.

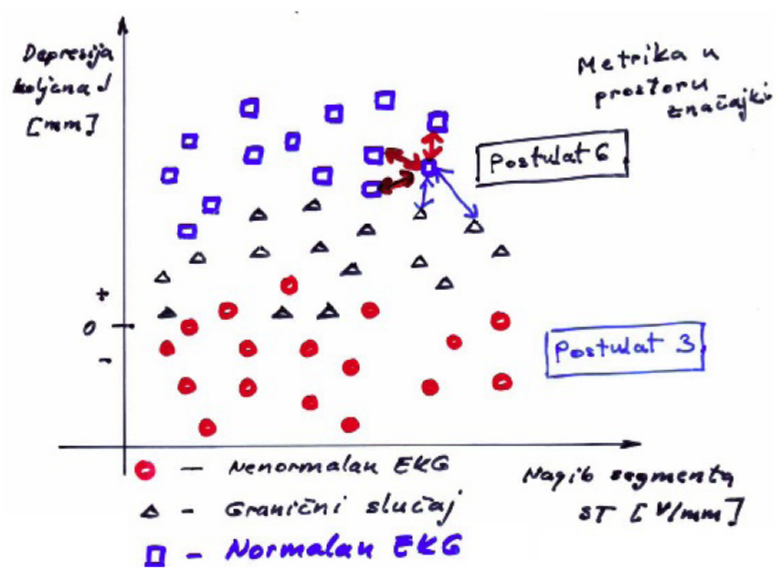
Značajke izlučujemo u skladu s Lesterovom dijagnostičkom metodom, te su nam značajke **koljeno *J*** i **spojnica *ST***. Značajku x_1 nam predtavlja nagib segmenta ST, a značajka x_2 je depresija koljena J.



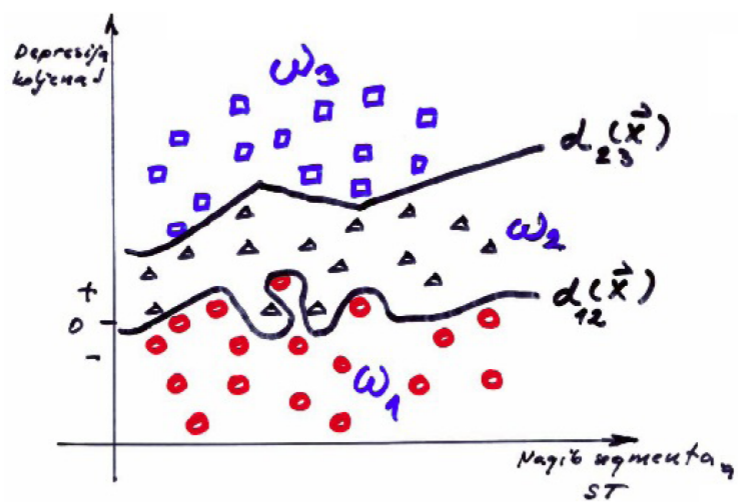
Slika 2.8: Prikaz značajki

Prilikom ekstrakcije značajki susrećemo se s dva problema. Prvi problem je što koljeno J ima različite oblike. Drugi problem je što se segment ST nalazi između kraja vala S i početka vala T te duljina segmenta ST ovisi o frekvenciji signala EKG-a, tj. broju otkucaja u minuti.

Tako dobiven prostor značajki prikazan je na slici 2.10., a primjer decizijskih funkcija za zadani prostor značajki prikazan je na slici 2.11.



Slika 2.10: Prostor značajki



Slika 2.11: Prostor značajki razdvojen decizijskim funkcijama

Poglavlje 3

Linearne funkcije odlučivanja

Funkcije odlučivanja

Prisjetimo se, osnovni zadatak sustava za raspoznavanje uzoraka je razvrstavanja uzoraka u razrede uzoraka. To izvršavamo pomoću funkcija koje dijele prostor uzoraka u područja tako da svako područje pripada samo jednom razredu uzoraka. Takve funkcije zovemo **funkcije odlučivanja** (eng. *decision functions, discriminant functions*).

Ako imamo N razreda uzoraka, moramo imati i jednako toliko funkcija odlučivanja $d_k(\vec{x}),\ k = 1, 2, \ldots, N$, tako da vrijedi:

$$d_i(\vec{x}) > d_j(\vec{x}),\ \forall \vec{x} \in \omega_i,\ i \neq j$$

To nam govori da funkcija odlučivanja $d_i(\vec{x})$ ima najveću vrijednost na području kojem pripadaju uzorci iz razreda ω_i. Za točke koje se nalaze na granici prostora ω_i i ω_j vrijedi $d_i(\vec{x}) = d_j(\vec{x})$. Prema tome zaključujemo da će nam za razdvajanje N razreda $\left[N, \frac{N(N-1)}{2}\right]$ granica. Treba primjetiti da izbor funkcije odlučivanja nije jedinstven te da se funkcije odlučivanja mogu zapisati u različitim oblicima.

3.1 Linearne funkcije odlučivanja

Linearna funkcija odlučivanja

Najjednostavniji oblik funkcije odlučivanja predstavlja **linearna funkcija**:

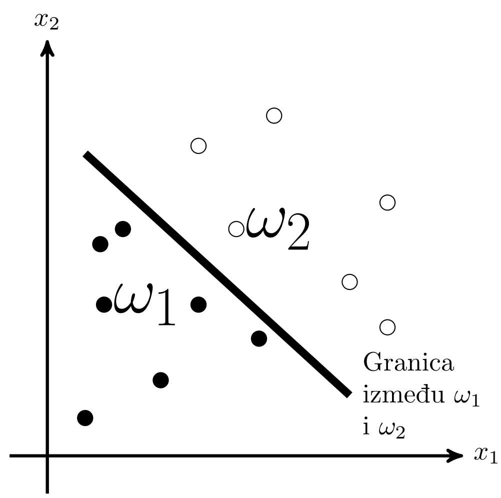
$$d(\vec{x}) = w_1 x_1 + w_2 x_2 + \ldots + w_n x_n + w_{n+1} = \vec{\mathbf{w}}^T \vec{\mathbf{x}} + w_{n+1}$$

Koeficijent praga

gdje je $\vec{\mathbf{w}}$ **vektor težinskih koeficijenata**, $\vec{x}$ **uzorak** a w_{n+1} **koeficijent praga** ili **pomak** (eng. *bias, threshold weight*).

Dimenzionalnost funkcije

Dimenzionalnost funkcije odlučivanja ovisi o broju varijabli koje čine ulazni prostor podataka. Za $n = 2$ funkcija odlučivanja je *pravac*, za $n = 3$ *ravnina*, a za $n > 3$ funkcija odlučivanja je *hiperravnina*.



Slika 3.1: Primjer prostora značajki razdvojenog linearnom funkcijom odlučivanja za vektor značajki $\vec{x}$

Linearne funkcije odlučivanja imaju nekoliko prednosti. Velika prednost ovih funkcija je to što ne ovise *statističkoj distribuciji* uzoraka po razredima. Iako često linearne funkcije *nisu* optimalne za rješavanje zadanog problema, često želimo *žrtvovati* dio performanse sustava zbog njihove jednostavnosti. Na kraju, linearne funkcije odlučivanja su jednostavne za izračunavanje te ako informacije o skupu ne sugeriraju drugačije, idealni su kandidati za inicijalne klasifikatore.

Pogledajmo prvo slučaj kada imamo dva razreda ($n = 2$).

3.1.1 Slučaj dva razreda

Za slučaj kada imamo samo dvije klase u 2D prostoru, funkcija odlučivanja izgleda:

$$d(\vec{x}) = w_1 x_1 + w_2 x_2 + w_3 = 0$$

$$d(\vec{x}) = \vec{\mathbf{w}}^T \vec{\mathbf{x}} + w_3 = 0$$

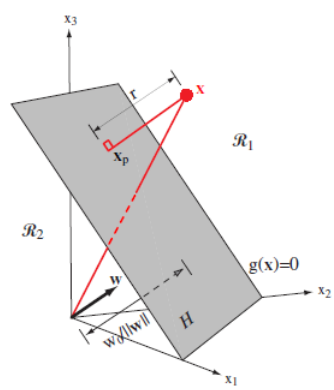
Pravilo odluke

Linearni klasifikator za dvije klase implementira slijedeće *pravilo odluke*:

$\vec{x}$ pripada klasi ω_1 ako $d(\vec{x}) > 0$, a ako je $d(\vec{x}) < 0$ $\vec{x}$ pripada klasi ω_2.

Prema definiranom pravilu, $\vec{x}$ će pripadati klasi ω_1 ako je skalarni produkt $\vec{\mathbf{w}}^T\vec{\mathbf{x}}$ veći od praga $-w_3$, inače pripada klasi ω_2. U slučaju $d(\vec{\mathbf{x}}) = 0$, $\vec{\mathbf{x}}$ može pripadati bilo kojoj od klasa, no mi ćemo se držati konvencije da je $\vec{x}$ u tom slučaju nedefiniran.

Pogledajmo geometrijsku interpretaciju ovog pravila.



Slika 3.2: Geometrijska interpretacija hiperravnine funkcije odlučivanja

Površina odluke

Jednadžba $d(\vec{x}) = 0$ definira **površinu odluke** koja razdvaja točke koje pripadaju razredu ω_1 od onih koje pripadaju razredu ω_2. Još jednom, ako je $d(\vec{\mathbf{x}})$ linearan, onda je površina odluke *hiperravnina*. Ako točke $\mathbf{x_1}$ i $\mathbf{x_2}$ obje leže na površini odluke, tada vrijedi:

$$\vec{\mathbf{w}}^T\mathbf{x_1} + w_0 = \vec{\mathbf{w}}^T\mathbf{x_2} + w_0$$

$$\vec{\mathbf{w}}^T(\mathbf{x_1} - \mathbf{x_2}) = 0,$$

iz čega slijedi da je $\vec{\mathbf{w}}$ normala na svaki vektor koji leži u hiperravnini. Općenito, hiperravnina H dijeli prostor značajki na dvije *poluravnine*, regiju R_1 za razred ω_1 te R_2 za razred ω_2. Iz toga slijedi da kada je $d(\vec{\mathbf{x}}) > 0$ i $\vec{\mathbf{x}}$ pripada regiji R_1, tada je vektor $\vec{\mathbf{w}}$ usmjeren prema regiji R_1. Često se kaže da je svaki $\vec{x}$ u R_1 na *pozitivnoj* strani H, a da je svaki $\vec{x}$ u R_2 na *negativnoj* strani H.

Nadalje, ako je $\mathbf{x}$ točka na hiperravnini, onda je $d(\vec{\mathbf{x}}) = 0$ i vrijedi:

$$\vec{\mathbf{w}}^T\mathbf{x} + w_3 = 0 \Longrightarrow \frac{\vec{\mathbf{w}}^T\mathbf{x}}{\|\vec{\mathbf{w}}\|} = -\frac{w_3}{\|\vec{\mathbf{w}}\|}.$$

Udaljenost hiperravnine od ishodišta

Vrijednost $\frac{\vec{\mathbf{w}}^T\mathbf{x}}{\|\vec{\mathbf{w}}\|}$ je skalarna projekcija vektora $\mathbf{x}$ na jedinični vektor $\frac{\vec{\mathbf{w}}}{\|\vec{\mathbf{w}}\|}$ i odgovara **udaljenosti** ravnine od ishodišta, odnosno udaljensot ravnine do ishodišta iznosi $d = \frac{w_{n+1}}{\|\mathbf{w}\|}$. Vidimo da parametar ω_3 određuje položaj hiperravnine u prostoru.

Orijentacija hiperravnine

Što još možemo saznati iz gornje relacije? Faktor $\frac{\vec{\mathbf{w}}}{\|\vec{\mathbf{w}}\|}$ je jedinični vektor s koeficijentima koje čine ravninu te na taj način određuje **orijentaciju** hiperravnine. Ako je neka komponenta jednaka 0 onda je hiperravnina paralelna s odgovarajućom koordinatnom osi.

Udaljenost točke od hiperravnine

Možemo još pokazati da je $d(\mathbf{x})$ proporcinalan udaljenosti d točke $\mathbf{x}$ od hiperravnine. Neka je $\mathbf{x}$ proizvoljno odabrana točka i neka je $\mathbf{x}_\perp$ njezina ortogonalna projekcija na hiperravninu. Tada vrijedi:

$$\mathbf{x} = \mathbf{x}_\perp + d\frac{\mathbf{w}}{\|\mathbf{w}\|}.$$

Ako pomnožimo obje strane s $\mathbf{w}^T$ te dodamo w_3 dobivamo

$$\mathbf{w}^T\mathbf{x} + w_3 = \mathbf{w}^T\mathbf{x}_\perp + w_3 + d\frac{\mathbf{w}^T\mathbf{w}}{\|\mathbf{w}\|}.$$

Primjetimo da je izraz $\mathbf{w}^T\mathbf{x} + w_3 = d(\mathbf{x})$, a $\mathbf{w}^T\mathbf{x}_\perp + w_3 = 0$. Iz toga jednostavno slijedi:

$$d = \frac{d(\mathbf{x})}{\|\mathbf{w}\|},$$

gdje d predstavlja udaljenost točke $\mathbf{x}$ od hiperravnine.

Iz gore navedena tri pravila, možemo lako uočiti neka svojstva hiperravnine. Promatranjem vektora težinksih faktora $\mathbf{w}$ moguće je utvrditi je li hiperravnina paralelna nekoj od koordinatnih osi, te ako je $w_{n+1} = 0$ onda hiperravnina prolazi kroz ishodište koordinatnog sustava.

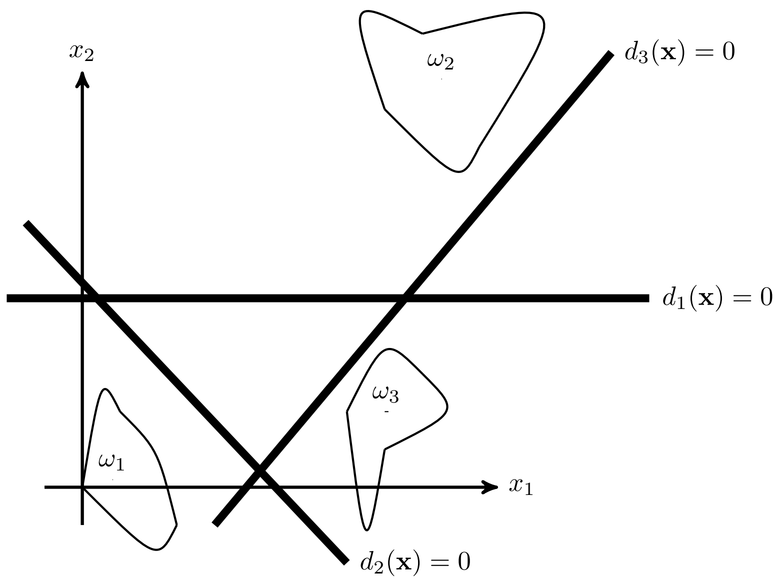
3.1.2 Slučaj više razreda

Višeklasna klasifikacija

Postoji više pristupa rješavanju višeklasne klasifikacije ($M > 2$) linearnim funkcijama odlučivanja.
Pri no što se upoznamo s različitim pristupima rješavanju ovog problema, pogledajmo s kakvim se situacijama sve možemo susresti:

1. slučaj: Svaki je razred uzoraka separabilan od ostalih razreda jednom decizijskom ravninom

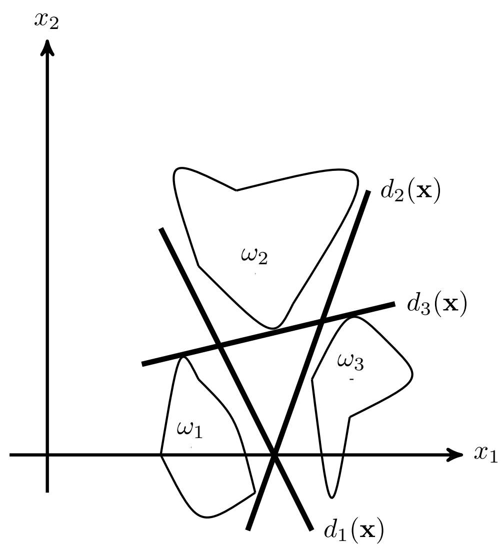
$$d_i(\mathbf{x}) = \mathbf{w}_i^T\mathbf{x} = \begin{cases} > 0 & \mathbf{x} \in \omega_1 \\ < 0 & \mathbf{x} \notin \omega_1 \end{cases}$$



2. slučaj: Svaki razred uzoraka je separabilan sa svakim pojedinim razredom i to jednom decizijskom ravninom, tj. razredi su separabilni po parovima. Za razdvajanje nam je potrebno $\frac{M(M-1)}{2}$ decizijskih ravnina.

Strategija rješavanja ovog slučaja je upotribiti $\frac{c(c-1)}{2}$ linearnih funkcija odlučivanja tak oda sa svakom funkcijom odvojimo par razreda. Granica između razreda ω_i i ω_j je zadana s:

$$d_{ij}(\mathbf{x}) = w_{ij1}x_1 + w_{ij2}x_2 + \ldots + w_{ijn}x_n + w_{ijn+1} = 0$$



3. slučaj: Postoji M decizijskih funkcija $d_k(\mathbf{x}) = \mathbf{w}_k^T\mathbf{x},\ k = 1, 2, \ldots, M$ sa svojstvom da $\mathbf{x}$ pripada razredu ω_i ako vrijedi

$$d_i(\mathbf{x}) > d_j(\mathbf{x})\ \forall i \neq j.$$

To je poseban slučaj **2. slučaja** zato što možemo definirati slijedeće:

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

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

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

ako je $d_i(\mathbf{x}) > d_j(\mathbf{x})\ \forall i \neq i$, tada je $d_{ij}(\mathbf{x}) > 0\ \forall j \neq i$, što znači da ako su srazredi separabilni za 3. slučaj, onda su automatski separabilni za 2.slučaj.

Klasifikator za ovaj slučaj je ravnina:

$$(\mathbf{w}_i - \mathbf{w}_j)^T\mathbf{x} + (w_{in+1} - w_{jn+1}) = 0.$$

Linearni klasifikator

Dobiveni klasifikator nazivamo **linearni klasifikator**.

3.2 Određivanje funkcije odlučivanja

Problem oblikovanja linearnog klasifikatora svodi se na određivanje koeficijenata linearne funkcije odlučivanja:

$$\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix},$$

te koeficijenta w_{n+1}. *Automatizaciju* postupka određivanja koeficijenata linearne funkcije odlučivanja provest ćemo iterativnim postupkom **učenja** koeficijenata linearne funkcije odlučivanja uporabom uzoraka iz skupa za učenje (eng. *training set*).

Postupak učenja

Skup za učenje sastoji se od N uzoraka $\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_n}$ razvrstanih u dva razreda ω_1 i ω_2. Vektori uzoraka $\mathbf{x_i}$ su **označeni** vektori (njihovu pripadnost razreda znamo) pa ćemo njih **upotrijebiti za učenje** $d(\mathbf{x})$.

Kako bismo olakšali postupak učenja, povećat ćemo dimenzionalnost vektora $\mathbf{w}$ za jedan:

$$\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \\ 1 \end{bmatrix}$$

radi elegantnijeg zapisa $d(\mathbf{x}) = \mathbf{w}^T\mathbf{x}$. Uspijemo li odrediti takav vektor težinskih koeficijenata $\mathbf{x}$ tako da s pomoću funkcije $f(\mathbf{x})$ pravilno razvrstamo sve uzorke iz skupa za učenje, kažemo da su ω_1 i ω_2 **linearno razdvojivi**.

Linearno razdvojivi razredi

Kažemo da je uzorak $\mathbf{x}$ pravilno razvrstan ako za sve $\mathbf{x}$ iz ω_1 vrijedi $\mathbf{w}^T\mathbf{x} > 0$ i ako za svaki $\mathbf{x}$ iz ω_2 vrijedi $\mathbf{w}^T\mathbf{x} < 0$. Gornje pravilo možemo zapisati kao jedinstven uvjet $\mathbf{w}^T\mathbf{x} > 0$ ako sve uzorke iz ω_2 pomnožimo s -1.

To nas vodi na redefiniciju početnog problema: Tražimo vektor koeficijenata $\mathbf{w}$ linearne funkcije odlučivanja tako da vrijedi

$$\mathbf{w}^T\mathbf{x} > 0$$

za svaki uzorak $\mathbf{x}$ iz skupa uzoraka za učenje, odnosno:

$$[\mathbf{x}]\mathbf{w} > 0 \qquad \forall \mathbf{x}$$

$$[\mathbf{x}] = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_n^T \end{bmatrix}$$

pri čemu je $[\mathbf{x}]$ matrica svih uzoraka iz skupa za učenje, s tim da su uzorci iz ω_2 pomnoženi s -1.

Primjer:

Zadani su slijedeći uzorci

$$\mathbf{x_1} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \mathbf{x_2} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

$$\mathbf{x_3} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad \mathbf{x_4} = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

pri čemu je $\mathbf{x_1} \in \omega_1,\ \mathbf{x_2} \in \omega_1$, a $\mathbf{x_3} \in \omega_2,\ \mathbf{x_4} \in \omega_2$.

Prvo što trebamo napraviti je povećati dimenzionalnost vektora:

$$\mathbf{x_1} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \qquad \mathbf{x_2} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

$$\mathbf{x_3} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \qquad \mathbf{x_4} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix},$$

Zatim, pomnožimo sve elemente razreda ω_2 s -1:

$$\mathbf{x_1} = \begin{bmatrix} -1 \\ 0 \\ -1 \end{bmatrix} \qquad \mathbf{x_2} = \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix}$$

Nakon toga, formirajmo matricu $[\mathbf{x}]$:

$$[\mathbf{x}] = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ -1 & 0 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

Nakon toga, potrebno je još samo riješiti slijedeći sustav nejednadžbi:

$$[\mathbf{x}] \cdot \mathbf{w} > 0$$

$$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ -1 & 0 & -1 \\ -1 & -1 & -1 \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} > \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Razdvojni vektor

Vektor **w** koji zadovoljava sustav linearnih nejednadžbi

$$[\mathbf{x}]\mathbf{w} > 0$$

nazivamo **razdvojni vektor**.

3.2.1 Gradijentni postupci određivanja razdvojnog vektora

Kriterijska funkcija

Problem pronalaska razdvojnog vektora često nije jednostavan te ga nije moguće riješiti analitički. Da bismo našli hiperravninu koja razdvaja zadane klase, moramo imati informaciju koliko je dobro naše trenutno rješenje te u kojem smjeru se naše rješenje poboljšava. Definirajmo kriterijsku funkciju $J(\mathbf{a})$ koja je minimizirana ako je **a** razdvojni vektor. Smjer poboljšanja u tom nam slučaju govori **gradijent funkcije**.

Gradijent funkcije više varijabli, $f(\mathbf{y})$, gdje je

$$\mathbf{y} = [x_1, x_2, \ldots, x_n]^T$$

definiran je kao

$$\mathbf{grad} f(\mathbf{y}) = \frac{df(\mathbf{y})}{d\mathbf{y}} = \begin{bmatrix} \frac{\partial f}{\partial y_1} \\ \frac{\partial f}{\partial y_2} \\ \vdots \\ \frac{\partial f}{\partial y_n} \end{bmatrix}.$$

Primjetimo da je gradijent skalarne funkcije vektorskog argumenta **vektor**. Svaka komponenta promjene gradijenta predstavlja veličinu promjene funkcije u smjeru komponente vektora te pri tome vrijedi:

- povećanje argumenata u smjeru pozitivnog gradijenta funkcije f dovodi do **maksimuma** funkcije f
- povećanje argumenata u smjeru negativnog gradiejnta funkcije f dovodi do **minimuma** funkcije f

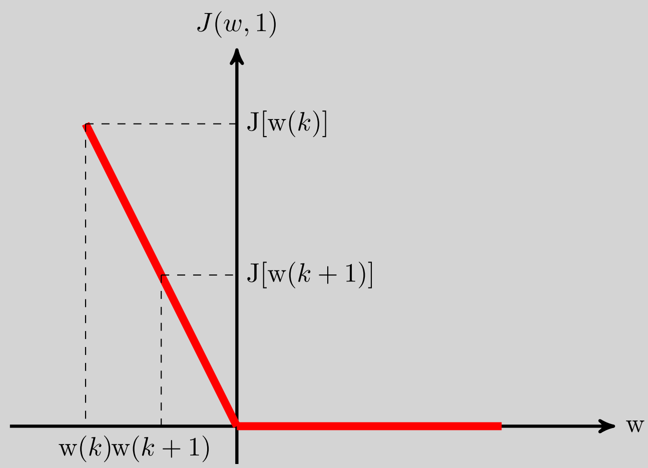
Primjer:

Uzmimo za primjer funkciju $J(w, 1) = (|w| - w)$.
Derivacije te funkcije je

$$\frac{\partial J(w,1)}{\partial w} = \mathbf{sgn}(w) - 1$$

gdje je

$$\mathbf{sgn}(x) = \begin{cases} 1, & za\ x > 0 \\ -1, & za\ x < 0 \end{cases}$$



Gradijent funkcije $J(w, 1)$ jednaka je

$$\frac{\partial J(w,1)}{\partial w} = \begin{cases} -2, & w \leq 0 \\ 0, & w > 0 \end{cases}$$

Ako promotrimo kretanje gradijenta funkcije $J(w, 1)$, vidimo da se pri kretanju $w(k) \mapsto w(k+1)$ povećavamo argument u smjeru negativne vrijednosti gradijenta, tj. krećemo se prema minimumu funkcije.

Osnovna zamisao je odabrati takvu funkciju koja će dostići minimum kad je ispunjen uvjet :

$$\mathbf{w}^T \mathbf{x}_i > 0, \qquad \forall i = 1, 2, \ldots, N.$$

Pri tome zahtjevamo da izabrana funkcija ima samo jedan minimum te da je funkcija $J(\mathbf{a})$ funkcija vektorskog argumenta $\mathbf{w}$.

Gradijentni postupak

Postupak pronalaženja minimuma funkcije pomoću gradijenta radimo tako da korak po korak povećavamo argumenta $\mathbf{w}$ u smjeru negativnog argumenta funkcije $J(\mathbf{w}, \mathbf{x})$ sve dok nije postignut minimum kriterijske funkcije $J(\mathbf{w}, \mathbf{x})$.

Ako je $\mathbf{w}(k)$ vrijednost $\mathbf{w}$ u k-tom koraku, tada vrijednost u koraku $(k+1)$ definiramo kao

$$\mathbf{w}(k+1) = \mathbf{w}(k) - c \cdot \left\{ \frac{\partial J(\mathbf{w}(k), \mathbf{x})}{\partial \mathbf{w}(k)} \right\},$$

gdje je c pozitivna konstanta različita od 0 koja određuje veličinu korekcije.

Korekcija se više ne izvodi kada je $\frac{\partial J(\mathbf{w}(k),\mathbf{x})}{\partial \mathbf{w}(k)} = 0$, što je i uvjet za minimum.

Podsjetnik o deriviranju funkcije

Funkcija	**Derivacija**
c (konst.)	0
x^n	nx^{n-1}
$\frac{1}{x}$	$-\frac{1}{x^2}$
$\frac{1}{x^n}$	$-\frac{n}{x^{n+1}}$
$\sqrt{x}$	$\frac{1}{2\sqrt{x}}$
$\sqrt[n]{x}$	$\frac{1}{n\sqrt[n]{x^{n-1}}}$
e^x	e^x
a^x	$a^x \mathbf{ln} a$
$\mathbf{ln} x$	$\frac{1}{x}$

Deriviranje linearnih funkcija

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

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

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

$$\frac{d}{d\mathbf{X}}(\mathbf{a}^T\mathbf{X}\mathbf{b}) = \mathbf{a}\mathbf{b}^T$$

$$\frac{d}{d\mathbf{X}}(\mathbf{a}^T\mathbf{X}\mathbf{a}) = \frac{d}{d\mathbf{X}}(\mathbf{a}^T\mathbf{X}^T\mathbf{a}) = \mathbf{a}\mathbf{a}^T$$

$$\frac{d}{d\mathbf{x}}(\mathbf{x}^T\mathbf{C}\mathbf{x}) = (\mathbf{C} + \mathbf{C}^T)\mathbf{x}$$

ako je $\mathbf{C} = \mathbf{C}^T$ onda je $\frac{d}{d\mathbf{x}}(\mathbf{x}^T\mathbf{C}\mathbf{x}) = 2\mathbf{C}\mathbf{x}$

$$\frac{d}{d\mathbf{x}}(\mathbf{x}^T\mathbf{x}) = 2\mathbf{x}$$

$$\frac{d}{d\mathbf{x}}(\mathbf{A}\mathbf{x} + \mathbf{b})(\mathbf{D}\mathbf{x} + \mathbf{e}) = \mathbf{A}^T(\mathbf{D}\mathbf{x} + \mathbf{e}) + \mathbf{D}^T(\mathbf{A}\mathbf{x} + \mathbf{b})$$

Nekoliko pravila

3.2.2 Perceptron

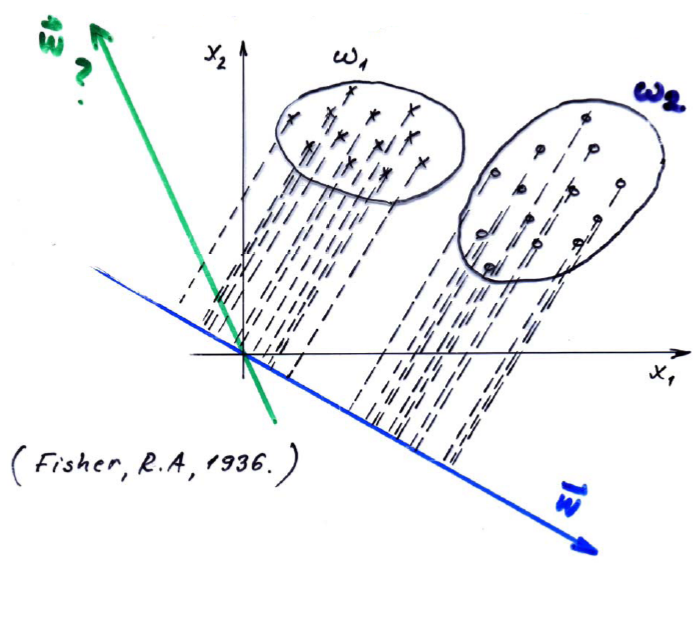
3.2.3 Ho-nešto algoritam

Poglavlje 4

Fisherova linearna diskriminanta

Fisherova analoza

Fisherova linearna diskriminanta je još jedan pristup linearnoj klasifikaciji koji se temelji na ideji da se početni d-dimenzionalan vektor značajki $\mathbf{x}$ reducira na jednu dimenziju ($\Re^n \mapsto \Re^1$) te ga tada upotrijebiti za klasifikaciju.

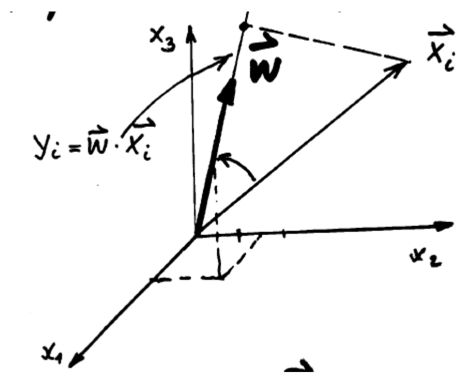


Cilj Fisherove linearne diskriminante je pronaći orijentaciju pravca na koji se projiciraju $d-$dimenzionalni uzorci $\mathbf{x}_i, i = 1, 2, \ldots, n$ ali tako da su projicirani uzorci separabilni. Upravo to je cilja *klasične* diskriminantne analize.

Za skup od n d-dimenzionalnih uzoraka, od kojih n_1 uzoraka čine podskup $\mathcal{D}_1$ i n_2 uzoraka koji čine podskup $\mathcal{D}_2$, tvorimo linearnu kombinaciju komponenti $\mathbf{x}$:

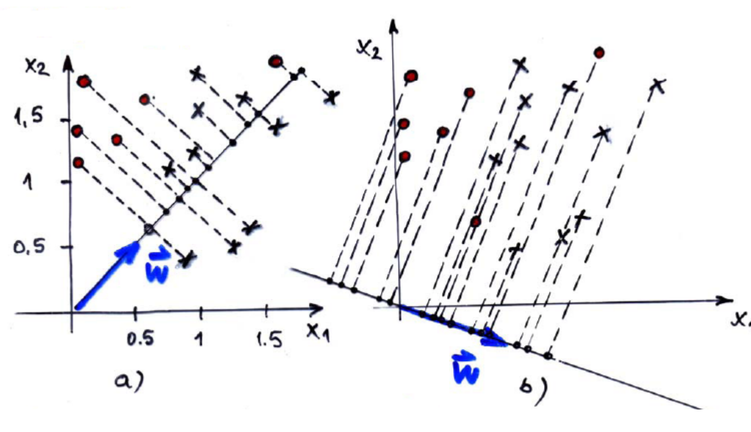
$$y_i = \mathbf{w}^T \mathbf{x}_i, \quad i = 1, 2, \ldots, n$$

Tako dobivamo odgovarajući skup od n uzoraka $y_1, y_2, \ldots, y_n$ koji su podjeljeni u podskupove $\mathcal{Y}_1$ i $\mathcal{Y}_2$. Ako je pri tome $\|\mathbf{w}\| = 1$, svaki y_i je projekcija odgovarajućeg vektora $\mathbf{x}_i$ na pravac u smjeru $\mathbf{w}$.



Veličina vektora $\mathbf{w}$ nema posebno značenje, ona samo skalira y_i. Važan je smjer od $\mathbf{w}$!

Ako zamislimo da uzorci iz ω_1 čine nakupinu, te da uzorci iz ω_2 čine drugu nakupinu, **želimo da projekcije na pravac određen s w budu takva da su dobro razdvojive**.



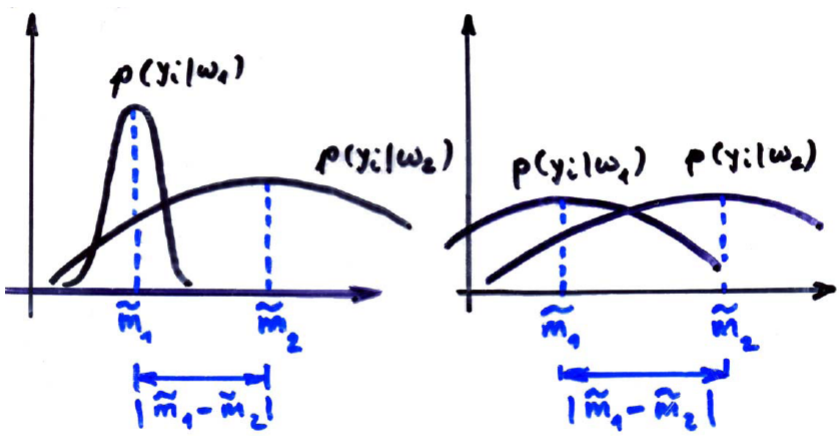
Uzmimo za mjeru odvojivosti između projiciranih točaka razliku srednjih vrijednosti projiciranih uzoraka. Neka je

$$\mathbf{m}_i = \frac{1}{n_i} \sum_{\mathbf{x} \in \mathcal{D}_i} \mathbf{x}; \quad i = 1, 2$$

Srednja vrijednosti projiciranih uzoraka:

$$\widetilde{m}_i = \frac{1}{n_i} \sum_{y \in \mathcal{Y}_i} y$$

Da li je udaljenost srednjoh vrijednosti projiciranih uzoraka *dobra mjera*?



Neka je $\widetilde{m}_i$ projekcija vektora $\mathbf{m}_i$:

$$\widetilde{m}_i = \frac{1}{n_i} \sum_{\mathbf{x} \in \mathcal{D}_i} \mathbf{w}^T \mathbf{x} = \mathbf{w}\mathbf{m}_i$$

tada je razlika između srednjih vrijednosti projiciranih uzoraka:

$$|\widetilde{m}_1 - \widetilde{m}_2| = |\mathbf{w}(\mathbf{m}_1 - \mathbf{m}_2)|$$

Primjetimo da razliku možemo učiniti proizvoljno velikom skaliranjem $\mathbf{w}$. da bismo dobili dobro odvajanje projiciranih podataka želimo da **udaljenost** između srednjih vrijednosti projiciranih točaka bude relativno velika u odnosu na **neku mjeru standardne devijacije za svaki razred.**

Sada kriterijsku funkciju možemo definirati kao

$$J(\mathbf{w}) = \frac{(razlika\ srednjih\ vrijednosti)^2}{varijanca\ uzoraka\ unutar\ razred} = \frac{(\widetilde{m}_1 - \widetilde{m}_2)^2}{\sigma^2_{\mathcal{Y}_1} + \sigma^2_{\mathcal{Y}_2}}.$$

Fisherova linearna diskriminanta određuje da linearna funkcija

$$\mathbf{w}^T \mathbf{x}$$

za koju je kriterijska funkcija $J(\mathbf{w})$ **maksimalna** vodi najboljem razdvajanju između projiciranih skupova.

Umjesto varijanci σ_i možemo definirati raspršenost projiciranih podataka unutar razreda C_i kao:

$$\widetilde{\mathbf{s}}_i^2 = \sum_{y \in \mathcal{Y}_i} (y - \widetilde{\mathbf{m}}_i)^2.$$

Tada je $\frac{1}{n}(\widetilde{\mathbf{s}}_1^2 + \widetilde{\mathbf{s}}_2^2)$ procjena varijance podataka a $\widetilde{\mathbf{s}}_1^2 + \widetilde{\mathbf{s}}_2^2$ je ukupna mjera raspršenosti projiciranih podataka unutar razreda (eng. *total within-class scatter*).

Kako želimo $J(\cdot)$ dobiti kao funkciju od vektora $\mathbf{w}$, definirajmo matricu:

$$\mathcal{S}_i = \sum_{\mathbf{x} \in \mathcal{D}_i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T$$

kao raspršenost unutar razreda ω_i. Dalje:

$$\mathcal{S}_W = \mathcal{S}_1 + \mathcal{S}_2$$

$$\widetilde{\mathbf{s}}_i^2 = \sum_{\mathbf{x} \in \mathcal{D}_i} (\mathbf{w}^T\mathbf{x} - \mathbf{w}^T\mathbf{m}_i)^2$$

$$= \sum_{\mathbf{x} \in \mathcal{D}_i} \mathbf{w}^T \underbrace{(\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T}_{\mathcal{S}_i} \mathbf{w}$$

$$\widetilde{\mathbf{s}}_i^2 = \mathbf{w}^T \mathcal{S}_i \mathbf{w}.$$

Suma $\widetilde{\mathbf{s}}_1^2 + \widetilde{\mathbf{s}}_2^2$ tada je jednaka

$$\mathbf{w}^T \mathcal{S}_w \mathbf{w}.$$

Matrica $\mathcal{S}_w$ simetrična je i pozitivno semidefinitna te je obično nesingularna ako je $n > d$.

Slično gornjemu, možemo definirati i slijedću vrijednost:

$$(\widetilde{\mathbf{m}}_1 - \widetilde{\mathbf{m}}_2)^2 = (\mathbf{w}^T\widetilde{\mathbf{m}}_1 - \mathbf{w}^T\widetilde{\mathbf{m}}_2)^2$$
$$= \mathbf{w}^T(\widetilde{\mathbf{m}}_1 - \widetilde{\mathbf{m}}_2)(\widetilde{\mathbf{m}}_1 - \widetilde{\mathbf{m}}_2)^T\mathbf{w}$$
$$= \mathbf{w}^T \mathcal{S}_B \mathbf{w}.$$

Matricu $\mathcal{S}_B$ predstavlja raspršenost projiciranih uzoraka između klasa (eng. *between-class scatter matrix*). Za matricu $\mathcal{S}_B$ vrijedi sve što vrijedi i za matricu $\mathcal{S}_W$. Sada kriterijsku funkciju možemo definirati kao:

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \mathcal{S}_B \mathbf{w}}{\mathbf{w}^T \mathcal{S}_W \mathbf{w}}.$$

Za određivanje matrica $\mathcal{S}_B$ i $\mathcal{S}_W$ koristimo uzorke iz skupova $\mathcal{D}_1$ i $\mathcal{D}_2$.

Odredimo maksimum $J(\mathbf{w})$:

$$\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}} = 0$$

$$= \frac{\partial J(\mathbf{w})}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \left(\frac{\mathbf{w}^T \mathcal{S}_B \mathbf{w}}{\mathbf{w}^T \mathcal{S}_W \mathbf{w}} \right) = 0$$

$$\frac{(\mathbf{w}^T \mathcal{S}_W \mathbf{w}) \frac{\partial}{\partial \mathbf{w}} (\mathbf{w}^T \mathcal{S}_B \mathbf{w}) - (\mathbf{w}^T \mathcal{S}_B \mathbf{w}) \frac{\partial}{\partial \mathbf{w}} (\mathbf{w}^T \mathcal{S}_W \mathbf{w})}{(\mathbf{w}^T \mathcal{S}_W \mathbf{w})^2} = 0$$

$$(\mathbf{w}^T \mathcal{S}_W \mathbf{w}) \cdot 2\mathcal{S}_B \mathbf{w} - (\mathbf{w}^T \mathcal{S}_B \mathbf{w}) \cdot 2\mathcal{S}_W \mathbf{w} = 0$$

$$\mathcal{S}_w \mathbf{w} (\mathbf{w}^T \mathcal{S}_B \mathbf{w})(\mathbf{w}^T \mathcal{S}_W \mathbf{w})^{-1} = \mathcal{S}_B \mathbf{w}$$

Ako to zapišemo u obliku:

$$\lambda \mathcal{S}_W \mathbf{w} = \mathcal{S}_B \mathbf{w}$$

gdje je λ skalar iznosa $(\mathbf{w}^T \mathcal{S}_B \mathbf{w})(\mathbf{w}^T \mathcal{S}_W \mathbf{w})^{-1}$, vidimo da se naš problem svodi na **generalizirani problem svojstvenog vektora**.

Ako $\mathcal{S}_W^{-1}$ postoji onda je **smjer w** jednak

$$\hat{\mathbf{w}} = (\mathcal{S}_W^{-1} \mathcal{S}_B) \mathbf{w}.$$

Za naš slučaj nije potrebno to rješavati na način da tražimo svojstvene vrijednosti i svojstvene vektore za $\mathcal{S}_W^{-1} \mathcal{S}_B)\mathbf{w}$ zato što je $\mathcal{S}_B \mathbf{w}$ **uvijek usmjeren kao** $\mathbf{m}_1 - \mathbf{m}_2$**!**

Budući da nas zanima samo smjer vektora $\mathbf{w}$, a ne faktor skaliranja, možemo napisati rješenje za $\mathbf{w}$:

$$\mathbf{m}_1 = \begin{bmatrix} m_{11} \\ m_{12} \\ m_{13} \end{bmatrix} \quad \mathbf{m}_2 = \begin{bmatrix} m_{21} \\ m_{22} \\ m_{23} \end{bmatrix}$$

$$\mathcal{S}_B = (\mathbf{m}_1 - \mathbf{m}_2) \cdot (\mathbf{m}_1 - \mathbf{m}_2)^T.$$

Pokažimo da je $\mathcal{S}_b \mathbf{w}$ zaista usmjeren u smjeru $\mathbf{m}_1 - \mathbf{m}_2$:

$$\mathcal{S}_B \mathbf{w} = (\mathbf{m}_1 - \mathbf{m}_2) \cdot \underbrace{(\mathbf{m}_1 - \mathbf{m}_2)^T \mathbf{w}}_{k}$$

$$\mathcal{S}_B \mathbf{w} = (\mathbf{m}_1 - \mathbf{m}_2) \cdot \mathbf{w}$$

Iz toga vidimo da je $\mathcal{S}_B \mathbf{w}$ usmjeren u smjeru $\mathbf{m}_1 - \mathbf{m}_2$.

Iz toga slijedi da

$$\hat{\mathbf{w}} = \mathcal{S}_W (\mathbf{m}_1 - \mathbf{m}_2),$$

te da $\mathbf{w}$ u skladu s Fisherovom linearnog diskriminantom određuje linearnu funkciju koja maksimizira omjer između raspršenja između razreda i raspršenja unutar razreda.