1) Raspoznavanje temeljeno na primjerima

Osnovna zamisao: na temelju (nekoliko) primjera (ili možda u kombinaciji s nekoliko kontraprimjera) uzoraka koji pripadaju nekom razredu ω_i sustav je u stanju odrediti da li novi (nepoznati) uzorak pripada tom razredu ω_i.

Proces raspoznavanja može se podijeliti u dvije faze:

I. faza učenja
(tutorial stage, learning stage)

II. faza odlučivanja
(decision stage, decision stage)

UČENJE S UČITELJEM
(Supervised learning)

FAZA UČENJA

U postupku učenja s učiteljem imamo skup uzoraka za vježbanje ili učenje. Učitelj označuje pripadnost uzorka razredu (ili pridružuje cijenu koštanja svakom uzorku) i traži način redukcije ukupne cijena koštanja za sve te uzorke, odnosno traži oblik decizijske funkcije koja će uz najmanju pogrešku (ili bez pogreške) klasificirati uzorke iz skupa za učenje.

Pitanja:

1) Kako možemo biti sigurni da je određeni algoritam dovoljno djelotvoran i da može "naučiti" rješenje za zadani problem?
2) Koliko je stabilan na promjene parametara?
3) Da li konvergira u konačnom vremenu?

FAZA ODLUČIVANJA

Nakon što je sustav naučio decizijsku funkciju on se primjenjuje na stvarnim (neoznačenim, nepoznatim) uzorcima te se verificiraju rezultati raspoznavanja

2) Raspoznavanje temeljeno na strukturi (složenih) uzoraka

- uzorak opisan primitivima (terminalima) i odnosom između tih primitiva

primitivi, produkcijska pravila,

(terminali neterminali)

gramatika, jezik

$G=(V_N, V_T, P, S)$ podloga su za jezični (sintaksni) pristup raspoznavanju

- svakom razredu uzoraka pridružen jezik, odnosno stroj koji raspoznaje pojedinu gramatiku

3) Grupiranje (engl. Clustering)

- učenje bez učitelja (engl. Unsupervised learning)

U sustavu nema učitelja, barata se s neoznačenim uzorcima (ne zna se njihova pripadnost niti broj razreda).

Sustav oblikuje grupe ili "prirodno grupira" ulazne uzorke. Korisnik obično eksperimentira s različitim postupcima grupiranja (svaki od njih daje drugačije "prirodne" grupe).

Na temelju generiranih grupa korisnik potvrđuje ili odbacuje hipoteze.

Još jedan postupak učenja

Učenje s kritikom (engl. Reinforcement learning, Learning with a critic)

Tipični postupci učenja klasifikatora su takvi da im se na ulazu postavi uzorak, zatim sustav određuje (tentativni) razred. On se uspoređuje s znanom stvarnom pripadnosti razreda i to se koristi da bi se poboljšao klasifikator (npr. ako je pogrešna klasifikacija koriqira se decizijska funkcija proporcionalno "veličini" uzorka)

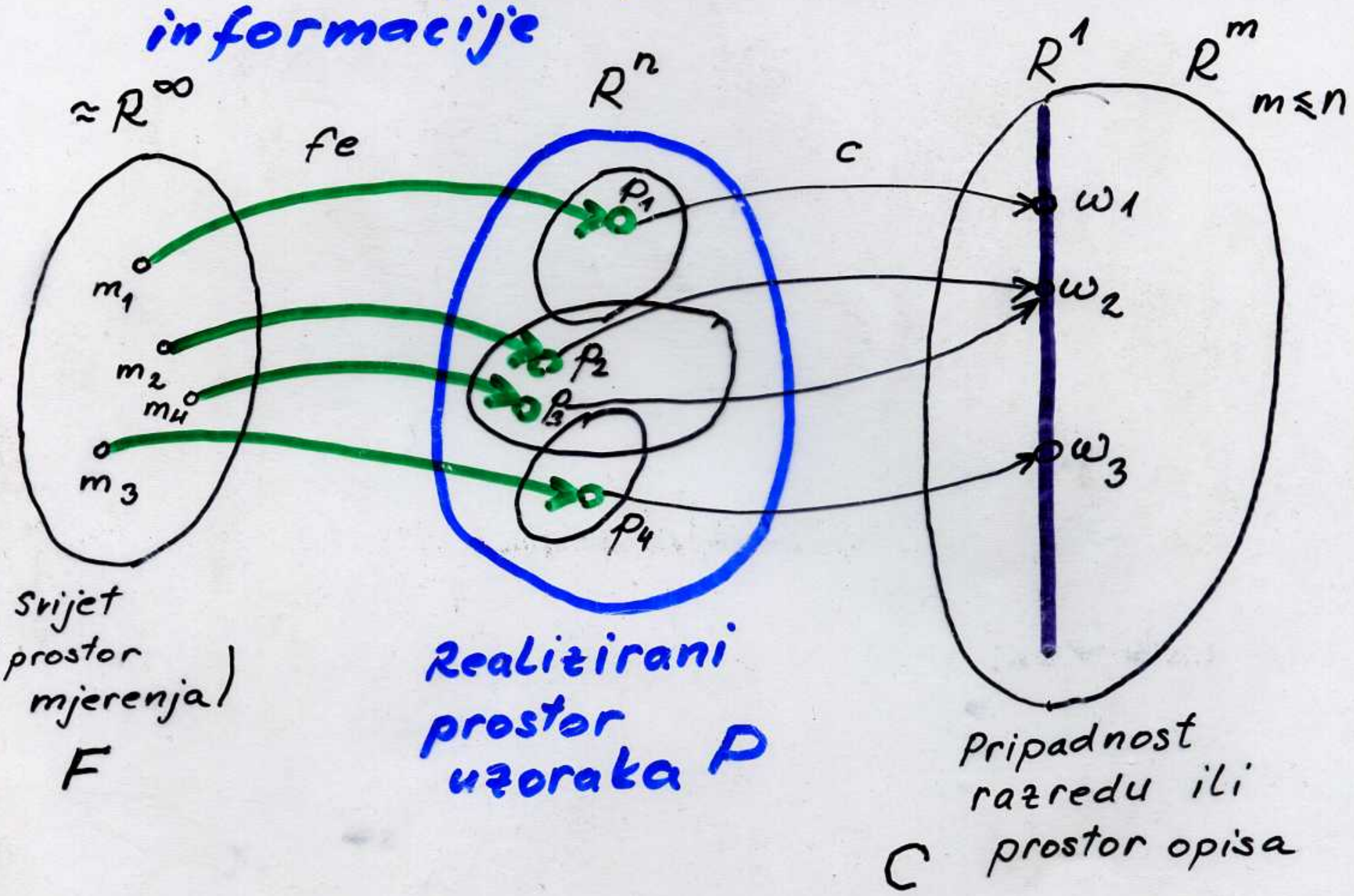
Učenje s kritikom ne dobiva signal o željenoj kategoriji, odnosno razredu, umjesto toga povratna korektivna veza nudi informaciju da li je nešto točno ili netočno.

NE SPECIFICIRA SE ZAŠTO, ODNOSNO KAKO JE TO POGREŠNO!

U RU obično se upotrebljava binarno učenje s kritikom

APSTRAKTNI PRIKAZ RASPOZNAVANJA UZORAKA

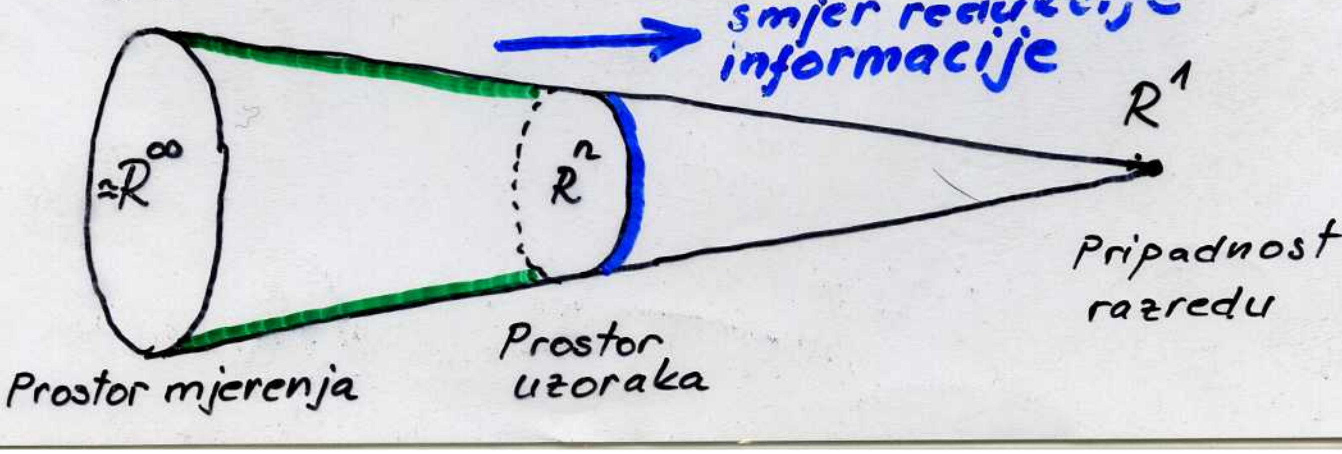
Raspoznavanje uzoraka – redukcija informacije, preslikavanje informacije i postupak označavanja informacije



$fe : F \rightarrow P$

$c : P \rightarrow R^1$
(C)

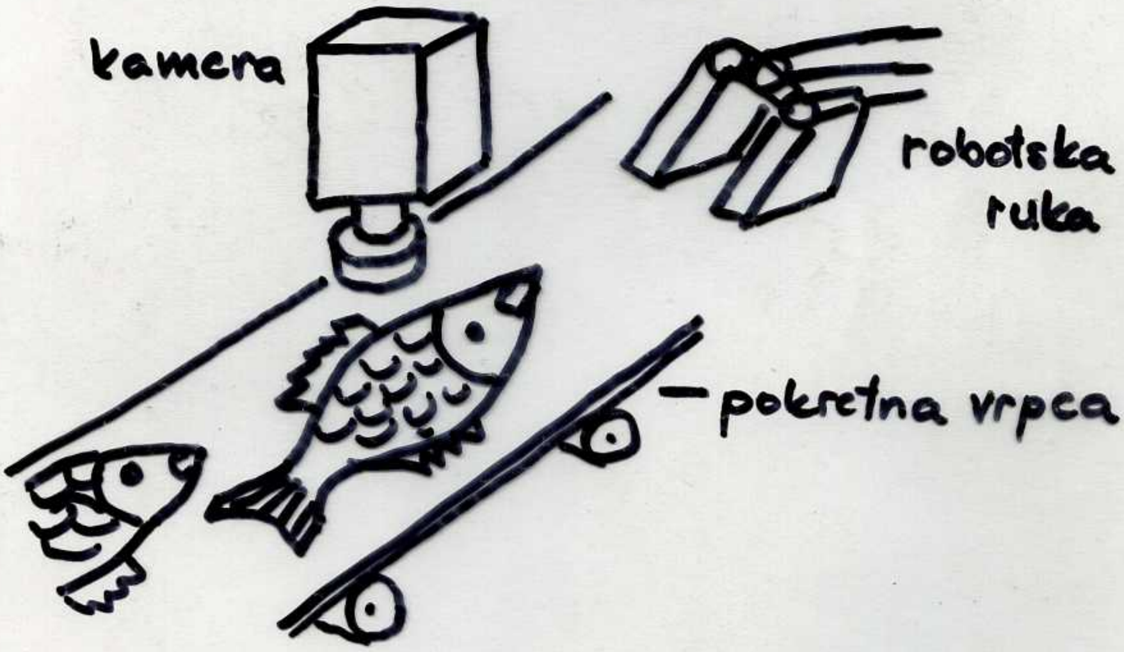
Uhr : stožac raspoznavanja



Primjer (R.O. Duda et al.):

Pretpostavimo da tvornica ribljih konzervi želi automatizirati proces razvrstavanja riba koje stižu na pokretnoj vrpci

Pilot-projekt : razdvajanje lososa i brancina uporabom kamere.



Instaliramo kameru, uzmemo nekoliko slika uzoraka (lososa i brancina) i za početak primjetimo neke fizičke razlike između tih dviju vrsta riba :

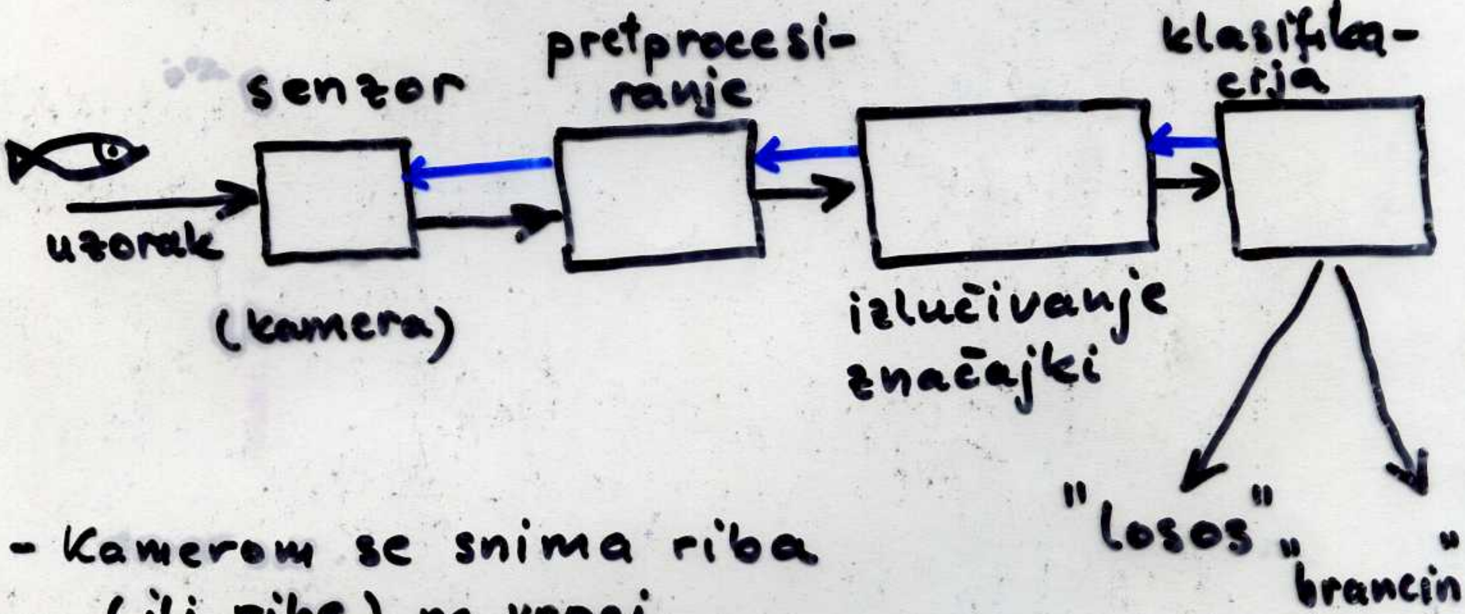
- veličina (duljina);
- svjetlina (ton);
- broj i oblik ljuski,
- položaj ustiju ;
- i.t.d.

značajke

Sve te značajke nude se u izvedbi klasifikatora.

Uočavamo i utjecaj šuma i varijacije u uzetim slikama – – svjetlosti, položaja i orijentacije ribe na pokretnoj vrpci...

Prototip sustava:



- Kamerom se snima riba (ili ribe) na vrpci
- slika se pretprocesira (uklanjanje šuma, poboljšanje kontrasta, segmentacija, lokalizacija)
- informacija koja se odnosi samo na jednu ribu šalje se podsustavu za izlučivanje značajki (engl. feature extractor).

U podsustavu za izlučivanje značajki reducira se količina informacije tako da se "mjere" određene značajke ili svojstva.

Vrijednost tih značajki se prosljeđuje klasifikatoru koji donosi konačnu odluku o pripadnosti uzorka razredu $\omega_1 =$ "losos" ili $\omega_2 =$ "brancin"

Pretpostavimo da nam je poznavalac riba (zaposlen u tvornici) rekao da je brancin općenito dulji od lososa.

Ta informacija će nam poslužiti za izgradnju <u>modela</u> riba:

m1: Brancin ima tipično veću duljinu od lososa.

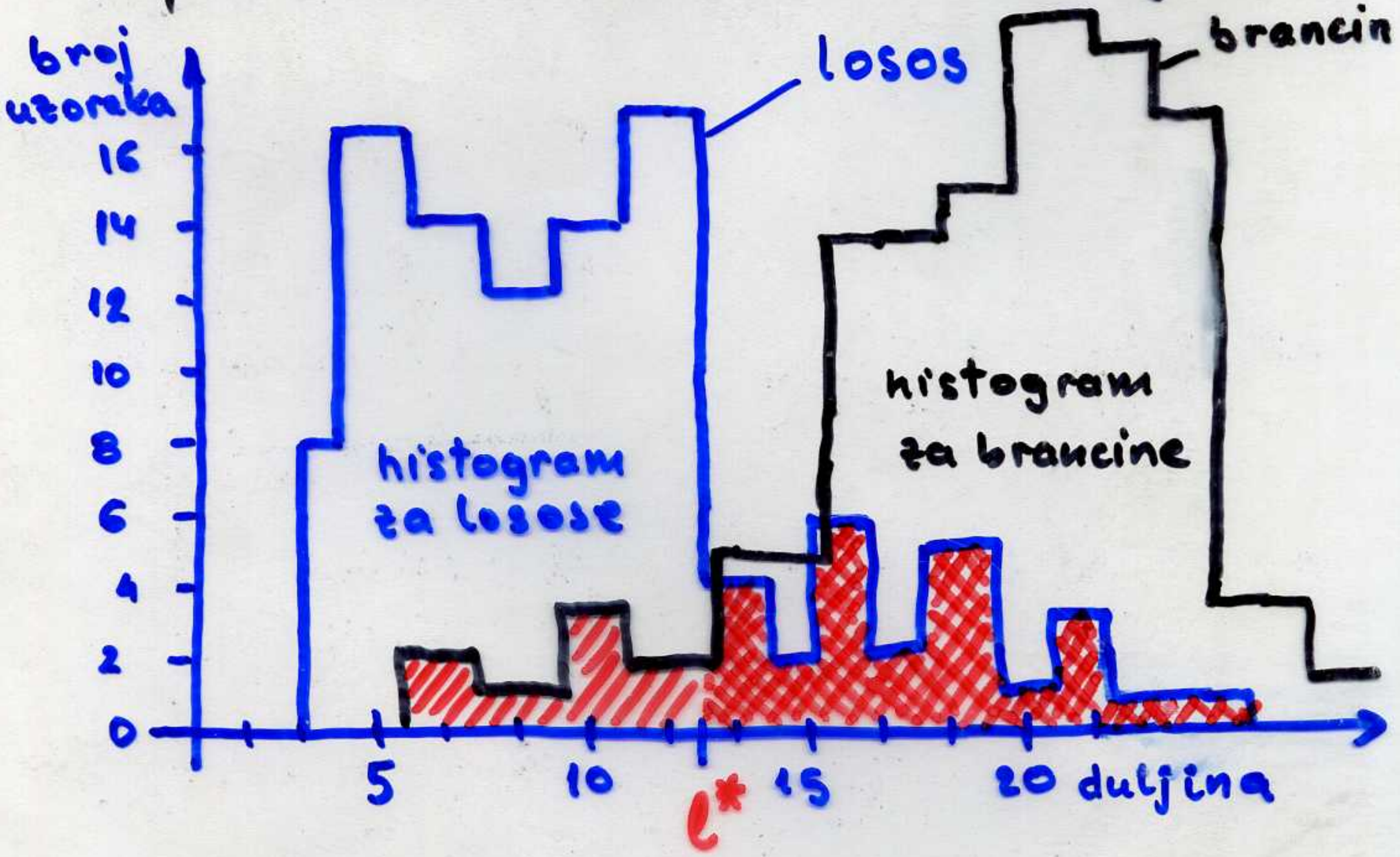
- Označimo duljinu s l i pretpostavimo da na temelju duljine ribe možemo obaviti klasifikaciju.
- Klasifikacijsko (jednostavno) pravilo: "Ako duljina ribe prelazi duljinu l^* onda je riba brancin.

l^* - kritična duljina.

Kako odrediti l^*?

l^* ćemo odrediti na temelju uzoraka iz skupa za učenje (engl. training samples): uzet ćemo veći broj slika riba, izmjeriti duljinu i provjeriti (uz pomoć znalca) rezultat.

Pretpostavimo da smo dobili histograme:



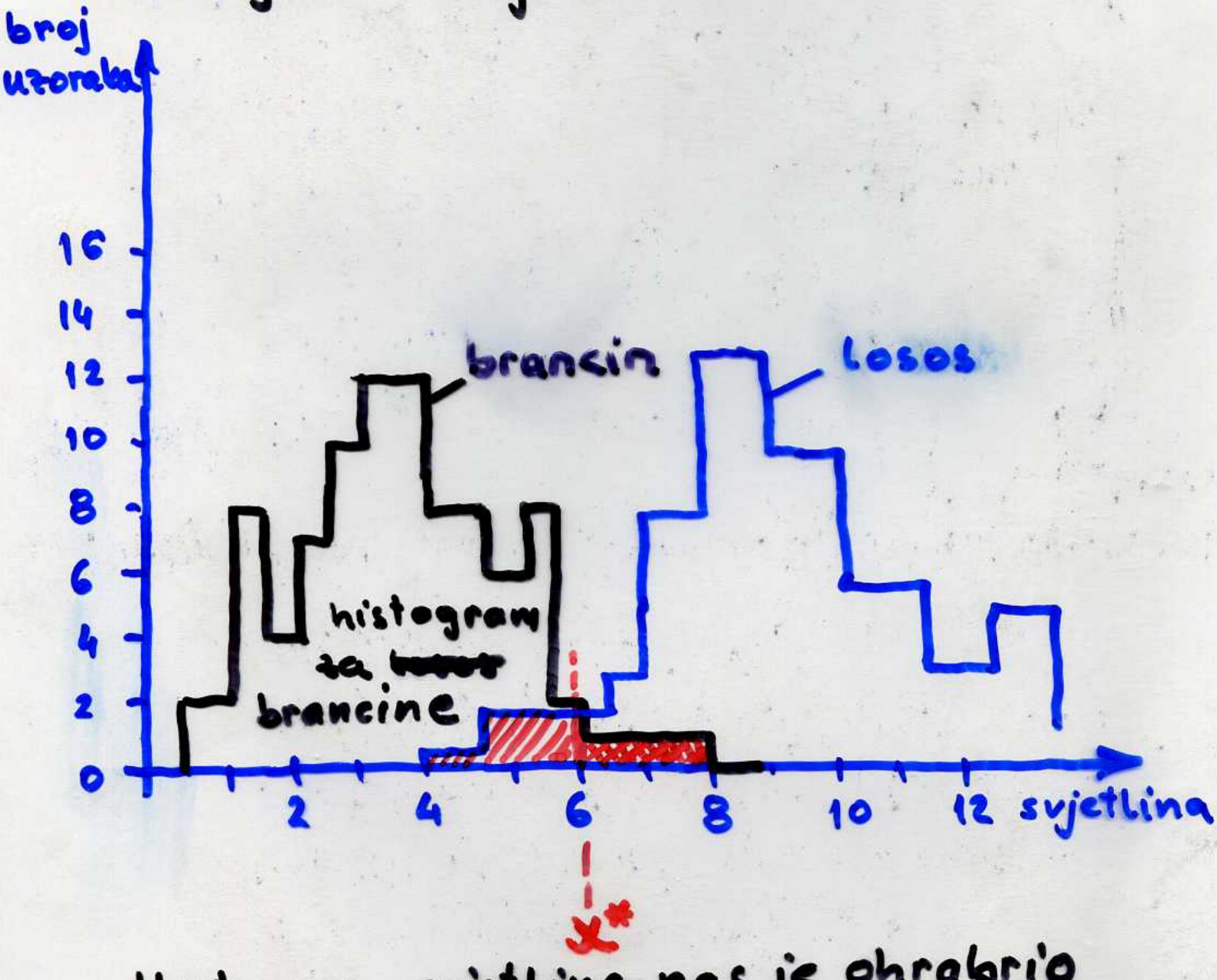
- duljina kao jedini kriterij nije dovoljna
- pogreška (vidi osjenčana područja!)

Pokušajmo sa značajkom :
prosječna svjetlina ribe

(pazi: moramo eliminirati varijacije osvjetljenja scene !)

m2 : Losos je prosječno svjetliji u odnosu na brancin

Histogrami svjetline



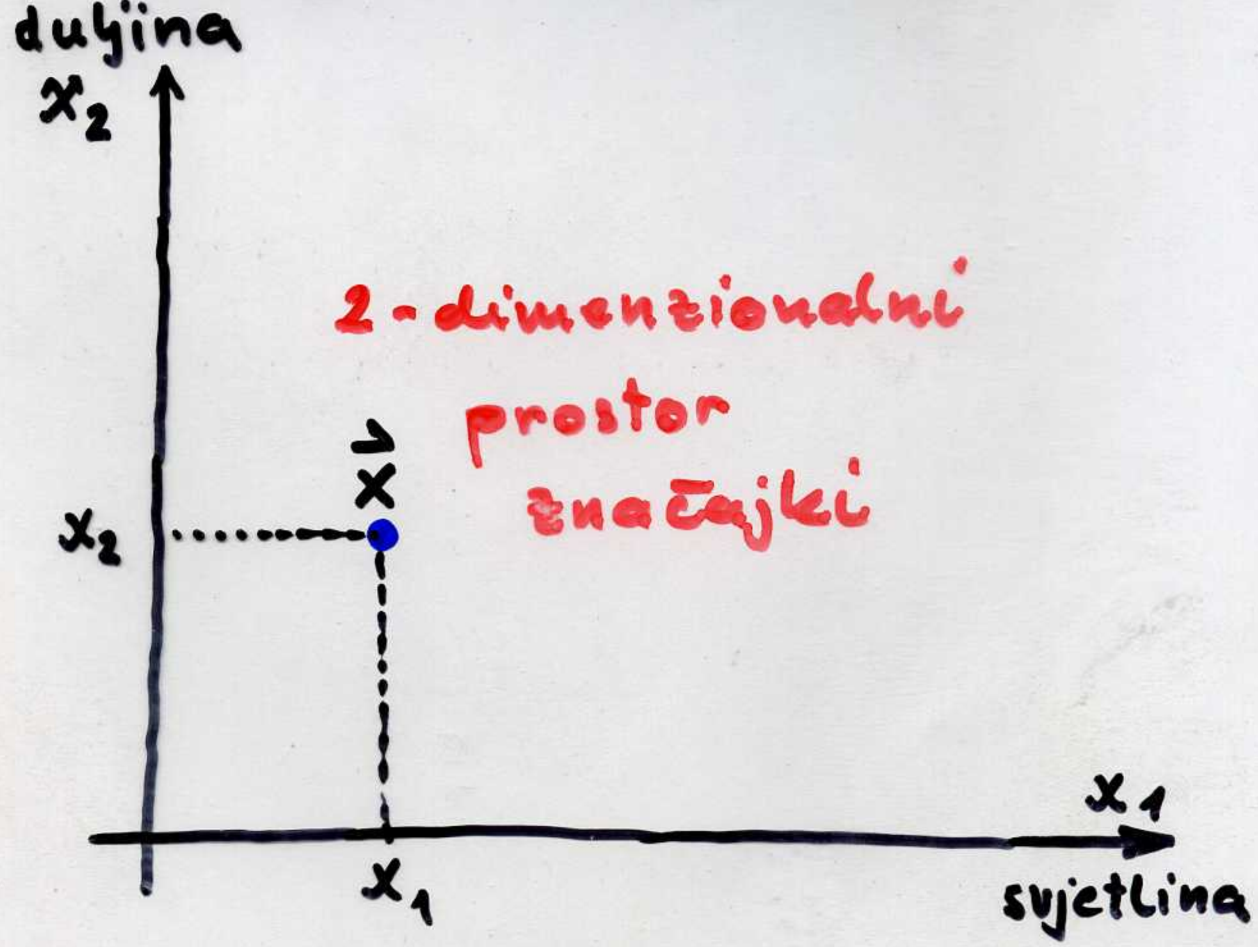
Histogram svjetline nas je ohrabrio
- x^* - optimalna (kritična) vrijednost praga daje manju pogrešku

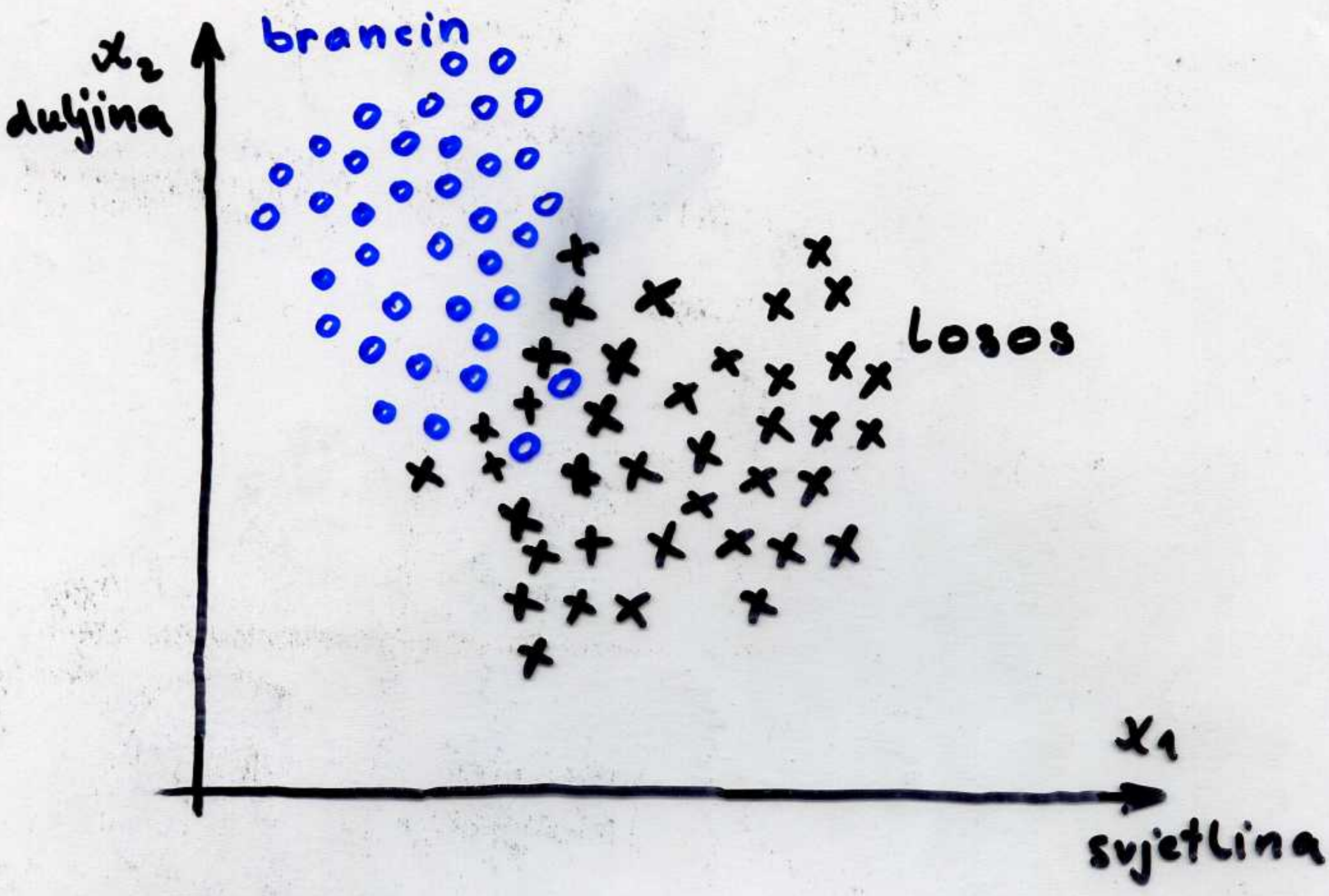
Upotrijebit ćemo obje značajke:
- duljinu l (x_2 - značajka)
- prosječnu svjetlinu (x_1 - značajka)

Oblikovali smo **vektor značajki:**

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

- dvodimenzionalni prostor značajki (duljina, svjetlina)



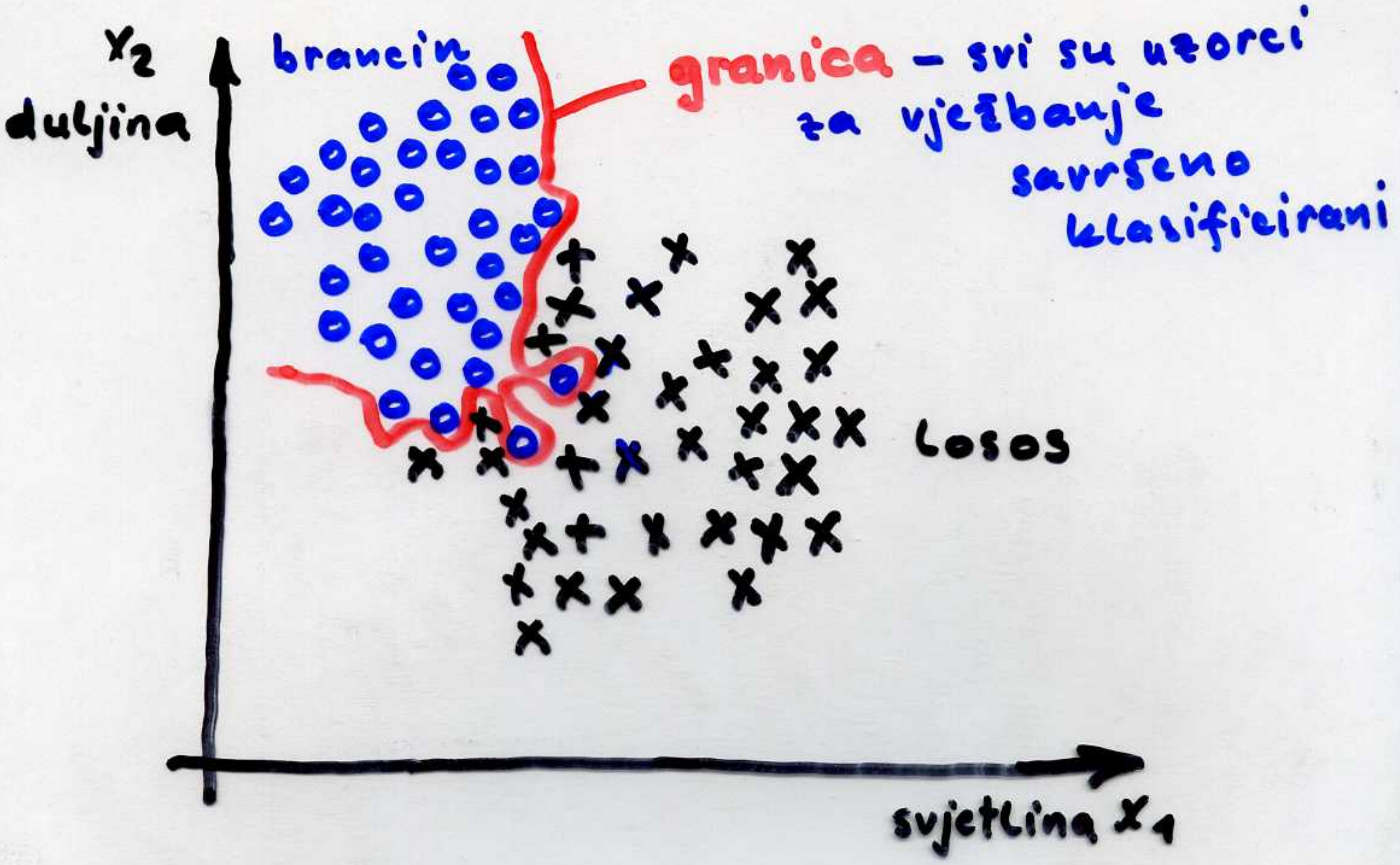


ω_1 = "losos" (×)

ω_2 = "brancin (o)

- generalizacija

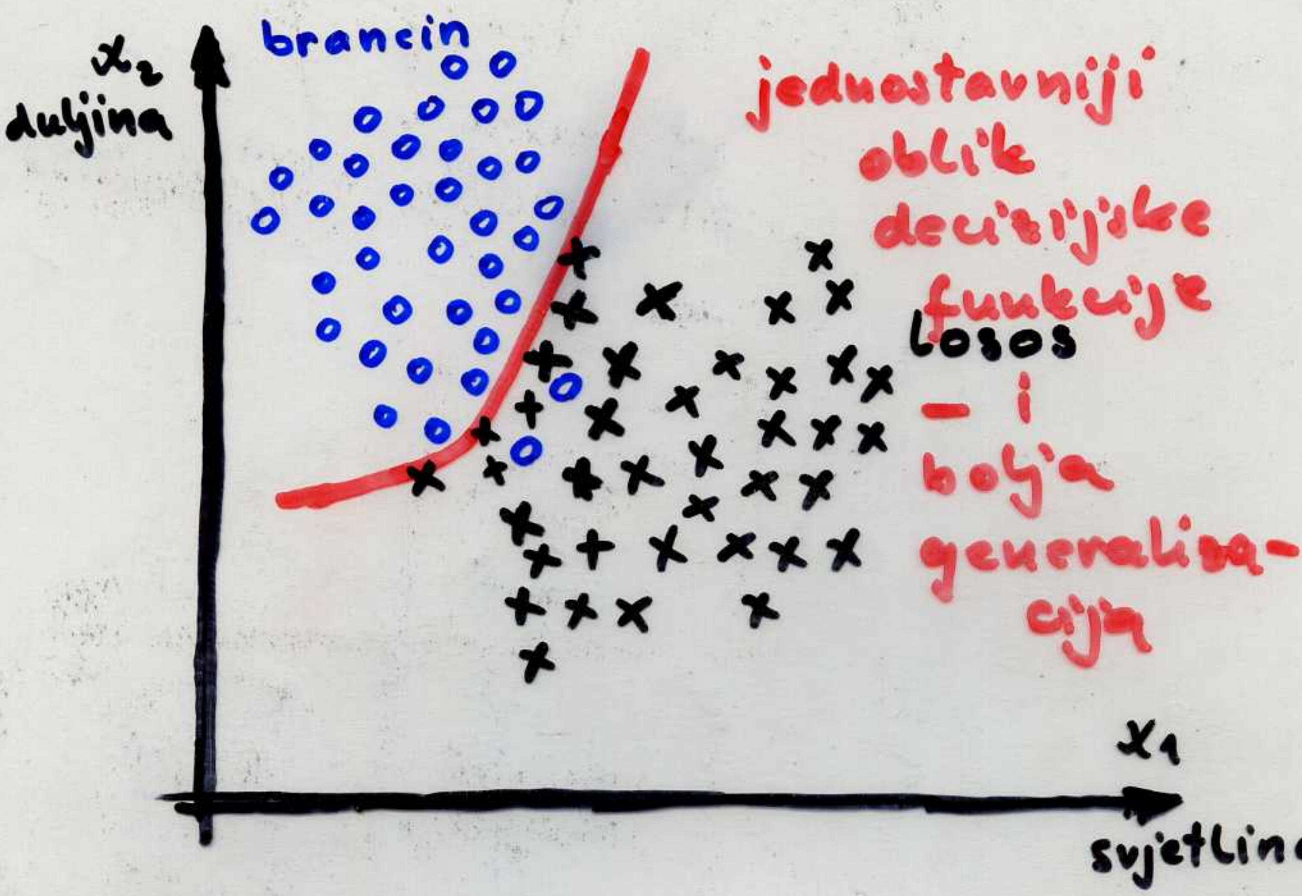
osnovna namjena dizajna sustava je uspješna klasifikacija uzoraka (engl. novel patterns) koje sustav nije "vidio".



Tako složena granica (koja savršeno dobro klasificira uzorke iz skupa za učenje) neće doprinositi generalizaciji

- u stvarnosti možemo se zadovoljiti s nešto skromnijom performansom sustava za uzorke iz skupa za učenje - a to može značiti da će sustav imati bolju performansu za nove uzorke.

Kako sustav automatski određuje decizijsku funkciju u n-dim. prostoru značajki ?

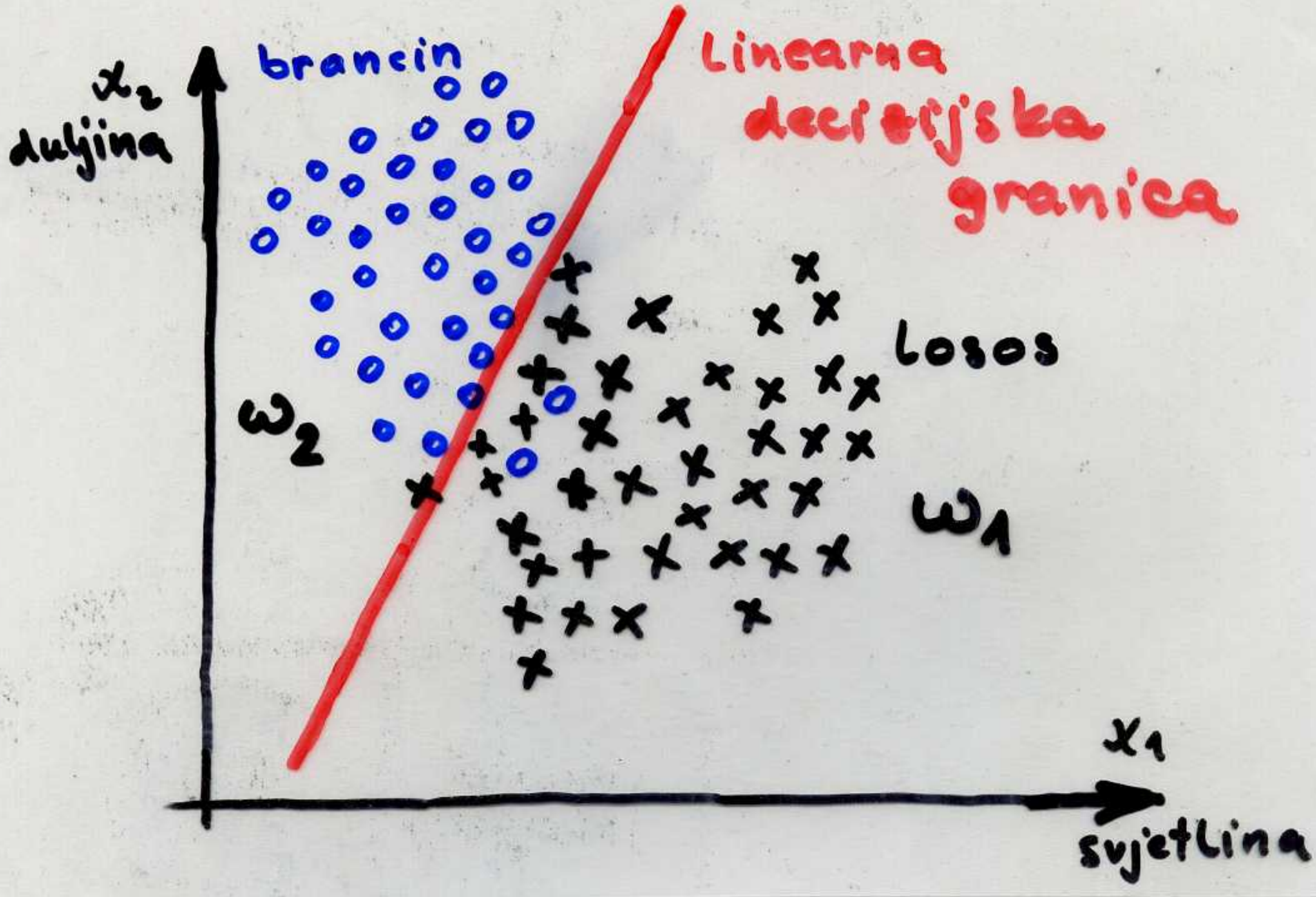


ω_1 = "losos" (×)

ω_2 = "brancin (o)

- generalizacija

  osnovna namjena dizajna sustava je uspješna klasifikacija uzoraka (engl. novel patterns) koje sustav nije "vidio".



ω_1 = "losos" (×)

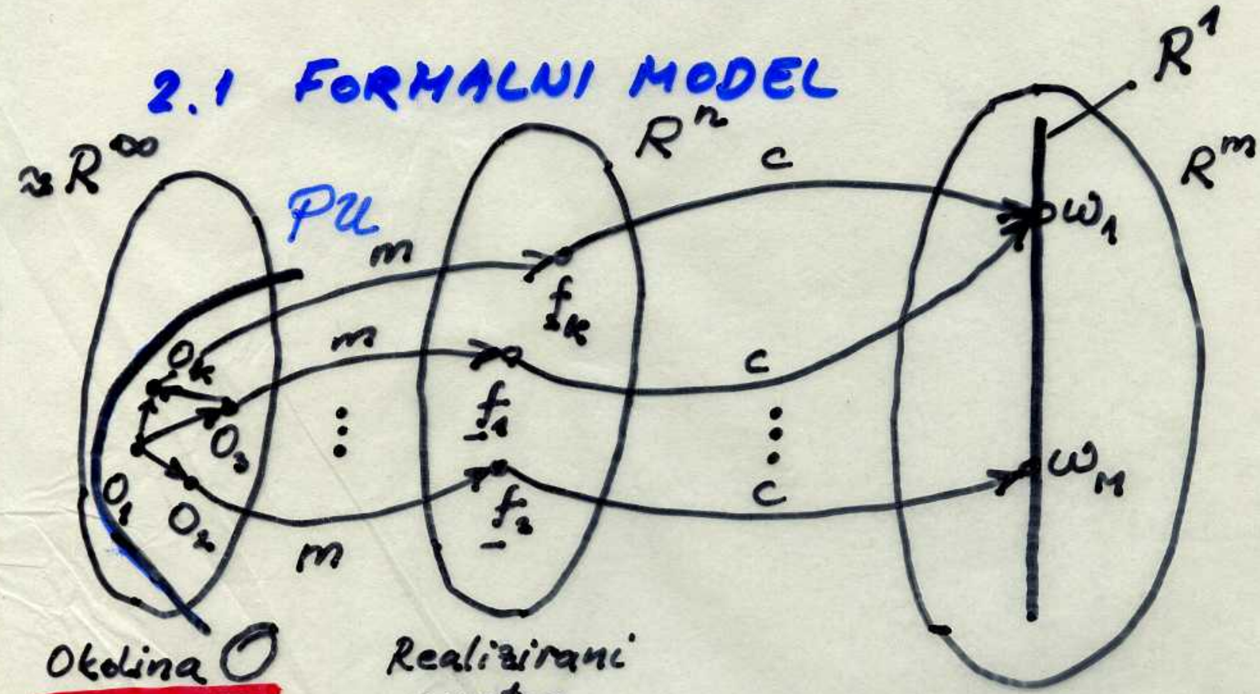
ω_2 = "brancin (∘)

- generalizacija

  osnovna namjena dizajna sustava je uspješna klasifikacija uzoraka (engl. novel patterns) koje sustav nije "vidio".

2. MODEL SUSTAVA ZA RASPOZNAVANJE UZORAKA

2.1 FORMALNI MODEL



Okolina $\mathcal{O}$

$o_k \in \mathcal{O}$

↳ objekt

$v_j \in V\mathcal{O}$

↳ veze

Realizirani prostor uzoraka

$\underline{f}_k$ - uzorci

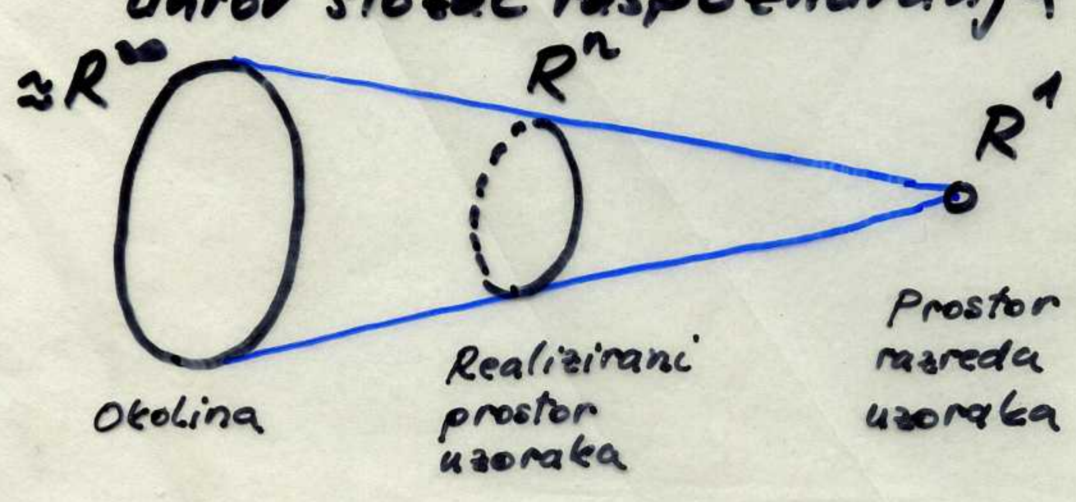
Prostor razreda uzoraka

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

$m: o_k \mapsto \underline{f}_k$ $\qquad$ $c: \underline{f}_k \mapsto \omega_i$

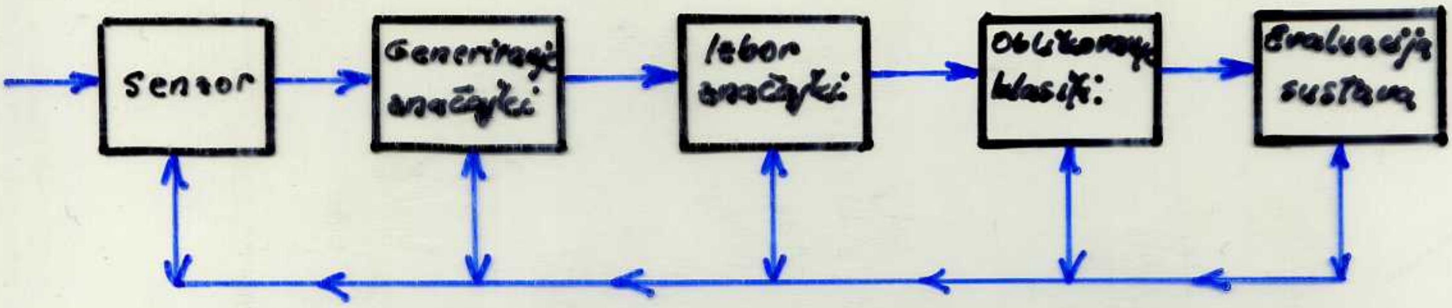
m - mjerenje

Uhrov stožac raspoznavanja - redukcija informacija



2.2. ZNAČAJKE, VEKTOR ZNAČAJKI I KLASIFIKATOR

osnovne faze/stupnjevi u postupku oblikovanja sustava RU



značajke – promatramo ih kao slučajne varijable : x_i

vektor značajki – slučajni vektori $\vec{x}$

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

- značajke koje predstavljaju razlike između razreda uzoraka nazivaju se <u>INTERSET</u> značajke

<u>INTRASET značajke</u> su zajedničke SVIM razredima iz PU (područja uporabe) i <u>ne nose</u> diskriminacijsku informaciju – TAKVE ZNAČAJKE MOGU SE ZANEMARITI

Izbor značajki – izlučiti i izabrati <u>INTERSET</u> značajke

- U većini slučajeva određivanje potpunog skupa diskriminacijskih značajki je IZNIMNO TEŠKO ILI ČAK NEMOGUĆE;

- Neke diskriminacijske značajke mogu se naći na temelju raspoloživih rezultata mjerenja (senzoriranja);

- Redukcija dimenzionalnosti vektora značajki uporabom transformacija uz minimalni gubitak informacije;

- Vektor značajki predočen kao točka u n-dimenzionalnom prostoru značajki;

- Obično definiramo i neku vrstu metrike u takvom prostoru značajki;

- Klasifikacija (razvrstavanje) uzorka temelji se na decizijskim funkcijama; PROBLEM: određivanje optimalne decizijske procedure;

- Problem klasifikacije može se promatrati kao razvrstavanje nepoznatog uzorka u potprostor prostora značajki na temelju decizijskih granica koje definiraju te potprostore!

- Decizijske granice određene su decizijskim funkcijama:

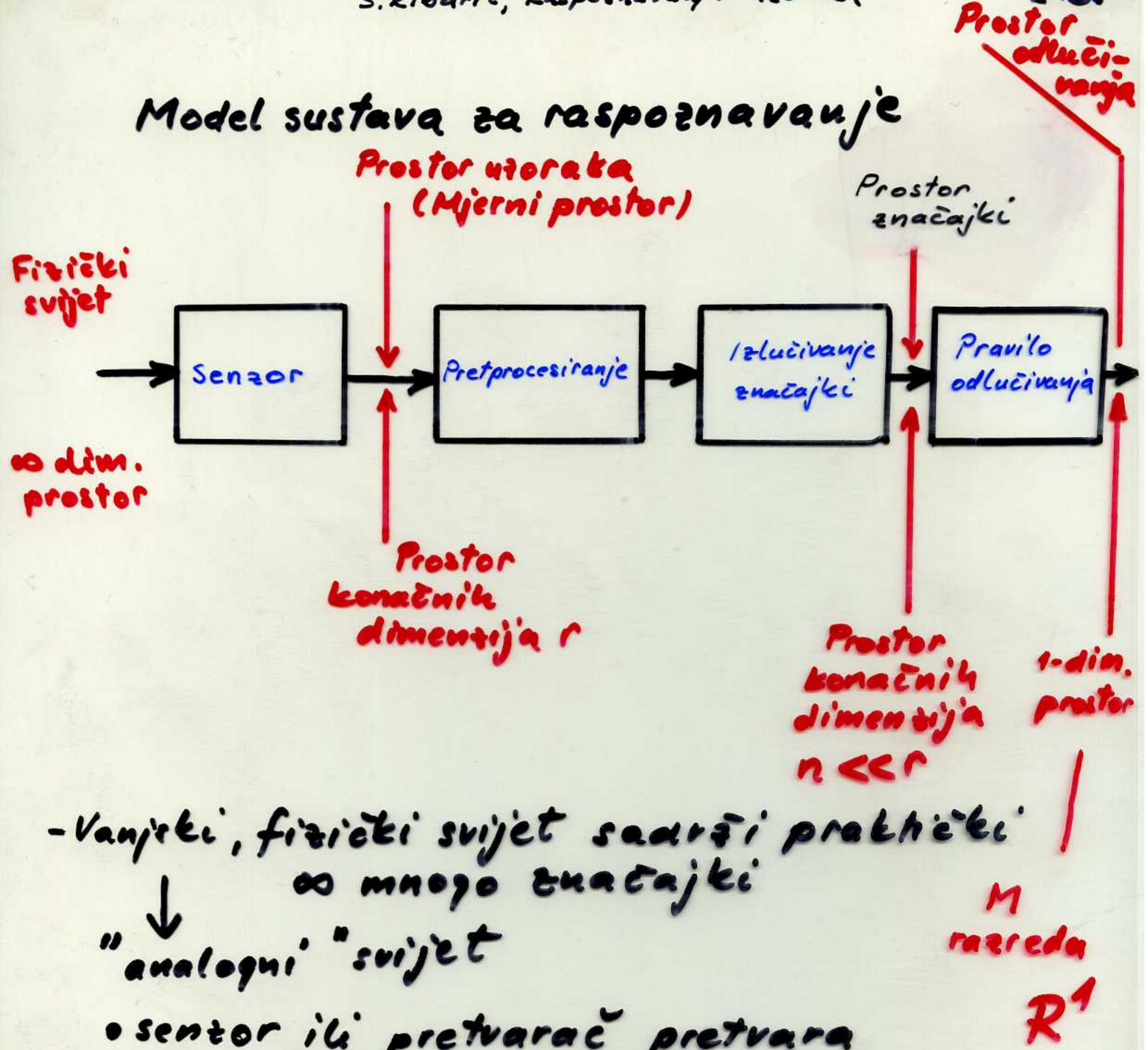
$$d_1(\vec{x}),\ d_2(\vec{x}),\ \dots,\ d_M(\vec{x})$$

/d_i je funkcija koja ima za argument VEKTOR a vraća SKALAR/ VAŽNO!

- Pravilo razvrstavanja:

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

Model sustava za raspoznavanje



- Vanjski, fizički svijet sadrži praktički ∞ mnogo značajki
  ↓
  "analogni" "svijet"
  - senzor ili pretvarač pretvara analogni svijet u zapis koji sadrži r (brojčanih) vrijednosti
- Pretprocesiranje: izlučivanje šuma, poboljšanje mjernog podatka
- Prostor značajki: $n << r$

  $\vec{X} = (x_1, x_2, x_3, \ldots, x_n)^T$
- Prostor odlučivanja – jednodim. prostor R^1

LINEARNE FUNKCIJE ODLUČIVANJA

(engl. Linear discriminant functions)

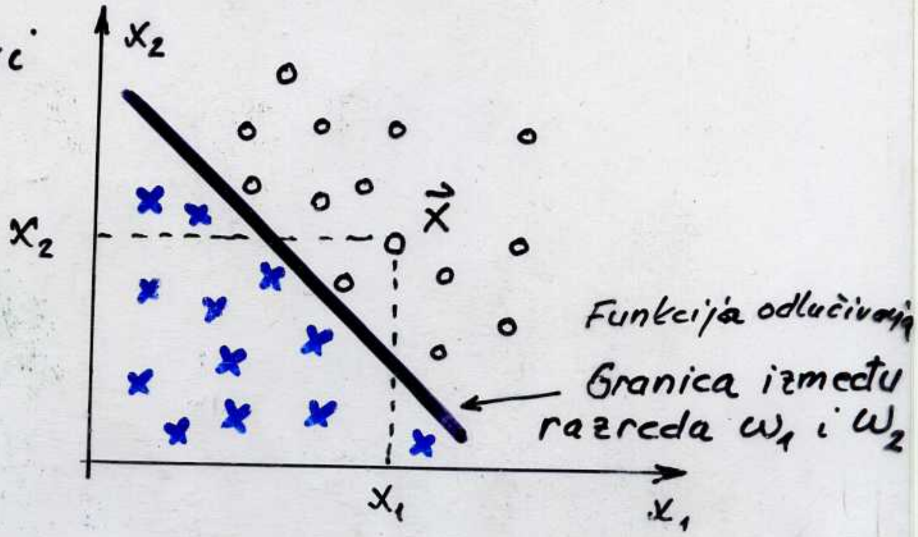
$\vec{X}$ - vektor značajki

$$\vec{X} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$\circ \in \omega_1$

$\times \in \omega_2$

$M = 2$



Funkcija odlučivanja kao linearna kombinacija komponenti vektora $\vec{X}$:

$$d(\vec{X}) = W_1 x_1 + W_2 x_2 + \dots + W_n x_n + W_{n+1}$$

za

$n = 2$ funkcija odlučivanja → jednadžba pravca

$n = 3$ jednadžba ravnine

$n > 3$ hiperravnina

W_i - težinski koeficijenti ; $i = 1, 2, \dots, n$

W_{n+1} - pomaknuće, utežnosni prag (engl. bias, threshold weight)

SLUČAJ: DVA RAZREDA (ω_1 i ω_2)

$M=2$

$$d(\vec{x}) = W_1 X_1 + W_2 X_2 + \dots + W_n \cdot X_n + W_{n+1} = 0$$

$M=2$; ω_1 i ω_2

Decizijsko pravilo:

ako $d(\vec{x}) > 0$ onda $\vec{X} \in \omega_1$

$d(\vec{x}) < 0$ onda $\vec{X} \in \omega_2$

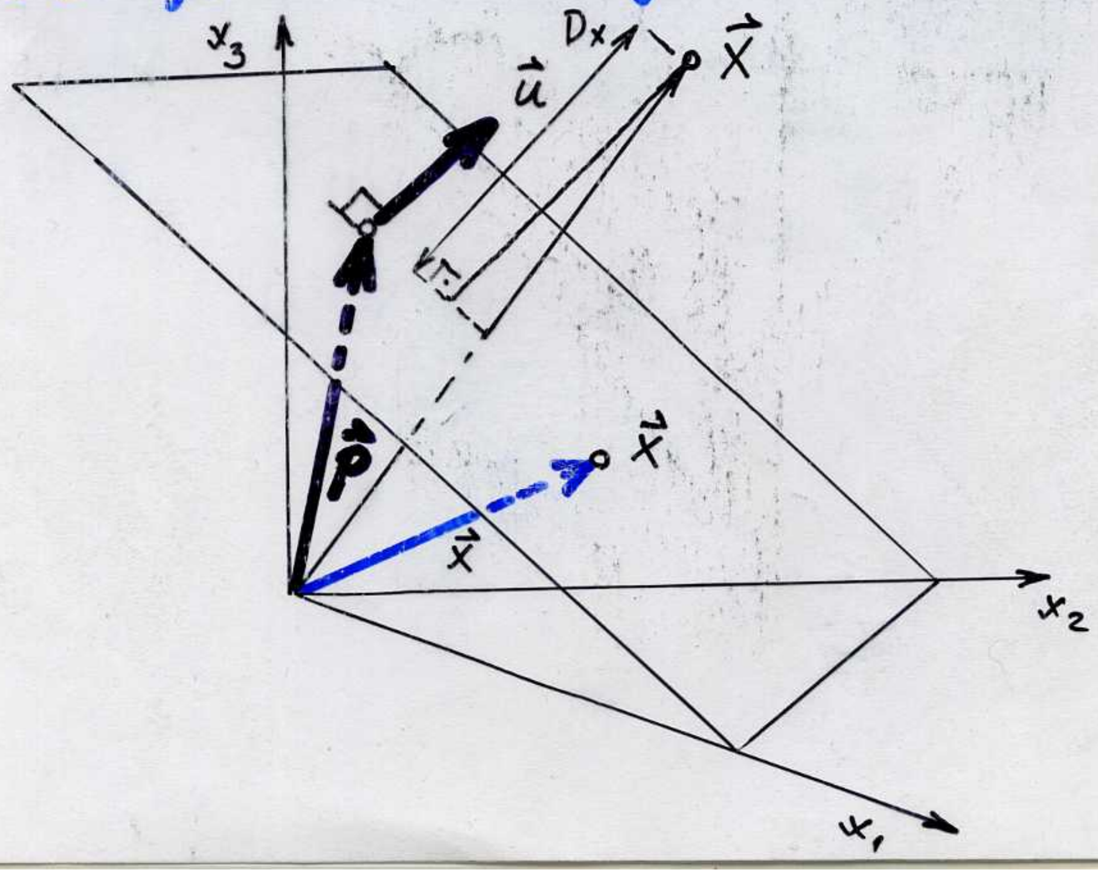
ako je $d(\vec{x}) = 0$

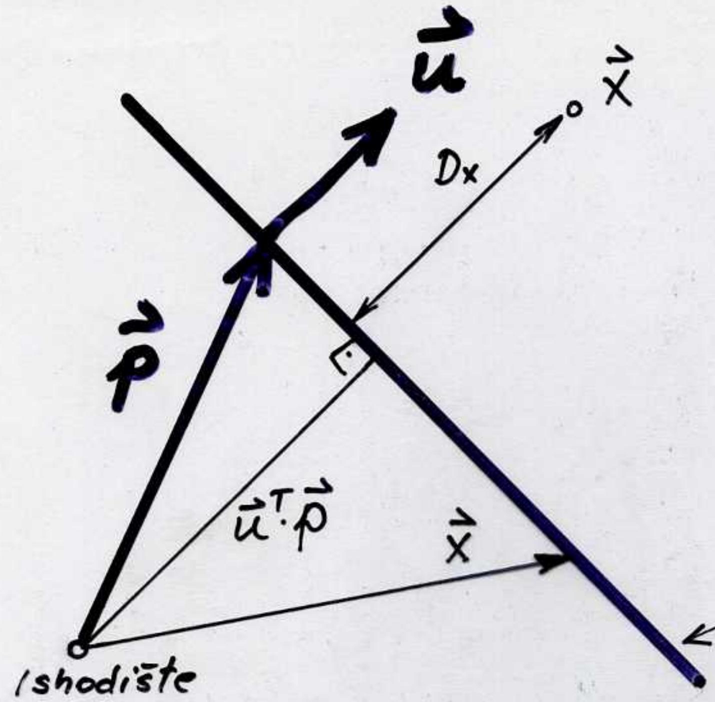
nedefinirano

$d(\vec{x})$ zapišimo u vektorskom obliku:

$$d(\vec{x}) = \vec{W}_0^T \vec{x} + W_{n+1} = 0$$

Geometrijska interpretacija funkcije odlučivanja:





$\vec{u}$ - jedinični vektor normalan na hiperravninu u točki $\vec{p}$

Jednadžba hiperravnine:

$$\vec{u}^T(\vec{x} - \vec{p}) = 0$$

$$\vec{W_0}^T \vec{X} + W_{n+1} = 0 \quad / \|\vec{W_0}\|$$

$$\|\vec{W_0}\| = \sqrt{W_1^2 + W_2^2 + \dots + W_n^2}$$

1x4 3x1

$$\frac{\vec{W_0}^T \vec{X}}{\|\vec{W_0}\|} = -\frac{W_{n+1}}{\|\vec{W_0}\|}$$

usporedimo!

$$\vec{u}^T \vec{x} = \vec{u}^T \cdot \vec{p}$$

$$\vec{u} = \frac{\vec{W_0}}{\|\vec{W_0}\|} \quad i \quad \vec{u}^T \cdot \vec{p} = -\frac{W_{n+1}}{\|\vec{W_0}\|}$$

$\vec{u} = \frac{\vec{W_0}}{\|\vec{W_0}\|}$ → pokazuje orijentaciju hiperravnine; ako je neka komponenta od $\vec{u}$ jednaka 0 onda je hiperravnina paralelna s odgovarajućom koordinatnom osi

Apsolutna vrijednost $\vec{u}^T \cdot \vec{p}$:

$|\vec{u}^T \cdot \vec{p}|$ **predstavlja** udaljenost hiperravnine od ishodišta

$$D_u = \frac{|W_{n+1}|}{\|\vec{W_0}\|}$$

Posljedice:

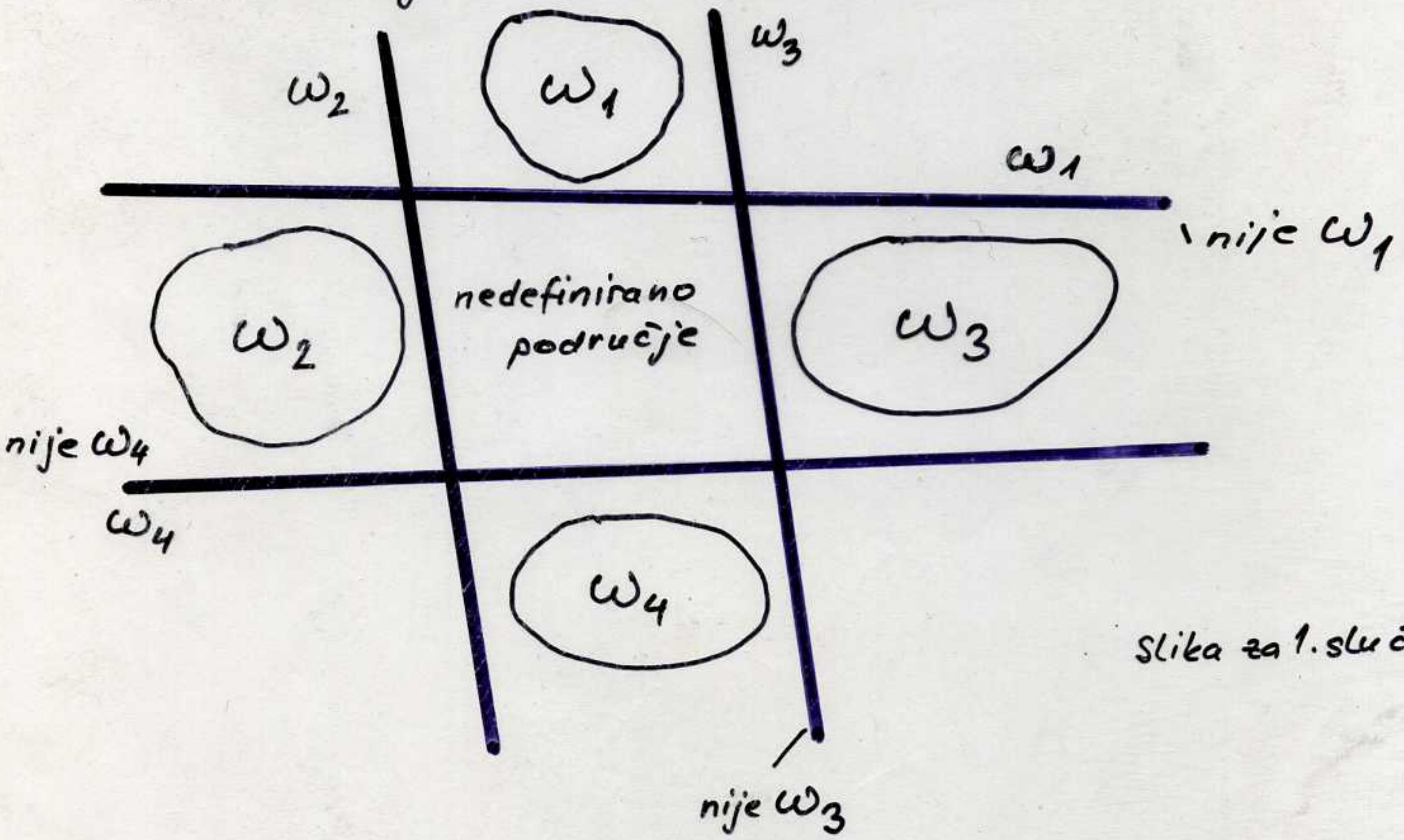
- Budući da je $\vec{u} = \frac{\vec{W_0}}{\|\vec{W_0}\|}$ ispitivanjem vektora težinskih koeficijenata $\vec{W_0}$ moguće je utvrditi da li je hiperravnina paralelna s bilo kojom koordinatnom osi.
- Ako je $W_{n+1} = 0$ onda hiperravnina prolazi kroz ishodište koordinatnog sustava
- Udaljenost točke $\vec{X}$ od hiperravnine je $D_X = |\vec{u}^T \cdot \vec{p} - \vec{u}^T \cdot \vec{X}|$
$$= \left| \frac{\vec{W_0}^T \cdot \vec{P}}{\|\vec{W_0}\|} - \frac{W_{n+1}\vec{X}}{\|\vec{W_0}\|} \right|$$

SLUČAJ: VIŠE RAZREDA
$M > 2$

- više pristupa rješavanju problema linearnog klasifikatora za $M > 2$

$M = c \qquad c > 2$

Primjer: problem se može reducirati na c problema klasifikacije u dva razreda u kojem se i-ti problem rješava linearnom funkcijom odlučivanja koja odvaja uzorke razreda ω_i od svih ostalih koji ne pripadaju razredu ω_i.



Slika za 1. slučaj

1. Slučaj: Granica između ω_i ; $i=1,2,\dots c$ i preostalih razreda

$$d_i(\vec{x}) = W_{i1}x_1 + W_{i2}x_2 + \dots + W_{in}x_n + W_{i,n+1} = 0$$

$$d_i(\vec{x}) = \vec{W_i}^T\vec{X} + W_{i\,n+1}$$

2. slučaj:

Upotrijebi $\frac{c(c-1)}{2}$ linearnih funkcija odlučivanja tako da sa svakom funkcijom odvojiš par razreda
$\overset{\shortparallel}{2}$

Granica između ω_i i ω_j je zadana s:

$$d_{ij}(\vec{x}) = W_{ij1} X_1 + W_{ij2} X_2 + \cdots + W_{ijn} X_n + W_{ijn+1} = 0$$

3. slučaj:

$$d_{ij}(\vec{x}) = d_i(\vec{x}) - d_j(\vec{x}) =$$
$$= (W_{i1} - W_{j1}) X_1 + (W_{i2} - W_{j2}) X_2 + \ldots +$$
$$(W_{in} - W_{jn}) \cdot X_n + (W_{in+1} - W_{jn+1}) = 0$$

$$d_i(\vec{x}) = \vec{W}_i^T \cdot \vec{X} + W_{in+1}$$

ako je $d_i(\vec{x}) > d_j(\vec{x})$; $j = 1, 2, \ldots$; $j \neq i$

onda $\vec{X} \in \omega_i$

klasifikator $\rightarrow$ Linearni stroj (engl. linear machine)

$$(\vec{W}_i - \vec{W}_j)^T \vec{X} + (W_{in+1} - W_{jn+1}) = 0$$

ODREĐIVANJE FUNKCIJE ODLUČIVANJA → UČENJE ILI VJEŽBANJE

Problem oblikovanja linearnog klasifikatora:
odrediti koeficijente linearne funkcije odlučivanja:

$$\vec{W} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} \quad i \quad w_{n+1}$$

?

$$\boxed{d(\vec{x}) = \vec{W}^T \vec{X} + w_{n+1}}$$

"Automatizirati" postupak određivanja koeficijenata linearne funkcije odlučivanja:
iterativni postupak **učenja** koeficijenata linearne funkcije odlučivanja uporabom uzoraka iz skupa za učenje (engl. training set).

N uzoraka: $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N$
razvrstani u dva razreda ω_1 i ω_2
Vektori uzoraka $\vec{x}_i$ $i = 1, 2, \dots N$
su "označeni" vektori tj. oni sa poznatom pripadnosti razredu (ω_1 ili ω_2). **UPOTRIJEBIT ĆEMO IH ZA UČENJE $d(\vec{x})$!**

Povećat ćemo dimenzionalnost vektora $\vec{x}$ za jedan

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \\ 1 \end{bmatrix}$$

radi elegantnijeg zapisa:

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

Uspijemo li odrediti takav vektor težinskih koeficijenata $\vec{w}$ tako da pomoću funkcije $d(\vec{x})$ pravilno razvrstamo sve uzorke (iz skupa za učenje), kažemo da su ω_1 i ω_2 LINEARNO RAZDVOJIVI.

$M = 2$

Uzorak $\vec{x}$ je pravilno razvrstan ako za sve $\vec{x}$ iz ω_1 vrijedi $\vec{w}^T\vec{x} > 0$ i ako za sve $\vec{x}$ iz ω_2 vrijedi $\vec{w}^T\vec{x} < 0$

Jedinstven uvjet: $\vec{w}^T\vec{x} > 0$ ako uzorke iz ω_2 pomnožimo s -1 !

Redefiniran problem:

Tražimo vektor koeficijenata $\vec{w}$ linearne funkcije odlučivanja tako da vrijedi

$$\vec{w}^T \vec{x} > 0 \quad \text{za sve uzorke}$$

$\vec{x}$ iz skupa uzoraka za učenje.

/ PAZI: uzorci $\vec{x} \in \omega_2$ su pomnoženi s -1 /

odnosno

$$[X]\, \vec{w} > \vec{0} \quad \text{za sve uzorke } \vec{x}$$

$$[X] = \begin{bmatrix} \vec{x}_1^T \\ \vec{x}_2^T \\ \vdots \\ \vec{x}_N^T \end{bmatrix}$$

$[X]$ je matrica svih uzoraka iz skupa za učenje, s tim da su uzorci iz ω_2 pomnoženi s -1.

$\vec{w} = (w_1, w_2, \dots, w_n, w_{n+1})^T$

$\vec{0}$ - nulti vektor

PRIMJER 1:

$$\vec{x}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} ; \quad \vec{x}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \qquad \begin{matrix} \vec{x}_1 \in \omega_1 \\ \vec{x}_2 \in \omega_1 \end{matrix}$$

$$\vec{x}_3 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} ; \quad \vec{x}_4 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \begin{matrix} \vec{x}_3 \in \omega_2 \\ \vec{x}_4 \in \omega_2 \end{matrix}$$

- POVEĆAJMO DIMENZIONALNOST VEKTORA :

$$\vec{x}_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} ; \quad \vec{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} ; \quad \vec{x}_3 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

$$\vec{x}_4 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

- POMNOŽIMO SVE UZORKE IZ RAZREDA ω_2 S -1

$$\vec{x}_3 = \begin{bmatrix} -1 \\ 0 \\ -1 \end{bmatrix} ; \quad \vec{x}_4 = \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix}$$

- Oblikujmo matricu $[X]$

$$[X]_{4\times 3} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ -1 & 0 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

-

$$[X]\,\vec{w} > \vec{0}$$

$$\underset{4\times 3}{\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ -1 & 0 & -1 \\ -1 & -1 & -1 \end{bmatrix}} \cdot \underset{3\times 1}{\begin{bmatrix} W_1 \\ W_2 \\ W_3 \end{bmatrix}} > \underset{4\times 1}{\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}}$$

vektor $\vec{w}$ koji zadovoljava sustav linearnih nejednadžbi

$$[X] \cdot \vec{w} > \vec{0}$$

nazivamo <u>razdvojni vektor</u>.

A) GRADIJENTNI POSTUPCI ODREĐIVANJA RAZDVOJNOG VEKTORA

$d(\vec{x})$ - funkcija vektorskog argumenta

vektor!

Općenito: $f(\vec{y})$

$$\vec{y} = (y_1, y_2, \dots, y_n)^T$$

Gradijent funkcije vektorskog argumenta:

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

- **GRADIJENT SKALARNE FUNKCIJE VEKTORSKOG ARGUMENTA JE <u>VEKTOR</u>**

- svaka komponenta gradijenta predstavlja veličinu promjene funkcije u smjeru komponente vektora

VRIJEDI :

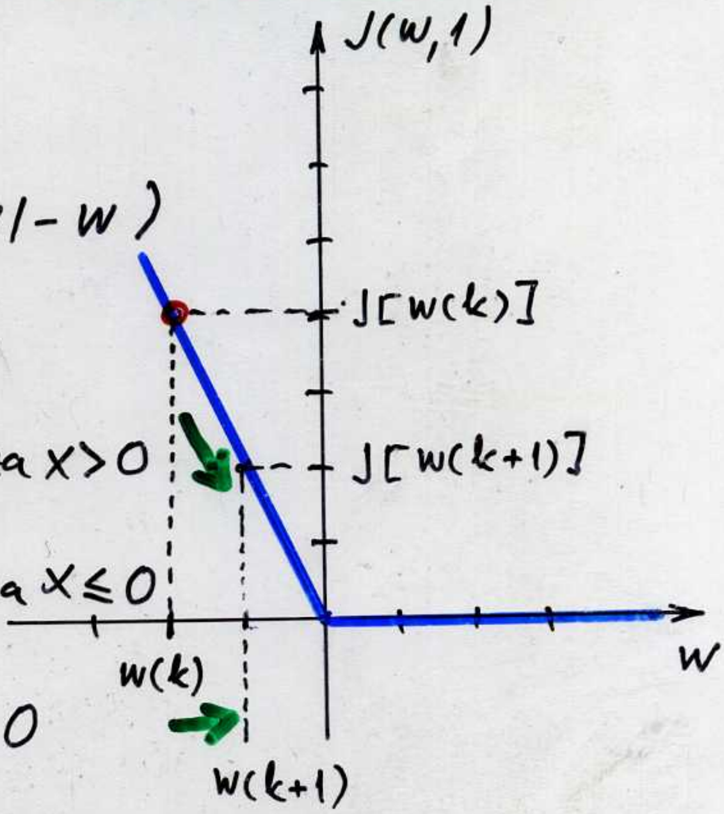
- Povećanje argumenta u smjeru pozitivnog gradijenta funkcije f dovodi nas do <u>maksimuma funkcije f</u>
- povećanje argumenta u smjeru negativnog gradijenta dovodi nas do <u>minimuma funkcije f</u>

PRIMJER **2** :

Funkcija $J(w,1) = (|w| - w)$

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

$$\mathrm{sgn}(x) = \begin{cases} 1 & \text{za } x > 0 \\ -1 & \text{za } x \leq 0 \end{cases}$$



$$\frac{\partial J}{\partial w} = -2 \quad \text{ako je } w \leq 0$$

$$= 0 \quad \text{ako je } w > 0$$

$w(k+1) > w(k)$

povećanje argumenta u smjeru negativne vrijednosti ~~komp.~~ gradijenta $\Rightarrow$ prema min.

$$\vec{w}^T \vec{x}_i > 0 \quad \text{za} \quad i = 1, 2, \ldots, N$$

$$\underset{[1\times3]}{\vec{w}^T} \cdot \underset{[3\times1]}{\vec{x}_i} \underset{[1\times1]}{>0}$$

gdje je $\vec{x}_i$ – i-ti (transponirani) redak matrice $[X]$ dimenzija $N \times (n+1)$

OSNOVNA ZAMISAO:

IZABRATI NEKU FUNKCIJU KOJA ĆE DOSTIĆI MINIMUM KAD JE ISPUNJEN UVJET:

$$\vec{w}^T \cdot \vec{x}_i > 0 \quad \text{za sve } i = 1, 2, \ldots N$$

- zahtijevamo da izabrana funkcija ima samo jedan minimum (i naravno, funkcija je funkcija vektorskog argumenta($\vec{w}$)).

UZMIMO FUNKCIJU:

$$J(\vec{w}, \vec{x}) = (\overbrace{|\vec{w}^T \vec{x}|}^{\text{apsolutna vrijednost!}} - \vec{w}^T \vec{x})$$

- minimum funkcije $J(\vec{w}, \vec{x}) = 0$
- trivijalan slučaj $\vec{w} = \vec{0}$

- Netrivijalan slučaj :

MINIMUM (KRITERIJSKE) FUNKCIJE $J(\vec{w},\vec{x})$ POSTIŽE SE PRI ISPUNJENJU UVJETA

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

GRADIJENTNI POSTUPAK :

KORAK PO KORAK POVEĆAVAMO ARGUMENT $\vec{w}$ U SMJERU NEGATIVNOG GRADIJENTA FUNKCIJE $J(\vec{w},\vec{x})$ SVE DOK NE POSTIGNEMO MINIMUM (KRITERIJSKE) FUNKCIJE $J(\vec{w},\vec{x})$.

$\vec{w}(k)$ - vrijednost $\vec{w}$ u k-tom koraku

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

$\vec{w}(k+1)$ - vrijednost "novog" vektora $\vec{w}$ (u k+1 koraku)

c - pozitivna konstanta, različita od 0, koja određuje veličinu korekcije

/ korekcija se više ne izvodi kada je

$$\frac{\partial J(\vec{w}, \vec{x})}{\partial \vec{w}} = \vec{0}$$

; što je uvjet za minimum /

PRIMJER **3**:

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

Derivacija funkcije s konst. faktorom

$$(cu)' = c \cdot u'$$

Derivacija produkta dviju ili nekoliko funkcija

$$(uv)' = uv' + u'v$$

$$(uvw) = uvw' + uv'w + u'vw$$

Derivacija razlomaka

$$\left(\frac{u}{v}\right)' = \frac{v \cdot u' - uv'}{v^2}$$

Derivacija funkcije od funkcije (složene funkcije)

$y = f(u)$ i $u = \varphi(x)$ tada

$$\frac{dy}{dx} = f'(u) \cdot \varphi'(x)$$

Derivacija složene funkcije od nekoliko varijabli: $u = f(x, y, \dots, t)$
gdje je $x = \varphi(\xi)$ $y = \psi(\xi) \dots, t = \chi(\xi)$

$$\frac{du}{d\xi} = \frac{\partial u}{\partial x}\frac{dx}{d\xi} + \frac{\partial u}{\partial y}\frac{dy}{d\xi} + \dots + \frac{\partial u}{\partial t}\frac{dt}{d\xi}$$

Deriviranje linearnih funkcija:

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

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

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

$$\frac{d}{dX}(\vec{a}^T X \vec{b}) = ab^T$$

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

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

ako je $C = C^T$ onda

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

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

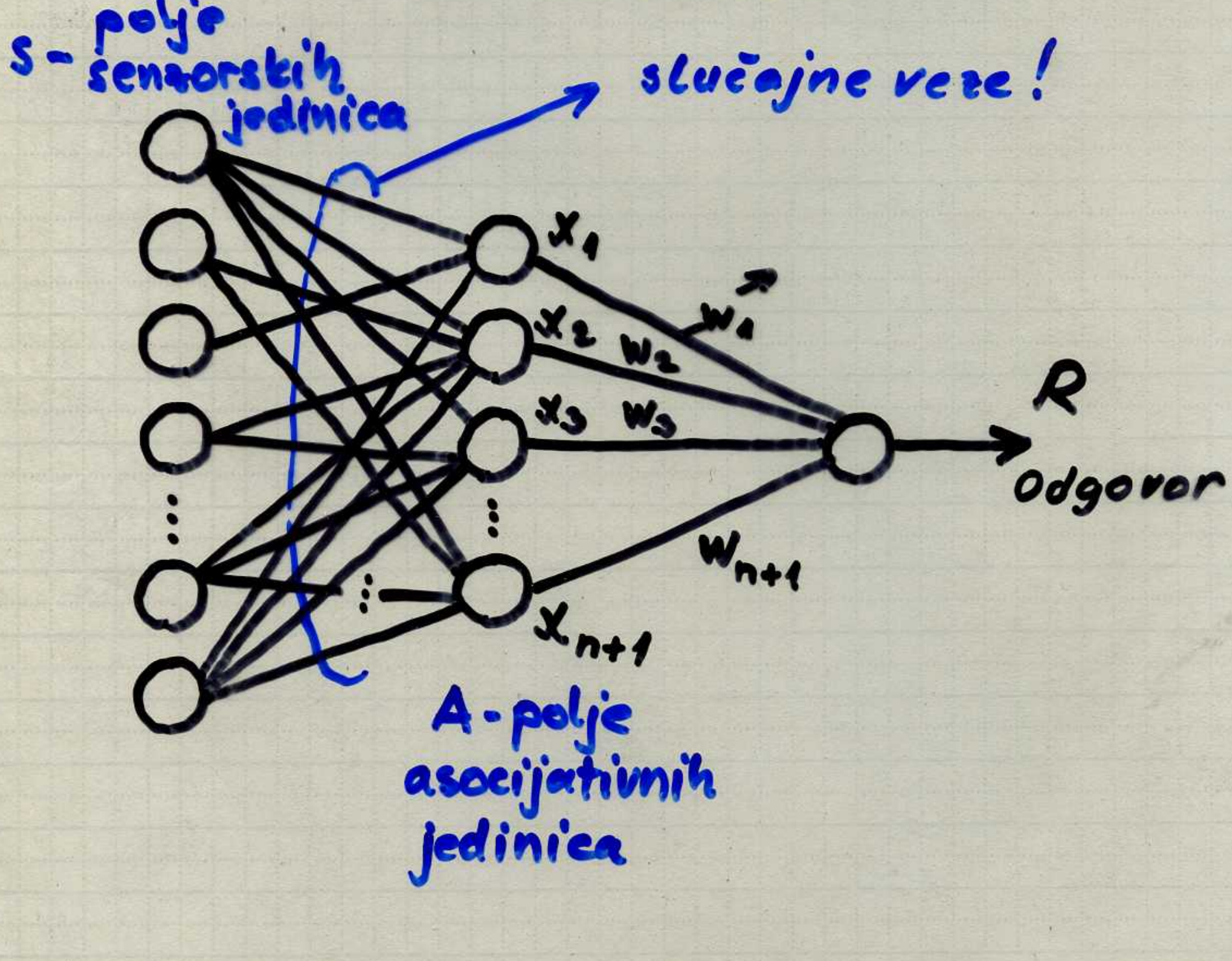
$$\frac{d}{d\vec{x}}(A\vec{x} + \vec{b})^T (D\vec{x} + \vec{e}) = A^T (D\vec{x} + \vec{e}) + D^T (A\vec{x} + \vec{b})$$

PERCEPTRON (Rosenblatt, 1957)

- bionika (biologija – elektronika): grana znanosti koja traži polazne točke za rješenje tehničkih problema u uzorima što ih čovjeku pruža sama priroda

  : primjena bioloških koncepata za izgradnju elektroničkih naprava

Perceptron – klasificira uzorke u jedan od dva razreda



Jedinica u polju A generira izlaz različit od 0 ako je dovoljan broj senzorskih jedinica priključenih na A jedinicu pobuđen (aktivan).

x_i - odgovor i-te asocijativne jedinice

W_i - težinska vrijednost

Odgovor R proporcionalan je sumi odgovora A jedinica; odgovori su pomnoženi težinskim vrijednostima W_i :

$$R = \sum_{i=1}^{n+1} w_i x_i = \vec{W}^T \cdot \vec{X}$$

Pravilo klasifikacije:

- ako je $R > 0$ nepoznati se uzorak razvrstava u ω_1
- ako je $R < 0$ nepoznati se uzorak razvrstava u ω_2

Perceptron – klasifikator uzoraka u dva razreda – ako su razredi linearno separatibilni

– modifikacija perceptrona za klasifikaciju uzoraka u $M>2$ razreda:

M – jedinica R

$R_1, R_2, \ldots, R_M$

Pravilo:

uzorak se razvrstava u razred ω_i akko $R_i > R_j$ za sve $i \neq j$.

Osnovni model perceptrona može se proširiti na nelinearne decizijske funkcije umetanjem nelinearnog pretprocesora između A i R polja

– Višeslojni perceptron MLP

ALGORITAM PERCEPTRONA

(Reward - Punishment Concept)

Algoritam učenja perceptrona:

- Zadana su dva skupa uzoraka za učenje koji pripadaju razredu ω_1 i ω_2

- $\vec{w}(1)$ početna vrijednost vektora težinskih koeficijenata - proizvoljno izabran

- k-ti korak učenja :

  - Ako $\vec{x}(k) \in \omega_1$ i

    $\vec{w}^T(k)\vec{x}(k) \leq 0$ zamijeni $\vec{w}(k)$ sa

    $$\vec{w}(k+1) = \vec{w}(k) + c\vec{x}(k),$$

    gdje je c korekcijski faktor

  - Ako $\vec{x}(k) \in \omega_2$ i

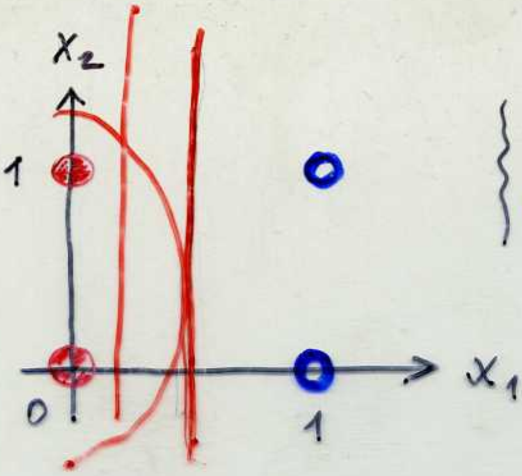
    $\vec{w}^T(k)\vec{x}(k) \geq 0$ zamijeni

    $$\vec{w}(k+1) = \vec{w}(k) - c\vec{x}(k)$$

  - U drugim slučajevima ostavi $\vec{w}(k) \Rightarrow$ $\vec{w}(k+1) = \vec{w}(k)$

- c mora biti pozitivan (i konstanta).

PRIMJER



$\circ \in \omega_1$

$\bullet \in \omega_2$

$\omega_1 : \{ (0,0,1)', (0,1,1)' \}$

$\omega_2 : \{ (1,0,1)', (1,1,1)' \}$

$c = 1 \quad ; \quad \vec{w}(1) = \vec{0}$

1.k.

$$\vec{w}^T(1) \cdot \vec{x}(1) = (0,0,0) \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = 0$$

$$\vec{w}(2) = \vec{w}(1) + \vec{x}(1) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

2.k.

$$\vec{w}^T(2) \cdot \vec{x}(2) = (0,0,1) \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = 1$$

$$\vec{w}(3) = \vec{w}(2) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

3.k.

$$\vec{w}^T(3) \cdot \vec{x}(3) = (0,0,1) \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = 1$$

$$\vec{w}(4) = \vec{w}(3) - \vec{x}(3) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}$$

4.k.

$$\vec{w}^T(4) \, \vec{x}(4) = (-1,0,0) \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = -1$$

$$\vec{w}(5) = \vec{w}(4) = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}$$

Zbog korekcija u 1. i 3. koraku postupak se ponavlja!

[RJEŠENJE SE DOBIVA SAMO KAD ALGORITAM DAJE POTPUNO ERROR - FREE ITERACIJU ZA SVE UZORKE U SKUPU ZA UČENJE]

$\vec{x}(5) = \vec{x}(1)$

$\vec{x}(6) = \vec{x}(2)$

$\vec{x}(7) = \vec{x}(3)$

$\vec{x}(8) = \vec{x}(4)$

DRUGA ITERACIJA DAJE:

$$\vec{w}^T(5) \cdot \vec{x}(5) = (-1,0,0)\begin{pmatrix}0\\0\\1\end{pmatrix} = 0$$

$$\vec{w}(6) = \vec{w}(5) + \vec{x}(5) = \begin{pmatrix}-1\\0\\0\end{pmatrix} + \begin{pmatrix}0\\0\\1\end{pmatrix} = \begin{pmatrix}-1\\0\\1\end{pmatrix}$$

$$\vec{w}^T(6) \cdot \vec{x}(6) = (-1,0,1)\begin{pmatrix}0\\1\\1\end{pmatrix} = 1$$

$$\vec{w}(7) = \vec{w}(6) = \begin{pmatrix}-1\\0\\1\end{pmatrix}$$

$$\vec{w}^T(7) \cdot \vec{x}(7) = (-1,0,1)\begin{pmatrix}1\\0\\1\end{pmatrix} = 0$$

$$\vec{w}(8) = \vec{w}(7) - \vec{x}(7) = (-1,0,1)^T - \begin{pmatrix}1\\0\\1\end{pmatrix} = \begin{pmatrix}-2\\0\\0\end{pmatrix}$$

$$\vec{w}^T(8) \cdot \vec{x}(8) = (-2,0,0)\begin{pmatrix}1\\1\\1\end{pmatrix} = -2$$

$$\vec{w}(9) = \vec{w}(8)$$

Dogodile su se dvije pogreške!

Nova iteracija

$\vec{x}(9) = \vec{x}(1)$

$\vec{x}(10) = \vec{x}(2)$

$\vec{x}(11) = \vec{x}(3)$

$\vec{x}(12) = \vec{x}(4)$

$\vec{w}^T(9)\vec{x}(9) = 0 \qquad \vec{w}(10) = \vec{w}(9) + \vec{x}(9) = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$

$\vec{w}^T(10) \cdot \vec{x}(10) = 1 \qquad \vec{w}(11) = \vec{w}(10) = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$

$\vec{w}^T(11) \cdot \vec{x}(11) = -1 \qquad \vec{w}(12) = \vec{w}(11) = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$

$\vec{w}^T(12) \cdot \vec{x}(12) = -1 \qquad \vec{w}(13) = \vec{w}(12) = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$

Dogodila se pogreška! Nova iteracija:

$\vec{x}(14) = \vec{x}(1) = (0,0,1)^T$

$\vec{x}(15) = \vec{x}(2) = (0,1,1)^T$

$\vec{x}(16) = \vec{x}(3) = (1,0,1)^T$

$\vec{x}(17) = \vec{x}(4) = (1,1,1)^T$

$\vec{w}^T(14) \cdot \vec{x}(14) = (-2,0,1) \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = 1 \qquad \vec{w}(15) = \vec{w}(14) = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$

$\vec{w}^T(15) \cdot \vec{x}(15) = (-2,0,1) \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = 1 \qquad \vec{w}(16) = \vec{w}(15)$

$\vec{w}^T(16) \cdot \vec{x}(16) = (-2,0,1) \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = -1 \qquad \vec{w}(17) = \vec{w}(16)$

$\vec{w}^T(17) \cdot \vec{x}(17) = (-2,0,1) \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = -1 \qquad \vec{w}(18) = \vec{w}(17)$

$\underline{d(\vec{x}) = -2x_1 + 1} \leftarrow \boxed{\vec{w} = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}}$

PRIMJER 4:

$\omega_1 = \{ (0,0)^T, (0,1)^T \}$ i

$\omega_2 = \{ (1,0)^T, (1,1)^T \}$

- povećajmo dimenzionalnost prostora značajki na dimenziju $n+1$.

$\{ (0,0,1)^T, (0,1,1)^T \}$

$\{ (1,0,1)^T, (1,1,1)^T \}$

- pomnožimo sve uzorke iz razreda ω_2 sa -1

$\{ (0,0,1)^T, (0,1,1)^T \}$

$\{ (-1,0,-1)^T (-1,-1,-1)^T \}$

- pretpostavimo da su skupovi uzoraka za učenje razdvojivi linearnom funkcijom odlučivanja

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

- postupak perceptrona sa stalnim prirastom

Neka je $C=1$ i neka je početni vektor koeficijenata $\vec{W}(1)$ proizvoljan, npr. $\vec{w}(1) = (-1,0,0)^T$.

- u prvom koraku učenja uzimamo uzorak $\vec{x}(1) = (0,0,1)^T$

Računamo:

$$\vec{W}^T(1) \cdot \vec{X}(1) = (-1,0,0) \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = 0,$$

zbog toga

$$\vec{W}(2) = \vec{W}(1) + \vec{x}(1) = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}.$$

U drugom koraku učenja uzimamo uzorak $\vec{X}(2) = (0,1,1)^T$

Računamo:

$$\vec{W}^T(2) \cdot \vec{X}(2) = (-1,0,1)\begin{pmatrix}0\\1\\1\end{pmatrix} = 1$$

Zato što je $\vec{W}^T(2) \cdot \vec{X}(2) > 0$, u skladu s postupkom učenja $\vec{W}(3) = \vec{W}(2)$

U trećem koraku upotrebljavamo uzorak $\vec{X}(3) = (-1,0,-1)^T$

Računamo:

$$\vec{W}^T(3) \cdot \vec{X}(3) = (-1,0,1)\begin{pmatrix}-1\\0\\-1\end{pmatrix} = 0$$

$$\vec{W}(4) = \vec{W}(3) + \vec{X}(3)$$

$$\vec{W}(4) = \begin{pmatrix}-2\\0\\0\end{pmatrix}$$

U četvrtom koraku učenja uzimamo uzorak $\vec{X}(4) = (-1,-1,-1)^T$

Računamo:

$$\vec{W}^T(4)\vec{X}(4) = (-2,0,0)\begin{pmatrix}-1\\-1\\-1\end{pmatrix} = 2$$

Budući da je $\vec{W}^T(4)\vec{X}(4) > 0$ vrijedi $\vec{W}(5) = \vec{W}(4)$

RJEŠENJE ĆEMO DOBITI KADA NA TEMELJU IZRAČUNATOG $\vec{W}$ PRAVILNO RAZVRSTAVAMO SVE UZORKE

– u ovom primjeru došlo je do pogrešnog razvrstavanja dva uzorka – POSLJEDICA → popravljanje vektora koef. funkcije odlučivanja $\vec{W}$.

Postupak učenja nastavljamo tako da pretpostavimo $\vec{X}(5) = \vec{X}(1)$, $\vec{X}(6) = \vec{X}(2)$ $\vec{X}(7) = \vec{X}(3)$ i $\vec{X}(8) = \vec{X}(4)$

Dobivamo:

$\vec{w}^T(5)\vec{x}(5) = 0$ slijedi $\vec{w}(6) = w(5) + \vec{x}(5)$

$$\vec{w}(6) = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$$

$w^T(6)\vec{x}(6) = 1$ slijedi $\vec{w}(7) = \vec{w}(6)$

$w^T(7)\vec{x}(7) = 1$ slijedi $\vec{w}(8) = \vec{w}(7) = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$

$w^T(8)\vec{x}(8) = 1$ slijedi $\vec{w}(9) = \vec{w}(8) = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$

I u drugom ponavljanju nisu bili razvrstani svi uzorci pravilno. Treća iteracija:

$\vec{x}(9) = \vec{x}(1)$ $\vec{x}(10) = \vec{x}(2)$ $\vec{x}(11) = \vec{x}(3)$

$\vec{x}(12) = \vec{x}(4)$

$$\vec{w}^T(9)\vec{x}(9) = (-2, 0, 1)\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = 1 \ ; \rightarrow \vec{w}^T(10) = \vec{w}^T(9)$$

$$\vec{w}^T(10)\vec{x}(10) = (-2, 0, 1)\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = 1 \rightarrow \vec{w}^T(11) = \vec{w}^T(10)$$

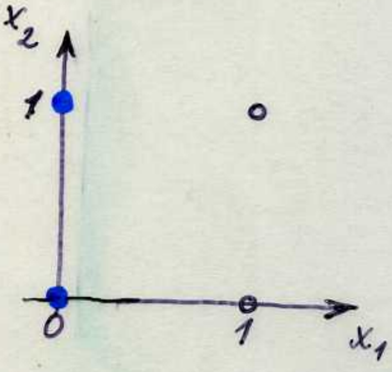
$$\vec{w}^T(11)\vec{x}(11) = (-2, 0, 1)\begin{pmatrix} -1 \\ 0 \\ -1 \end{pmatrix} = 3 \qquad \vec{w}^T(12) = \vec{w}^T(11)$$

$$\vec{w}^T(12)\vec{x}(12) = (-2, 0, 1)\begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix} = 1 \qquad \vec{w}^T(13) = \vec{w}^T(12)$$

$$\vec{w} = (-2, 0, 1)^T$$

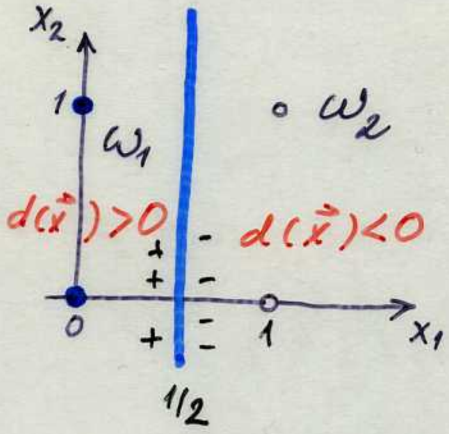
Decizijska funkcija $d(\vec{x}) = \underline{\vec{w}^T\vec{x} = -2x_1 + 1}$

Uzorci za vježbanje:



Naučena funkcija odlučivanja:

$$d(\vec{x}) = -2x_1 + 1 = 0$$



• $\in \omega_1$

o $\in \omega_2$