Inlärningens geometri
Populärföreläsning SF1611, Introduktionskurs i matematik.
Timo Koski

TK

22.09.2014

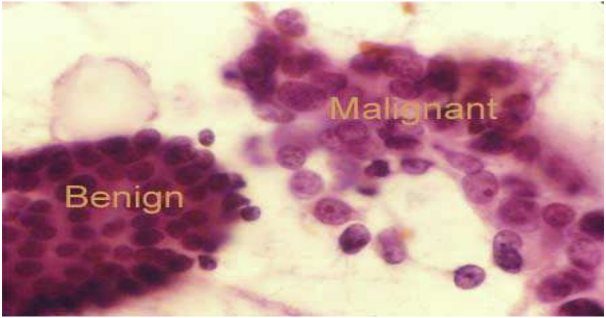
KTH Matematik

En grundläggande form av inlärning är att inhämta kunskap ur valda exempel. Med ett exempel avses information om ett föremål och en klassificering av detsamma. Det kan handla om t.ex. handskrivna siffor. Förmågan att igenkänna handskrivna siffror är på något sätt inbyggd i den mänskliga hjärnan, förutsatt förtrogenhet med siffrorna. En dator behöver emellertid tydliga regler för hur den skall kunna utöva igenkänningens konst.

Detta föredrag presenterar en euklidisk geometrisk modell för inlärning. Denna matematiska modell utgör grunden för många av dagens bästa högteknologiska verktyg för igenkänning med datorns hjälp.

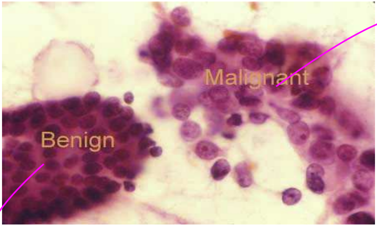
(Bröst)cancer eller inte ? Vävnadsprov

TVÅ KLUMPAR AV CELLER (FNA)

 [1]

[1] malignant= elakartad, benign= välartad

Cancer eller inte ? Mätdata



x_1
x_2
.
.
x_9
(M)

x_1
x_2
.
.
x_9
(B)

KTH Matematik

Inlärning från data & diagnos

Talen x_1, $x_2 \ldots, x_9$ i vektorn $\mathbf{x}$ representerar var sitt numeriskt givna drag hos cellklumparna: t.e.x

1. Clump thickness 2. Uniformity of cell size 3. Uniformity of cell shape 4. Marginal Adhesion 5. Single epithelial cell size 6. Bare nuclei 7. Bland chomatin 8. Normal nucleoli 9. Mitoses

$$\mathbf{x} = (x_1, x_2, \ldots, x_9)$$

KTH Matematik

Inlärning från data & diagnos

Vi har ett stort antal mätvektorer $\mathbf{x}$ av vävnadsprov, d.v.s. exempel, som av experter klassificerats som fall av elakartade eller välartade klumpar av celler.

- FRÅGA: Hur kan vi få en dator att utifrån dessa exempel lära sig att diagnosticera (rätt) dessa och nya fall av bröstcancer?

Svar

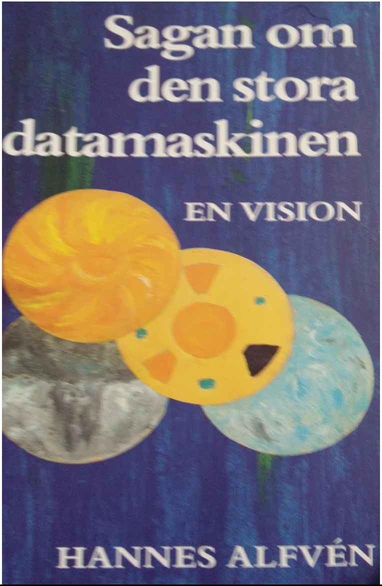
Frank Rosenblatt, amerikansk forskare 1928–1969, tog fram en allmän lösning (i själva verket en maskin) på problemet (ej inskränkt på automatisk diagnos av bröstcancer), och kallade sin *inlärningsmaskin* för **perceptron**.

- Rosenblatts lösning: Datorn skall tänka i en euklidisk geometri, eller precisare sagt, hitta ett *hyperplan* som skiljer vektorerna med elakartade fall från vektorerna med välartade fall. Det är givetvis inte alltid möjligt att göra så.

Ur Wikipedia:

In a 1958 press conference organized by the US Navy, Rosenblatt made statements about the perceptron that caused a heated controversy; based on Rosenblatt's statements, The New York Times reported the perceptron to be **"the embryo of an electronic computer that [the Navy] expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence."**

Hannes Alfvén, 1908—1995, professor i bl.a. plasmafysik vid KTH, Nobelpristagare i fysik 1970



KTH Matematik

R^n, hyperplan

R^n: reella vektorer med n (=dimension) komponenter.

$$\mathbf{x} = (x_1, x_2, \ldots, x_n) .$$

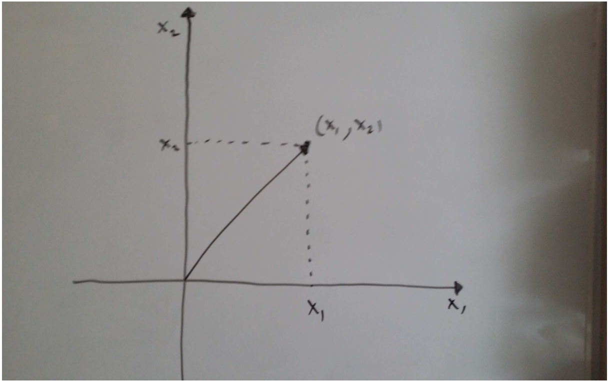
R^3 svarar mot det tredimensionella rummet som vi upplever genom sinnena. Hyperplan är ett vanligt plan i R^3.

R^2 är det bekanta (cartesiska) talplanet. En rät linje är ett hyperplan i R^2.

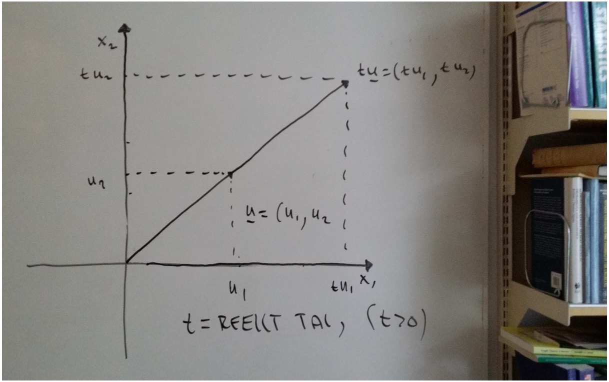
Låt

$$\mathbf{x} = (x_1, x_2)$$

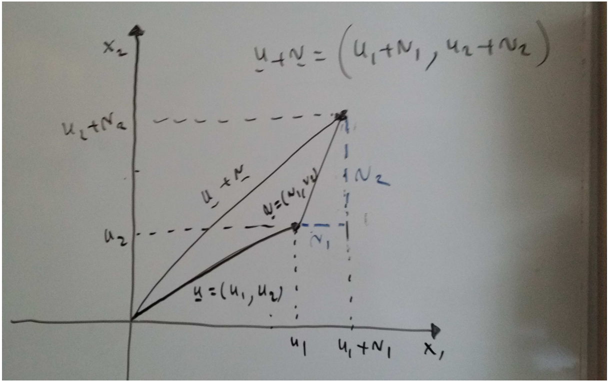
beteckna ett talpar (en vektor) i R^2.



t är ett reellt tal, $t\mathbf{u} = t(u_1, u_2) = (tu_1, tu_2)$.



$$\mathbf{u} = (u_1, u_2), \mathbf{v} = (v_1, v_2)\ , \mathbf{u} + \mathbf{v} = (u_1 + v_1, u_2 + v_2)$$



Geometri: $\mathbf{u} - \mathbf{v}$

$$\mathbf{u} = (u_1, u_2),\ \mathbf{v} = (v_1, v_2)\ ,$$

$$\mathbf{u} - \mathbf{v} = \mathbf{u} + (-1)\mathbf{v} = (u_1 - v_1, u_2 - v_2)$$

KTH Matematik

Rät linje i R^2

En rät linje är

$$w_2 x_2 + w_1 x_1 + b = 0$$

ty vi har ($w_2 \neq 0$)

$$x_2 = -\frac{w_1}{w_2} x_1 - \frac{b}{w_2}$$

Vi sätter

$$f(\mathbf{x}) \stackrel{\text{def}}{=} w_2 x_2 + w_1 x_1 + b.$$

Då definierar även

$$f(\mathbf{x}) = 0$$

en rät linje.

Geometri: 'dotprodukt'

$$\mathbf{w} = (w_1, w_2).$$

Dotproduktren är ett tal, som ges för ett par av vektorer enligt

$$\mathbf{w} \bullet \mathbf{x} \stackrel{\text{def}}{=} w_1 x_1 + w_2 x_2$$

Då kan man skriva

$$f(\mathbf{x}) = \mathbf{w} \bullet \mathbf{x} + b.$$

Vi har även (från definitionen) $\mathbf{w} \bullet (\mathbf{x} + \mathbf{z}) = \mathbf{w} \bullet \mathbf{x} + \mathbf{w} \bullet \mathbf{z}$.

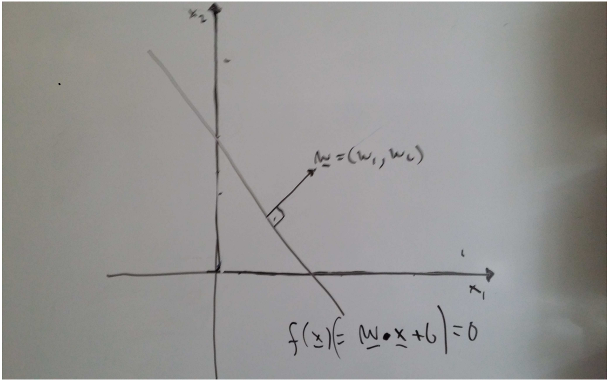
Anta att den räta linjen i planet går genom **z** och **x**, d.v.s.

$$f(\mathbf{z}) = \mathbf{w} \bullet \mathbf{z} + b = 0, f(\mathbf{x}) = \mathbf{w} \bullet \mathbf{z} + b = 0.$$

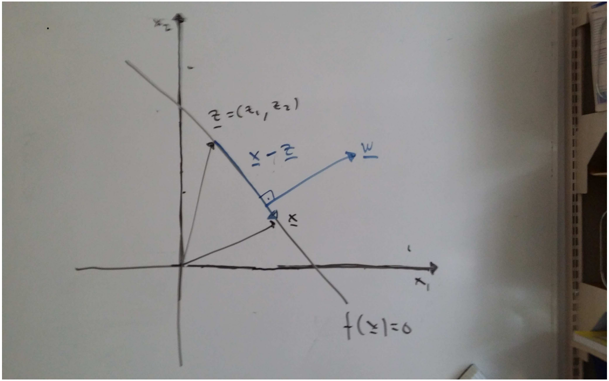
Då fås

$$\mathbf{w} \bullet (\mathbf{x} - \mathbf{z}) = \underbrace{\mathbf{w} \bullet \mathbf{x}}_{=-b} - \underbrace{\mathbf{w} \bullet \mathbf{z}}_{=-b} = -b + b = 0.$$

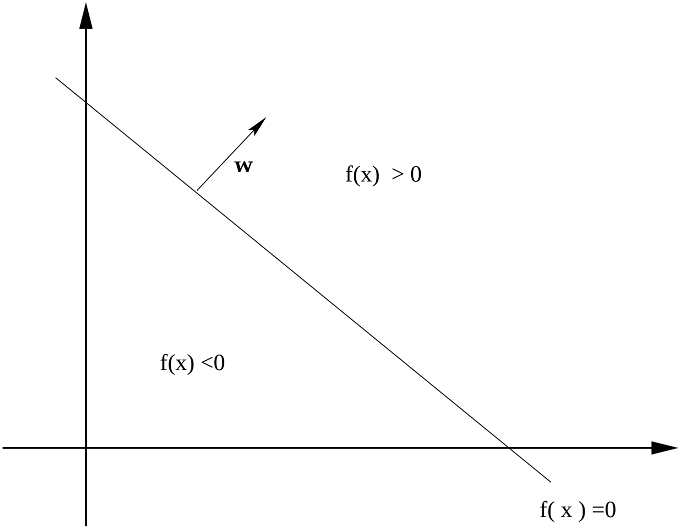
P.g.a. att $\mathbf{w} \bullet (\mathbf{x} - \mathbf{z}) = 0$ säger vi att **w** är **normalvektorn** till linjen $f(\mathbf{x}) = 0$ och ortogonal mot varje vektor som ligger i $f(\mathbf{x}) = 0$.


x_2
$w = (w_1, w_2)$
x_1
$f(x) (= w \cdot x + b) = 0$

Geometri: vektorerna i $f(\mathbf{x}) = 0$.



vi har en **positiv sida** och en **negativ sida** av linjen $f(\mathbf{x}) = 0$:



om

$$\mathrm{sign}(x) \stackrel{\mathrm{def}}{=} \begin{cases} +1 & 0 < x, \\ -1 & x < 0 \end{cases}$$

så har vi med

$$y = \mathrm{sign}\left(f\left(\mathbf{x}\right)\right)$$

en matematisk formel för att tillordna antingen -1 eller $+1$ som etikett till $\mathbf{x}$, beroende på om $\mathbf{x}$ ligger på den negativa sidan eller positiva sidan av linjen $f(\mathbf{x}) = 0$, resp.. Rosenblatts perceptron förverkligar denna idé, liksom skall ses.

Hjälpmedel: ländgen av en vektor, avståndet mellan vektorer

Längden av $\mathbf{x}$ betecknas med $\|\mathbf{x}\|$ och definieras som

$$\|\mathbf{x}\| \stackrel{\text{def}}{=} \sqrt{\mathbf{x} \bullet \mathbf{x}} = \sqrt{x_1^2 + x_2^2}$$

- En vektor $\mathbf{x}$ med längd $\|\mathbf{x}\| = 1$ kallas enhetsvektor.
- $\|\alpha\mathbf{x}\| = \sqrt{\alpha\mathbf{x} \bullet \alpha\mathbf{x}} = \sqrt{\alpha^2 \left(x_1^2 + x_2^2\right)} = |\alpha|\|\mathbf{x}\|$.
- $\|\mathbf{x} - \mathbf{z}\|$ är **avståndet** mellan $\mathbf{x}$ och $\mathbf{z}$,

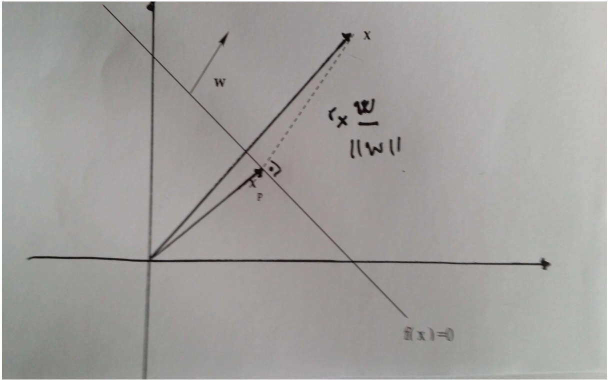
$$\|\mathbf{x} - \mathbf{z}\| = \sqrt{(x_1 - z_1)^2 + (x_2 - z_2)^2}$$

KTH VETENSKAP OCH KONST

KTH Matematik

$$\mathbf{x} = \mathbf{x}_p + r_x \cdot \frac{\mathbf{w}}{\|\mathbf{w}\|},$$

där r_x är ett tal som vi kommer att bestämma och $\frac{\mathbf{w}}{\|\mathbf{w}\|}$ är en enhetsvektor i $\mathbf{w}$:s riktning. $\mathbf{x}_p$ är den vinkelräta (=ortogonala) projektionen av $\mathbf{x}$ på $f(\mathbf{x}) = 0$.



Avstånd med tecken

$\mathbf{x} = \mathbf{x}_p + r_x \cdot \frac{\mathbf{w}}{\|\mathbf{w}\|}$. Vi beräknar det kvadrerade avståndet

$$\|\mathbf{x} - \mathbf{x}_p\|^2 = \|r_x \cdot \frac{\mathbf{w}}{\|\mathbf{w}\|}\|^2 = \left(\frac{r_x}{\|\mathbf{w}\|} w_1\right)^2 + \left(\frac{r_x}{\|\mathbf{w}\|} w_2\right)^2$$

$$= \frac{r_x^2}{\|\mathbf{w}\|^2}\left(w_1^2 + w_2^2\right)$$

vilket är liktydigt med att

$$\|\mathbf{x} - \mathbf{x}_p\|^2 = \frac{r_x^2}{\|\mathbf{w}\|^2}\|\mathbf{w}\|^2 = r_x^2$$

$$d.v.s.$$

$$\|\mathbf{x} - \mathbf{x}_p\| = |r_x|$$

Således är r_x **avståndet med tecken** $\pm$ från $\mathbf{x}$ till linjen $f(\mathbf{x}) = 0$.

KTH Matematik

Ytterligare, med

$$f(\mathbf{x}) = \mathbf{w} \bullet \mathbf{x} + b =$$

$$= \mathbf{w} \bullet \left(\mathbf{x}_p + r_x \cdot \frac{\mathbf{w}}{\|\mathbf{w}\|} \right) + b = \mathbf{w} \bullet \mathbf{x}_p + \mathbf{w} \bullet \left(\frac{r_x}{\|\mathbf{w}\|} \cdot \mathbf{w} \right) + b$$

$$= \underbrace{\mathbf{w} \bullet \mathbf{x}_p + b}_{=f(\mathbf{x}_p)} + \mathbf{w} \bullet \left(\frac{r_x}{\|\mathbf{w}\|} \cdot \mathbf{w} \right)$$

$$= \underbrace{f(\mathbf{x}_p)}_{=0} + \mathbf{w} \bullet \left(\frac{r_x}{\|\mathbf{w}\|} \cdot \mathbf{w} \right)$$

Avstånd med tecken

$$= \mathbf{w} \bullet \left(\frac{r_x}{\|\mathbf{w}\|} \cdot \mathbf{w} \right)$$

$$= \frac{r_x}{\|\mathbf{w}\|} \cdot \underbrace{\mathbf{w} \bullet \mathbf{w}}_{=\|\mathbf{w}\|^2} = r_x \|\mathbf{w}\|.$$

KTH Matematik

Avstånd med tecken

Med andra ord

$$f(\mathbf{x}) = r_x \|\mathbf{w}\|$$

d.v.s.

$$r_x = \frac{f(\mathbf{x})}{\|\mathbf{w}\|}$$

Därtill, om $\mathbf{0} = (0, 0)$, $f(\mathbf{0}) = \mathbf{w} \bullet \mathbf{0} + b = 0 + b = b$, d.v.s.,

$$r_0 = \frac{f(\mathbf{0})}{\|\mathbf{w}\|} = \frac{b}{\|\mathbf{w}\|}$$

är avståndet med tecken från planets origo till linjen $f(\mathbf{x}) = 0$.

KTH VETENSKAP OCH KONST

KTH Matematik

Avstånd med tecken

$$f(\mathbf{x}) = r_x \|\mathbf{w}\|$$

D.v.s $f(\mathbf{x}) > 0$ om $r_x > 0$ (den positiva sidan av $f(\mathbf{x}) = 0$) och $f(\mathbf{x}) < 0$ om $r_x < 0$ (den negativa sidan av $f(\mathbf{x}) = 0$).

Avstånd med tecken

- $$r_x = \frac{f(\mathbf{x})}{\|\mathbf{w}\|}$$

  m.a.o.

  $$f(\mathbf{x}) = r_x \|\mathbf{w}\|$$

- $$r_0 = \frac{f(\mathbf{0})}{\|\mathbf{w}\|} = \frac{b}{\|\mathbf{w}\|}$$

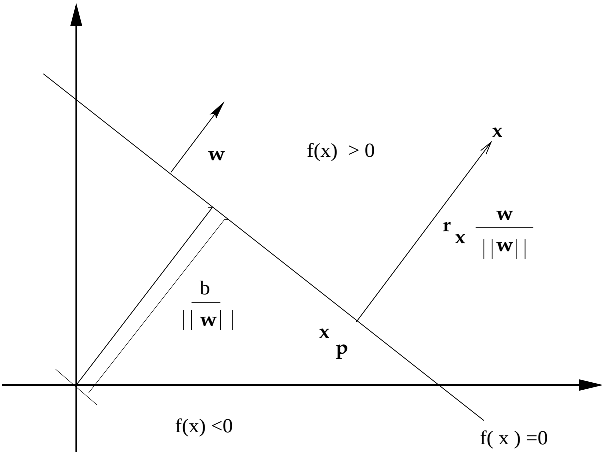
Parallellförflyttning av ett hyperplan

r_0 är avståndet med tecken från origo till linjen, och

$$r_0 = \frac{b}{\|\mathbf{w}\|}$$

Vi ser således att när b varieras (med fixt $\mathbf{w}$), så kommer hyperplanet att förflyttas parallellt. Om $b > 0$, då ligger $\mathbf{0}$ i positiva sidan, och om $b < 0$, så ligger $\mathbf{0}$ i den negativa sidan.

Inlärningens geometri



Exempel: En mängd av exempel i R^2

Mätdata & etiketterna

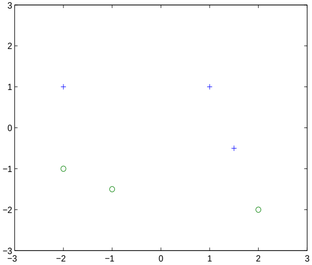
x_1	x_2	y
1	-0.5	+1
1	1	+1
-2	1	+1
-1	-1.5	-1
2	-2	-1
-2	-1	-1

KTH Matematik

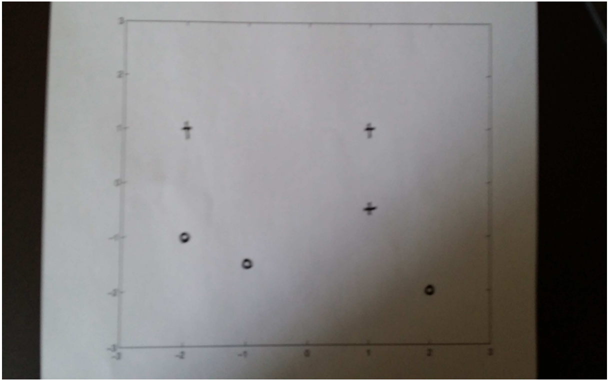
Exempel i R^2: $+1 \mapsto +$, -1 $\mapsto o$

Nummer l Mätdata $\mathbf{x}(l) = (x_1(l), x_2(l))$ etiketterna y

l	$x_1(l)$	$x_2(l)$	$y(l)$
1	1.5	-0.5	+1
2	1	1	+1
3	-2	1	+1
4	-1	-1.5	-1
5	2	-2	-1
6	-2	-1	-1



Da Capo



KTH Matematik

Vi vill nu hitta w_1, w_2 och b så att

$$f(\mathbf{x}(l)) = w_2 x_2(l) + w_1 x_1(l) + b$$

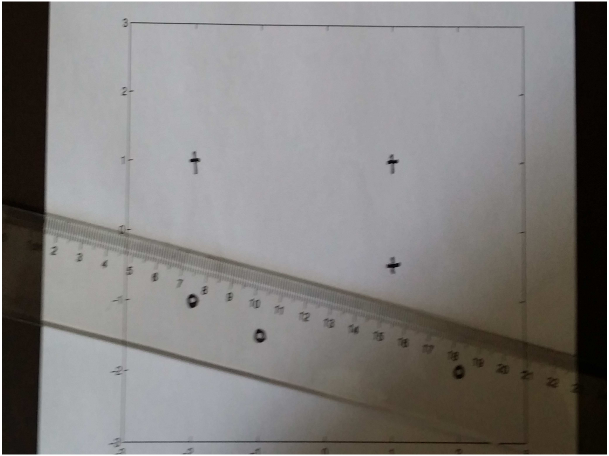
har rätt förtecken (och därmed rätt etikett) för varje $\mathbf{x}(l)$, d.v.s.

$$\operatorname{sign}(f(\mathbf{x}(l))) = y(l)$$

bör gälla för varje datapunkt i de exempel vi har.

Detta kallas *övervakad inlärning*.

En korkad frågeställning?

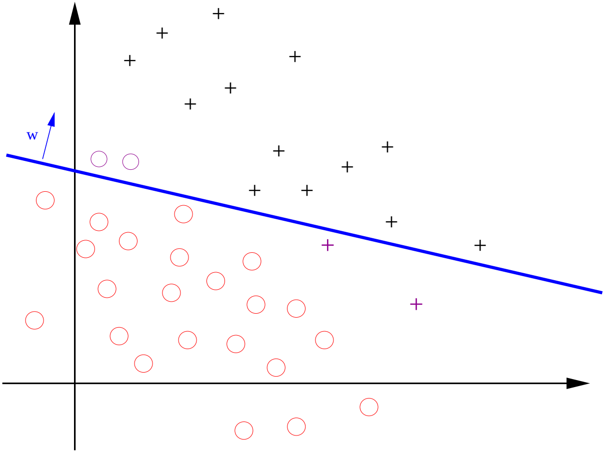


En korkad frågeställning?

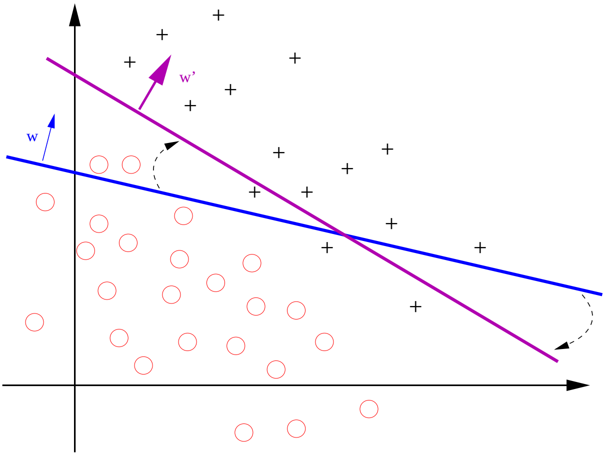
Jodå (kanske), men vi vill hitta en **algoritm** för placering av linjalen !

Algoritm: varje räkneförfarande efter ett givet schema kallas algoritm

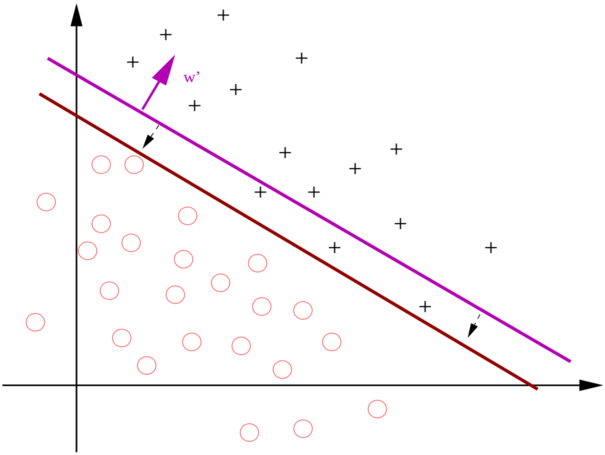
$\bigcirc \mapsto -1$ och $+ \mapsto +1$



Linjen ändrar riktning



Parallell förflyttning av linjen



En algoritm som gör dessa steg

- Välj startvärdena w_1, w_2 och b på måfå.
- För varje l: om $f(\mathbf{x}(l))$ har rätt förtecken med dessa vikter och bias, gör ingenting.
- Om $\text{sign}(f(\mathbf{x}(l))) \neq y(l)$, uppdatera bias och vikterna enligt följande (vi tar $\eta = 0.2$):

$$\begin{aligned} b &\leftarrow b + \eta y(l) \\ w_1 &\leftarrow w_1 + \eta y(l) x_1(l) \\ w_2 &\leftarrow w_2 + \eta y(l) x_2(l) \end{aligned}$$

- Fortsätt att gå genom $\mathbf{x}(l)$ tills alla har fått korrekt etikett.

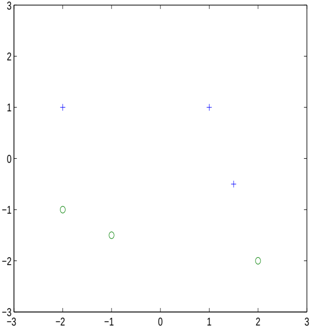
En allmän princip om uppdatering (numerisk metodik, statistik, signalbehandling m.m.)

$$\text{nya värdet} \leftarrow \text{ nuvarande värdet} + \text{korrektionen}$$

Korrektionen beror på data.

Geometri i verket

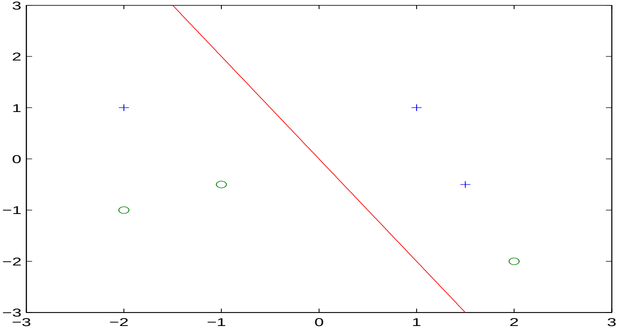
Vi väljer (på måfå) $b = 0, w_1 = 1, w_2 = 0.5$, d.v.s.,



$$f(\mathbf{x}) = 0.5x_2 + x_1$$

Exempel i R^2

Vi plottar datapunkterna samt den räta linjen $0.5x_2 + x_1 = 0$ d.v.s. $x_2 = -2x_1$



KTH Matematik

Exempel i R^2: uppdatering

Vi har fel i $l = 5$ med $\mathbf{x}(5) = (2, -2)$, $y(5) = -1$ Regeln

$$\begin{aligned} b &\leftarrow b + \eta y(5) \\ w_1 &\leftarrow w_1 + \eta y(5) x_1(5) \\ w_2 &\leftarrow w_2 + \eta y(5) x_2(5) \end{aligned}$$

ger med $b = 0, w_1 = 1, w_2 = 0.5, \eta = 0.2$

$$\begin{aligned} -0.2 &= 0 + 0.2 \cdot (-1) \\ 0.6 &= 1 + 0.2 \cdot (-1) \cdot 2 \\ 0.9 &= 0.5 + 0.2 \cdot (-1) \cdot (-2) \end{aligned}$$

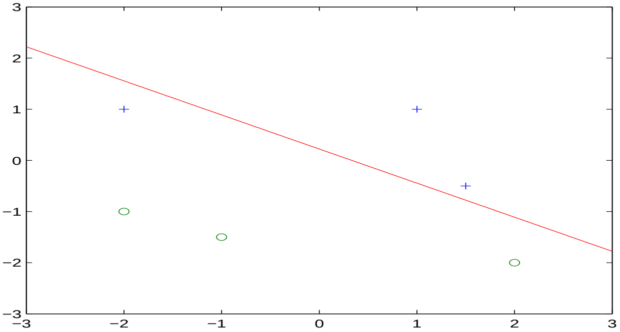
d.v.s. vi har en ny bias och de nya vikterna

$$f(\mathbf{x}) = 0.9 x_2 + 0.6 x_1 - 0.2$$

d.v.s. $x_2 = -\frac{0.6}{0.9} x_1 + \frac{0.2}{0.9}$.

Exempel i R^2: efter en första uppdatering

Datapunkterna och $x_2 = -\frac{0.6}{0.9}x_1 + \frac{0.2}{0.9}$



Exempel i R^2: uppdatering

Vi har fel i $I = 3$ med $\mathbf{x}(3) = (-2, 1)$, $y(3) = 1$ och med $b = -0.2, w_1 = 0.6, w_2 = 0.9, \eta = 0.2$

$$\begin{aligned} 0 &= 0.2 + (-0.2) \cdot 1 \\ 0.2 &= 0.6 + 0.2 \cdot 1 \cdot (-2) \\ 1.1 &= 0.9 + 0.2 \cdot 1 \cdot 1 \end{aligned}$$

d.v.s. vi har en ny bias och de nya vikterna

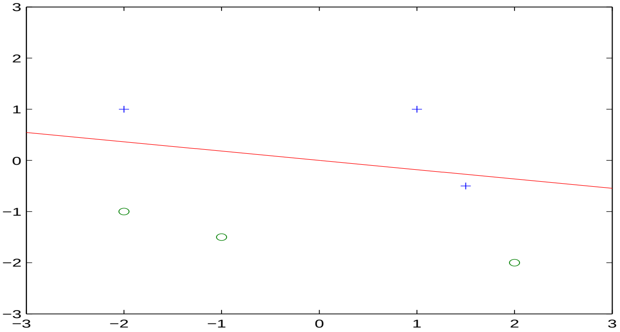
$$f(\mathbf{x}) = 1.1x_2 + 0.2x_1$$

d.v.s. $x_2 = -\frac{0.2}{1.1}x_1$.

KTH VETENSKAP OCH KONST
KTH Matematik

Exempel i R^2: andra uppdateringen

Datapunkterna och $x_2 = -\frac{0.2}{1.1}x_1$.



Exempel i R^2: uppdatering

Nu blev det fel i $l = 1$ med $\mathbf{x}(1) = (1.5, -0.5)$, $y(1) = 1$. Med $b = 0, w_1 = 0.2, w_2 = 1.1, \eta = 0.2$

$$\begin{aligned} 0.2 &= 0 + 0.2 \cdot 1 \\ 0.5 &= 0.2 + 0.2 \cdot 1 \cdot 1.5 \\ 1.0 &= 1.1 + 0.2 \cdot 1 \cdot (-0.5) \end{aligned}$$

d.v.s. vi har en ny bias och de nya vikterna

$$f(\mathbf{x}) = 1.0x_2 + 0.5x_1 + 0.2$$

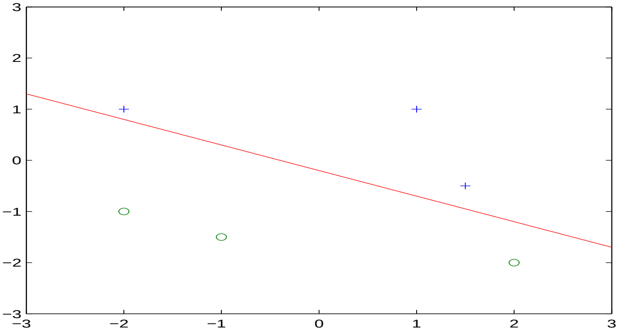
d.v.s. $x_2 = -0.5x_1 - 0.2$.

KTH VETENSKAP OCH KONST

KTH Matematik

Exempel i R^2:tredje uppdateringen

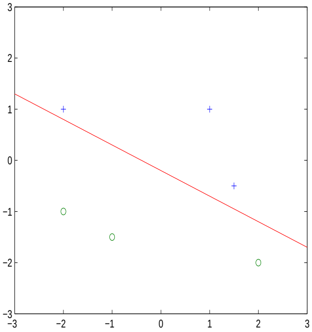
Datapunkterna och $x_2 = -0.5x_1 - 0.2$.

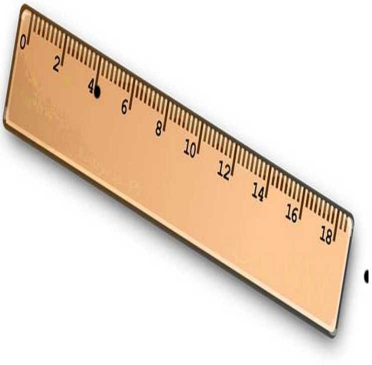


KTH VETENSKAP OCH KONST
KTH Matematik

Exempel i R^2: felfritt ! Algoritmen stannar !

Datapunkterna och $x_2 = -0.5x_1 - 0.2$.



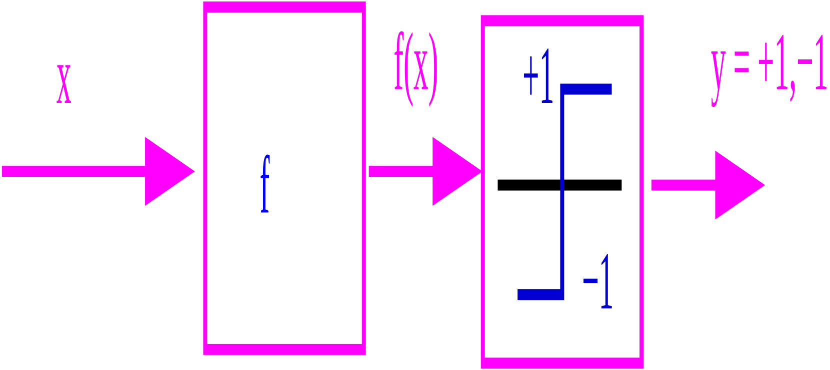


Algoritmen stannar ? En matematisk fråga

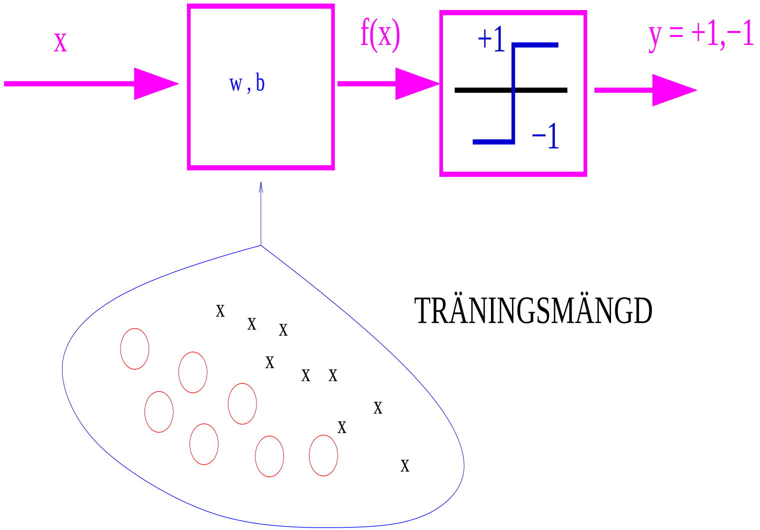
- Hur är det ? Kommer den ovan applicerade algoritmen alltid att stanna eller var detta en tillfällighet ?
- Innan algoritmen stannat, kan vi uppskatta hur länge vi måste hålla p å tills den stannar?

SVAREN ges i slutet av föredraget.

Systemdiagram



INLÄRNING



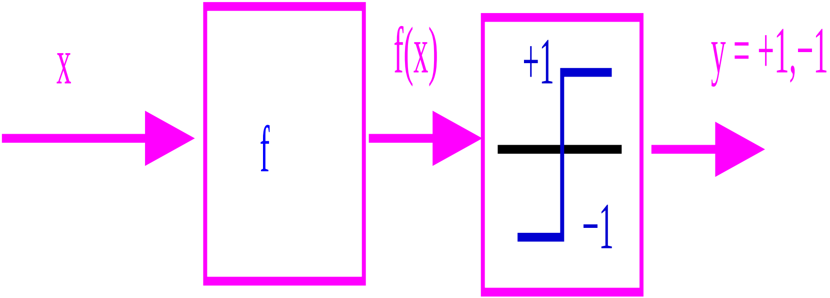
Exemplen

Nummer l Mätdata $\mathbf{x}(l) = (x_1(l), x_2(l))$ etiketterna y

l	$x_1(l)$	$x_2(l)$	$y(l)$
1	1.5	-0.5	+1
2	1	1	+1
3	-2	1	+1
4	-1	-1.5	-1
5	2	-2	-1
6	-2	-1	-1

har ovan **lagrats** med 'trial and error' i
$b = 0.2, w_1 = 0.5, w_2 = 1.0.$

I DRIFT EFTER INLÄRNING



Vi matar in en **ny** vektor **x** av mätdata, ut kommer en diagnos: t.ex., elakartad, säg $+1$, eller välartad, -1

LINEAR LEARNING MACHINE a.k.a. PERCEPTRON (ur Wikipedia)

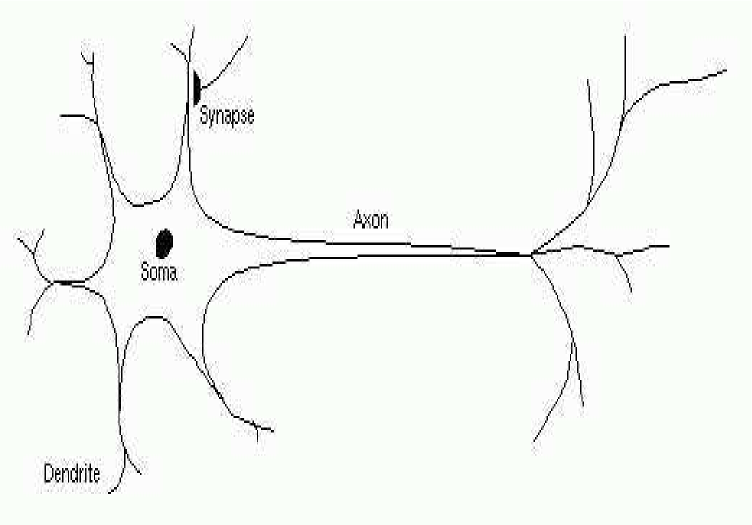
Frank Rosenblatt developed in 1957 the perceptron, i.e.,

$$y = \operatorname{sign}\left(f\left(\mathbf{x}\right)\right), \quad y \in \{+1, -1\}$$

as a model & and learning algorithm to understand human memory, learning, and cognition. The perceptron was based on biological ideas about networks of neurons in the brain.

KTH Matematik

Bioneuron



Mark I Perceptron (loc.cit)

In 1960 Rosenblatt demonstrated at Cornell University a piece of electromechanical hardware, named **Mark I Perceptron**, the first machine that could learn to recognize and identify optical patterns. This was the first computer that could learn new skills by trial and error, using a type of neural network that simulates human thought processes.

Hannes Alfvéns superdator

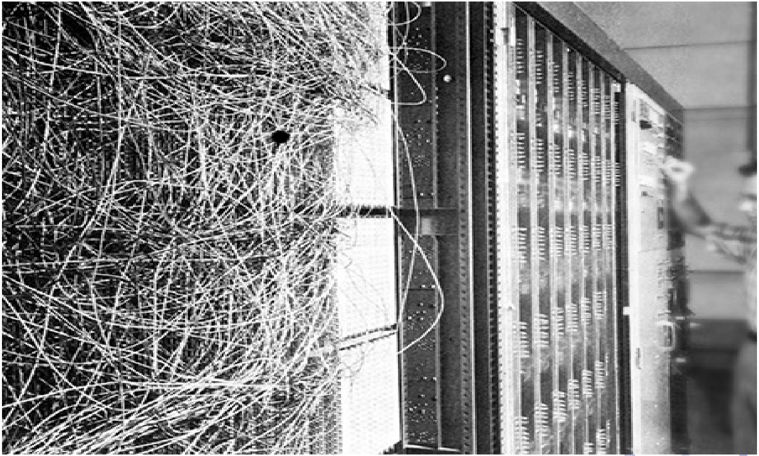
Hannes Alfvén *Sagan om den stora datamaskinen. En vision.* (1966). Pilgrim press Stockholm, 1987. En författarröst från en avlägsen framtid berättar om superdatorernas utveckling:

> Men allteftersom neurofysiologien har lärt oss uppbyggnaden och funktionen av (neuronerna) har kretsar av liknande typ införts i (super)datorerna. sid. 127
> ... människornas nervsystem och hjärna ... dess mest förnämliga egenskaper (inspirerar) till förbättring av (super)datorernas struktur.   sid. 132

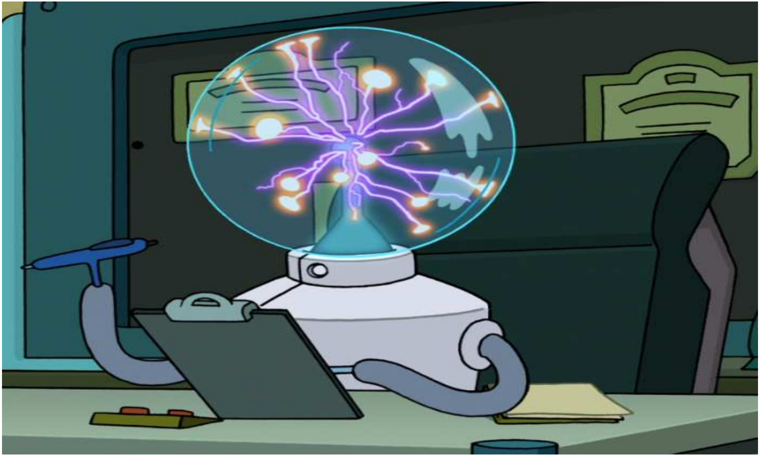
KTH Matematik

The Mark 1 Perceptron

This machine was designed for image recognition: it had an array of 400 photocells, randomly connected to the "neurons". Weights were encoded in potentiometers, and weight updates during learning were performed by electric motors.

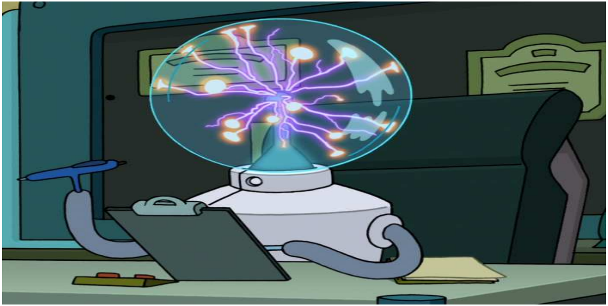


Dr. Perceptron



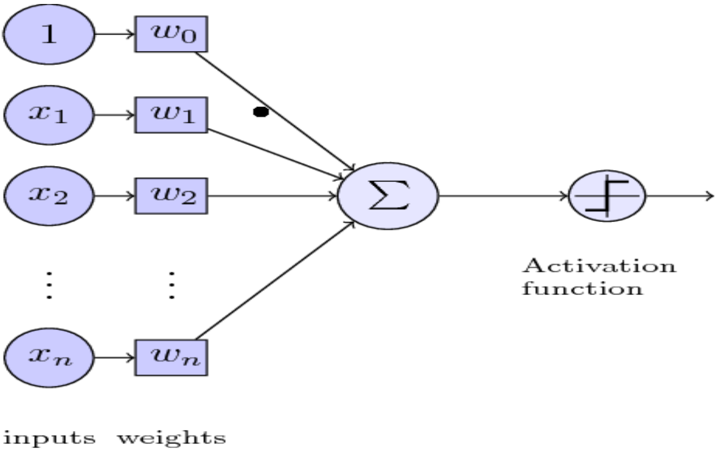
Dr. Perceptron is a doctor of Freudian Circuit Analysis, and works at the HAL Institute For Criminally Insane Robots.

Dr. Perceptron

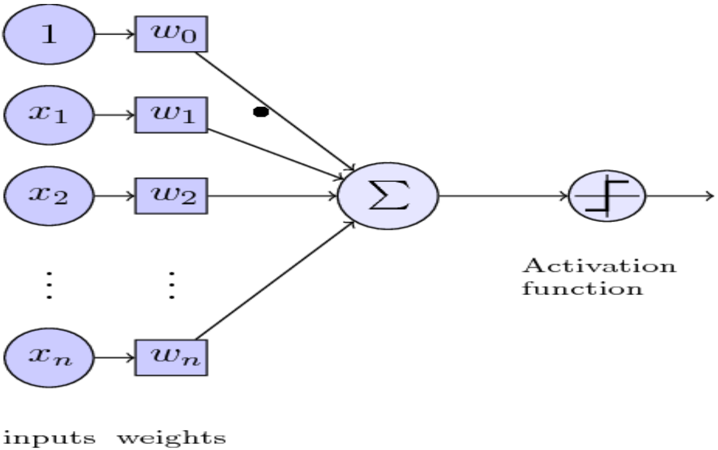


Freudian Circuit Analysis = freudiansk kretsanalys $\leftrightarrow$ psykoanalys

Dr. Perceptron: det allmänna fallet



Dr. Perceptron: det allmänna fallet



1. Clump thickness 2. Uniformity of cell size 3. Uniformity of cell shape 4. Marginal Adhesion 5. Single epithelial cell size 6. Bare nuclei 7. Bland chomatin 8. Normal nucleoli 9. Mitoses

0.2000	0.2000	0.5000	0.5000	0.5000	0.2000	0.3000	1.0000	1.0000
0.4000	0.1000	0.1000	0.1000	0.4000	0.3000	0.3000	0.5000	0.5000
0.9000	0.1000	0.1000	0.1000	0.1000	0.6000	0.3000	0.1000	0.7000
0.6000	0.8000	0.1000	0.1000	0.1000	0.1000	0.8000	0.1000	0.1000
0.8000	1.0000	0.7000	0.1000	0.2000	0.2000	0.2000	0.4000	0.2000
0.3000	0.8000	0.6000	0.6000	0.2000	0.1000	0.1000	0.1000	0.1000
0.1000	0.1000	0.9000	1.0000	0.4000	0.1000	0.2000	0.3000	0.2000
0.8000	0.2000	0.1000	0.7000	0.7000	0.7000	0.3000	0.1000	0.1000
0.1000	1.0000	0.1000	0.1000	1.0000	0.7000	1.0000	0.1000	0.1000
0.1000	0.1000	0.1000	0.1000	0.1000	0.7000	1.0000	0.3000	0.1000

Övergång till R^n

$$\mathbf{w} = (w_1, w_2, \ldots, w_n), \mathbf{x} = (x_1, x_2, \ldots, x_n)$$

$$\mathbf{w} \bullet \mathbf{x} \stackrel{\text{def}}{=} w_1 x_1 + w_2 x_2 + \ldots + w_n x_n$$

$$f(\mathbf{x}) = \mathbf{w} \bullet \mathbf{x} + b$$

KTH Matematik

Dotprodukt i R^n

$$\mathbf{x} \bullet \mathbf{y} = x_1 y_1 + \ldots + x_n y_n = \sum_{i=1}^{n} x_i y_i$$

Matchning av två vektorer.

R^n med dotprodukt kan åskådliggöras med bekanta geometriska ideer, som linje, punkt, plan, vinkel och kallas därför ett euklidiskt rum.

Euklides av Alexandria
f. ung. 325 f.Kr.
d. ung. 265 f.Kr.

Norm

- **längden** av $\mathbf{x}$ är

$$\|\mathbf{x}\| = \sqrt{\mathbf{x} \bullet \mathbf{x}} = \sqrt{x_1^2 + \ldots + x_n^2}$$

KTH Matematik

Hyperplan

Ett *hyperplan* är ett affint[2] underrum av dimension $n-1$, som delar upp R^n i två halvrum. Vi betraktar följande *hyperplan* $D \subset R^n$:

$$D = \{\mathbf{x} \in R^n \mid f(\mathbf{x}) = 0\}$$

w är igen ortogonal mot varje vektor in D.

[2]=parallel till ett linjärt underrum, ett linjärt underrum innehåller **0**

KTH VETENSKAP OCH KONST
KTH Matematik

Halvrum

$$f(\mathbf{x}) = \mathbf{w} \bullet \mathbf{x} + b$$

$$D = \{\mathbf{x} \in R^n \mid f(\mathbf{x}) = 0\}$$

Hyperplanet D delar upp R^n i två halvrum $\mathcal{R}_1$ och $\mathcal{R}_2$:

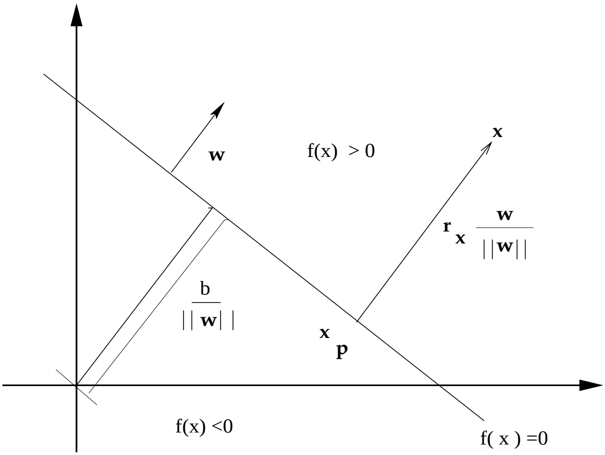
$$\mathcal{R}_1 = \{\mathbf{x} \in R^n \mid f(\mathbf{x}) > 0\}$$

$$\mathcal{R}_2 = \{\mathbf{x} \in R^n \mid f(\mathbf{x}) < 0\}$$

Som ovan är **w** ortogonal mot varje vektor i D och pekar ut mot $\mathcal{R}_1$.

KTH Matematik

Inlärningens geometri är densamma !



$$f\left(\mathbf{x}\right) = \mathbf{w} \cdot \mathbf{x} + b$$

där **w** kallas igen en *viktvektor*, och b är ett reellt tal som igen kallas *bias* (eller tröskel).

TRÄNING AV EN PERCEPTRON

- Låt $\mathbf{x}_i \in R^n$, $y_i \in Y = \{+1, -1\}$, $i = 1, \ldots, m$ vara m par av *exempel* . Vi kallar

  $$\mathcal{S} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_m, y_m)\}$$

  en *träningsmängd*.
- Nu gäller det uppenbarligen att bestämma vikterna $\mathbf{w}$ och bias b utifrån exemplen i träningsmängden.
- 'självprogrammerande dator' ? Datorn lär sig regelbundenheterna i träningsmängden och lagrar dessa i $\mathbf{w}$ och b.

Övervakad inlärning med perceptroner

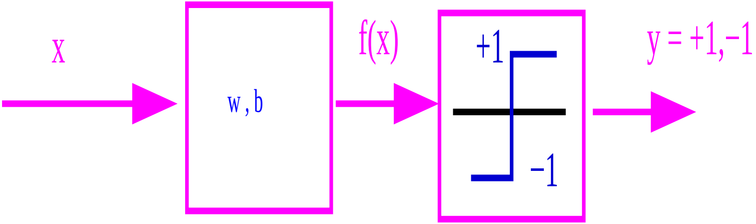
Bestäm $(\mathbf{w}, b)$ m.h.a. $\mathcal{S}$ så att $f(\mathbf{x}_i) = y_i$ för all i, d.v.s., varje vektor i träningsmängden är rätt klassificerad.

1. Finns det en lösning till denna uppgift ?
2. Om en lösning existerar, finns det en algoritm som hittar lösningen i ett ändligt antal steg ?

EN PERCEPTRON

En perceptron är därmed en enhet som tar en vektor $\mathbf{x}$ och beräknar $y \in \{+1, -1\}$ enligt regeln $y = \mathrm{sign}(f(\mathbf{x}))$, d.v.s.

- Om $\mathbf{x} \in \mathcal{R}_1$ så är $y = +1$
- Om $\mathbf{x} \in \mathcal{R}_2$ så är $y = -1$



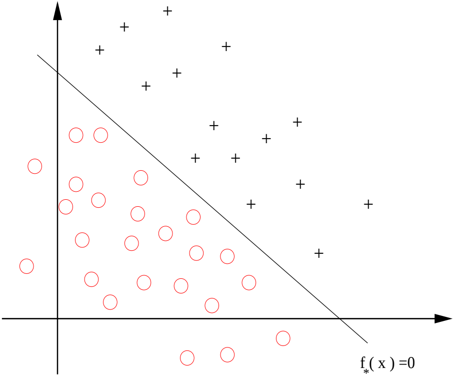
Vi antar att det finns (ett för oss tillsvidare okänt) hyperplan sådant att det finns $(\mathbf{w}_*, b_*)$ sådana att om

$$f_*(\mathbf{x}) = \mathbf{w}_* \bullet \mathbf{x} + b_*$$

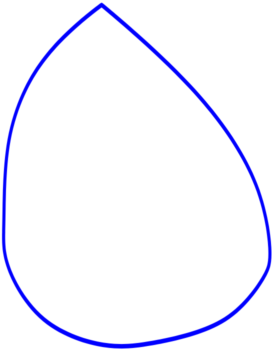
så gäller det för alla $(\mathbf{x}_i, y_i) \in \mathcal{S}$ att $f_*(\mathbf{x}) > 0$ om och endast om $y_i = +1$.

När detta antagande är sant, kallas $\mathcal{S}$ *linjärt separabel*.

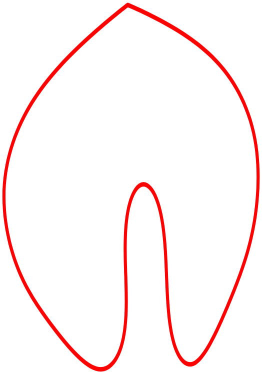
Exemplen representeras av punkter med två etiketter $\circ$ och $+$. Alla $+$ ligger i $\mathcal{R}_1$ och alla $\circ$ ligger i $\mathcal{R}_2$ med avseende på hyperplanet $f_*(\mathbf{x}) = 0$.



Konvexitet



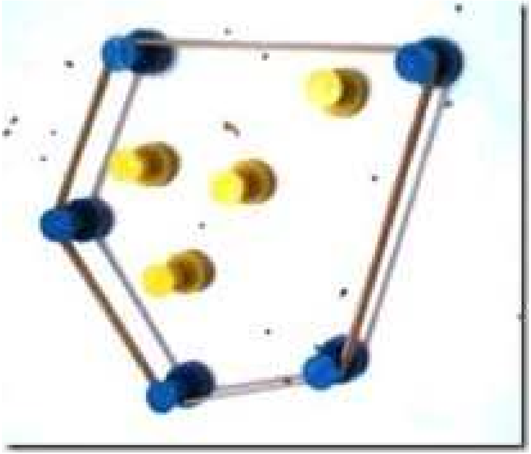
KONVEX



ICKE-KONVEX

Konvext hölje

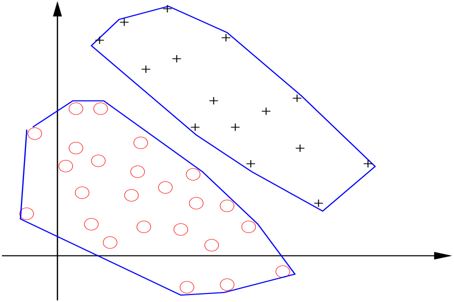
Vi erhåller ett konvext område (=konvext hölje för ett ändligt antal punkter) genom att spänna ett gummiband runt punktmängden.



Vi erhåller två konvexa områden genom att spänna två gummiband i blått runt två punktmängder. De **konvexa höljena** är i bilden disjunkta. Vi har följande matematiska sats:

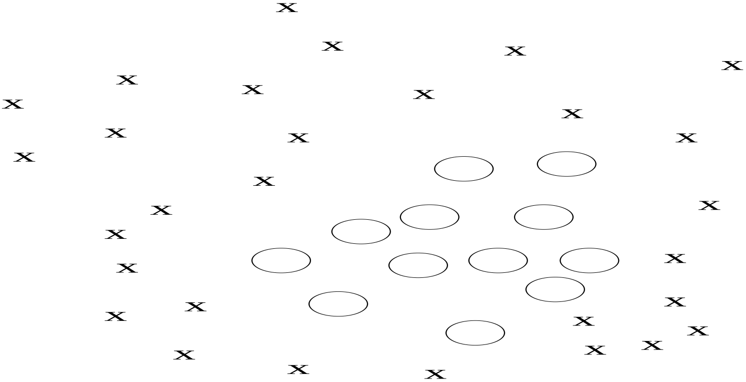
Två disjunkta konvexa mängder är linjärt separabla.

Alltså råder linjär separabilitet för de konvexa höljena i bilden, och därför finns ett hyperplan som separerar punktmängderna i bilden.



ETT ICKE-SEPARABELT FALL

Det konvexa höljet av punkterna med etiketterna x innehåller det konvexa höljet av punkterna med etiketterna ∘. Denna träningsmängd kan inte separeras av ett hyperplan utan att ett antal fel uppstår.



Rosenblatts perceptron algoritm

- Vi tar på måfå en viktvektor $\mathbf{w}$ och en bias b.
- Sedan sveper vi genom exemplen i träningsmängden med den motsvarande perceptronen.
- För varje felaktigt klassificerat exempel ändras hyperplanets riktning och/eller hyperplanet förflyttas parallellt.

Varje gång något exempel felklassificeras under ett svep, uppdateras viktvektorn genom att addera en term som är proportionell mot felmarginalen (med tecken).

- Om $\mathbf{w}_k$ och b_k är sådana att

$$y_i \left(\mathbf{w}_k \bullet \mathbf{x}_i + b_k \right) = y_i f \left(\mathbf{x}_i \right) > 0$$

  så är $\mathbf{x}_i$ rätt klassificerad.

Således ändras hyperplanets riktning genom att addera till viktvektorn en vektor som är proportionell mot $y_i f \left(\mathbf{x}_i \right)$, så snart som $y_i f \left(\mathbf{x}_i \right)$ är negativt.

Perceptron: konvergensteoremet

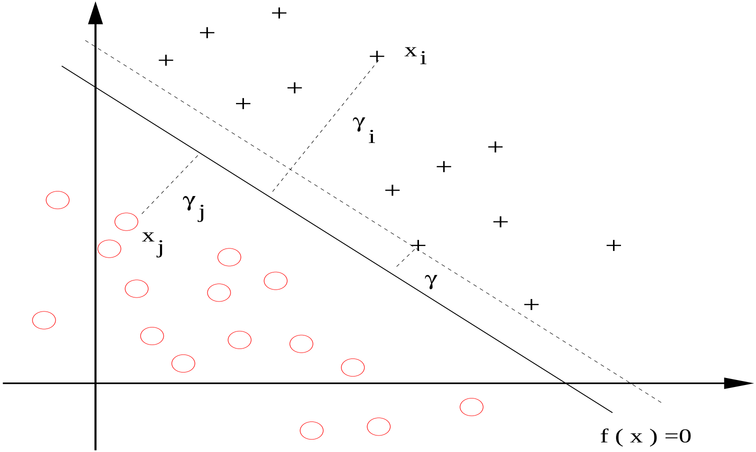
Algoritmen körs (svepet upprepas) tills inga exempel i träningsmängden är felklassificerade.

Perceptronkonvergens Det kan bevisas att algoritmen alltid kommer att stanna efter ett ändligt antal steg, förutsatt att vi har en linjärt separabel träningsmängd.

Rosenblatts perceptron algoritm

$\mathbf{w}_0 \leftarrow \mathbf{0},\ b_0 \leftarrow 0,\ k \leftarrow 0\ \mathcal{R} \leftarrow \max_{1 \le i \le l} \|\mathbf{x}_i\|$
repeat
    for $i = 1$ to l **do**
        if $y_i \left(\mathbf{w}_k \bullet \mathbf{x}_i + b_k\right) \le 0$ **then**
            $\mathbf{w}_{k+1} \leftarrow \mathbf{w}_k + y_i \mathbf{x}_i,$
            $b_{k+1} \leftarrow b_k + y_i \mathcal{R}^2,$
            $k \leftarrow k + 1$
        end if
    end for
until no mistakes made in the loop
return $\mathbf{w}_k, b_k$, where k is the number of mistakes.

Geometrisk marginal



Konvergensteoremet för perceptroner

Låt $\mathcal{S}$ vara en icke-trivial linjärt separabel träningsmängd och låt för alla i

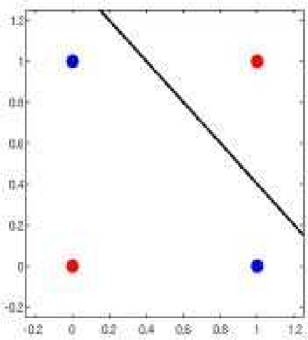
$$\gamma_i = y_i \mathbf{w}^* \bullet \mathbf{x}_i + b^* \geq \gamma$$

där $\|\mathbf{w}^*\| = 1$.
Då är antalet 'repeats' i Rosenblatts perceptron algoritm högst

$$\left(\frac{\mathcal{R}^2}{\gamma}\right)$$

KTH Matematik

Ett enkelt fall där en perceptron inte kan fungera



Vikterna efter träningen

Algoritmen arbetade genom att addera till en tidigare viktvektor en felklassficerad träningsvektor $\mathbf{x}_i$ multiplicerad med motsvarande y_i. Efter att algoritmen har stannat har vi således

$$\mathbf{w}^* = \sum_{i=1}^{m} a_i y_i \mathbf{x}_i$$

där a_i är antalet gånger felklassificering av x_i har föranlett en ändring av **w** under inlärningen. Vi kan skriva

$$\mathbf{w}^* = \sum_{i=1|y_i=+1}^{m} a_i \mathbf{x}_i - \sum_{i=1|y_i=-1}^{m} a_i \mathbf{x}_i$$

Vikterna efter träningen

Med

$$\mathbf{w}^* = \sum_{i=1|y_i=+1}^{m} a_i \mathbf{x}_i - \sum_{i=1|y_i=-1}^{m} a_i \mathbf{x}_i$$

fås

$$f^*(\mathbf{x}) = \mathbf{w}^* \bullet \mathbf{x} + b^*$$

$$= \sum_{i=1|y_i=+1}^{m} a_i (\mathbf{x}_i \bullet \mathbf{x}) - \sum_{i=1|y_i=-1}^{m} a_i (\mathbf{x}_i \bullet \mathbf{x}) + b^*$$

KTH Matematik

Associativt minne

Med andra ord, träningsmängden har lagrats i vikterna, enligt

$$f^*(\mathbf{x}) = \sum_{i=1|y_i=+1}^{l} a_i (\mathbf{x}_i \bullet \mathbf{x}) - \sum_{i=1|y_i=-1}^{l} a_i (\mathbf{x}_i \bullet \mathbf{x}) + b^*$$

Varje ny mätvektor $\mathbf{x}$ **matchas** med $\mathbf{x}_i$ i varje exempel, och matchningarna viktas och summeras över alla exempel. Jämförelsen av skillnaden mellan de positiva och negativa exemplen med bias b^* (tröskel) avgör resultatet.

Associativt minne ?

Matchning mot de lagrade vektorerna gör att vi vill tala om ett *associativt minne*

$$f^{*}(\mathbf{x}) = \sum_{i=1 | y_i=+1}^{l} a_i \left(\mathbf{x}_i \bullet \mathbf{x}\right) - \sum_{i=1 | y_i=-1}^{l} a_i \left(\mathbf{x}_i \bullet \mathbf{x}\right) + b^{*}$$

KTH Matematik

Support Vector Machines m.m.

Det finns nyare, kraftfullare utvecklingar och utvidgningar av perceptronen, som klarar av icke-separabla problem. En av dessa, som kallas på engelska *support vector machines*, tar avstamp i en viss mening på formeln

$$f^*(\mathbf{x}) = \sum_{i=1|y_i=+1}^{l} a_i(\mathbf{x}_i \bullet \mathbf{x}) - \sum_{i=1|y_i=-1}^{l} a_i(\mathbf{x}_i \bullet \mathbf{x}) + b^*$$

De nya inlärningsmaskiner som förbättrar perceptronernas egenskaper kallas neurala nätverk,

KTH Matematik

Megatron (senare Galvatron)

Visioner

”the embryo of an electronic computer that will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.”

Hannes Alfvén: ” (Människans) verkliga storhet ligger i att hon är den levande varelse som var intelligent nog för att inse att (evolutionens) mål är datorn. ” sid. 17 i a.a.

KTH Matematik

Slut på föreläsningen - tack !