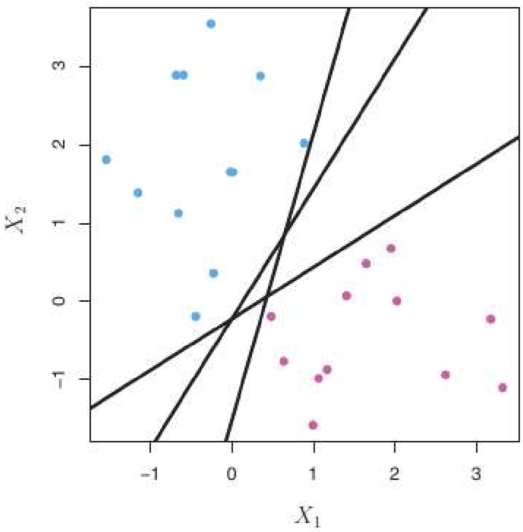
5 Metoda podpornih vektorjev

(angl. Support Vector Machines, SVM)

Predpostavimo, da imamo dvo-razsežne linearno ločljive podatke: atributa x_1 in x_2, ter razred $y \in \{-1, 1\}$. Obstaja veliko načinov na katere lahko ločimo oba razreda. Tri možnosti so narisane na sliki [1]Fig. 8 (modre in rdeče točke predstavljajo $y = 1$ in $y = -1$).



Slika 8: Vizualizacija učne množice z dvema razredoma, ($y = 1$ modre in $y = -1$ rdeče točke)

Zgornje odločitveno pravilo drži za poljubno ločnico; za klasifikator si želimo najboljšo od možnih ločnic. Intuitivno je najboljša ločnica tista, od katere so tako negativni kot pozitivni primeri kar najbolj oddaljeni.

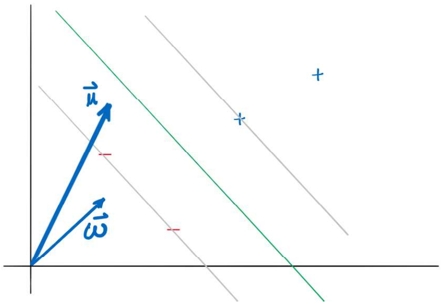
5.1 Klasifikator z največjim razmikom

Decision boundaries (naveži na LDA, QDA)
Vapnik:

- pozitivne in negativne primere želimo ločiti z ravno črto
- tako, da naredimo kar se da širok pas med - in +
- naredimo vektor $\bar{\omega}$, ki naj bo pravokoten na sredinsko črto našega pasu
- in vektor $\bar{u}$ (unknown), do neklasificiranega primera

[1]Povzeto po An Introduction to Statistical Learning, James G. et. al.

- zanima nas, ali $\bar{u}$ pripada točkam na desni ali na levi strani pasu 9



Slika 9: Ločnica z najširšim pasom med pozitivnimi in negativnimi primeri.

Na kateri strani leži $\bar{u}$?
Pogledamo skalarni produkt z $\bar{\omega}$: $\bar{\omega} \cdot \bar{u} \geq c$
To bo naše odločitveno pravilo, ki ga lahko brez izgube za splošnost nekoliko preoblikujemo, pri čemer velja ($c = -b$):

$$\bar{\omega} \cdot \bar{u} + b \geq 0 \Longrightarrow + \tag{1}$$

Ne poznamo vrednosti b in $\bar{\omega}$.
Lahko rečemo, da za vsak pozitivni primer velja:

$$\bar{\omega} \cdot \bar{x}_+ + b \geq 1$$

in za vsak negativni primer velja:

$$\bar{\omega} \cdot \bar{x}_- + b \leq 1$$

S tem smo skalirali širino pasu med -1 in 1.
Matematično je priročno, da uvedemo novo spremenljivko y_i, tako da velja:

- $y_i = +1$ za pozitivne primere
- $y_i = -1$ za negativne primere

Zgornji enačbi na levi pomnožimo z y_i:

$$y_i(\bar{\omega} \cdot \bar{x}_i + b) \geq 1$$

$$y_i(\bar{\omega} \cdot \bar{x}_i + b) \geq 1$$

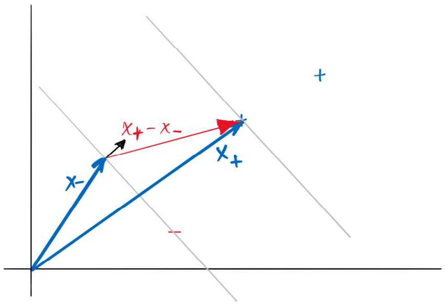
In dobimo isto enačbo:

$$y_i(\bar{\omega} \cdot \bar{x}_i + b) - 1 \geq 0$$

Za primere na robu tega pasu bo veljalo:

$$y_i(\bar{\omega} \cdot \bar{x}_i + b) - 1 = 0 \tag{2}$$

Kaj želimo doseči? Določiti želimo tako ločnico med pozitivnimi in negativnimi primeri, da bo pas okrog nje čim širši.



Slika 10: Ločnica z najširšim pasom med pozitivnimi in negativnimi primeri.

Širina pasu je (slika 10):

$$(x_+ - x_-) \cdot \frac{\bar{\omega}}{||\bar{\omega}||}$$

Iz enačbe 2 za pozitivne primere x_+ in $y_i = 1$ sledi:

$$x_+ \cdot \bar{\omega} = 1 - b$$

in podobno za negativne ($y_i = -1$) sledi:

$$-x_- \cdot \bar{\omega} = 1 + b$$

Če tako izražena x_+ in x_- nesemo v izraz za širino pasu, dobimo

$$\frac{2}{||\omega||}.$$

To želimo **maksimizirat** in to je isto kot:

$$\max \frac{1}{||\omega||},$$

kar je isto kot

$$\min ||\omega||$$

in isto, a matematično bolj priročno:

$$\min \frac{1}{2}||\omega||^2. \tag{3}$$

Kako minimiziramo izraz (3) pri omejitvah (2)?
Če iščemo ekstrem funkcije z omejitvami, uporabimo Lagrangeove multiplikatorje; konstruiramo nov izraz, katerega ekstreme poiščemo brez omejitev:

$$L = \frac{1}{2}||\bar{\omega}||^2 - \sum \alpha_i[y_i(\bar{\omega} \cdot \bar{x}_i + b) - 1] \tag{4}$$

Kaj se v L lahko spreminja? Spreminjata se lahko $\bar{\omega}$ in b.
Odvajajmo po obeh:

$$\frac{\partial L}{\partial \bar{\omega}} = \bar{\omega} - \sum \alpha_i y_i \bar{x}_i = 0 \Longrightarrow \bar{\omega} = \sum \alpha_i y_i \bar{x}_i \tag{5}$$

$$\frac{\partial L}{\partial b} = -\sum \alpha_i y_i = 0 \Longrightarrow \sum \alpha_i y_i = 0 \tag{6}$$

Vstavimo izračunan $\bar{\omega}$ v L:

$$L = \frac{1}{2}(\sum \alpha_i y_i \bar{x}_i)(\sum \alpha_j y_j \bar{x}_j) - (\sum \alpha_i y_i \bar{x}_i) \cdot (\sum \alpha_j y_j \bar{x}_j) - \sum \alpha_i y_i b + \sum \alpha_i$$

$$L = \sum \alpha_i - \frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j (\bar{x}_i \cdot \bar{x}_j) \tag{7}$$

Vsota $\sum \alpha_i y_i b$ je enaka 0 (glej 6).
Minimizirat poskušamo ta izraz, ki je očitno **odvisen samo od skalarnega produkta $\bar{x}_i \cdot \bar{x}_j$**.
In tudi **odločitveno pravilo je odvisno samo od skalarnega produkta**:

$$\sum \alpha_i y_i (\bar{x}_i \cdot \bar{u}) + b \geq 0 \Longrightarrow + \tag{8}$$

Če primeri niso linearno ločljivi v originalnem prostoru, jih morda lahko preslikamo v drug prostor, kjer bodo.
Rabimo preslikavo, npr. $\Phi(\bar{x})$, ki $\bar{x}$ preslika v nov prostor.
In tam potrebujem samo skalarni produkt $\Phi(\bar{x}) \cdot \Phi(\bar{y})$.
Za klasifikacijo prav tako. Če imam torej funkcijo

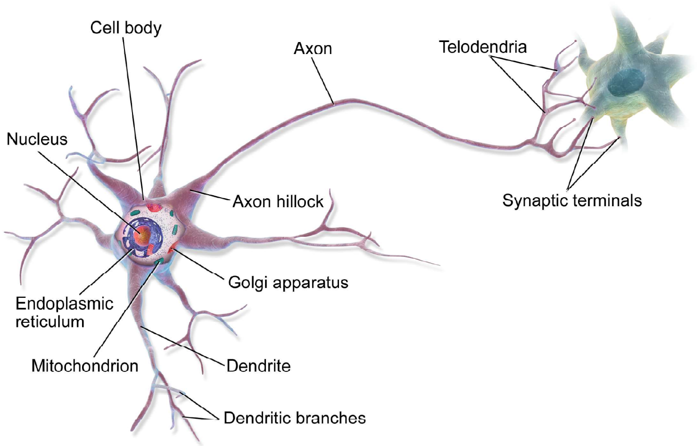
$$K(x_i, x_j) = \Phi(\bar{x}) \cdot \Phi(\bar{y}),$$

ki ji rečem jedro (*angl. kernel*), funkcije Φ sploh ne rabimo.
Popularna jedra:

- linearno jedro: $(\bar{u} \cdot \bar{v} + 1)^n$
- radial basis: $\exp(-\frac{||x_i - x_j||}{\sigma})$

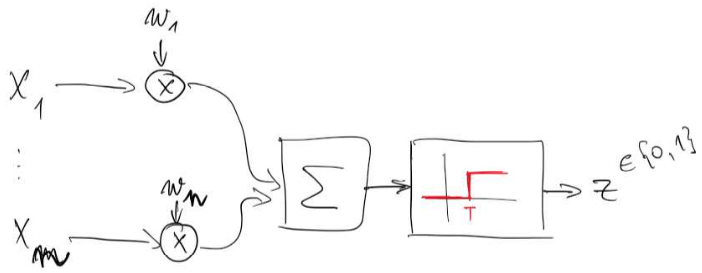
6 Nevronske mreže

Ideja simulacije človeških možganov za izdelavo inteligentnih strojev sega v same začetke področja umetne inteligence. Nevron (slika 11) oz. živčna celica je osnovni gradnik možganov. Glavno vprašanje začetnikov umetne inteligence je bilo: Kako modelirati nevron v računalniku oz. kako posnemati njegovo delovanje in narediti umetne možgane?



Slika 11: Shema nevrona[2].

Vhodne binarne spremenljivke $x_1, \ldots, x_n$ pomnožimo z utežmi $w_1, \ldots, w_n$, ki modelirajo moč sinaptične povezave, nato pa seštejemo skupen vpliv vhodnih spremenljivk. Ali je skupen vpliv dovolj močan, da nevron prevaja signal? Signal pošljemo čez stopničasto funkcijo, ki ima vrednost 0 za vrednosti $\leq T$ in 1 za vrednosti nad pragom T. Izhod je $z \in \{0, 1\}$.



Slika 12: Umetni nevron.

Nevroni (umetni) se med seboj povezujejo v strukturo, ki ji pravimo **nevronska**

[2]Vir: Bruce Blaus (CC 3.0), en.wikipedia.org/wiki/Neuron.

mreža. Lahko si jo predstavljamo kot škatlo, polno nevronov, uteži in pragov; na eni strani imamo vhode v škatlo x_i, na drugi strani pa izhode z_i. Nevronska mreža je pravzaprav funkcijski aproksimator:

$$\bar{z} = f(\bar{x}, \bar{w}, \bar{\tau})$$

Na vzorcu podatkov dobimo želeni izhod $\bar{d}$ kot funkcijo vhoda: $\bar{d} = g(\bar{x})$.
Konstruirajmo cenilno funkcijo

$$P = -||\bar{d} - \bar{z}||^2.$$

Uteži $\bar{w}$ in pragove $\bar{\tau}$) želimo tako, da bo cenilna funkcija dosegla maksimum.
Gradient ascent:
Parcialno odvajajmo P po utežeh in izračunajmo najboljšo smer premika:

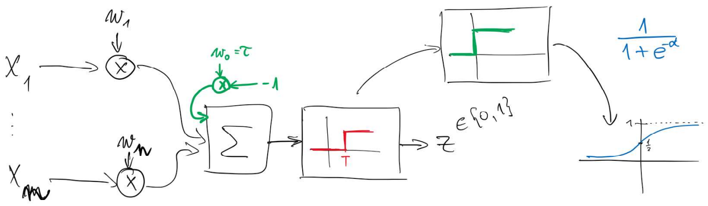
$$\Delta\bar{w} = r(\frac{\partial P}{\partial w_1} i + \frac{\partial P}{\partial w_2} j).$$

Zakaj tole ne bo šlo:

- lokalni ekstremi
- nezvezna funkcija (stopnica); 25 let nihče ni vedel, kako odpraviti to pomanjkljivost (šele leta 1974 ideja o sigmoidni funkciji)

Dodajmo utež, ki je kar konstanta -1 pomnožena z $w_0 = \tau$.13. To stopnico pomakne v levo, do vrednosti 0. Na ta način smo se znebili pragov in imamo samo še opravka z utežmi:

$$\bar{z} = f(\bar{x}, \bar{w})$$

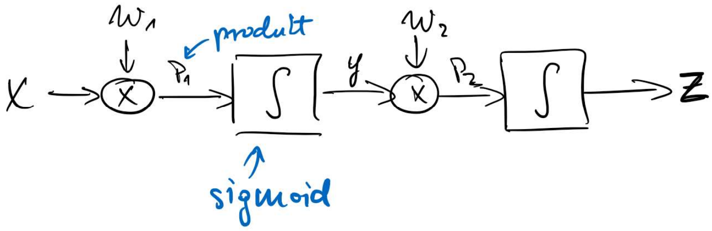


Slika 13: Umetni nevron.

$$P = -\frac{1}{2}||\bar{d} - \bar{z}||^2$$

$$\frac{\partial P}{\partial w_2} = \frac{\partial P}{\partial z}\frac{\partial z}{\partial P_2}\frac{\partial P_2}{\partial w_2}$$

$$\frac{\partial P}{\partial w_1} = \frac{\partial P}{\partial z}\frac{\partial z}{\partial P_2}\frac{\partial P_2}{\partial y}\frac{\partial y}{\partial P_1}\frac{\partial P_1}{\partial w_1}$$



Slika 14: Umetni nevron.