

Perceptron

The New Yorker, December 6, 1958 P. 44

Talk story about the perceptron, a new electronic brain which hasn't been built, but which has been successfully simulated on the I.B.M. 704. Talk with Dr. Frank Rosenblatt, of the Cornell Aeronautical Laboratory, who is one of the two men who developed the prodigy; the other man is Dr. Marshall C. Yovits, of the Office of Naval Research, in Washington. Dr. Rosenblatt defined the perceptron as the first non-biological object which will achieve an organization o its external environment in a meaningful way. It interacts with its environment, forming concepts that have not been made ready for it by a human agent. If a triangle is held up, the perceptron's eye picks up the image & conveys it along a random succession of lines to the response units, where the image is registered. It can tell the difference betw. a cat and a dog, although it wouldn't be able to tell whether the dog was to the left or right of the cat. Right now it is of no practical use, Dr. Rosenblatt conceded, but he said that one day it might be useful to send one into outer space to take in impressions for us.

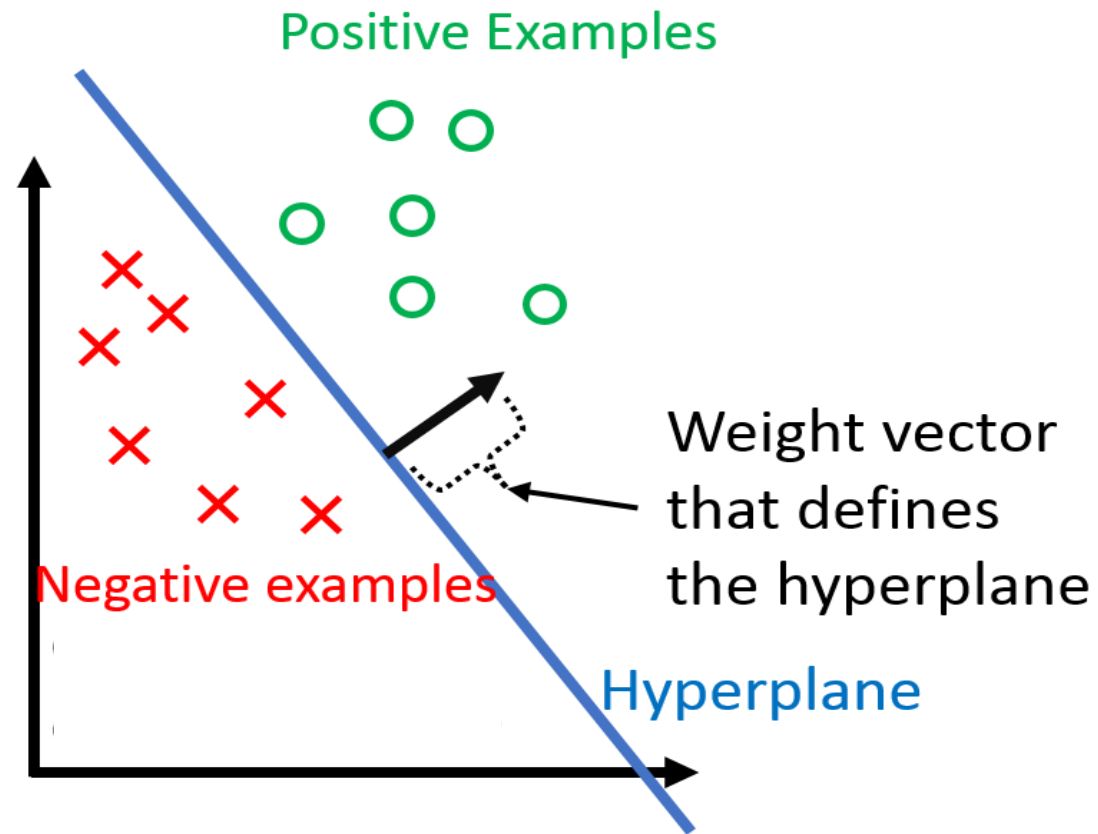
F. Rosenblatt. The perceptron : a probabilistic model for information storage and organization in the brain. *Psychological Reviews*, 65:386–408, 1958.
Wiederabdruck in [AR88], S. 92–114

In 1969, Minsky and Papert published a book called “Perceptrons” that analysed what they could do and showed their limitations.

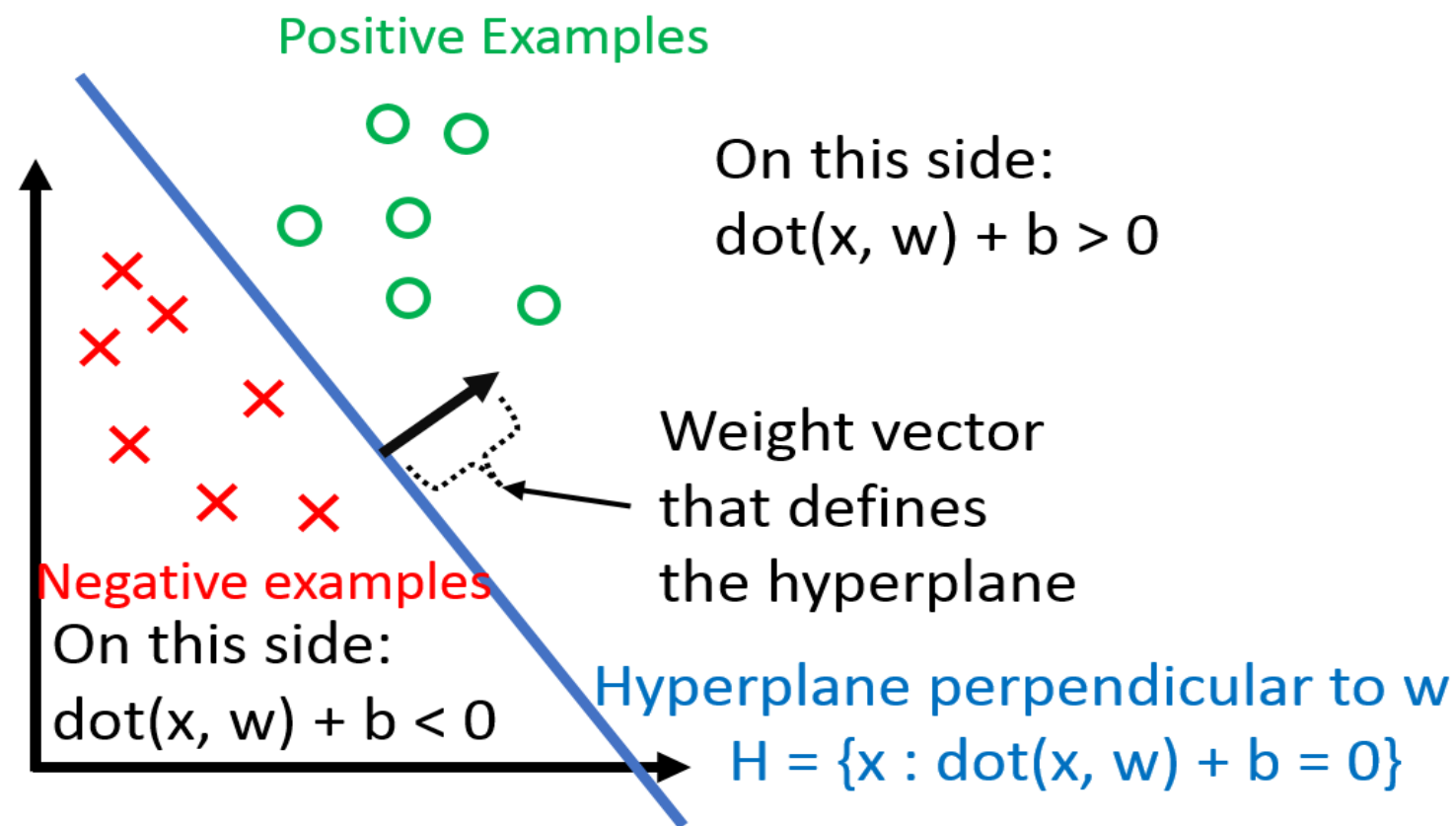
Many people thought these limitations applied to all neural network models.

Hiperravan

$$\mathcal{H} = \{\mathbf{x} \mid \mathbf{w}^T \mathbf{x} + b = 0\}$$



Hiperravan



$$\mathcal{H} = \{\mathbf{x} \mid \mathbf{w}^T \mathbf{x} + b = 0\}$$

Određivanje klase nove
tačke:

$$\mathbf{w}^T \mathbf{x}_i + b > 0 \rightarrow y_i = +1$$

$$\mathbf{w}^T \mathbf{x}_i + b < 0 \rightarrow y_i = -1$$

Obučavanje perceptrona

- Obučavanje perceptrona se svodi na određivanje jednačine hiperravni.
- Određujemo vektor težina w i bias b .
- Učenje bajasa istovremeno sa učenjem težina?

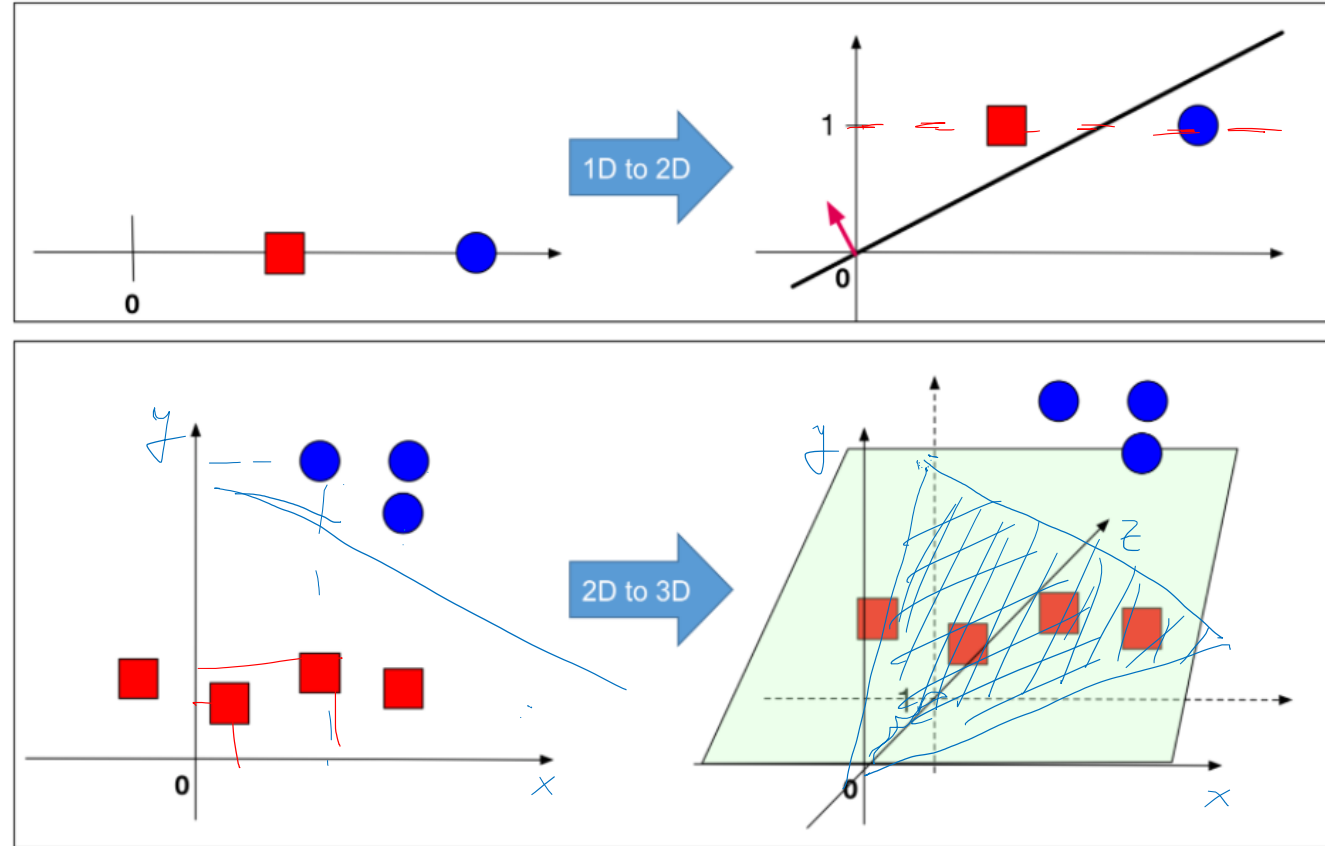
Apsorpcija bajasa u vektor težina

$\mathbf{x}$ postaje $\begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix}$, a $\mathbf{w}$ postaje $\begin{bmatrix} \mathbf{w} \\ b \end{bmatrix}$

$$\begin{bmatrix} \mathbf{w} \\ b \end{bmatrix}^T \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix} = \mathbf{w}^T \mathbf{x} + b$$

$$\mathcal{H} = \{ \mathbf{x} \mid \mathbf{w}^T \mathbf{x} = 0 \}$$

Promena dimenzija prostora podataka



$$\begin{bmatrix} x \\ 1 \end{bmatrix}$$

(Left:) The original data is 1-dimensional (top row) or 2-dimensional (bottom row). There is no hyper-plane that passes through the origin and separates the red and blue points.
(Right:) After a constant dimension was added to all data points such a hyperplane exists.

Algoritam perceptrona

$$w^T x$$

- Uzimamo primer po primer iz trening skupa.
 - Ako perceptron daje očekivani izlaz, težine veza se ne menjaju.
 - Ako je dobijen izlaz manji od 0, a očekivan je pozitivan primer, vektor težina se menja tako što mu se dodaje vektor ulaza.

$$w_{new} = w_{old} + x$$

- Ako je dobijen izlaz veći od nule, a očekivan je negativan primer, vektor težina se menja tako što se umanjuje za vektor ulaza.

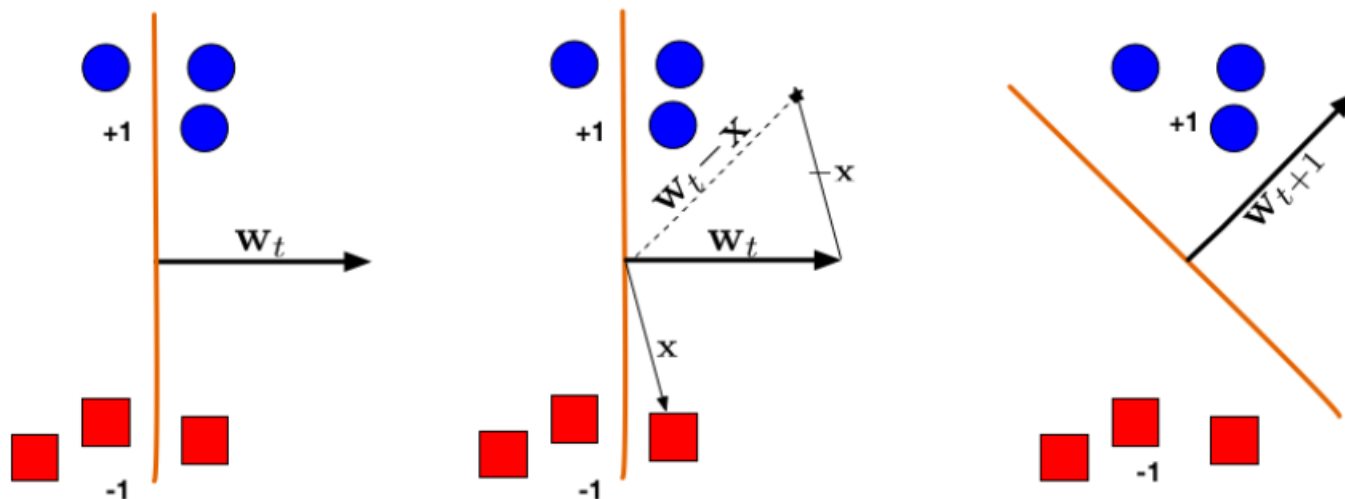
$$w_{new} = w_{old} - x$$

- U slučaju da su primeri iz trening skupa linearno separabilni, prethodna procedura će sigurno naći rešenje

Perceptron Algorithm

```
Initialize  $\vec{w} = \vec{0}$  // Initialize  $\vec{w}$ .  $\vec{w} = \vec{0}$  misclassifies everything.
while TRUE do // Keep looping
     $m = 0$  // Count the number of misclassifications,  $m$ 
    for  $(x_i, y_i) \in D$  do // Loop over each (data, label) pair in the dataset,  $D$ 
        if  $y_i(\vec{w}^T \cdot \vec{x}_i) \leq 0$  then // If the pair  $(\vec{x}_i, y_i)$  is misclassified
             $\vec{w} \leftarrow \vec{w} + y\vec{x}$  // Update the weight vector  $\vec{w}$ 
             $m \leftarrow m + 1$  // Counter the number of misclassification
        end if
    end for
    if  $m = 0$  then // If the most recent  $\vec{w}$  gave 0 misclassifications
        break // Break out of the while-loop
    end if
end while // Otherwise, keep looping!
```

Vizuelizacija algoritma



*Illustration of a Perceptron update. (Left:) The hyperplane defined by $\mathbf{w}_t$ misclassifies one red (-1) and one blue (+1) point. (Middle:) The red point $\mathbf{x}$ is chosen and used for an update. Because its label is -1 we need to **subtract** $\mathbf{x}$ from $\mathbf{w}_t$. (Right:) The updated hyperplane $\mathbf{w}_{t+1} = \mathbf{w}_t - \mathbf{x}$ separates the two classes and the Perceptron algorithm has converged.*

Geometrijska interpretacija algoritma perceptrona preko ugla između vektora $\mathbf{x}$ i $\mathbf{w}$

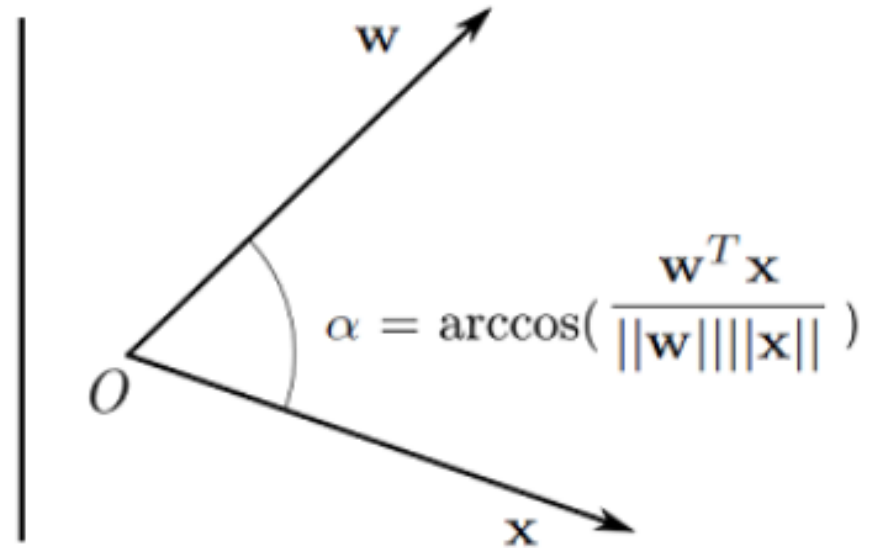
$$\mathbf{w} = [w_0, w_1, w_2, \dots, w_n]$$

$$\mathbf{x} = [1, x_1, x_2, \dots, x_n]$$

$$\mathbf{w} \cdot \mathbf{x} = \mathbf{w}^T \mathbf{x} = \sum_{i=0}^n w_i * x_i$$

$$\mathbf{w}^T \mathbf{x} = \|\mathbf{w}\| \|\mathbf{x}\| \cos \alpha$$

$$\cos \alpha = \frac{\mathbf{w}^T \mathbf{x}}{\|\mathbf{w}\| \|\mathbf{x}\|}$$

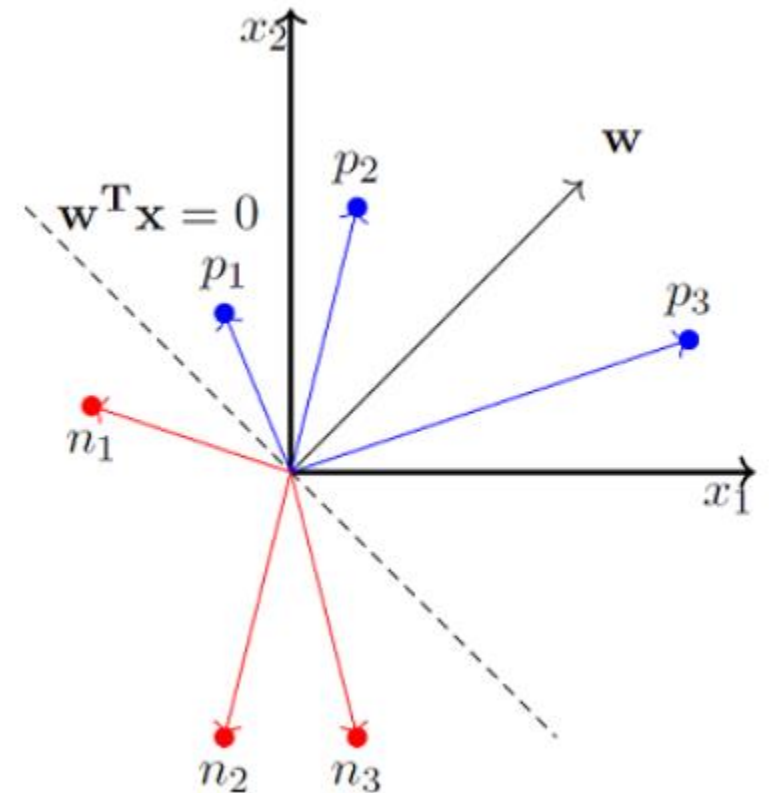


Geometrijska interpretacija algoritma perceptrona preko ugla između vektora $\mathbf{x}$ i $\mathbf{w}$

$$\cos \alpha = \frac{\mathbf{w}^T \mathbf{x}}{\|\mathbf{w}\| \|\mathbf{x}\|} \quad \left| \quad \cos \alpha \propto \mathbf{w}^T \mathbf{x} \right.$$

So if $\mathbf{w}^T \mathbf{x} > 0 \Rightarrow \cos \alpha > 0 \Rightarrow \alpha < 90$

Similarly, if $\mathbf{w}^T \mathbf{x} < 0 \Rightarrow \cos \alpha < 0 \Rightarrow \alpha > 90$



Zašto algoritam za obučavanje perceptrona funkcioniše?

Ako je x pozitivan primer, koji je pogrešno klasifikovan:

- Dodavanjem x na w kosinus ugla α se povećava, što znači da se ugao α smanjuje.
- Cilj je da $\alpha < 90^\circ$, pa nas smanjivanje ugla vodi ka ostvarenju cilja.

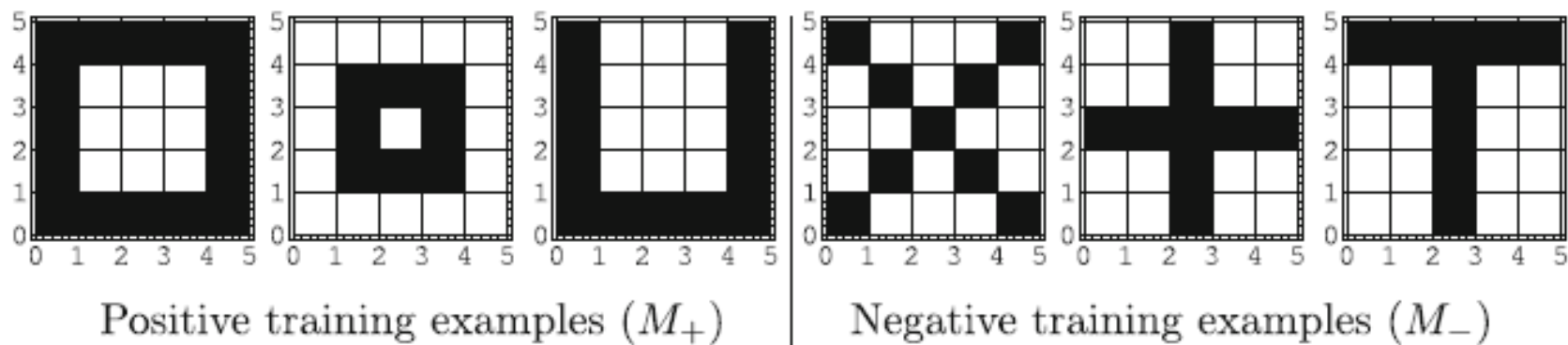
(α_{new}) when $\mathbf{w}_{new} = \mathbf{w} + \mathbf{x}$

$$\begin{aligned}\cos(\alpha_{new}) &\propto \mathbf{w}_{new}^T \mathbf{x} \\ &\propto (\mathbf{w} + \mathbf{x})^T \mathbf{x} \\ &\propto \mathbf{w}^T \mathbf{x} + \mathbf{x}^T \mathbf{x} \\ &\propto \cos\alpha + \mathbf{x}^T \mathbf{x} \\ \cos(\alpha_{new}) &> \cos\alpha\end{aligned}$$

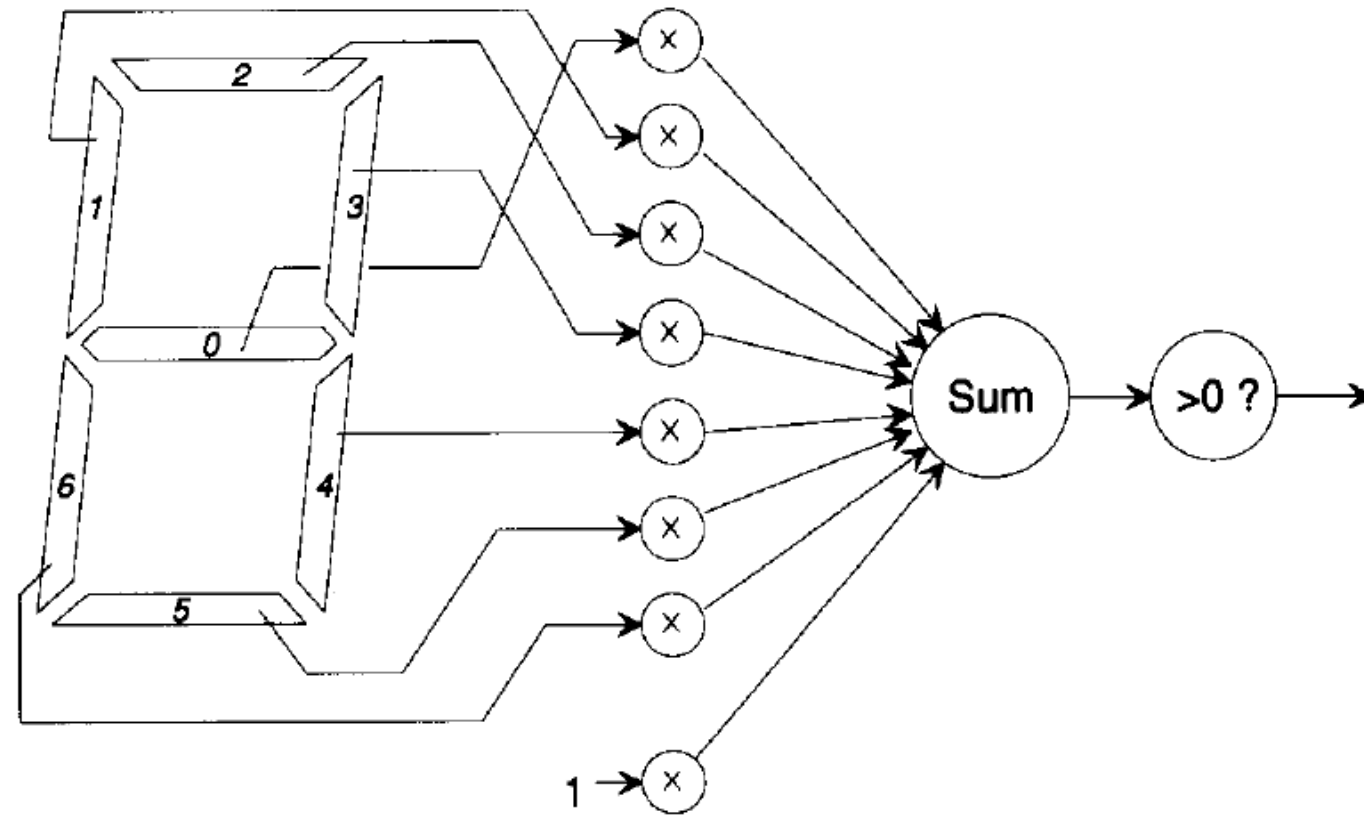
(α_{new}) when $\mathbf{w}_{new} = \mathbf{w} - \mathbf{x}$

$$\begin{aligned}\cos(\alpha_{new}) &\propto \mathbf{w}_{new}^T \mathbf{x} \\ &\propto (\mathbf{w} - \mathbf{x})^T \mathbf{x} \\ &\propto \mathbf{w}^T \mathbf{x} - \mathbf{x}^T \mathbf{x} \\ &\propto \cos\alpha - \mathbf{x}^T \mathbf{x} \\ \cos(\alpha_{new}) &< \cos\alpha\end{aligned}$$

Primer



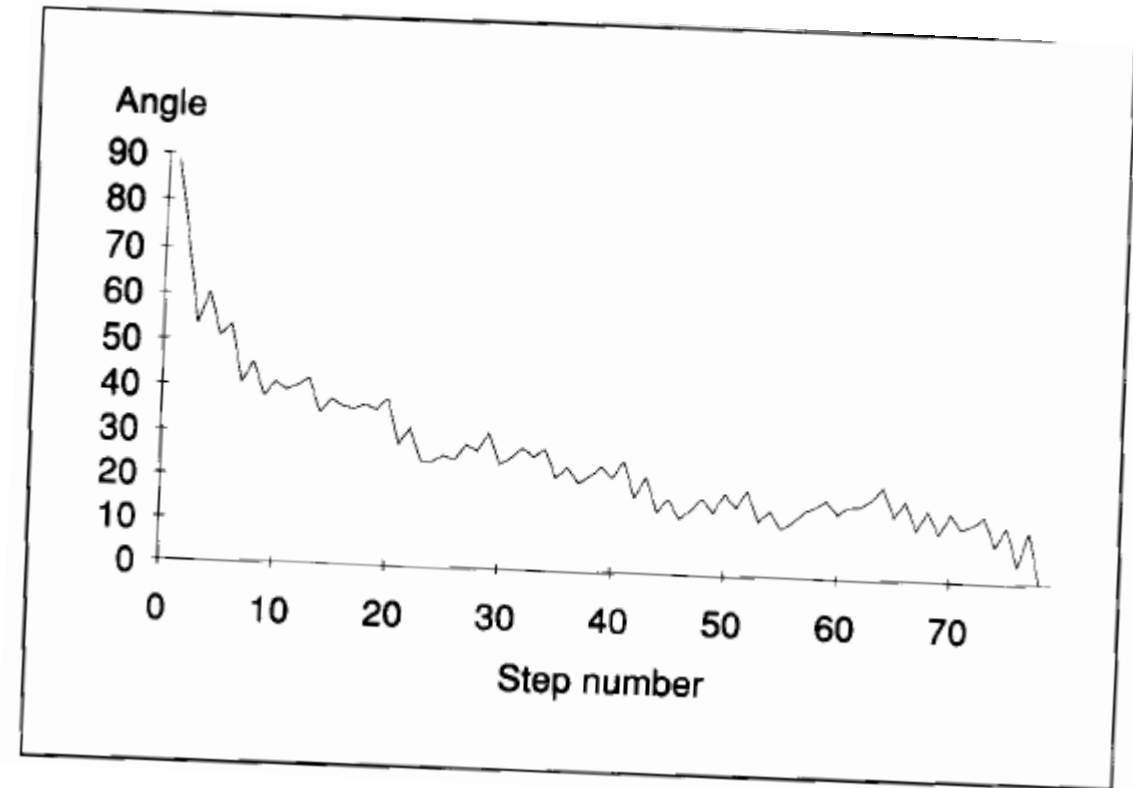
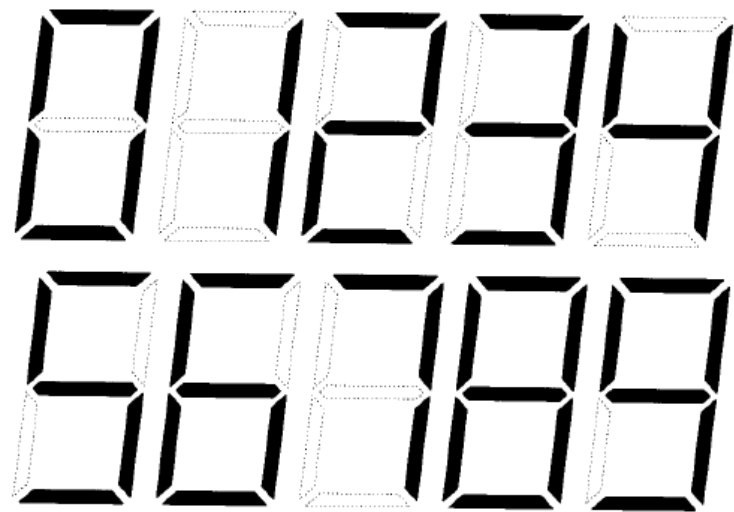
Primer



Primer

Digit	x_0	x_1	x_2	x_3	x_4	x_5	x_6
0	0	1	1	1	1	1	1
9	1	1	1	1	1	1	0
8	1	1	1	1	1	1	1
7	0	0	1	1	1	0	0
6	1	1	1	0	1	1	1
5	1	1	1	0	1	1	0
4	1	1	0	1	1	0	0
3	1	0	1	1	1	1	0
2	1	0	1	1	0	1	1
1	0	0	0	1	1	0	0

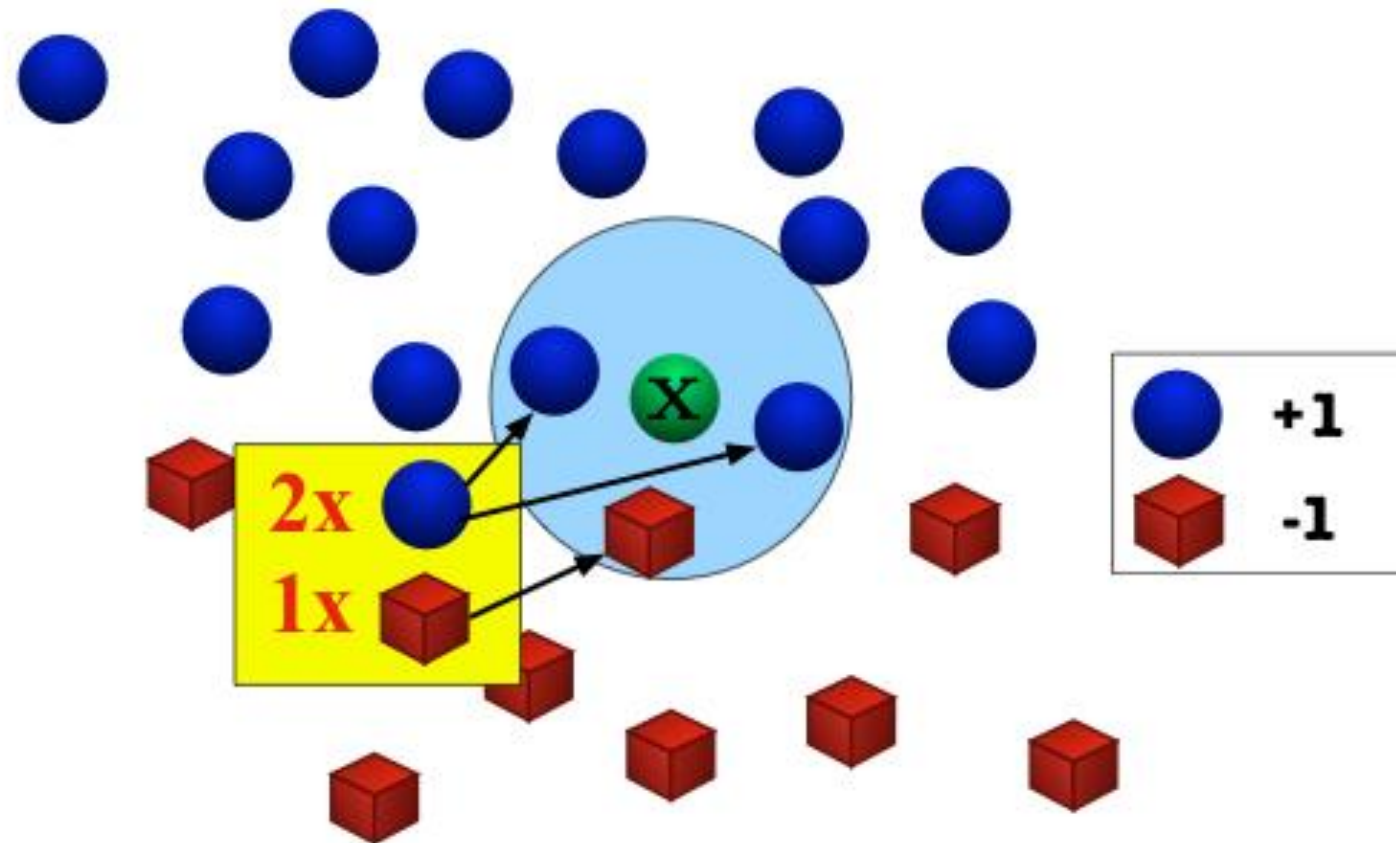
Primer



k najbližih suseda

kNN

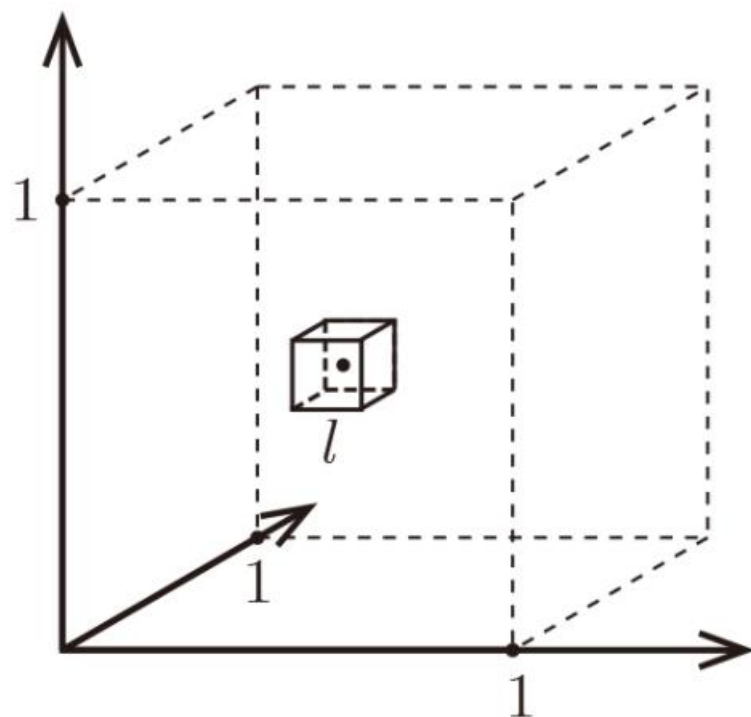
kNN



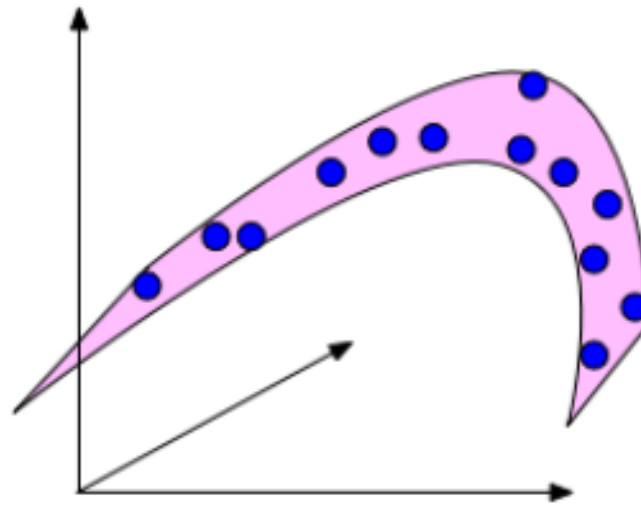
Rastojanje Minkovskog

$$\text{dist}(\mathbf{x}, \mathbf{z}) = \left(\sum_{r=1}^d |x_r - z_r|^p \right)^{1/p}.$$

Prokletstvo dimenzionalnosti



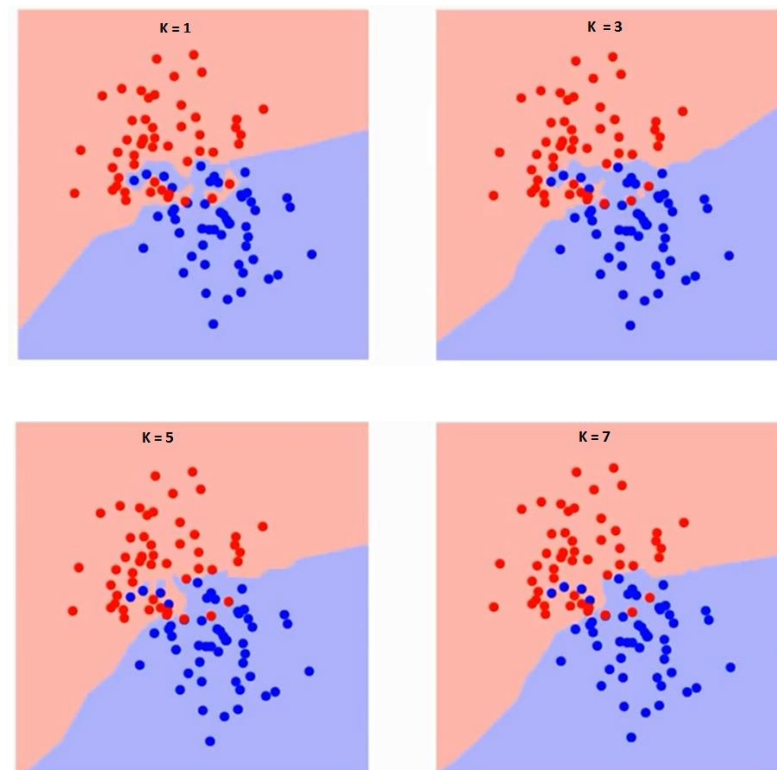
Prokletstvo dimenzionalnosti



An example of a data set in 3d that is drawn from an underlying 2-dimensional manifold. The blue points are confined to the pink surface area, which is embedded in a 3-dimensional ambient space.

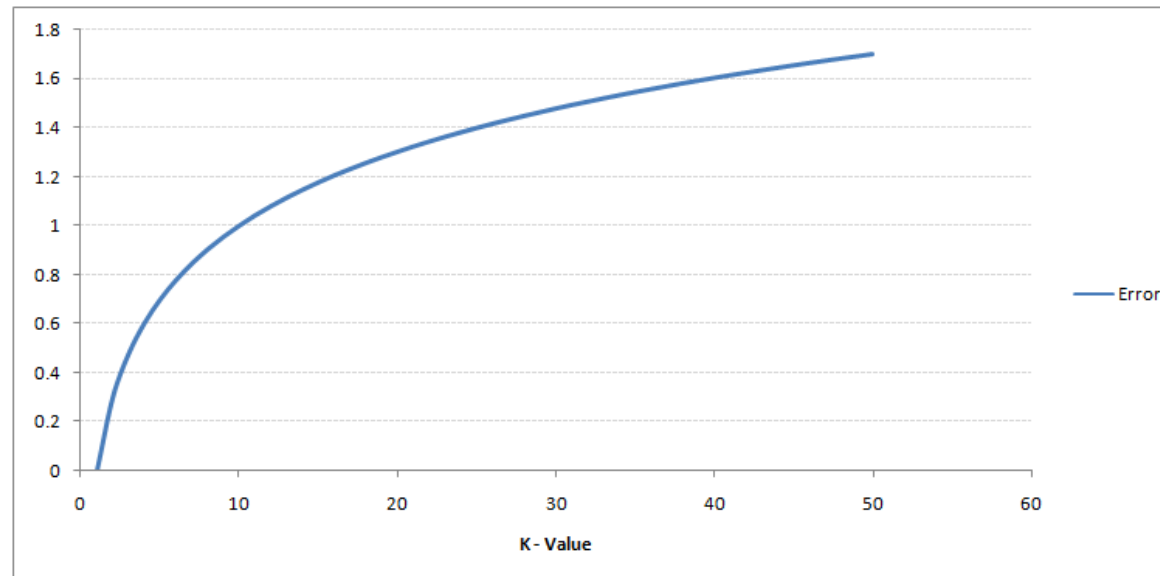
Uticaj faktora k

- Granice između klasa zavise od faktora k
- Kako faktor k raste, granica postaje sve više glatka
- Ukoliko faktor k ima vrednost 1, onda je greška 0. Sa povećanjem vrednosti faktora k raste i stopa greške
- Treba odrediti optimalna vrednost za k .



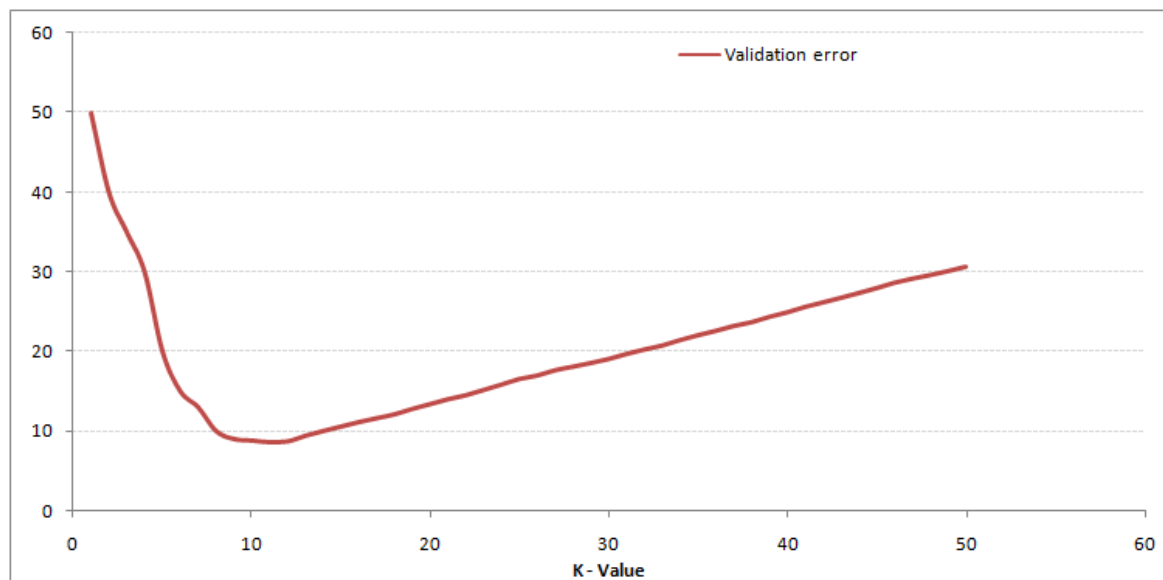
Stopa greške u obuci

- Procenat uzoraka iz trening podataka koji za neko k ne bivaju dobro klasifikovani
- Ukoliko faktor k ima vrednost 1, onda je greška 0
- Sa povećanjem vrednosti faktora k raste i stopa greške



Stopa greške validacije

- Skup za validaciju služi da se pronade optimalna vrednost za k uz pomoć stope greške u obuci



Izbor vrednosti faktora k

Kako bi se izabrala vrednost faktora k , potrebno je izvršiti sledeće korake:

1. Podeliti početni skup podataka na podatke za trening i podatke za validaciju
2. Izvršiti treniranje klasifikatora nad trening skupom za razne vrednosti $k=1,3,5,\dots$
3. Oceniti svaki od klasifikatora nad podacima za validaciju
4. Odabrati parametar k na osnovu klasifikatora koji ima najbolju ocenu, odnosno najbolje performanse

Primena

- Kreditni rejting - ljudi sa sličnim finansijskim detaljima dobijaju slične kreditne rejtinge
- Politika - klasifikovanje potencijalnih birača na one koji će glasati za određenu stranku i one koji neće

Prednosti i mane

Prednosti

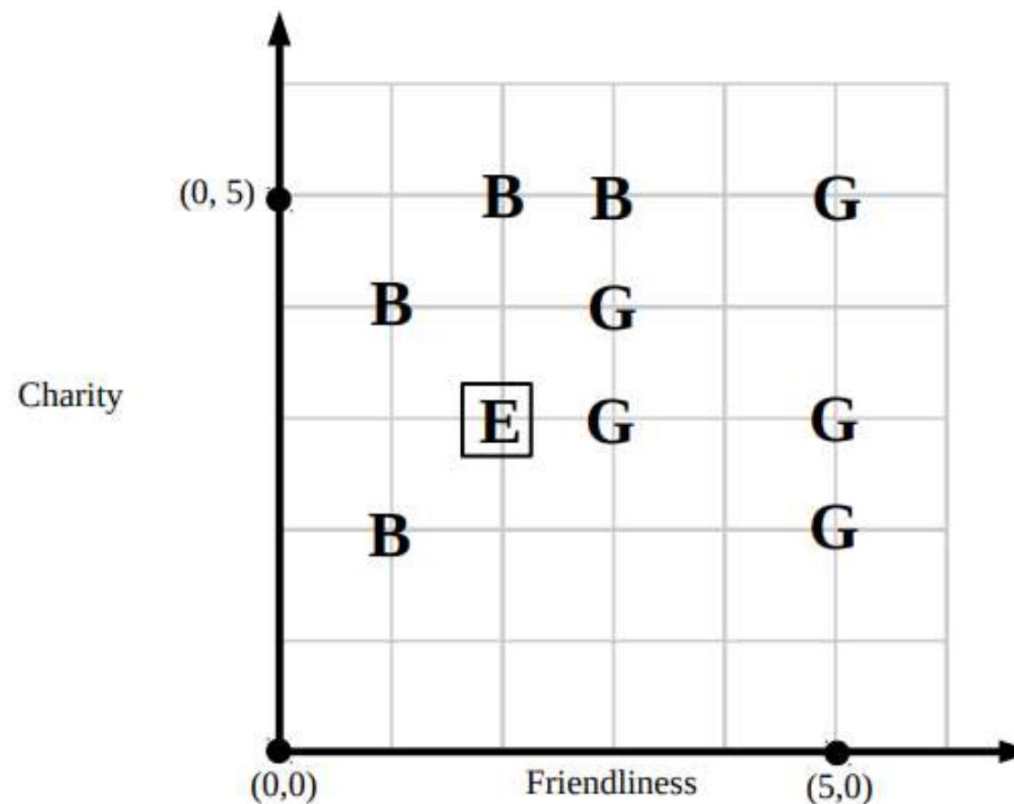
- Jednostavan za razumevanje
- Lak za implementaciju
- Malo ili nimalo vremena obuke
- Funkcioniše jednako lako sa skupovima podataka sa više klasa

Mane

- Računski skupa faza testiranja
- Asimetrična (*skewed*) distribucija klasa (česta klasa dominira većinskim glasanjem)
- Tačnost može biti ozbiljno degradirana sa podacima koji imaju veliki broj atributa

Primer

- Ljudi se mogu klasifikovati na dobre (good – G) i loše (bad – B) na osnovu dve karakteristike: koliko su prijateljski naklonjeni prema drugim ljudima i koliko daju u dobrotvorne svrhe.
- Šta se, na osnovu slike desno, može zaključiti o osobi Eri Eriću (predstavljen slovom E na grafikonu) korišćenjem kNN algoritma i Euklidskog rastojanja, u svakom od sledećih slučajeva:
- $k=1$, $k=5$, $k=8$ i $k=9$



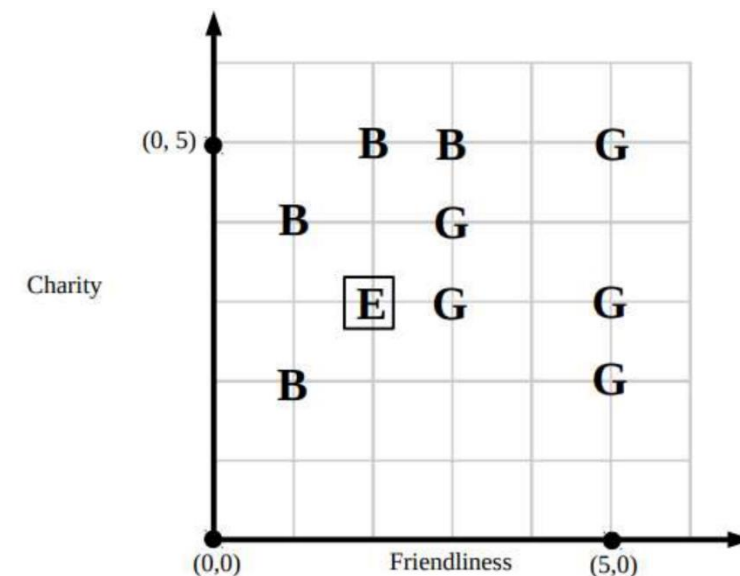
Rešenje

k=1	Dobar	Loš	ne može se zaključiti ništa	U slučaju kada je $k = 1$ traži se jedan najbliži sused što je u slučaju sa slike tačka G sa koordinatama (3,3).
------------	-------	-----	-----------------------------	--

k=5	Dobar	Loš	ne može se zaključiti ništa	Za $k = 5$ najbliži susedi su: 1. B (1,2) 2. G (3,3) 3. B (1,4) 4. G (3,4) 5. B (2,5) Pošto među 5 najbližih suseda preovladavaju susedi sa oznakom B klasa tačke E će biti B(Loš).
------------	-------	-----	-----------------------------	--

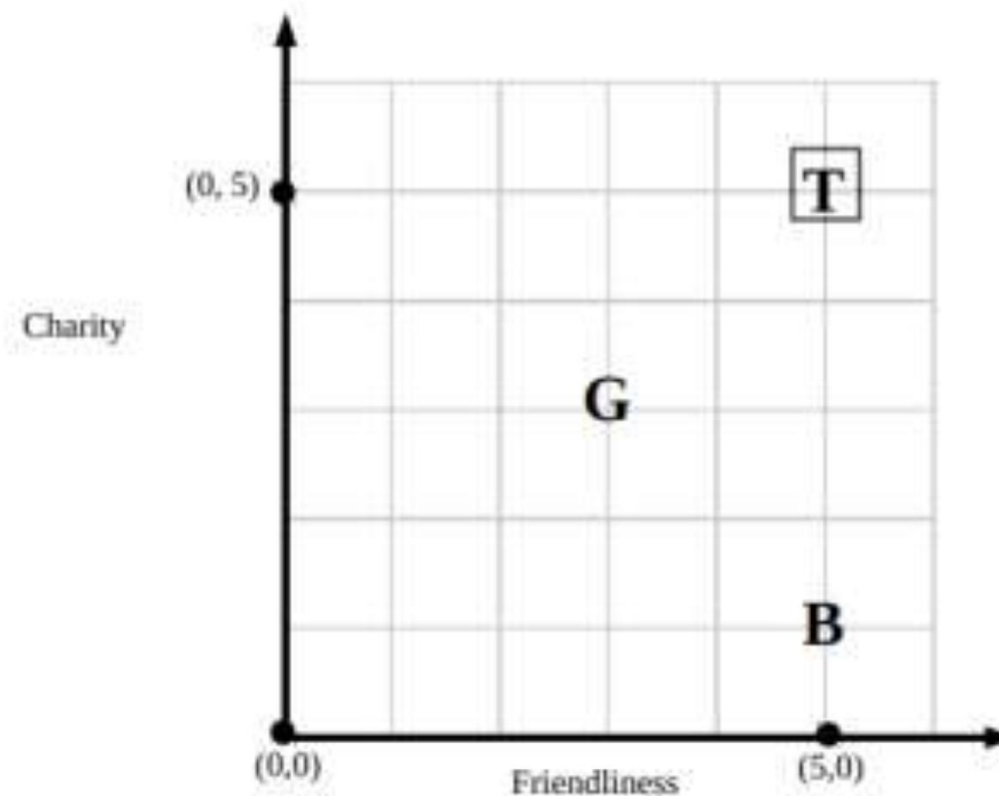
k=8	Dobar	Loš	ne može se zaključiti ništa	Za $k = 8$ sve tačke ulaze u najbliže susede sem tačke G(5,5). Kako je među najbližim susedima odnos klasa G i B izjednačen, o uzorku se ne može ništa zaključiti. Zaključak: U slučaju kada postoje dve klase, k ne bi trebalo da uzima parne vrednosti.
------------	-------	-----	-----------------------------	---

k=9	Dobar	Loš	ne može se zaključiti ništa	Kada je $k = 9$, sve tačke ulaze u grupu najbližih suseda. Klasa tačke E se može ustanoviti na osnovu prebrojavanja svih tačaka. Kako tačaka sa klasom G ima 5, a sa klasom B ima 4, klasa tačke E će biti G (Dobar).
------------	-------	-----	-----------------------------	---



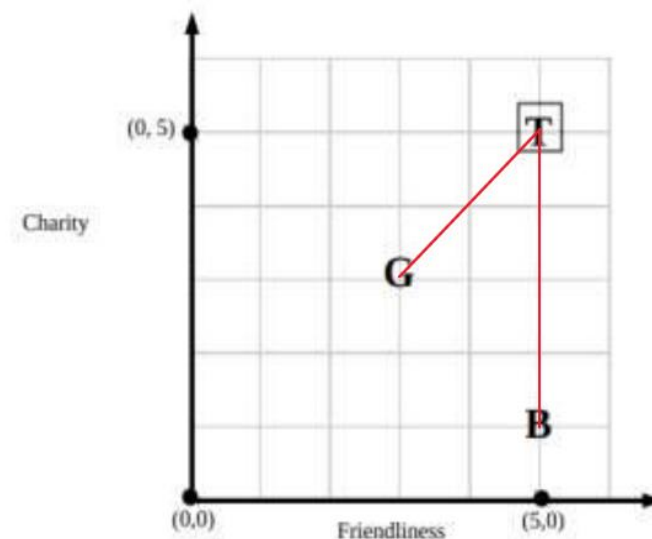
Primer

- Potrebno utvrditi da li je Teodora dobra ili loša na osnovu grafikona desno
- Korišćenjem 1-najbližeg suseda, klasifikuj Teodoru (predstavljena slovom T na grafikonu) kao dobru ili lošu, korišćenjem različitih metrika



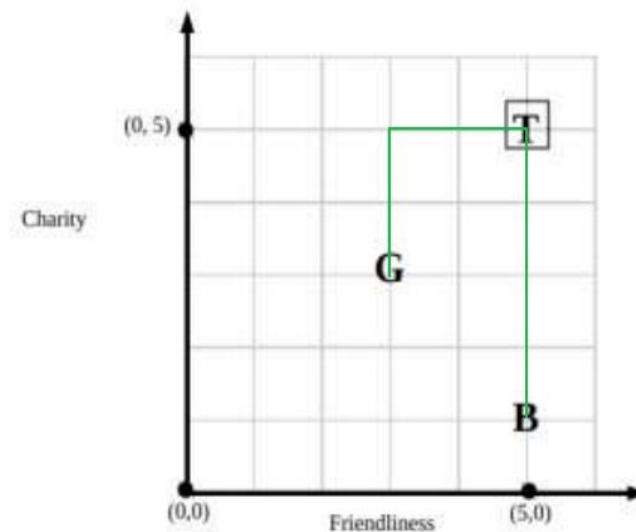
Rešenje

Tačka G je bliža od tačke B.			
Euklidsko rastojanje	Dobra	Loša	ne može se zaključiti ništa



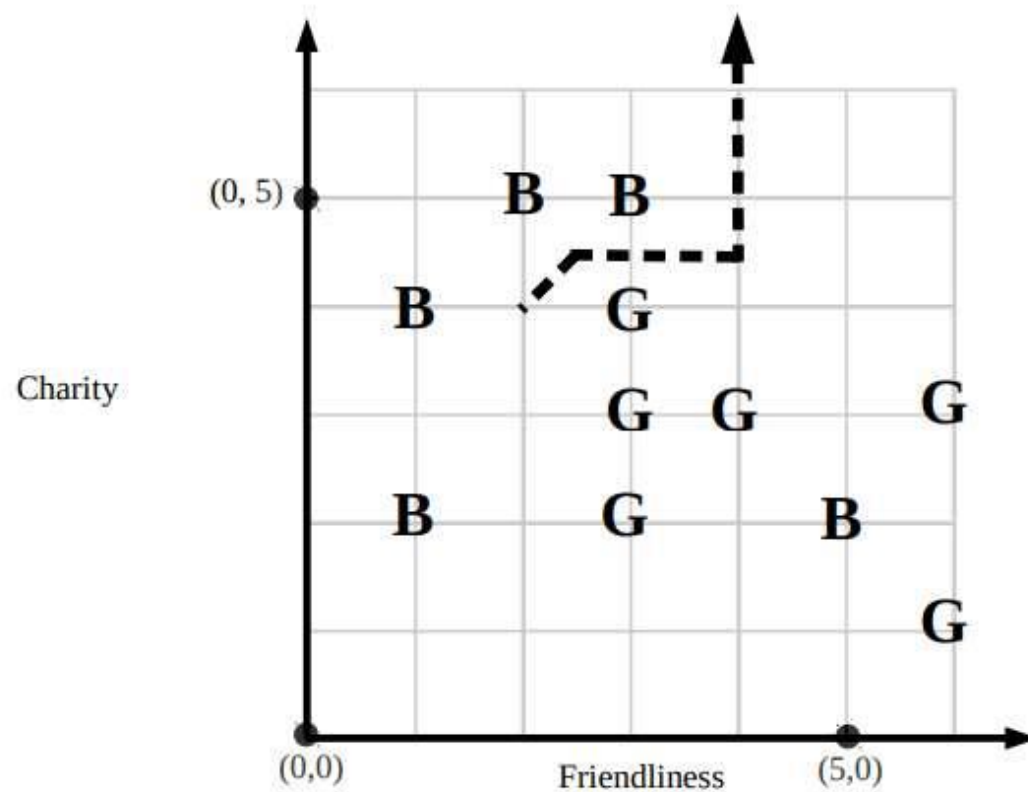
Menhetn rastojanje	Dobra	Loša	ne može se zaključiti ništa
-----------------------	-------	------	--------------------------------

Hamingovo rastojanje ima smisla kod kategorijskih promenljivih. Kako su vrednosti promenljivih u primeru kontinualne, nema smisla raditi Hamingovo rastojanje.			
Hamingovo rastojanje	Dobra	Loša	ne može se zaključiti ništa



Primer

- Korišćenjem algoritma 1-najbliži sused sa Euklidskim rastojanjem razgraničiti primere sa grafikona na dobre i loše. Deo granice je već napravljen i prikazan isprekidanom linijom



UNI 2020

