

# Vještačka inteligencija

## Predavanje 6: Funkcije gubitka

*Učenje nije produkt nastavnčkog predavanja.  
Učenje je produkt aktivnosti onih koji uče.  
~John Holt*

Odgovorna nastavnica: Vanr. prof. dr Amila Akagić

Univerzitet u Sarajevu



# Uvodne informacije

- This work is licensed under a Creative Commons 'Attribution-NonCommercial-ShareAlike 4.0 International' license. EN: <https://creativecommons.org/licenses/by-nc-sa/4.0/>

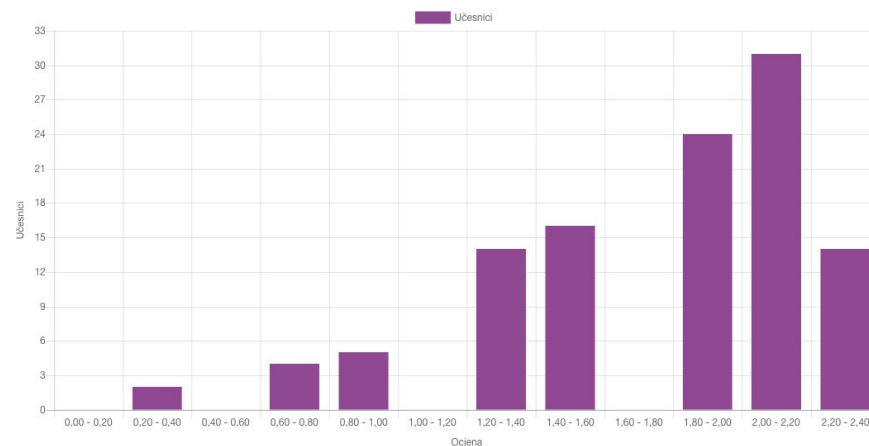


- Ovaj rad je licenciran pod međunarodnom licencom 'Imenovanje-Nekomercijalno-Dijeli pod istim uvjetima 4.0' od strane Creative Commons. HR: <https://creativecommons.org/licenses/by-nc-sa/4.0/deed.hr>

# Najave

- ❑ Danas u 14:00 je četvrti kviz u okviru realizacije vježbi.
  - ❑ 8 pitanja, 5 minuta vremena.
  - ❑ Automatsko ocjenjivanje nakon isteka vremena.
- ❑ Sa predavanjima nastavljamo u 14:10+.
- ❑ Ovo je već VI sedmica nastave. 29.04. je prvi kviz u okviru realizacije predavanja.
- ❑ Sljedeće sedmice nema nove lab. vježbe, ali ima kviz 5.

Ukupan broj studenata koji ostvaruju raspon ocjena



# Aktuelnosti / Novosti

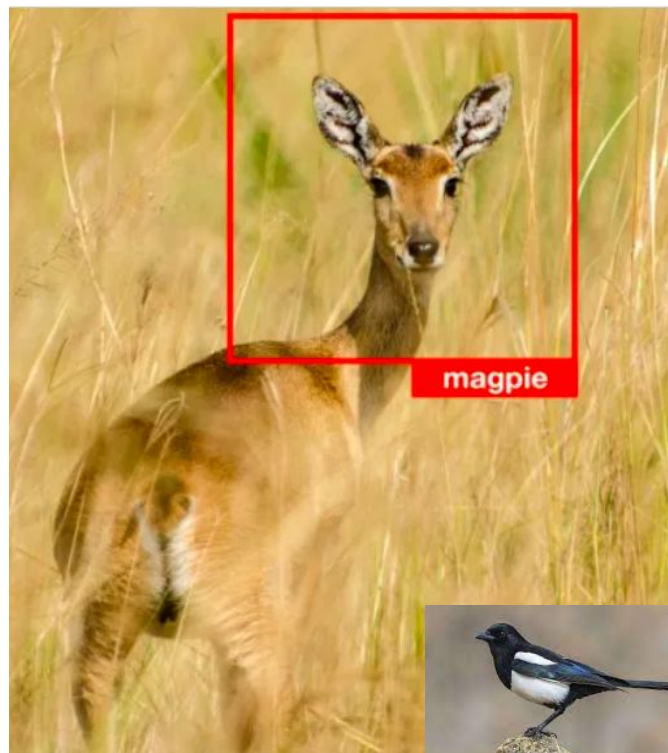
Artificial intelligence / Machine learning

## Error-riddled data sets are warping our sense of how good AI really is

Our understanding of progress in machine learning has been colored by flawed testing data.

by **Karen Hao**

April 1, 2021



<https://www.technologyreview.com/2021/04/01/1021619/ai-data-errors-warp-machine-learning-progress/>

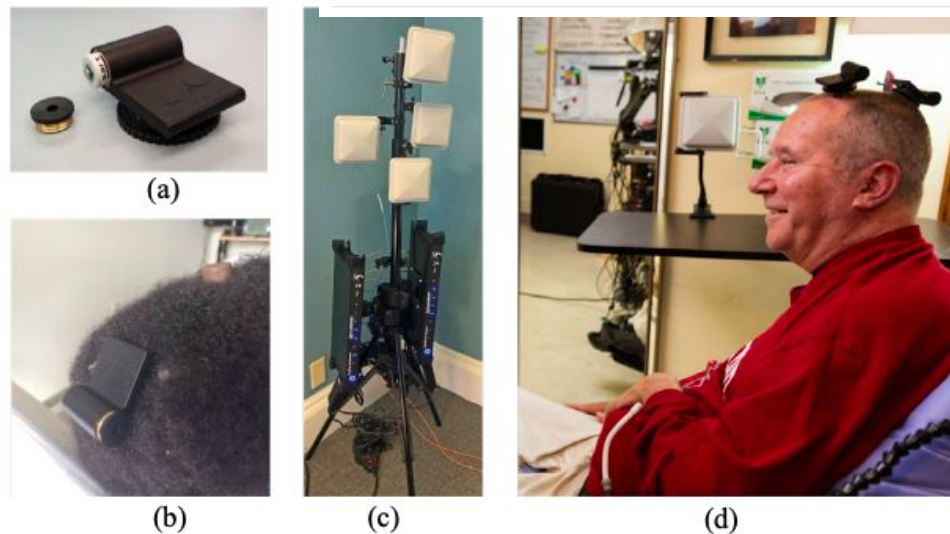


Fig. 2. Some components of the wireless system. **(a)** BWD transmitter (52 mm x 44 mm) showing battery compartment. Turn-screw disc is used to attach the device onto a percutaneous pedestal. **(b)** The BWD connected to T10's posterior pedestal (here, the anterior pedestal is covered by a protective cap). **(c)** A two-frequency wireless receiver system in a four-antenna configuration as deployed for T10. The output optical fibers (orange) connect to downstream NSPs. **(d)** T5 in his home with two transmitters. The antenna in the background was one of four mounted around the room. Photos used with permission.

<https://www.unilad.co.uk/technology/human-brain-connected-wirelessly-to-computer-for-first-time-ever/>

<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9390339>

# Kako koristiti podatke za proces učenja?

- ❑ Na prethodnim predavanjima vidjeli smo različite vrste podataka koje koristimo za treniranje različitih modela za učenje.
- ❑ Koji je najbolji način da ih koristimo? Koje su to postojeće prakse podjele jednog skupa podataka (*dataset*)?
- ❑ U slučaju K-NN, kako izabrati K i **distancu**?
  - ❑ K i **distanca** su primjeri hiperparametara koje bираmo prije treniranja!
  - ❑ Ove parametre ne učimo iz podataka, nego ih *a priori* određujemo.
  - ❑ Izbor ovih parametara će uveliko zavisiti od problema koji pokušavamo riješiti.
  - ❑ Najbolje rješenje: Isprobati sve varijacije i odlučiti se za najbolju.

# Različiti načini podjele podataka

- ❑ Prva ideja: cijeli skup podataka se koristi za treniranje.

Cilj: postići najveću tačnost na skupu podataka.

Cijeli skup podataka

Ako izaberemo  $K = 1$  sistem će savršeno raditi na cijelom skupu podataka (tačnost 100%), ali ne možemo znati kako će se ponašati nad podacima koje nikad nije imao priliku da "vidi".

- ❑ Druga ideja: podijeli podatke na skup za treniranje i testiranje.

Treniranje

Testiranje

Nad kojim skupom ćemo birati hiperparametre? Skupu za treniranje ili skupu za testiranje?

Svrha algoritama mašinskog učenja je generalizacija nad "neviđenim" podacima. Da li to znači da je hiperparametre bitno izabrati na osnovu testnih podataka?

- ❑ Treća ideja: podijeli podatke na skup za treniranje, validaciju i testiranje.

Treniranje

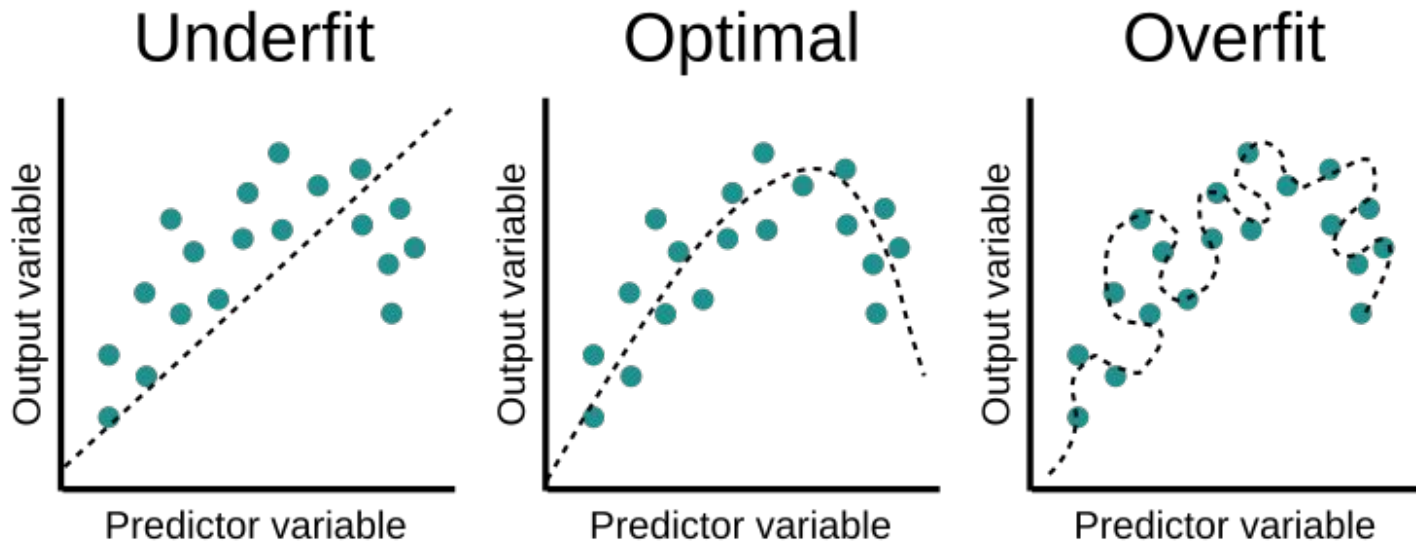
Validacija

Testiranje

Treniraj model na skupu za treniranje, izaberi parametre na osnovu skupa za validaciju, a evaluaciju napravi na osnovu skupa za testiranje!

Hold-out strategija

# Problem *overfitting*-a

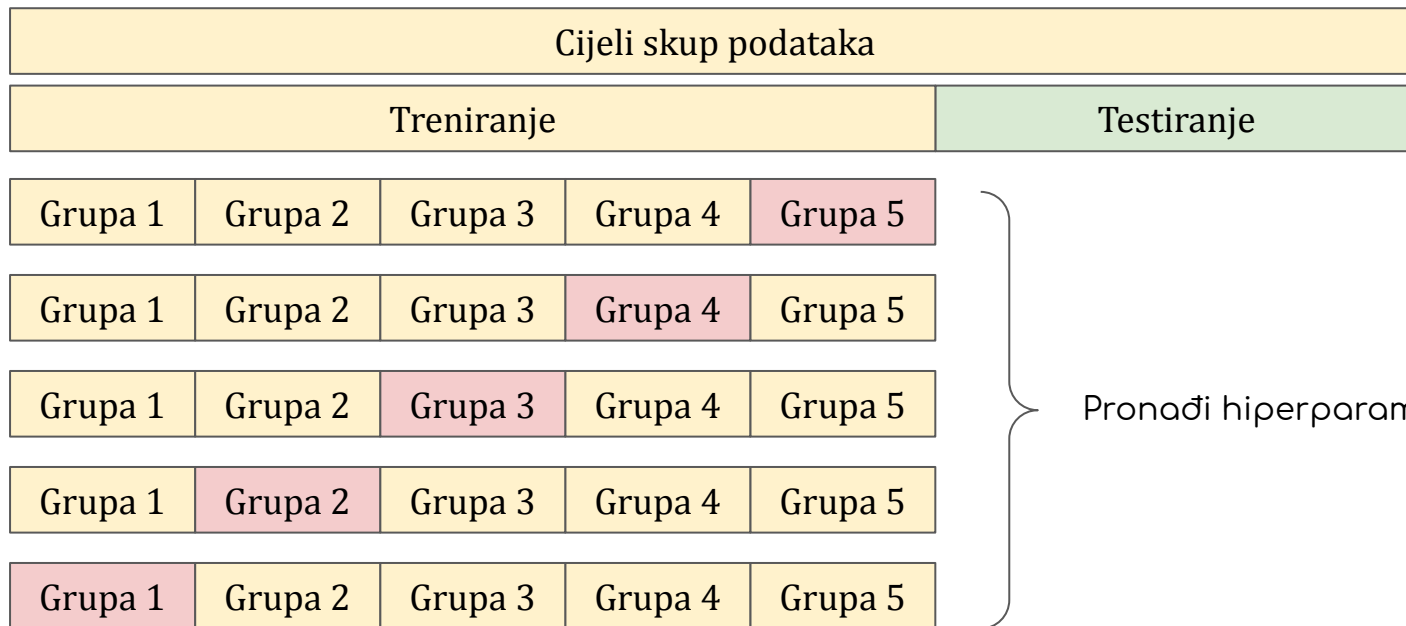


- ❑ Underfit model rezultira visokim pogreškama u predviđanju kako za trening tako i za test podatke. Relacija između podataka nije dovoljno dobro naučena.
- ❑ Overfit model daje vrlo nisku pogrešku predviđanja na podacima za trening, ali vrlo visoku na podacima testa. **Model je memorizirao podatke i nema mogućnost generaliziranja.**



# K-fold strategija

- ❑ Četvrta ideja: unakrsna validacija (*cross-validation*), tj. podijeli podatke na manje grupe od kojih je jedna grupa u jednom vremenskom periodu grupa za validaciju, a ostale grupe se koriste za treniranje. Pronađi najbolje hiperparametre na način da izračunaš srednju vrijednost nad svim mogućim kombinacijama.



Ova tehnika je više pogodnija za manje skupove podataka.

Fold = grupa

# CIFAR10

Avion



Automobil



Ptica



Mačka



Srna



Pas



Žaba



Konj



Brod



Kamion



10 klasa

50k slika za treniranje (5k po klasi)

10k slika za testiranje (1k po klasi)

32x32 RGB slika

<https://www.kaggle.com/c/cifar-10/overview>

<https://www.cs.toronto.edu/~kriz/cifar.html>

# Metode sa parametarskim pristupom

- ❑ Linearni klasifikatori spadaju u grupu tzv. Metoda sa parametarskim pristupom: postoji skup fiksnih parametara (srednja vrijednost i standardna devijacija) koji mogu **modelirati vjerovatnoću** nad nekim skupom podataka.



$32 \times 32 \times 3$

3072

Kako definisati funkciju  $f()$ ?  
Vektor ulaznih značajki  
(input feature vector)

$f(x, W)$

Funkcija  $f$  mapira vektor značajki (feature vector) u labele (numeričke vrijednosti.)

Parametri  
ili težine  
(weights)

Avion  
Automobil  
Ptica  
Mačka  
Srna  
Pas  
Žaba  
Konj  
Brod  
Kamion

Oznaka, labela (label)

# Linearni klasifikator

- ❑ Linearni klasifikatori spadaju u grupu tzv. Metoda sa parametarskim pristupom.

$$f(x, W) = Wx + b$$



$32 \times 32 \times 3$

3072

→  $f(x, W)$  →

Kako predstavljamo ulaze u ovakav sistem?

Koliko ima ulaza?

Koliko ima izlaza?

Koja je dimenzionalnost  $W$ ?

3072

10

$10 \times 3072$

Avion

Automobil

Ptica

Mačka

Srna

Pas

Žaba

Konj

Brod

Kamion

# Linearni klasifikator

- ❑ Linearni klasifikatori spadaju u grupu tzv. Metoda sa parametarskim pristupom.

$$f(x, W) = Wx + b$$

Bias ili otklon

Opisuje koliko dobro se model priklanja skupu podataka. Model sa visokim otklonom neće se previše prikloniti skupu podataka, tj. sa niskim otklonom model će se uskladiti/prikloniti sa skupom podataka.



$32 \times 32 \times 3$

3072

Može se ukloniti iz jednačine ukoliko se doda kao zadnja kolona u matricu  $W$ , te se modifikuje  $x$  dodajući još jedan element čija je vrijednost 1.

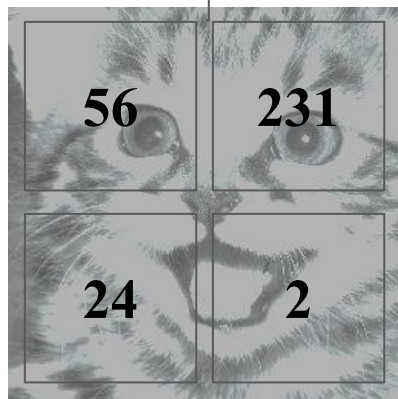
$$f(x, W) = Wx$$

Pojednostavljeni model

# Jednostavan primjer

$$f(x, W) = Wx + b$$

3 klase:  
Mačka  
Pas  
Brod



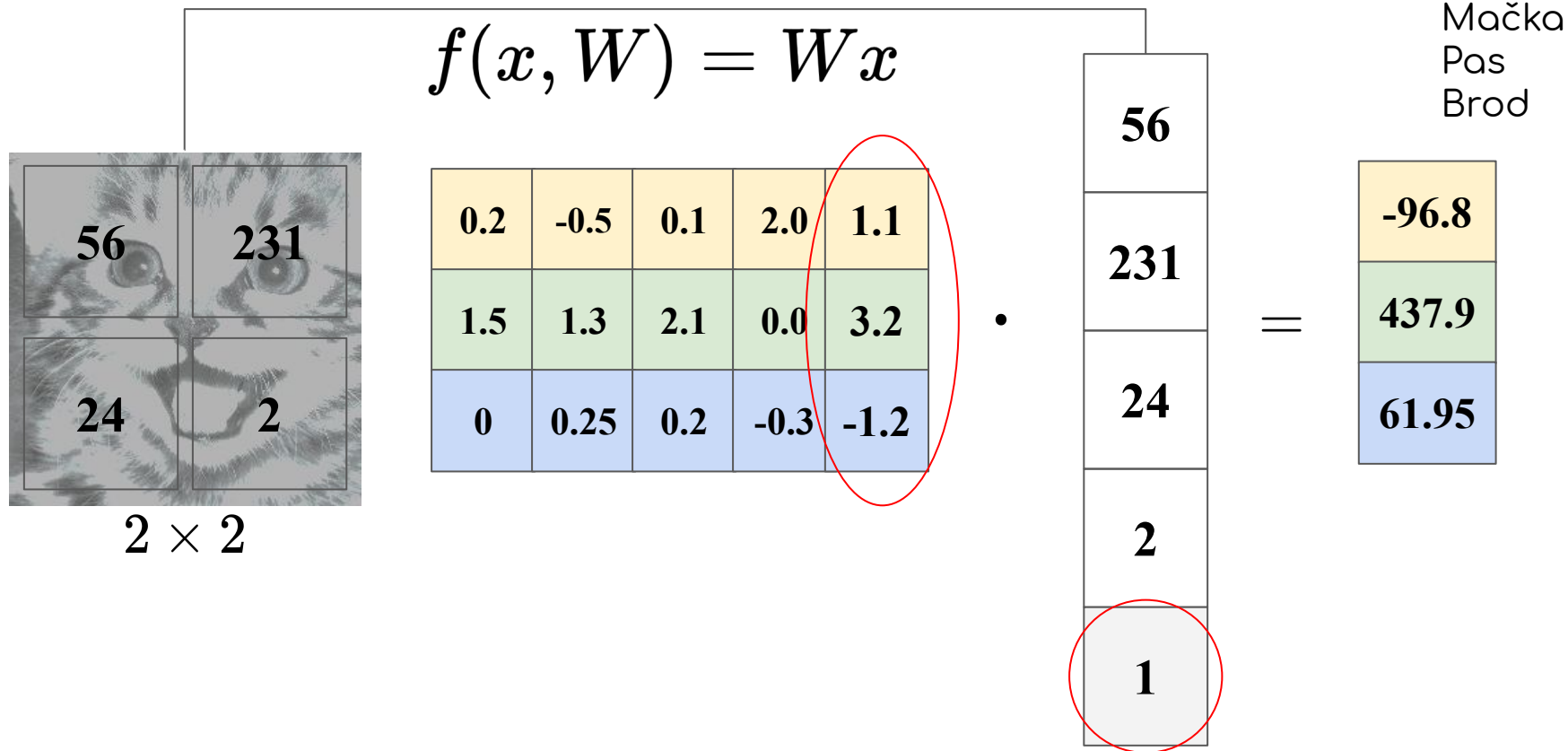
$2 \times 2$

0.2	-0.5	0.1	2.0
1.5	1.3	2.1	0.0
0	0.25	0.2	-0.3

Neka proizvoljno data matrica  $W$ .  
Vrijednosti matrice formiraju se  
nakon procesa učenja, te su ovdje  
date samo kao ilustrativni primjer.

•	56	+	1.1	=	-96.8
	231		3.2		437.9
	24		-1.2		61.95
	2				

# Jednostavan primjer

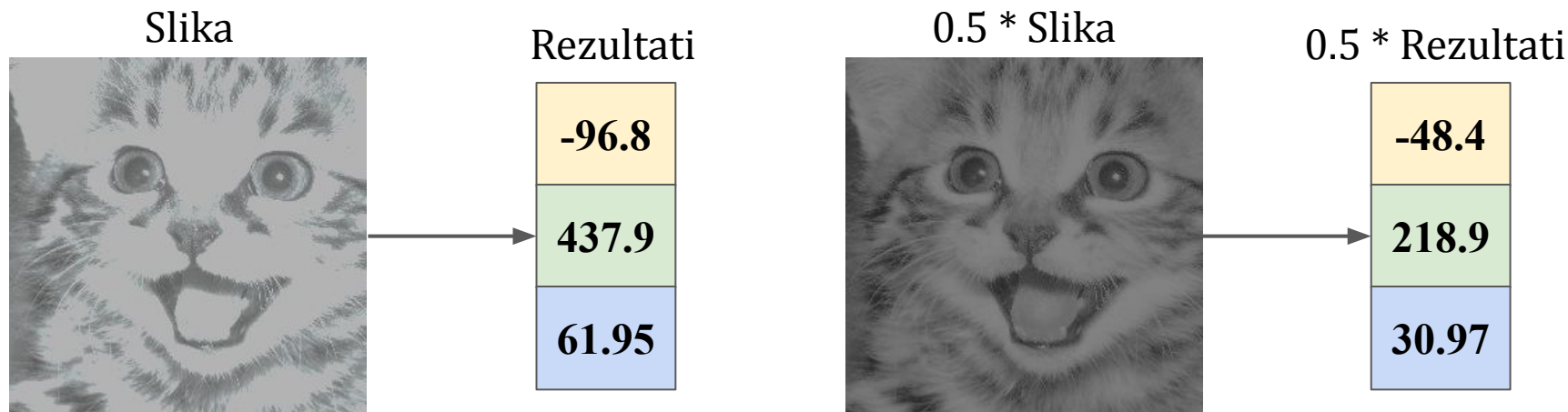




# I predikcije su linearne!

$$f(x, W) = Wx + b$$

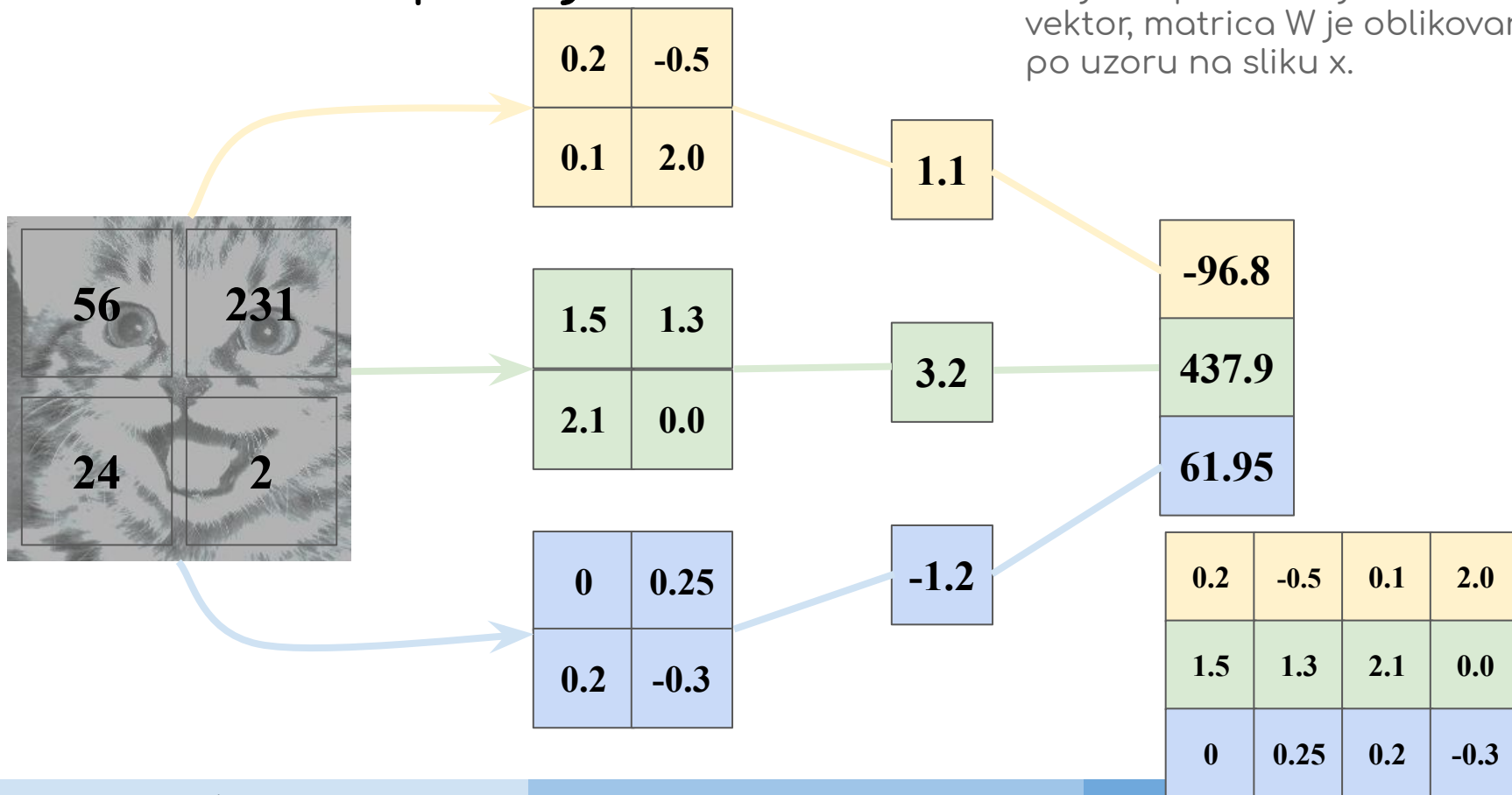
$$f(cx, W) = W(cx) = c * f(x, W)$$



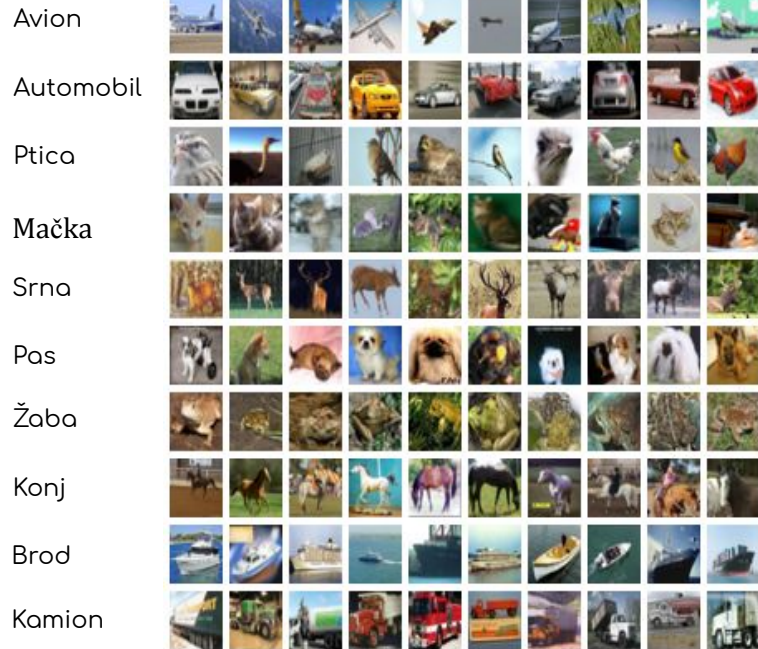


# Jednostavni primjer

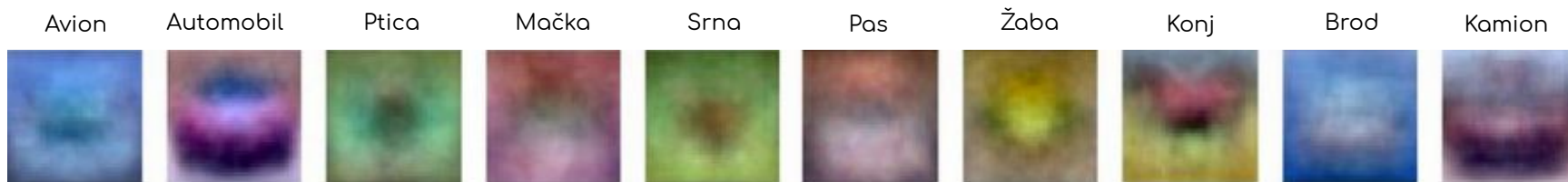
Umjesto pretvaranja slike u vektor, matrica  $W$  je oblikovana po uzoru na sliku  $x$ .



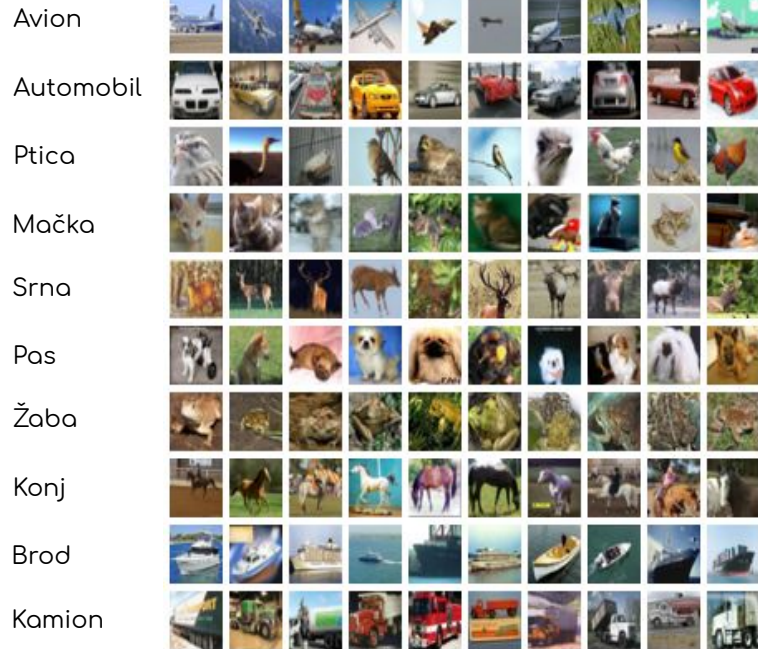
# Grafička reprezentacija matrice W



- Na osnovu težina matrice  $W$ , možemo “vidjeti” na koji način linearni klasifikator donosi odluke o predikciji.
- Vidi se da pozicija objekta na slici igra značajnu ulogu, kao i okolina objekta, što može biti problem (npr. šta ako na slici imamo dva objekta koja pripadu dvije različite klase?)
- U nekim klasama boja objekta igra značajnu ulogu (npr. crveno auto)



# Grafička reprezentacija matrice $W$



Avion

Automobil

Ptica

Mačka

Srna

Pas

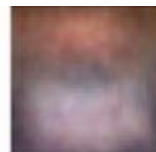
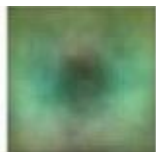
Žaba

Konj

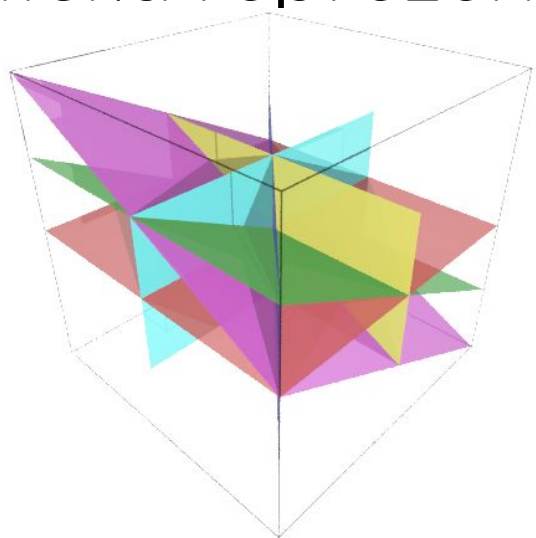
Brod

Kamion

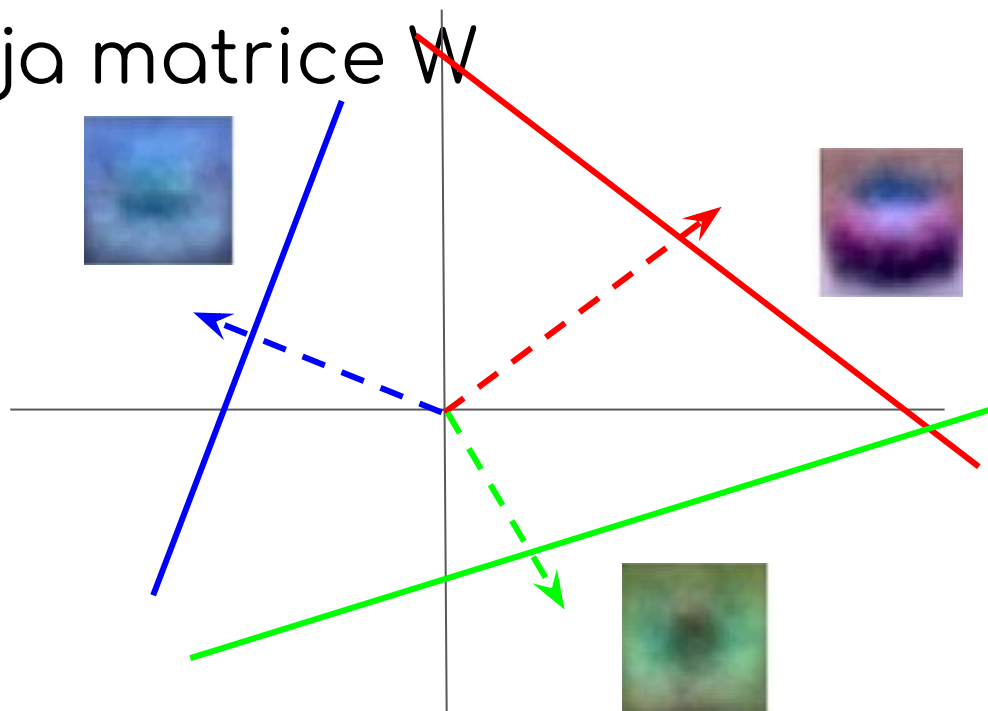
Pojednostavljeni prikaz  
u 2D Euklidskom  
prostoru. Pokušajte  
zamisliti **n-prostor**!



# Grafička reprezentacija matrice $W$



U stvarnosti u  $n$ -prostoru svaka klasa je predstavljena preko hiperravnini koja ta prostor dijeli na dva dijela.



Avion

Automobil

Ptica

Mačka

Srna

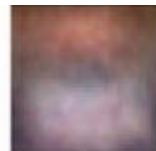
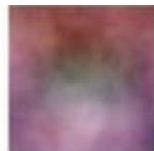
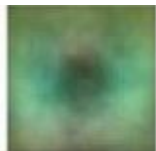
Pas

Žaba

Konj

Brod

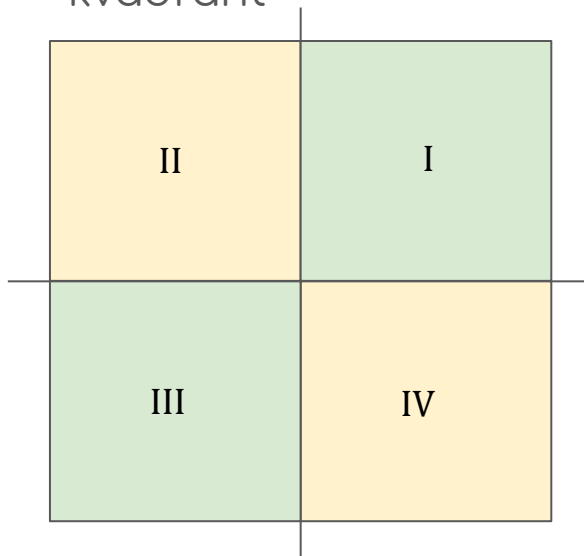
Kamion



# Ograničenja linearnih klasifikatora

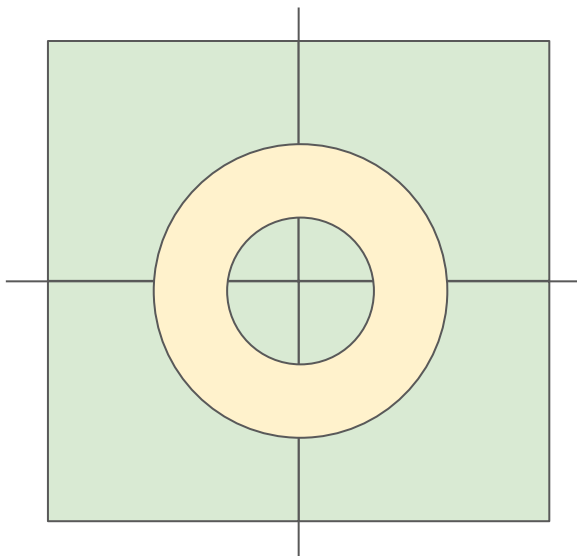
Klasa 1: I i III  
kvadrant

Klasa 2: II i IV  
kvadrant



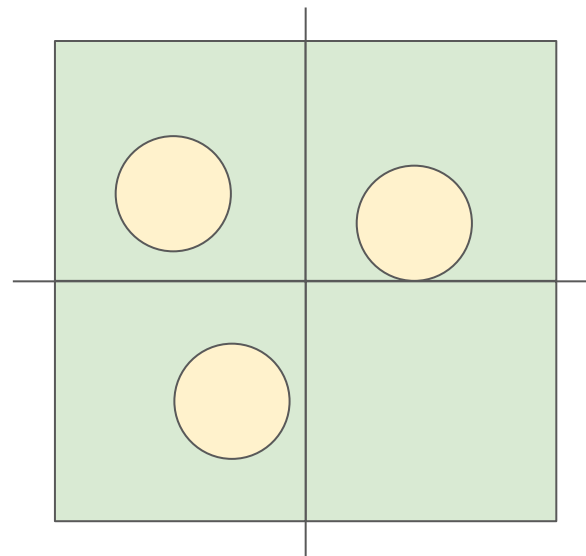
Klasa 1:  
 $1 \leq L2 \text{ norma} \leq 2$

Klasa 2: ostalo



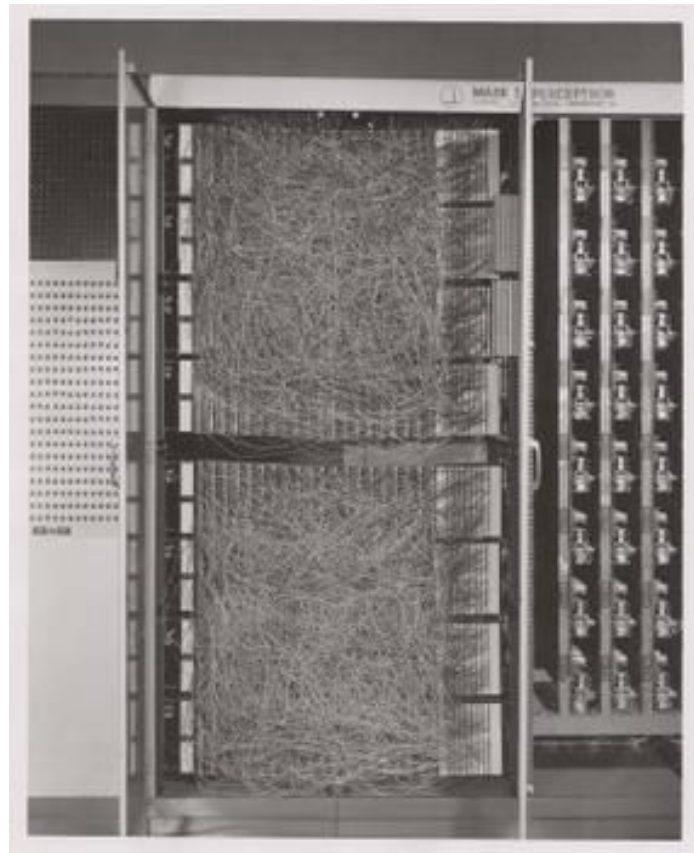
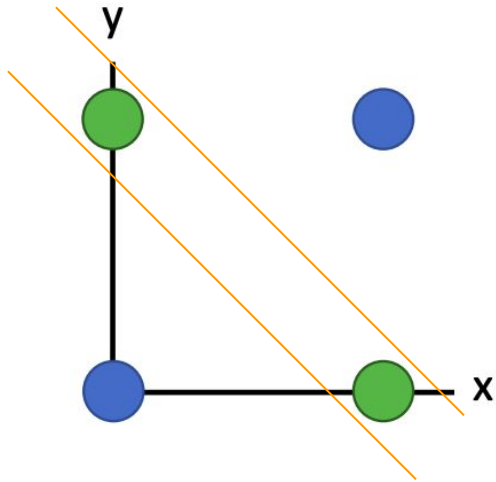
Klasa 1: tri regije

Klasa 2: ostalo



# Ograničenja linearnih klasifikatora

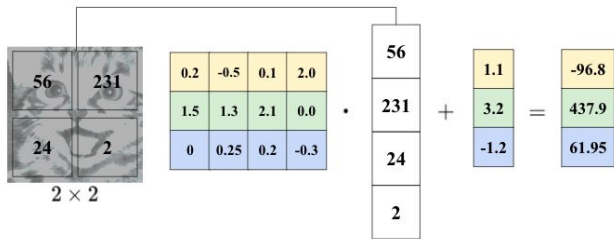
X	Y	F(x,y)
0	0	0
0	1	1
1	0	1
1	1	0





# Tri različita pogleda na linearne klasifikatore

## Algebarski pogled

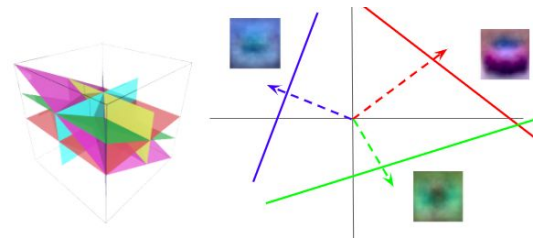


$$f(x, W) = Wx + b$$

## Vizuelni pogled



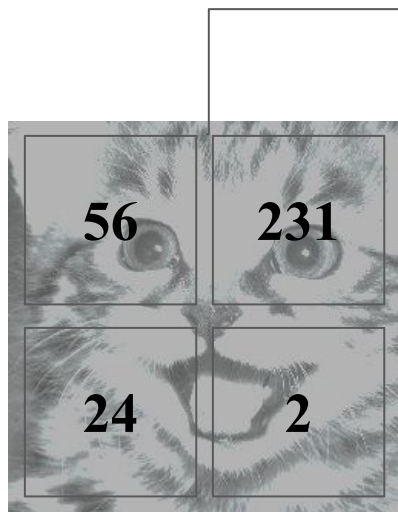
## Geometrijski pogled



# Jednostavan primjer

$$f(x, W) = Wx + b$$

3 klase:  
Mačka  
Pas  
Brod



$2 \times 2$

Testna slika

0.2	-0.5	0.1	2.0
1.5	1.3	2.1	0.0
0	0.25	0.2	-0.3

•

56
231
24
2

+

1.1
3.2
-1.2

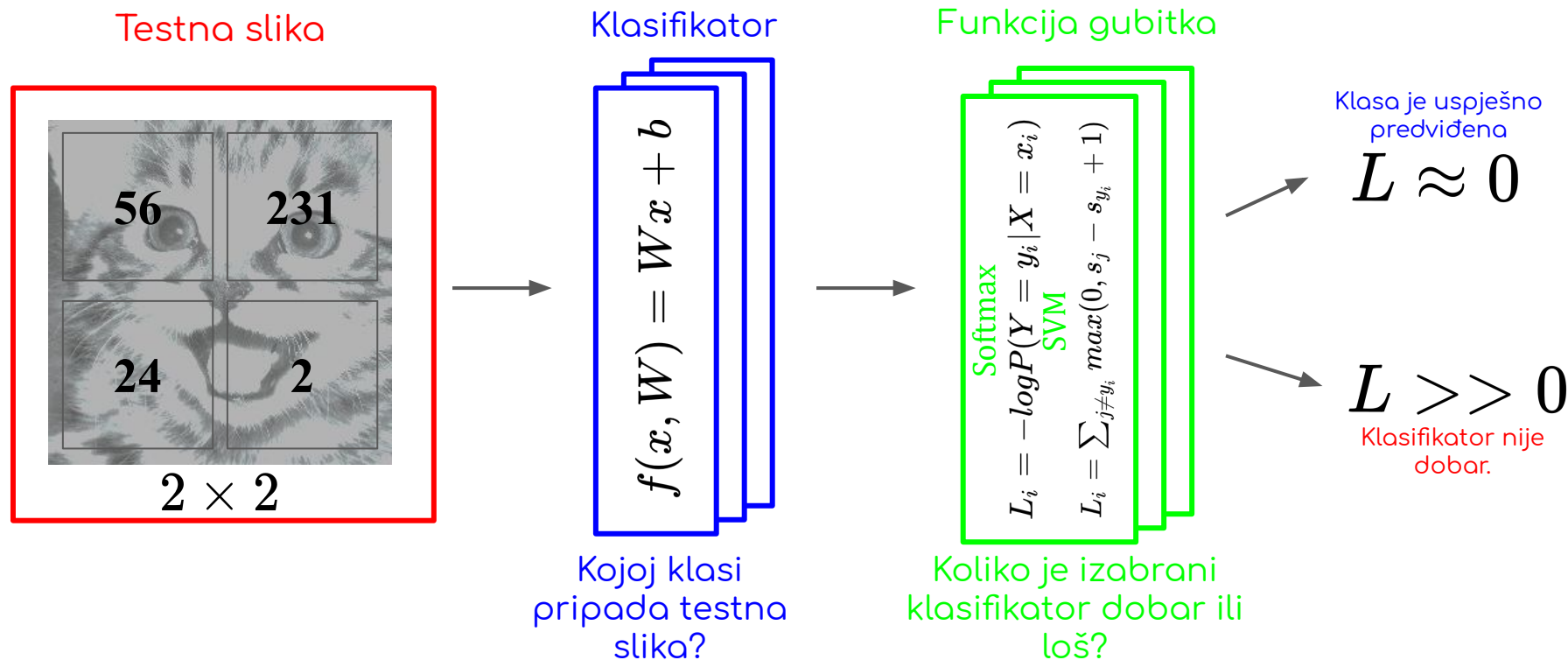
=

-96.8
437.9
61.95

1. Kako tumačiti dobijene rezultate (scores)?
2. Kako izabrati  $W$ ?

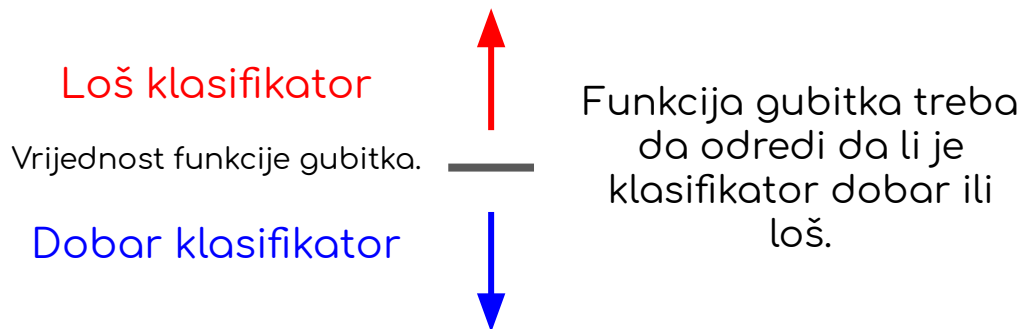


# Kako znati da li je klasifikator dobar ili loš?

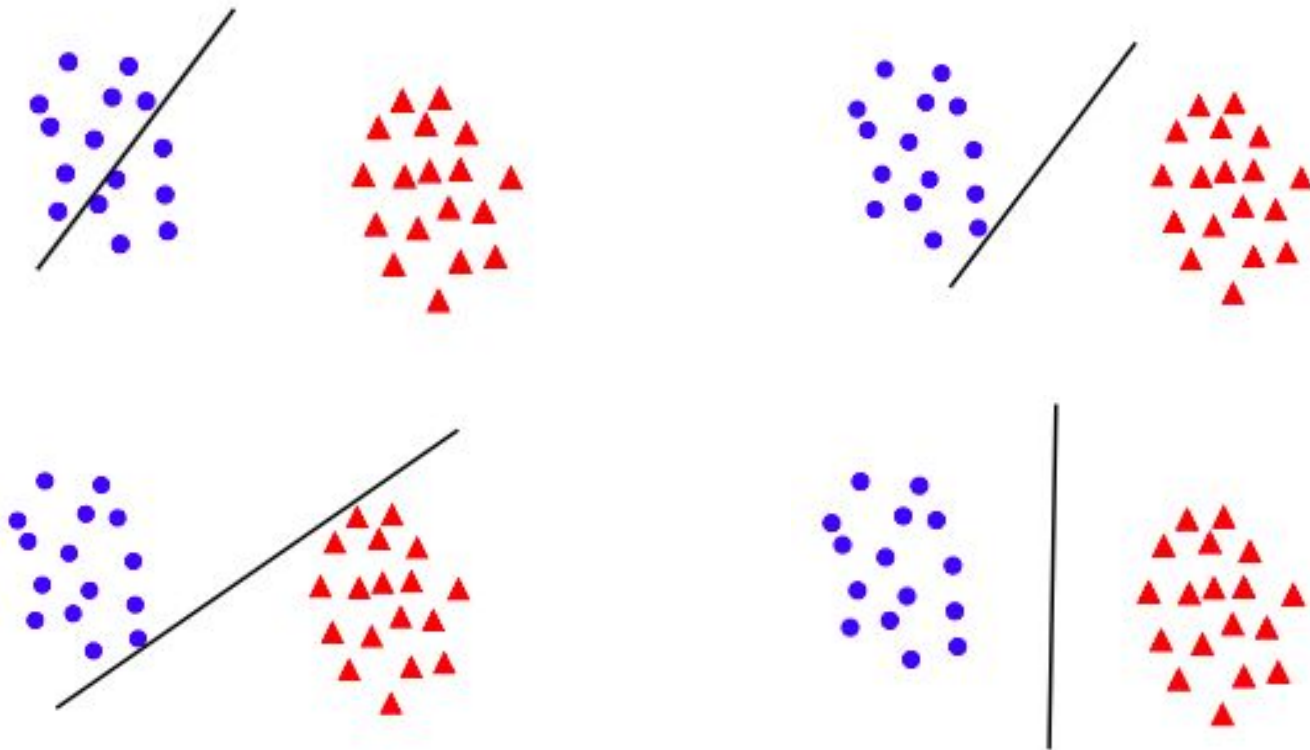


# Izbor matrice $W$

- ❑ Potrebna je neka funkcija koja će omogućiti evaluaciju različitih matrica  $W$  i pomoći nam da odlučimo koja od matrica je pogodna za dati dataset.
- ❑ Ta funkcija se naziva **funkcija gubitka** (*loss function*). Njen cilj je pronaći  $W$  na način da se **minimizira funkciju gubitka** (optimizacija). Naziva se još **funkcija cilja** ili **objektivna funkcija**.
- ❑ Funkcija gubitka govori koliko je pogodan ili dobar klasifikator za dati problem.
  - ❑ Što je manja vrijednost funkcije gubitka to znači da je klasifikator dobar.
  - ❑ Što je veća vrijednosti funkcije gubitka to znači da je klasifikator loš.



# Izbor matrice $W$



Kako izabrati najbolje  $W$ ?

Rješenje maksimalne margine (maximum margin): najstabilnije pod pomjeranja ulaza.

# Funkcija gubitka

Skup podataka

$$(x_i, y_i)_{i=1}^N$$

Apstraktna funkcija  
gubitka

$$L_i(f(x_i, W), y_i)$$

Za jednu sliku (i).

$$\left( \begin{array}{c} \text{M} \\ \text{A} \\ \text{Č} \\ \text{K} \\ \text{A} \end{array}, \right)_{i=1}^N$$



Obično se kodira preko integer vrijednosti.

$$L = \frac{1}{N} \sum_i L_i(f(x_i, W), y_i)$$

Za cijeli skup podataka.

U praksi postoji nekoliko različitih vrsta funkcija gubitka.

Na ovom času upoznat ćemo se sa **unakrsnom entropijom** (koristi se za softmax klasifikator) i **višeklasnom SVM funkcijom**.

# Funkcija gubitka

- ❑ U praksi postoji nekoliko različitih vrsta funkcija gubitka.
- ❑ Na ovom času upoznat ćemo se sa **unakrsnom entropijom** (koristi se za softmax klasifikator) i **više-klasnom SVM** funkcijom.

```
1 model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])
```

```
1 opt = keras.optimizers.Adam(learning_rate=0.01)  
2 model.compile(loss='categorical_crossentropy', optimizer=opt, metrics=['accuracy'])
```

# Unakrsna entropija (cross-entropy loss)

- ❑ Unakrsna entropija je funkcija gubitka koju se primjenjuje na vektoru rezultata (*score vector*).
- ❑ Problem koji rješavamo je višeklasna klasifikacija objekata (npr. slika) (eng. *multiclass classification*).
- ❑ Rezultat koji želimo postići je izračunavanje vjerovatnoće pripadnosti testne slike u odnosu na predefinisane klase.
- ❑ **Softmax regresija** je primjer generalizirane logističke regresije koja se može koristiti za višeklasnu klasifikaciju objekata. Sinonimi koju se mogu naći u literaturi **Multinomial Logistic**, **Maximum Entropy Classifier**, **Multi-class Logistic Regression**.
- ❑ Veoma popular u primjeni neuronskih mreža.



$$\rightarrow f(x, W) = Wx + b =$$

Ovo je samo ilustracija.

-96.8	Mačka	5%
437.9	Pas	85%
61.95	Brod	10%

Kako mapirati rezultate u vjerovatnoće?

# Softmax funkcija

$$P(Y = k | X = x_i) = \frac{e^{s_k}}{\sum_j e^{s_j}}$$



Mačka	2.6
Pas	-1.3
Brod	3.5

$e^{s_k}$

Nenormalizirane  
vjerovatnoće

13.46
0.27
33.12

$\frac{e^{s_k}}{\sum_j e^{s_j}}$

Normalizirane  
vjerovatnoće

0.29
0.00
0.71

Suma vrijednosti  
mora biti = 1.

## Softmax funkcija

Razlog zašto se ova funkcija naziva **softmax** je zbog toga što je to **diferencijabilna** (razlikovanje) funkcija **maksimizacije**. Primjetite koja vrijednost vektora rezultata je dobila najveću normalizovanu vrijednost vjerovatnoće. Šta bi onda mogla biti **hardmax** funkcija?

Funkcija  
rezultata (score  
function).

$$s = f(x_i, W)$$

$s_j$

Funkcija rezultata za klasu j.

$$\sum_j e^{s_j} = 46.85$$

# Sljedeći korak: izračunavanje funkcije gubitka!

Softmax funkcija

$$P(Y = k | X = x_i) = \frac{e^{s_k}}{\sum_j e^{s_j}}$$

Funkcija gubitka

$$L_i(f(x_i, W), y_i)$$

$$L_i = -\log P(Y = y_i | X = x_i)$$



Nenormalizirane  
vjerovatnoće (logits)

Normalizirane  
vjerovatnoće

Mačka  
Pas  
Brod

2.6
-1.3
3.5

$e^{s_k}$

13.46
0.27
33.12

$\frac{e^{s_k}}{\sum_j e^{s_j}}$

0.29
0.00
0.71

$$L_1 = -\log(0.29) = 1.79$$

$$L_2 = -\log(0.005) = 7.64$$

$$L_3 = -\log(0.71) = 0.49$$

$$s = f(x_i, W) \quad \sum_j e^{s_j} = 46.85$$

Suma vrijednosti  
mora biti = 1.

$$L = \frac{1}{N} \sum_i L_i$$



# Kullback-Leibler divergencija

P - vjerovatnoća u odnosu na podatke, dok je Q aproksimacija u odnosu na model.

- ❑ Kullback-Leibler divergencija (relativna entropija): jedan od načina da se mjeri razlika između dvije distribucije vjerovatnoće.

$$D_{KL}(P||Q) = \sum_y P(y) \log \frac{P(y)}{Q(y)}$$

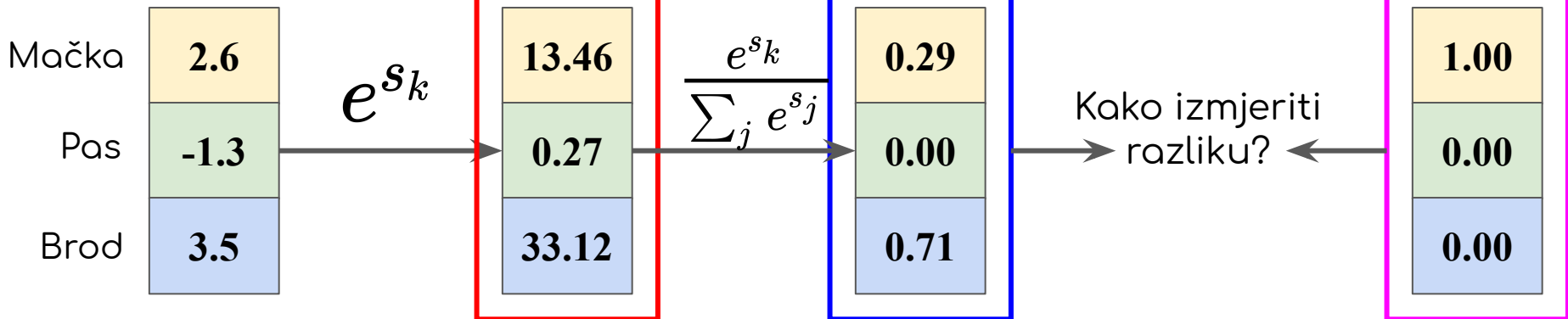
$$= -\log P(Y = y_i | X = x_i)$$



Nenormalizirane  
vjerovatnoće (logits)

Normalizirane  
vjerovatnoće

Poželjna  
vjerovatnoća



$$s = f(x_i, W) \sum_j e^{s_j} = 46.85$$

# Unakrsna entropija (cross-entropy loss)

- ❑ Koja je minimalna i maksimalna vrijednost funkcije gubitka Li?

$$\min = 0$$

$$\max = +\infty$$

$$P(Y = k|X = x_i) = \frac{e^{s_k}}{\sum_j e^{s_j}}$$

$$L_i = -\log P(Y = y_i|X = x_i)$$

- ❑ Ako su svi vektori rezultata (score) veoma male *random* vrijednosti, koja je vrijednost funkcije gubitka Li?

$$-\log \frac{1}{C}$$

C je broj klasa

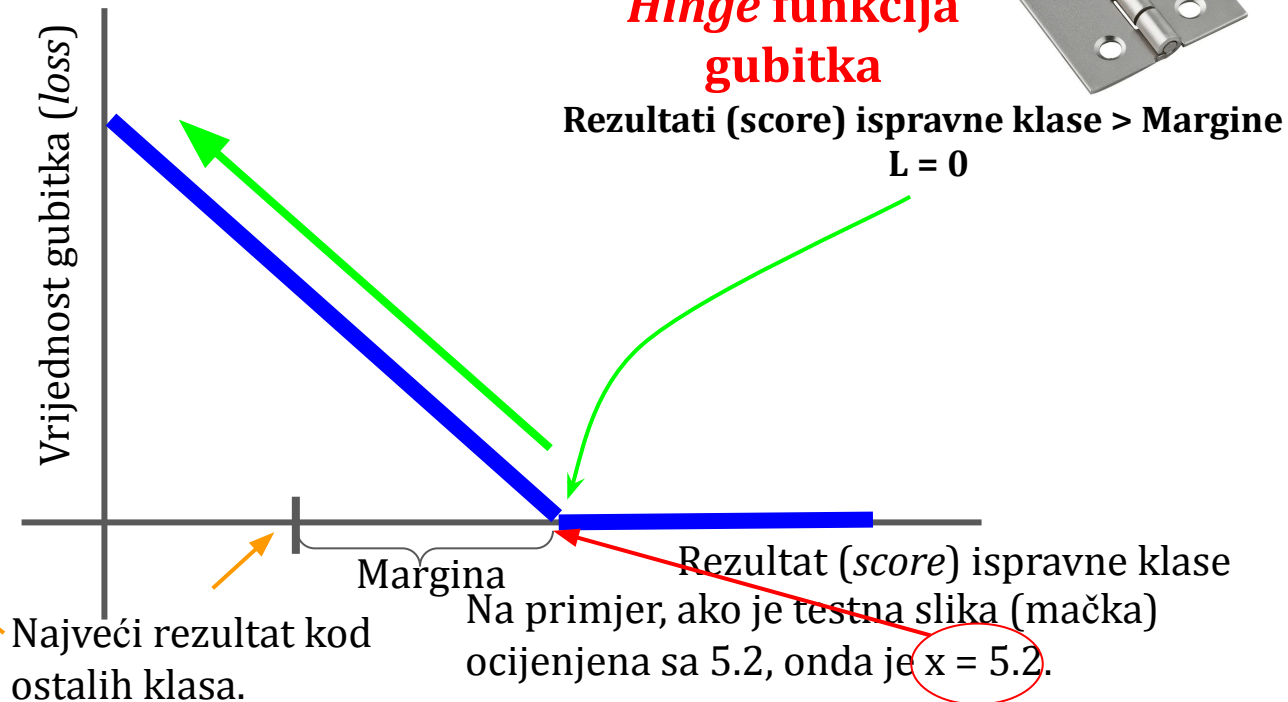
$$-\log_{10} \frac{1}{10} = 3.3$$

# Višeklasna SVM funkcija gubitka

- ❑ Rezultat tačne klase trebao bi biti veći od svih ostalih rezultata.
- ❑ Da li je to slučaj kod primjera koji pratimo?



Mačka	<del>2.6</del>	5.2 ✓
Pas	<del>-1.3</del>	-3.3
Brod	<del>3.5</del>	1.5



# Funkcija gubitka

Skup podataka

$$(x_i, y_i)_{i=1}^N$$

Apstraktna funkcija  
gubitka

$$L_i(f(x_i, W), y_i)$$

Za jednu sliku (i).

$$\left( \text{MAČKA} \right)_{i=1}^N$$

Obično se kodira preko integer vrijednosti.

$$L = \frac{1}{N} \sum_i L_i(f(x_i, W), y_i)$$

Za cijeli skup podataka.

U praksi postoji nekoliko različitih vrsta funkcija gubitka.

Na ovom času upoznat ćemo se sa unakrsnom entropijom (koristi se za softmax klasifikator) i višeklasnom SVM funkcijom gubitka.

# Višeklasna SVM funkcija gubitka

Apstraktna funkcija  
gubitka

$$s = f(x_i, W)$$

$$L_i(f(x_i, W), y_i)$$

Za jednu sliku (i).

$$s_j = f(x_i, W)_j$$

Rezultat (score) koji  
smo dobili iz  
klasifikatora za  
netačnu klasu.

$$L = \frac{1}{N} \sum_i L_i(f(x_i, W), y_i)$$

Za cijeli skup podataka.

$$L_i = \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + 1)$$

SVM funkcija gubitka

# Višeklasna SVM funkcija gubitka

## SVM funkcija gubitka

$$L_i = \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + 1)$$

Margina



Rezultat (score) koji smo dobili iz klasifikatora za netačnu klasu.

Rezultat (score) koji smo dobili iz klasifikatora za tačnu klasu.

Pas

2.14

Mačka

5.76

Konj

-1.34

# Višeklasna SVM funkcija gubitka



Pas

**-1.32**

2.14

**-3.2**

Mačka

0.98

**5.76**

1.3

Konj

3.72

**-1.34**

**-2.5**

Gubitak  
(*Loss*)

**9.34**

**0**

**5.1**

$$L_i = \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + 1)$$

```
>>> max(0, 0.98 - (-1.32) + 1) + max(0, 3.72 - (-1.32) + 1)
```

```
9.34
```

```
>>> max(0, 2.14 - 5.76 + 1) + max(0, -1.34 - 5.76 + 1)
```

```
0
```

```
>>> max(0, -3.2 - (-2.5) + 1) + max(0, 1.3 - (-2.5) + 1)
```

```
5.1
```

$$L = \frac{9.34+0+5.1}{3} = 4.46$$

# Višeklasna SVM funkcija gubitka



Pas	<b>-1.32</b>	2.14	-3.2
Mačka	0.98	<b>5.76</b>	1.3
Konj	3.72	-1.34	<b>-2.5</b>
Gubitak (Loss)	<b>9.34</b>	<b>0</b>	<b>2.8</b>

$$L_i = \sum_{j \neq y_i} \max(0, s_j - s_{y_i} + 1)$$

1. Šta će se desiti ukoliko se rezultat za klasu "mačka" promijeni malo?
  - a. Obzirom da je score značajno veći u odnosu na druge klase, funkcija gubitka će i dalje imati vrijednost 0.
2. Koja je maksimalna i minimalna vrijednost za gubitak (loss)?
  - a. Min je 0, a max je beskonačno!
3. Ako se matrica W inicijalizira na neke male nasumične vrijednosti (npr. sa Gausovom distribucijom), šta možemo očekivati kao vrijednost gubitka?
  - a.  $L = C - 1$ , C je broj klasa/kategorija.
4. Šta bi se desilo kada bismo računali sumi svih kategorija (tačnih i netačnih)?