

Vještačka inteligencija

Predavanje 4: Uvod u neuronske mreže

*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

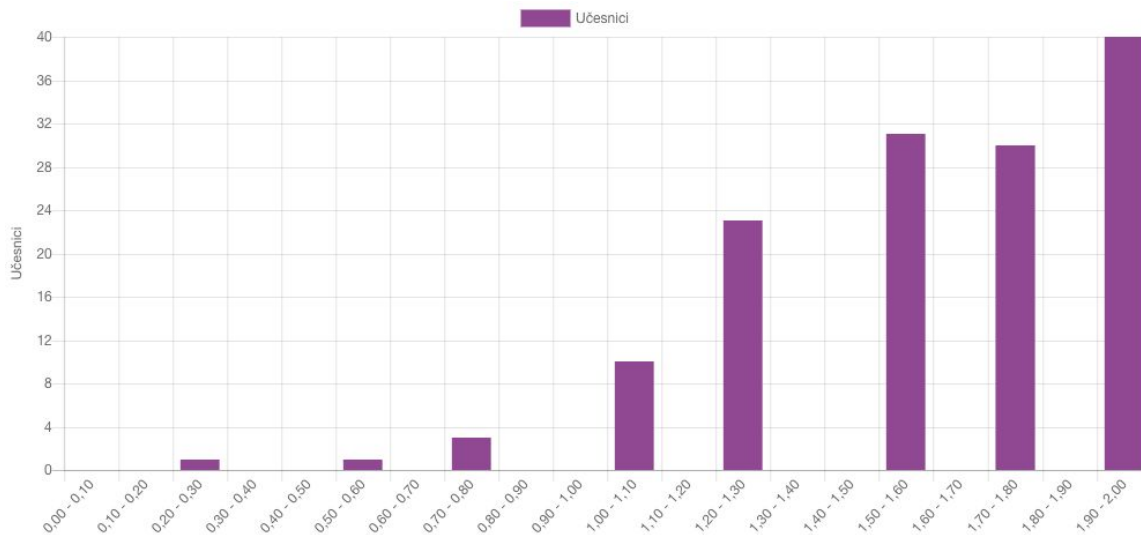
- This work is licensed under a Creative Commons 'Attribution-NonCommercial-ShareAlike 4.0 International' license. EN: <https://creativecommons.org/licenses/by-nc-sa/4.0/>



- 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 drugi kviz u okviru realizacije vježbi.
 - ❑ 8 pitanja, 5 minuta vremena.
 - ❑ Automatsko ocjenjivanje nakon isteka vremena.
- ❑ Sa predavanjima nastavljamo u 14:10+.
- ❑ Prva verzija skripte objavljena na Google Drive-u.

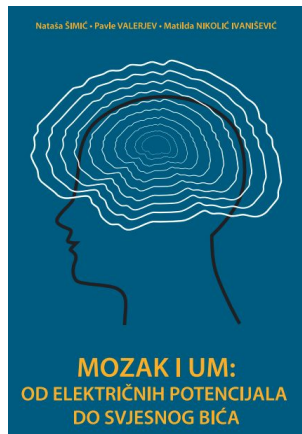


Motivacija

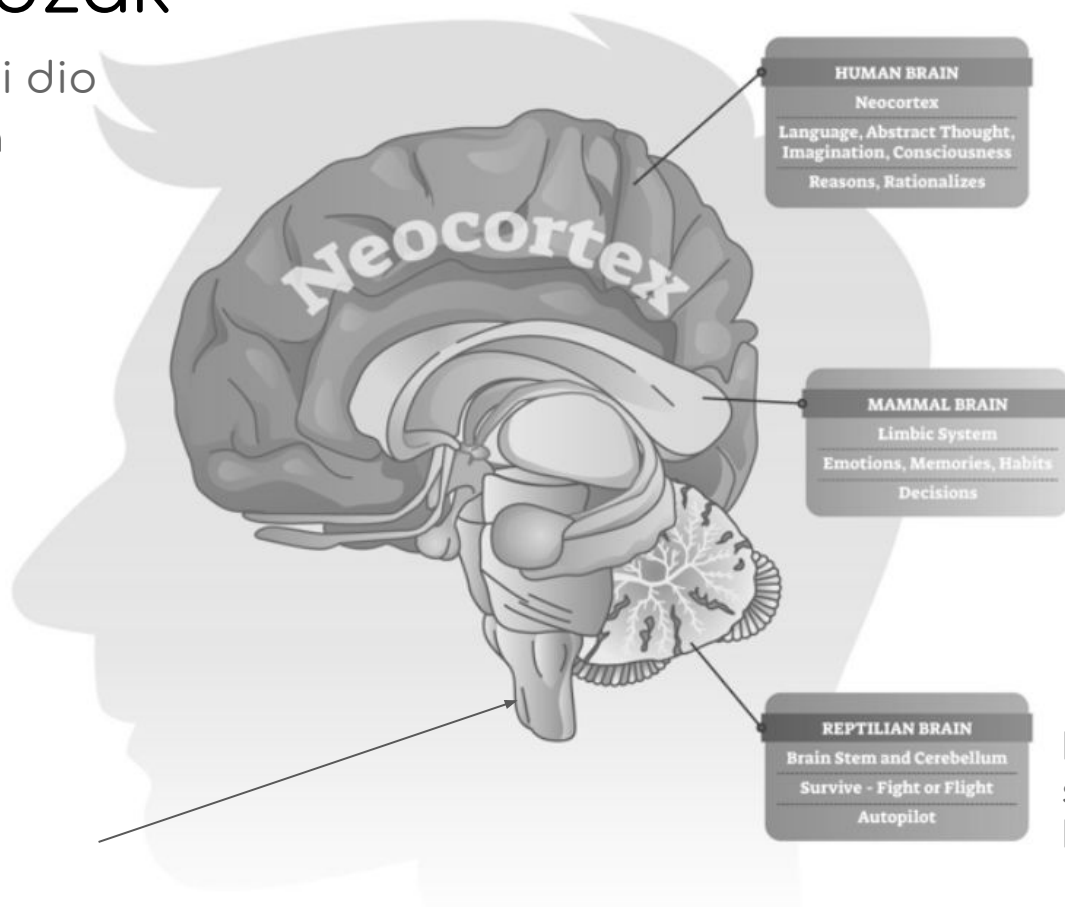
- ❑ Neke aktivnosti se mogu automatizirati i njih danas pretežno realiziraju računari.
- ❑ Ali, šta je sa onim aktivnostima koje još nisu automatizirane?
- ❑ **Mentalna aktivnost** se sastoji od elektrohemijskih aktivnosti **moždanih ćelija** koje se nazivaju **neuroni**.
- ❑ Inspirisani ovom hipotezom, naučnici su pokušali na naprave **vještačke neuronske mreže**, koje mogu na isti ili sličan način procesirati informacije i donositi zaključke.
- ❑ Sistem vještačke inteligencije koji uspješno oponaša rad mozga naziva se **intelligentnim sistemom**.
 - ❑ Drugi nazivi koji se još mogu naći u praksi: **konektivizam**, **paralelno distribuirano procesiranje** i **neuronsko računanje**.

Ljudski mozak

Mozak je središnji dio živčanog sistema čovjeka.



Kičmena
moždina



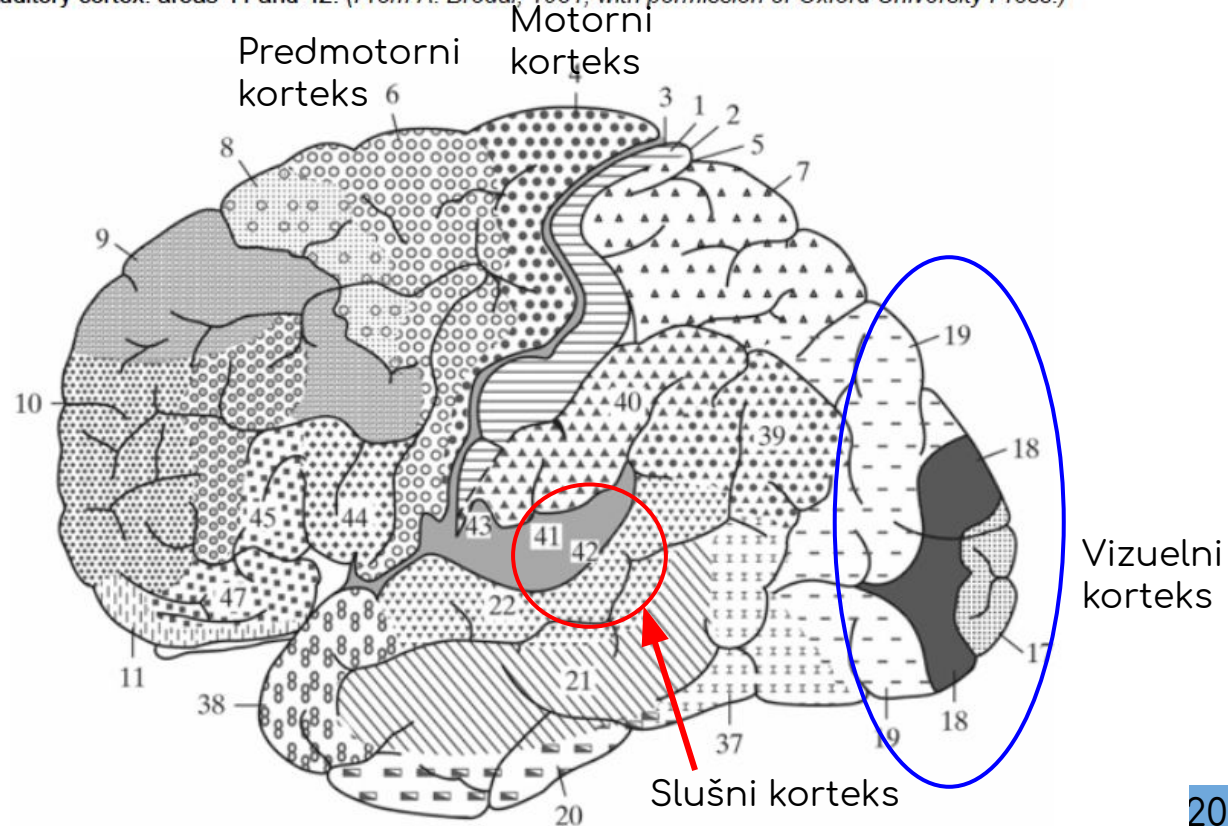
Neokorteks
može biti
sjedište
svijesti.

Limbički sistem
Mammal (sisavac)

Moždano
stablo
Mali mozak

Ljudski mozak

Figure 4 Cytoarchitectural map of the cerebral cortex. The different areas are identified by the thickness of their layers and types of cells within them. Some of the key sensory areas are as follows: Motor cortex: motor strip, area 4; premotor area, area 6; frontal eye fields, area 8. Somatosensory cortex: areas 3, 1, and 2. Visual cortex: areas 17, 18, and 19. Auditory cortex: areas 41 and 42. (From A. Brodal, 1981; with permission of Oxford University Press.)



Motivacija

- ❑ Znamo da je mozak sačinjen od velikog broja neurona koji rade paralelno.
- ❑ Poznate su i sljedeće činjenice:
 - ❑ Postoji više od 100 vrsta različitih neurona.
 - ❑ Svaka vrsta radi vrlo jednostavnu obradu podataka.
 - ❑ Vrijeme obrade unutar neurona: 2 ms.
 - ❑ Broj neurona u ljudskom mozgu: 10^{11} .
 - ❑ Svaki neuron u prosjeku dobija informacije od 10^3 do 10^4 drugih neurona.
 - ❑ Informacije se obrađuju i serijski i paralelno.
 - ❑ Informacije su analogne.
 - ❑ Obrada je tolerantna na pogreške.

Vještačke neuronske mreže

- ❑ Nova paradigma: vještačke neuronske mreže (eng. *Artificial Neural Networks* ili ANN).
- ❑ Nova grana računarskih nauka: **neuro-računarstvo**.
- ❑ Postoje dva pristupa razvoja vještačke inteligencije:
 - ❑ **Simbolički pristup**: znanje iz neke domene nastoji se obuhvatiti skupom atomičkih semantičkih objekata (simbola) i zatim manipulirati tim simbolima pomoću **algoritamskih pravila**.
 - ❑ Nije ispunio početna očekivanja, mada je imao određene uspjehe u oblasti ekspertnih sistema. Sa ovim uspjesima upoznat ćemo se na nekom od narednih predavanja.
 - ❑ Nije svako znanje moguće formulizirati formalnim pravilima.
 - ❑ **Konektivistički pristup**: zasniva se na izgradnji sistema arhitekture slične arhitekturi mozga koji, umjesto da se programira, **uči samostalno** na bazi iskustva.

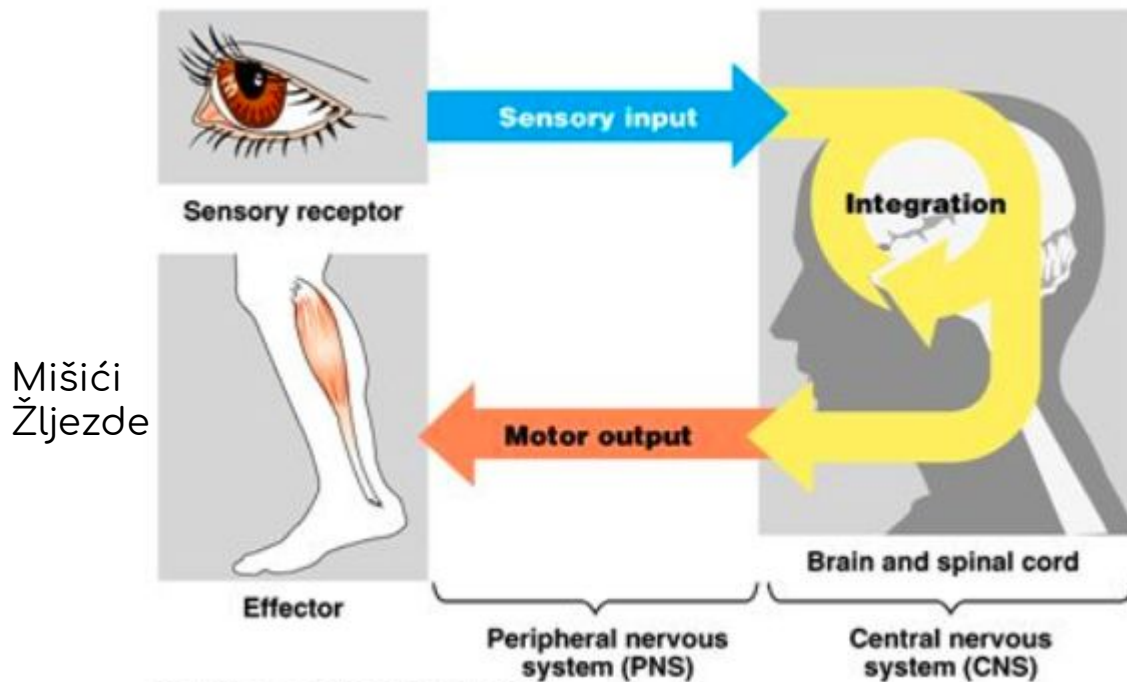
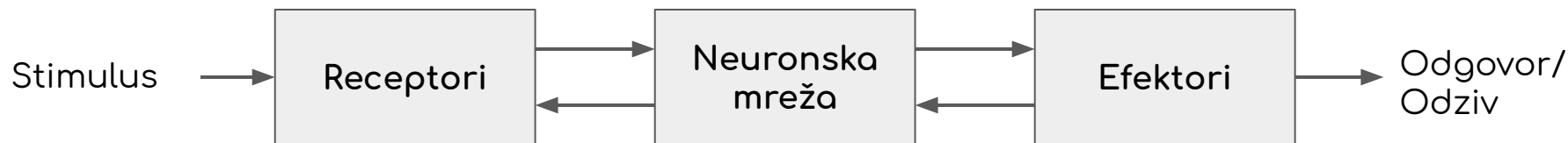
Zašto *neuronska mreža*?

- ❑ Mozak je izuzetno složen, nelinearan i paralelan računar.
- ❑ Funkcija **vizualnog sistema** je da pruži reprezentaciju okoline oko nas i informacije potrebne za interakciju sa okolinom.
- ❑ Zadaci prepoznavanja ili percepcije rutinski se izvode za otprilike 100-200ms.



- ❑ Plastičnost mozga omogućava nervnom sistemu da se prilagodi svojoj okolini.

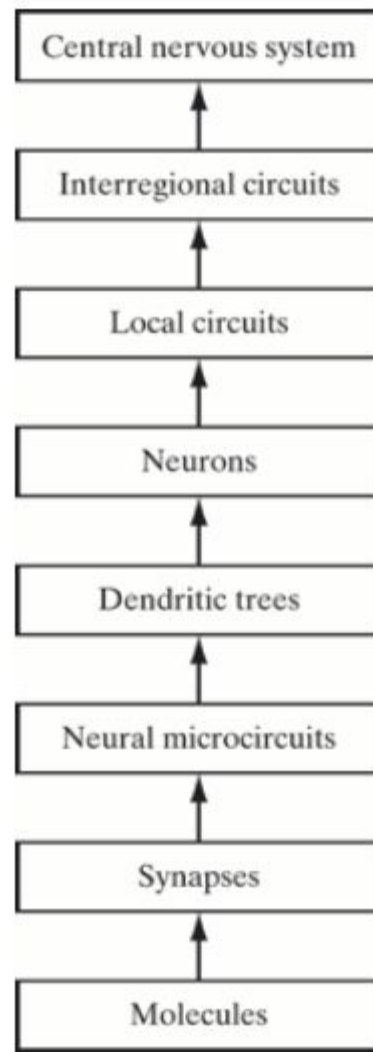
Zašto *neuronska mreža*?



Copyright © Pearson Education, Inc., publishing as Benjamin Cummings.

Zašto *neuronska mreža*?

- ❑ Vještačka neuronska mreža je mašina koja je dizajnirana da modelira način na koji mozak izvršava određene zadatke ili funkcije od interesa.
- ❑ U istraživanju se fokusiramo na neuronske mreže koje izvedu korisne proračune kroz proces učenja.
- ❑ Algoritam učenja je postupak koji se koristi za izvođenje procesa učenja.
- ❑ Zanimljiva je činjenica je da biološka neuronska mreža može modificirati vlastitu topologiju, koja je motivirana činjenicom da ljudske moždane ćelije (neuroni) mogu “*umrijeti*”, te da se mogu formirati nove sinaptičke veze koje ne uključuju ove neurone.
- ❑ Računska snaga neuronske mreže:
 - ❑ Masivno je distribuirana i paralelna struktura.
 - ❑ Ima sposobnost učenja, pa stoga i generaliziranja (*generalizacija*).



Proces učenja

- ❑ Postoje dvije faze rada vještačke neuronske mreže:
 - ❑ Faza učenja (treniranja) i
 - ❑ Faza zaključivanja ili obrade podataka (*inference* ili zaključivanje).
- ❑ **Učenje** je iterativan pristup predstavljanja ulaznih primjera (uzoraka, iskustva) i očekivanog izlaza pri čemu dolazi do postupnog prilagođavanja koeficijenata (veza) između neurona.
- ❑ Razlikujemo tri vrste učenja:
 - ❑ Pojedinačno učenje (eng. *online*): učenje se dešava nakon svakog predstavljenog uzorka.
 - ❑ Učenje s minigrupama (eng. *mini-batches*): učenje se događa nakon više predočenih uzoraka.
 - ❑ Grupno učenje (eng. *batch*): učenje se događa tek nakon svih predočenih uzoraka.
- ❑ **Epochom** nazivamo jednu fazu ili stadij predstavljanja ulaznih podataka.

Proces učenja

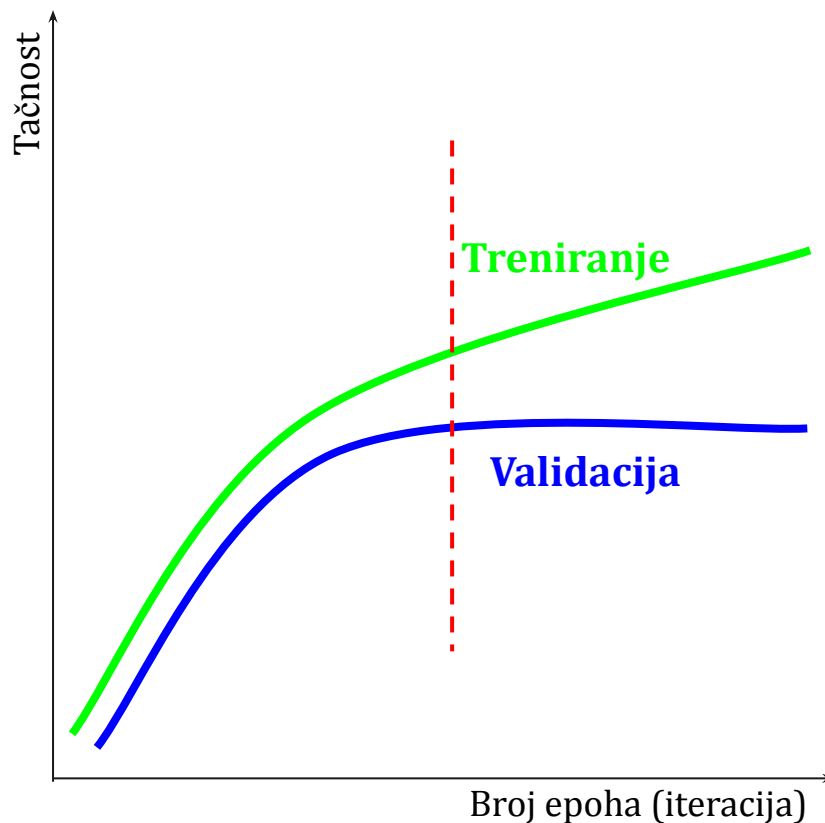
- ❑ Proces učenja omogućava odlučivanje o nekim budućim događajima na osnovu posmatranja svoje okoline ili pravljenja opservacija o svijetu.
- ❑ Učenje može biti trivijalno, kao što je biranje broja na telefonu, ili složeno, kao ono koje je pokazao Albert Einstein - nova teorija u Univerzumu.
- ❑ Zašto je bitno imati mogućnost učenja?
 - ❑ Nemoguće je predvidjeti sve moguće scenarije u kojima će se program/agent naći.
 - ❑ Nemoguće je predvidjeti sve promjene u vremenu koje će se desiti. Program koji treba da predviđa stanje na tržištu dionica mora imati mogućnost da se prilagodi na nove promjene onda kada one nastaju.
 - ❑ Ponekad programeri ne znaju kako isprogramirati rješenje za neki problem. Primjer: Kako napisati program za prepoznavanje lica?
- ❑ **Induktivno učenje:** sposobnost pronalaženja generalne funkcije ili pravila na osnovu tačno određenih ulazno/izlaznih parova podataka.

Proces učenja

- ❑ Postoji nekoliko različitih načina učenja u odnosu na povratnu spregu (*feedback*) u procesu učenja:
 - ❑ **Nadzirano učenje** (*supervised learning*): u procesu učenja postoje ulazni/izlazni parovi koji pomažu programu/agentu da uči iz podataka.
 - ❑ **Nenadzirano učenje** (*unsupervised learning*): ne postoji povratna sprega u procesu učenja.
 - ❑ **Podržano učenje** (*reinforcement learning*): učenje na bazi nagrada i kazni.
- ❑ Proces učenja izvodi se nad **skupom podataka** (*dataset*). Tok procesa se stalno prati kroz **provjeru tačnosti** nad skupom podataka za treniranje (*learning accuracy*), zatim nad skupom za validaciju i skupom za testiranje.
- ❑ Mreža mora imati sposobnost **generalnog zaključivanja** nad skupom za testiranje - podaci koje mreža nikad nije vidjela.
- ❑ Bitno je da ne dođe do pojave **pretreniranosti** (*overfitting*) nad podacima za treniranje.

Proces učenja

- ❑ Proces treniranja potrebno je zaustaviti u trenutku kada tačnost počinje stagnirati ili opadati na skupu za validaciju.



Nadzirano učenje

- ❑ Zadatak nadziranog učenja je sljedeći:

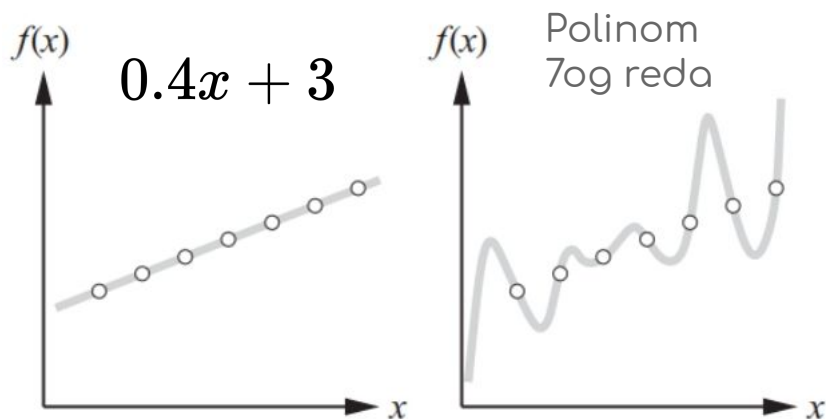
- ❑ Poznat je skup podataka za treniranje

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_N, y_N)$$

- ❑ Svaki y_i generisan je nepoznatom funkcijom $y = f(x)$.
- ❑ Zadatak je naći funkciju h koja vrši aproksimaciju funkcije f .
- ❑ Funkcija h se naziva hipoteza.
- ❑ Proces učenja je pretraživanje prostora svih hipoteza H i pronalaženje one hipoteze koja će imati zadovoljavajuće rezultati i nad podacima izvan skupa za treniranje.
- ❑ Definiše se i testni skup podataka nad kojim se mjeri tačnost hipoteze.
- ❑ Kažemo da hipoteza ima sposobnost generaliziranja ukoliko ispravno izvrši predikciju vrijednosti y nad testnim skupom podataka.

Nadzirano učenje

- ❑ **Klasifikacija**: izlaz se može definisati kao jedna vrijednost od konačnog skupa mogućih vrijednosti (npr. sunčano, oblačno, kišovito, itd.).
- ❑ Razlikujemo dvije vrste klasifikacije: **binarnu** (samo dvije vrijednosti) i **višeklasnu klasifikaciju** (više vrijednosti).
- ❑ **Regresija**: izlaz je broj (npr. temperatura)
- ❑ Prvi primjer:

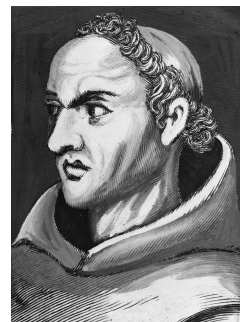


Ukoliko funkcija h odgovara svim podacima, onda se ona naziva **konzistentom hipotezom**.

Problem s kojim se često susrećemo: kako izabrati između dvije ili više konzistentnih hipoteza?

Ockham's razor

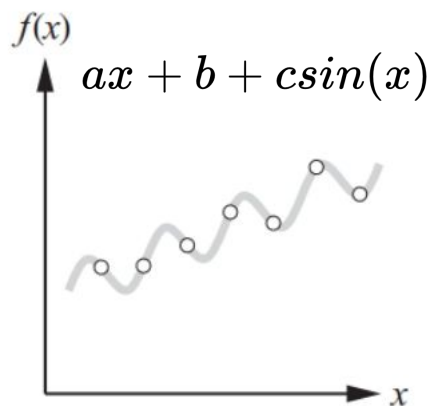
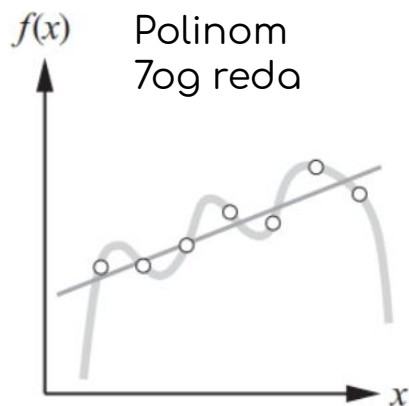
- ❑ Problem s kojim se često susrećemo: kako izabrati između dvije ili više konzistentnih hipoteza?
- ❑ Jedno od mogućih rješenja je da bismo najjednostavnije pravilo/funkciju u odnosu na konzistentnost sa podacima.
- ❑ Ovaj princip nazivamo "Ockham's razor".
- ❑ Ime je dobilo po engleskom filozofu William of Ockham iz 14. vijeka.



William of Ockham

Nadzirano učenje

- ❑ Drugi primjer predstavlja drugi skup podataka.
- ❑ U prvom slučaju predstavljeno je rješenje u dva oblika: polinom 7og reda i ravna crta. Koje rješenje je bolje?
- ❑ Potrebno je napraviti **kompromis** između složenosti funkcije (polinom) i mogućnosti generalizacije (ravna crta).
- ❑ U drugom slučaju pronađena je konzistentna hipoteza sa jednostavnom funkcijom!



Kako znati da li je moguće pronaći funkciju koju je moguće realizirati?

Istinska ili prava funkcija nije uvijek poznata, pa je teško reći da li je proces učenja moguće realizirati.

Brzina izračunljivosti funkcije h igra važnu ulogu u izboru funkcije.

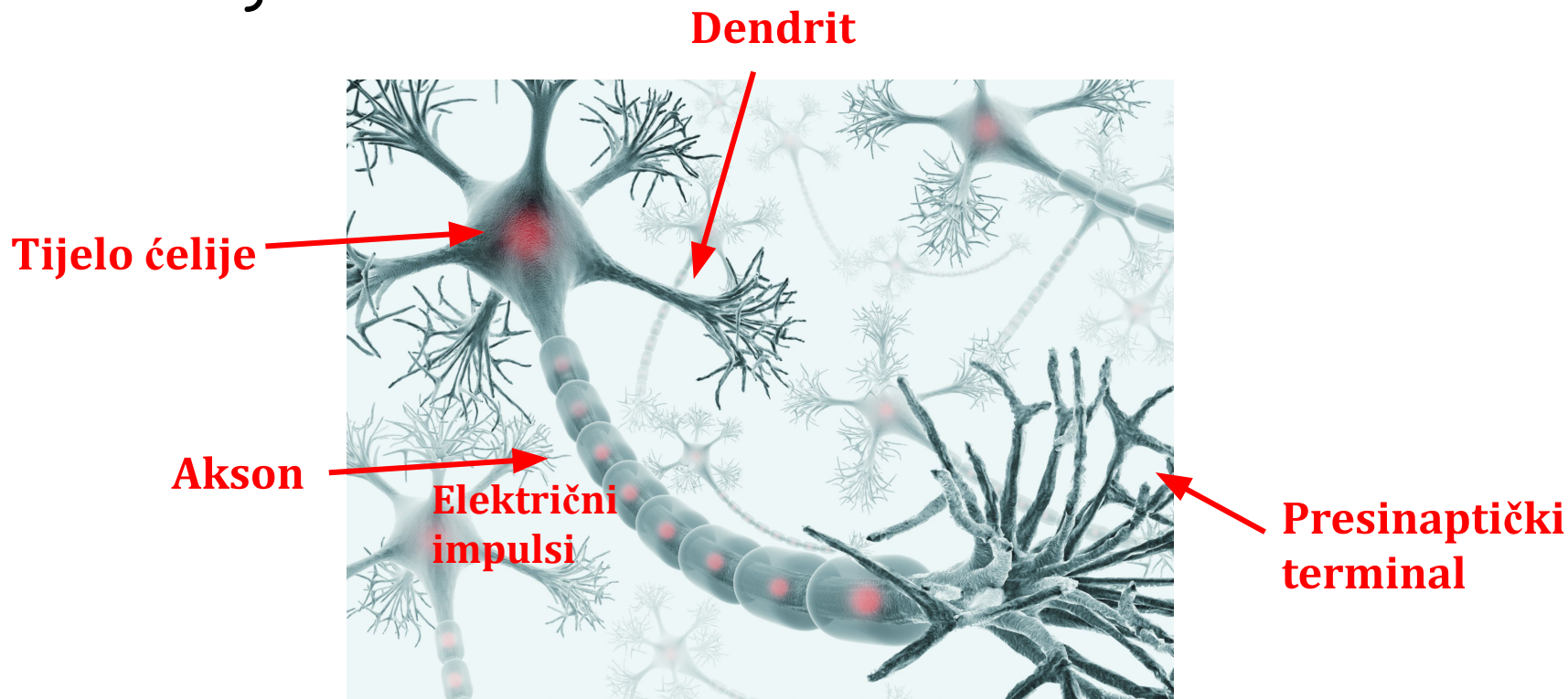
Izbor modela

- ❑ Polinomi većeg reda mogu značajno bolje opisati skup podataka za treniranje, međutim bitnije je naći funkciju koja će se pokazati zadovoljavajuće na validacijskom i testnom skupu podataka.
- ❑ **Izbor modela** je proces kojim se bira red polinoma, tačnije bira se prostor hipoteze, a zatim se procesom **optimizacije** pronalazi funkcija najbolja hipoteza unutar tog prostora.
- ❑ Definišemo **funkciju gubitka** L (od eng. *loss*).

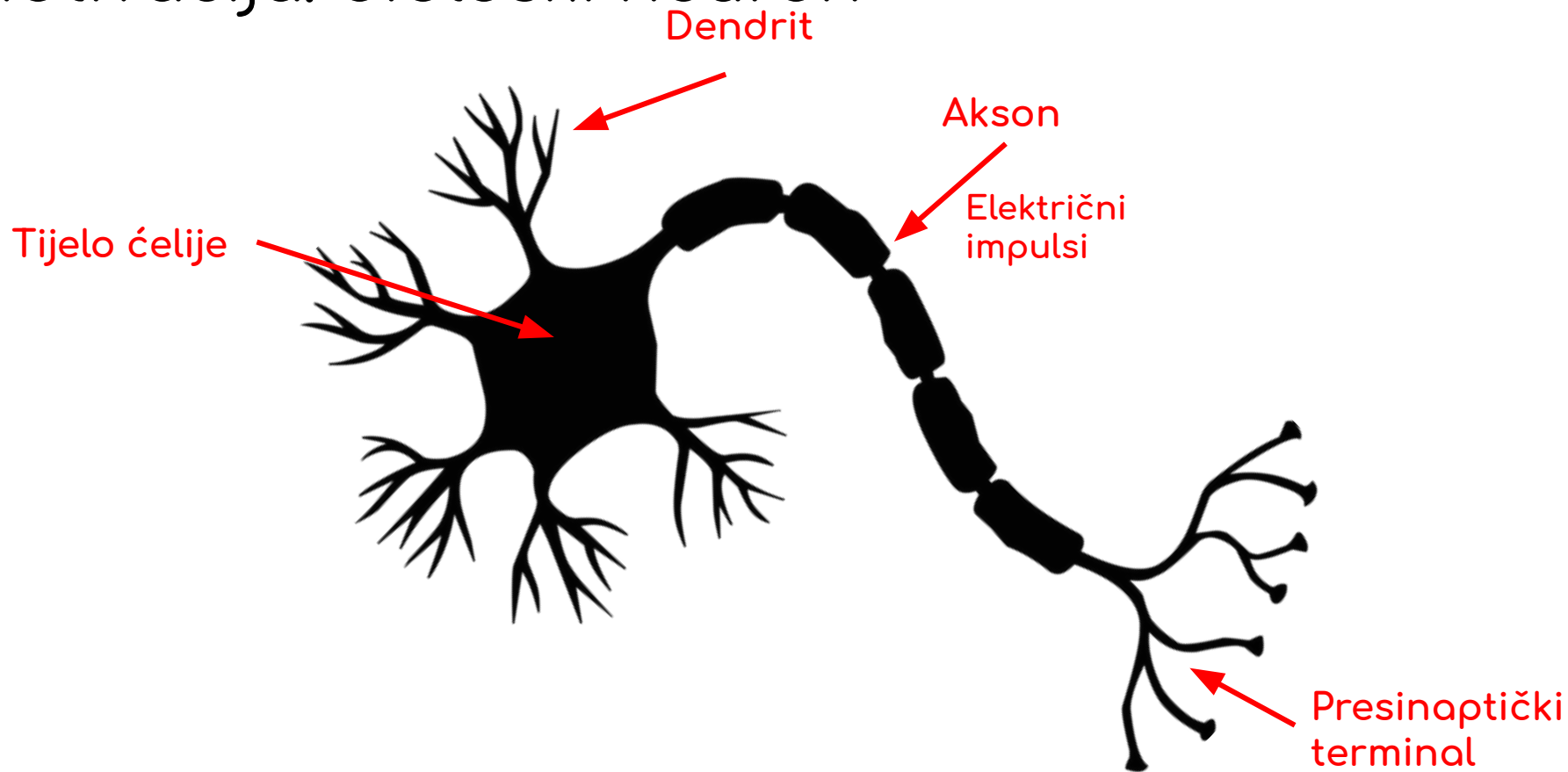
$$\begin{array}{ll} h(x) = \hat{y} & h(x) = 137.036 \\ f(x) = y & y(x) = 137.035999 \end{array} \quad L(x, y, \hat{y})$$

- ❑ Apsolutna vrijednost gubitka (L1): $L_1(y, \hat{y}) = |y - \hat{y}|$
- ❑ Kvadratna vrijednost greške (L2): $L_2(y, \hat{y}) = (y - \hat{y})^2$

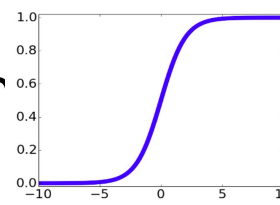
Motivacija: biološki neuron



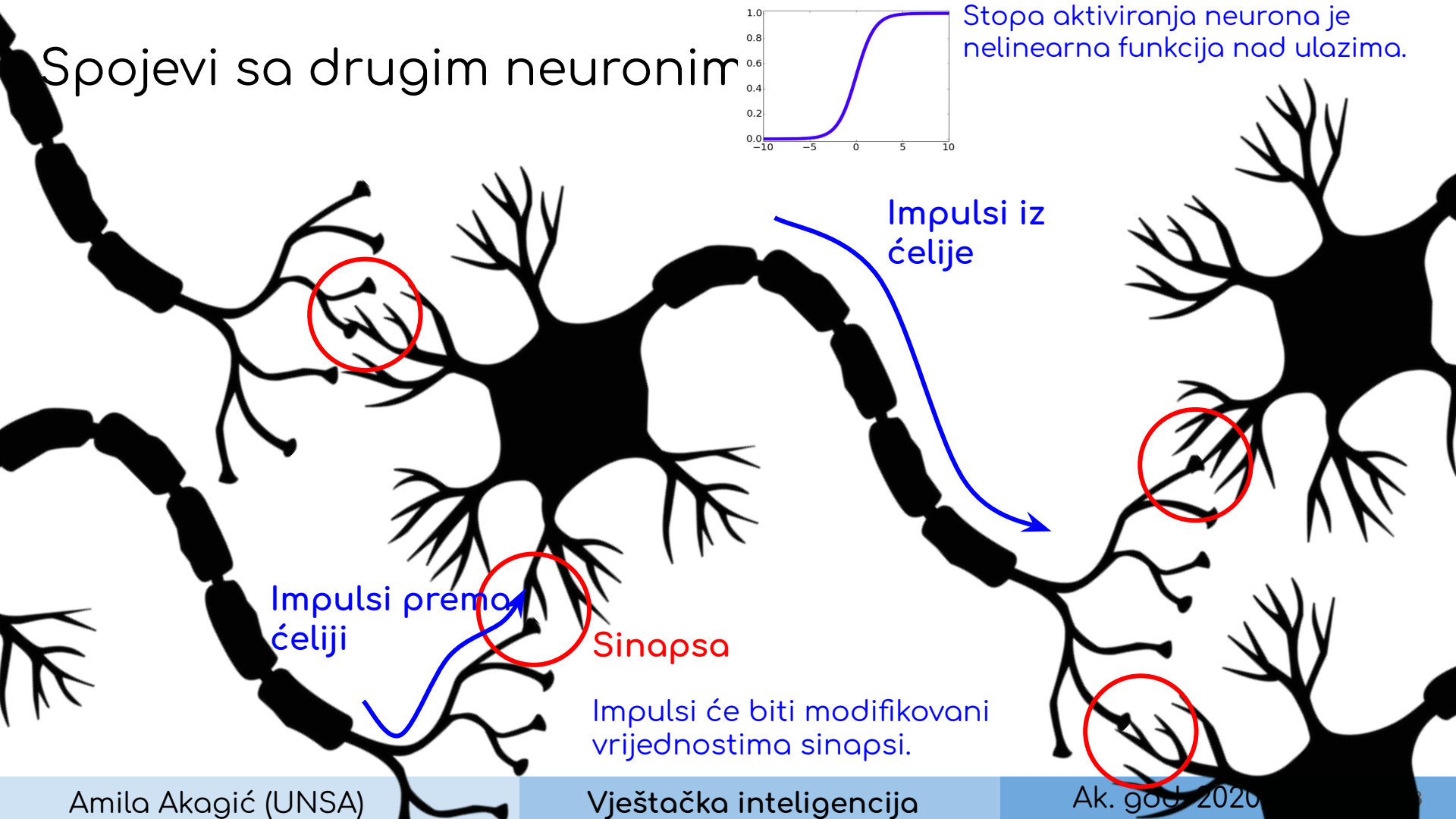
Motivacija: biološki neuron



Spojevi sa drugim neuronim

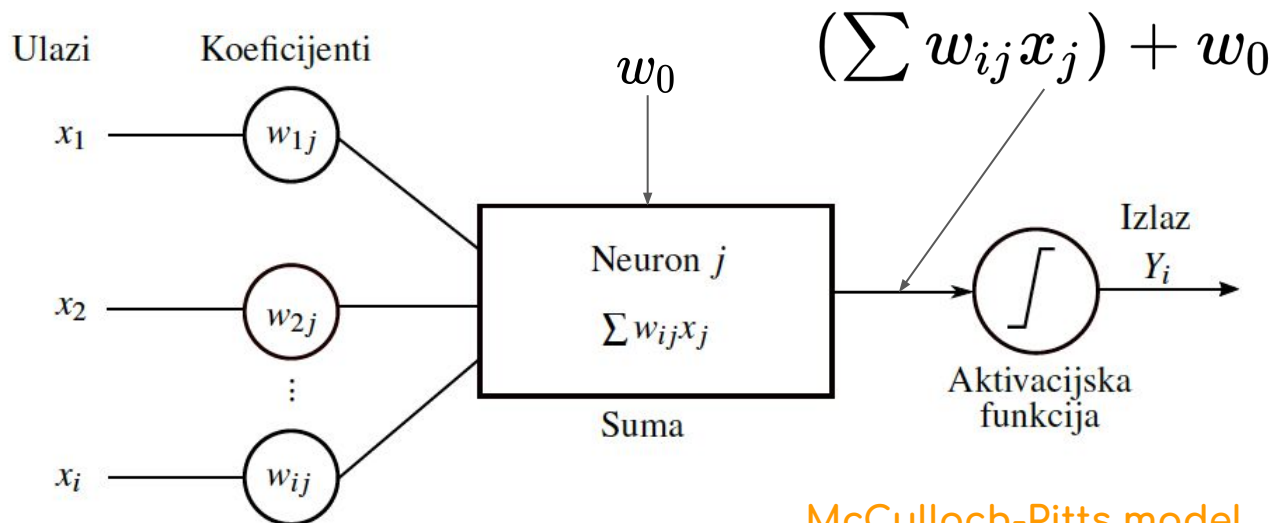


Stopa aktiviranja neurona je nelinearna funkcija nad ulazima.



Prvi model biološkog neurona

- ❑ Threshold Logic Unit ili TLU-perceptron predstavljen je od strane Warren McCulloch i Walter Pitts, 1943 godine.
- ❑ Predstavlja prvi model biološkog neurona.

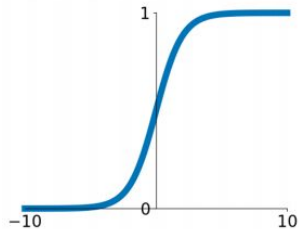


McCulloch-Pitts model

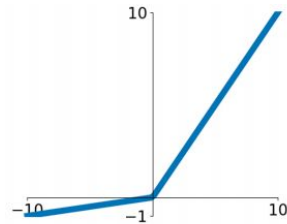
Aktivacijske funkcije: koju funkciju izabrati?

Sigmoid

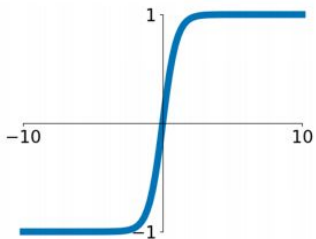
$$\sigma(x) = \frac{1}{1+e^{-x}}$$



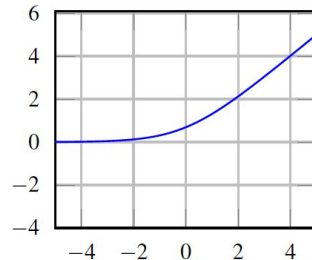
Leaky ReLU
 $\max(0.01x, x)$



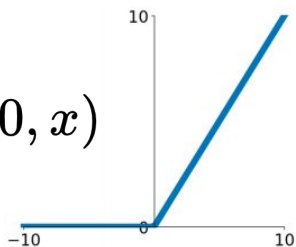
$$\tanh(x) = \frac{e^{2x}-1}{e^{2x}+1}$$



Softplus
 $\log(1 + e^x)$

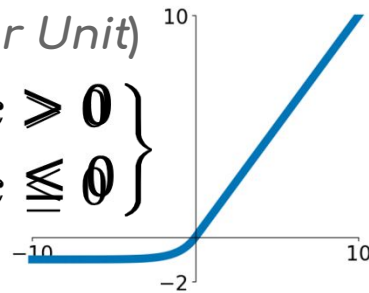


$$\text{ReLU}(x) = \max(0, x)$$

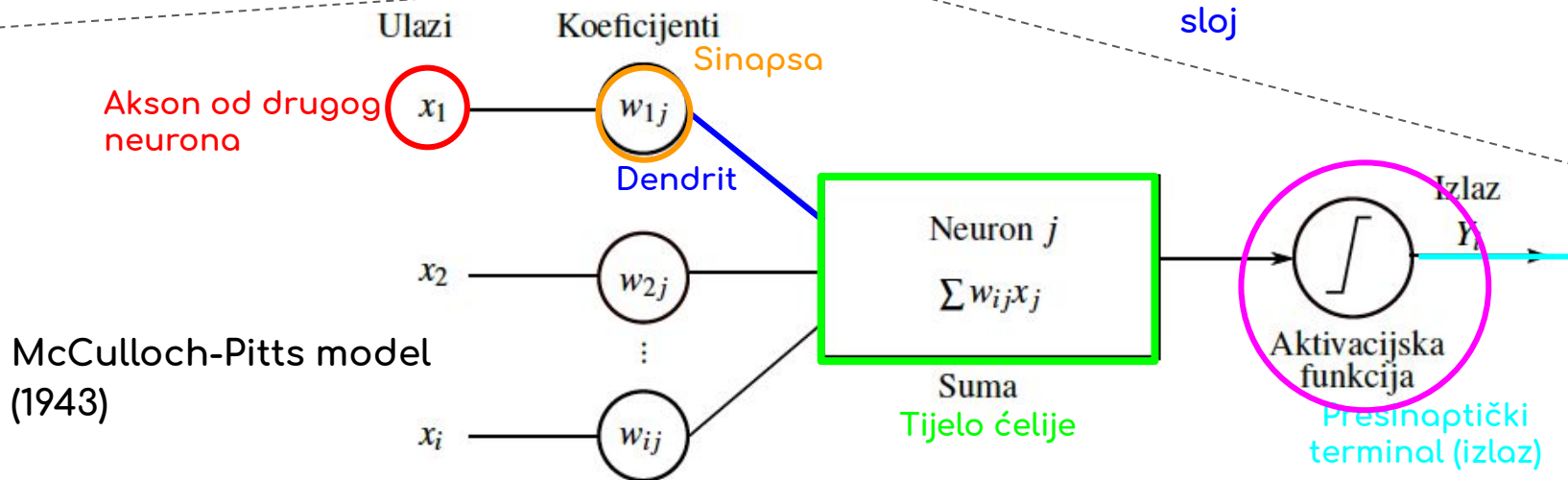
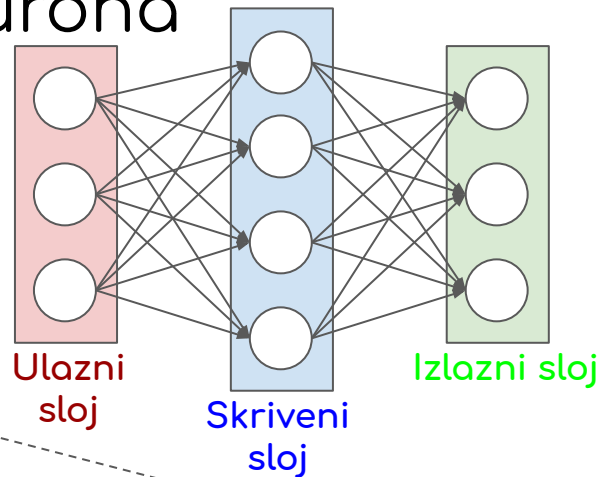
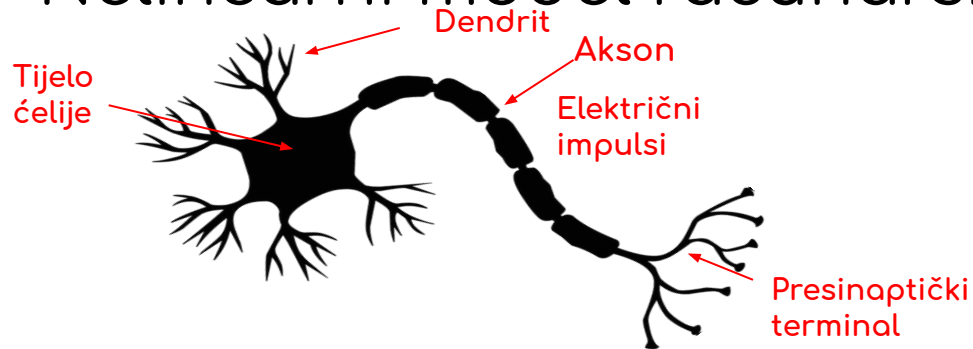


ELU (*Exponential Linear Unit*)

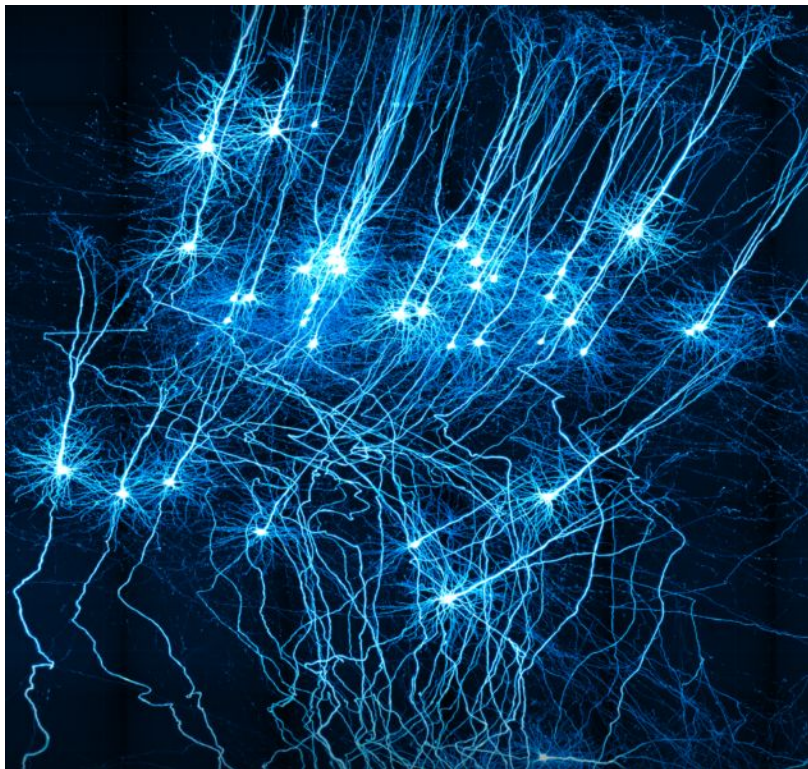
$$f(x) = \begin{cases} x, & x \geq 0 \\ \alpha(e^x - 1), & x \leq 0 \end{cases}$$



Nelinearni model računarskog neurona



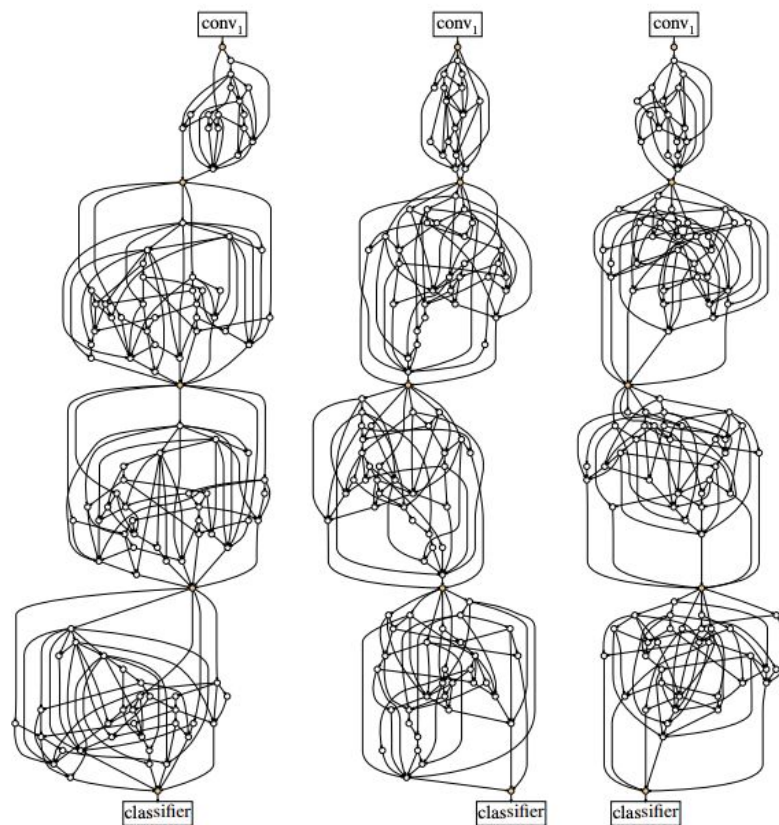
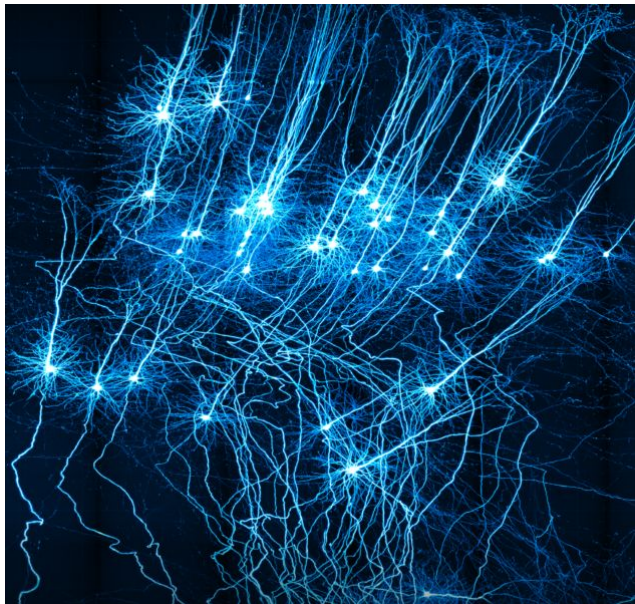
Razlika između biološkog i računarskog neurona



- ❑ Biološki neuroni imaju izuzetno kompleksnu topologiju (vezu između neurona).
- ❑ Postoji različit broj vrsta bioloških neurona.
- ❑ Dendriti mogu izvršavati složene nelinearne proračune.
- ❑ Sinapse nisu pojedinačne težine već složeni nelinearni dinamički sistem.
- ❑ Računarski neuroni su organizovani u slojeve regularnog formata što omogućava efikasno izračunavanje različitih operacija (primarno sabiranje i množenje).
- ❑ Sloj je apstraktna vještačka tvorevina koja treba da preslika strukturu bioloških neurona.

Treba biti oprezan kada se pravi analogija između biološkog i računarskog neurona jer su sličnosti simbolične, tj. računarski neuron ne predstavlja stvarnu reprezentaciju biološkog neurona.

Razlika između biološkog i računarskog neurona



Exploring Randomly Wired Neural Networks for Image Recognition

Saining Xie Alexander Kirillov Ross Girshick Kaiming He

Facebook AI Research (FAIR)

<https://arxiv.org/pdf/1904.01569.pdf>

1949: Hebbovo pravilo učenja

❑ Hebb-ovo pravilo učenja:

"Kad je akson neurona A dovoljno blizu da aktivira neuron B i to ponavlja veći broj puta dolazi do metaboličkih promjena tako da se povećava efikasnost neurona A u aktiviranju neurona B."

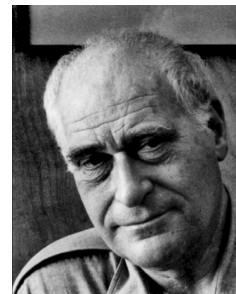
'neurons that fire together wire together'



Donald O. Hebb

❑ Stent je 1973. godine proširio Hebb-ovo pravilo:

"Ako jedan neuron ne utiče na drugog (tj. neuroni se aktiviraju asinhrono) onda sinapsa među njima postaje slabija ili se potpuno elimiše."



Gunther S. Stent

Pravilo učenja perceptrona

- ❑ 1958. godine Rosenblatt spaja Hebb-ovu ideju i McCulloch-Pitts model te definiše pravilo učenja perceptrona, koje glasi:
 1. Ciklično se prolazi kroz svih N uzoraka za učenje, jedan po jedan.
 2. Izvrši se klasifikacija trenutnog uzorka.
 - a. Ako je klasifikacija ispravna, ne mijenjaj koeficijente i
 - i. ako je to N-ti uzastopni uzorak klasificiran ispravno, prekini proces učenja,
 - ii. inače pređi na sljedeći uzorak.
 - b. Ako je klasifikacija neispravna, izvrši korekciju vrijednosti koeficijenta perceptrona prema sljedećoj formuli:

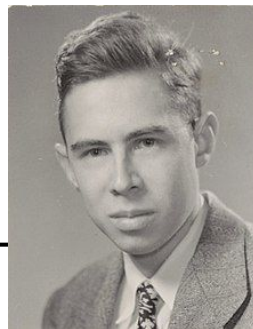
$$w_i(k + 1) = w_i(k) + \eta(y - \hat{y})x_i$$

Stopa učenja

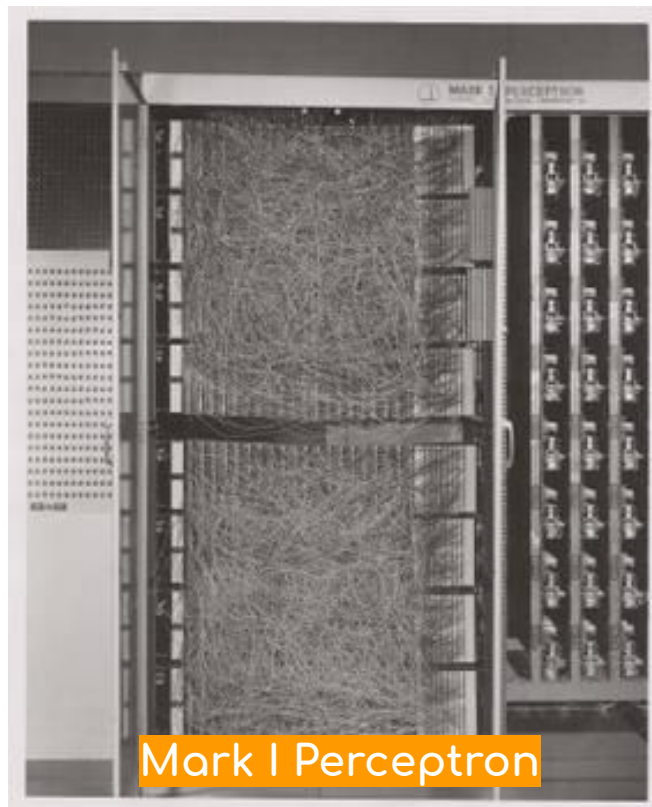
- ❑ Parametar η se naziva **stopa učenja**, koji može da poprими različite vrijednosti, kao npr.
 - ❑ Može biti pozitivan veoma mali broj (npr. 0.0001 ili 0.5), koji određuje koliko treba ažurirati trenutnu vrijednost koeficijenta.
 - ❑ Ako je broj premali, postupak učenja će napredovati veoma sporo.
 - ❑ Ako je broj preveliki, postupak može divergirati u neželjenom smjeru.

1958: Perceptron

- ❑ Jedan od prvih algoritama koji je imao mogućnost učenja iz podataka.
- ❑ Bio je implementiran u hardveru!
 - ❑ Vrijednosti koeficijenta su čuvani u potencijometarima, a električni motori su korišteni za njihovo ažuriranje!!!
- ❑ Rad perceptrona je demonstriran na prepoznavanju alfabeta sa 20x20 slike. Kamera je koristila fotočelije kadmijum sulfida.
- ❑ Po današnjoj klasifikaciji metoda, ova metoda bi spadala u metode linearnih klasifikatora.



Frank Rosenblatt

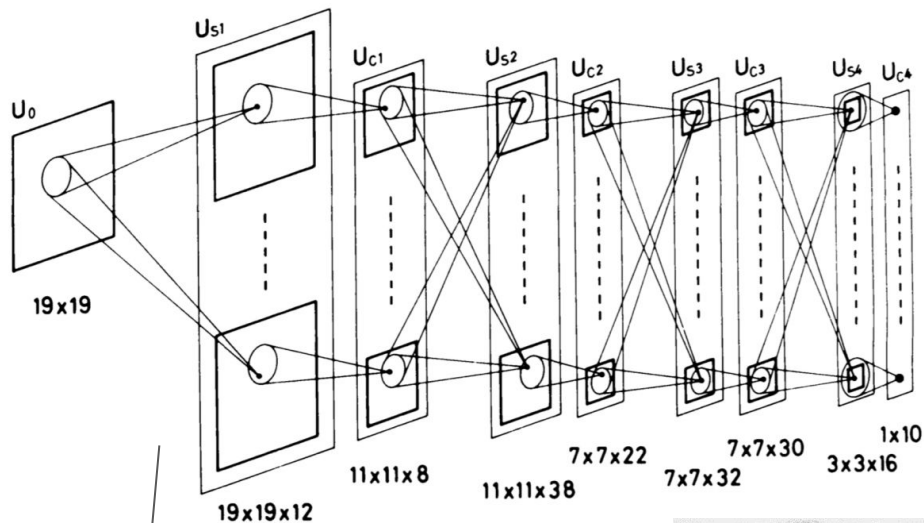


Mark I Perceptron

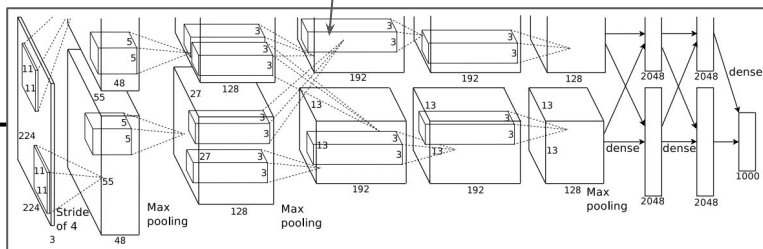
1958
Perceptron

1980: Neocognitron

- 1980: Novi algoritam za prepoznavanje šablona (*pattern recognition*) baziran na modelu vizuelnog sistema.
- Višeslojna neuronska mreža.
- Neocognitron je samoorganizirajuća mreža nastala učenjem bez nadzora.
- Osnovne karakteristike: hijerarhija zasnovana na jednostavnim (**konvolucija**) i složenim (**pooling**) ćelijama.



32 godine!



Kunihiko Fukushima

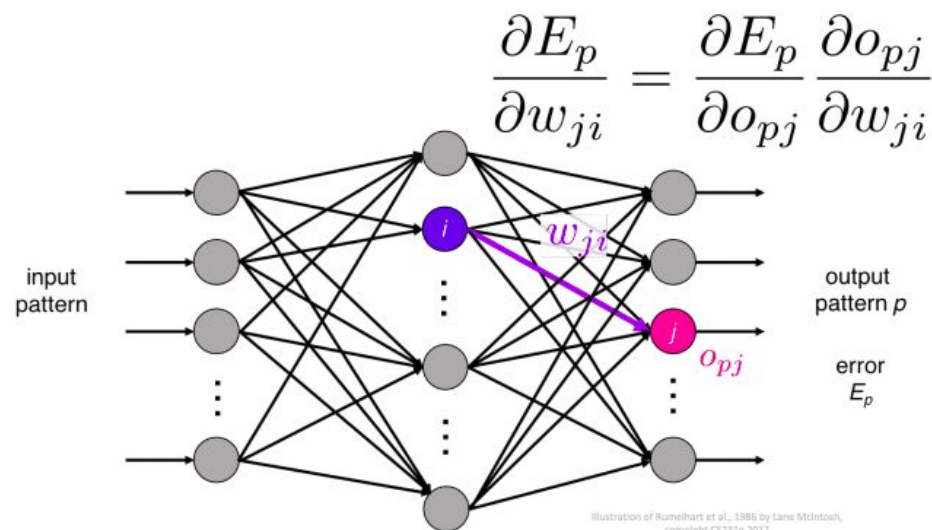
1958
Perceptron

1969
Minsky & Papert

1980
Fukushima

1986: Backprop / Rumelhart, Hinton i Williams

- ❑ Uvođenjem backpropagacije omogućeno je uspješno treniranje višeslojnih perceptrona.
- ❑ Koristi se kod izračunavanja gradijenta u neuronskim mrežama.



Learning representations by back-propagating errors

David E. Rumelhart*, Geoffrey E. Hinton†
& Ronald J. Williams*

* Institute for Cognitive Science, C-015, University of California,
San Diego, La Jolla, California 92093, USA

† Department of Computer Science, Carnegie-Mellon University,
Pittsburgh, Philadelphia 15213, USA



1958
Perceptron

1969
Minsky & Papert

1980
Fukushima

1986
Backprop