nedelja informatike v7.0

Neuralne mreže u mozgu: zašto nas miris torte podseća na detinjstvo

Anastasija Ilić

Matematička gimnazija

12. 04. 2022.

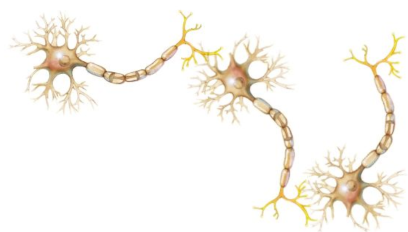
Agenda

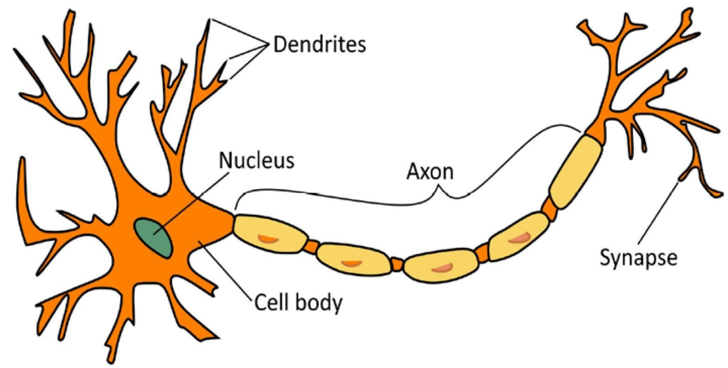
1. Uvod u računarsku neuronauku
2. Autoasocijativno sećanje
3. Hopfildova mreža

Moždane neuralne ćelije - neuroni

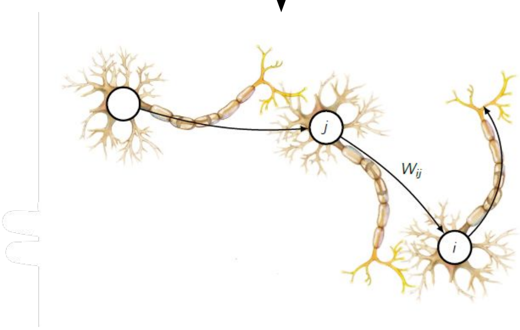
- neuroni = neuralne ćelije u mozgu
- osnovne jedinice u mozgu
- **električni inputi i hemijski signali** za razmenu informacija u mozgu i ostatku nervnog sistema

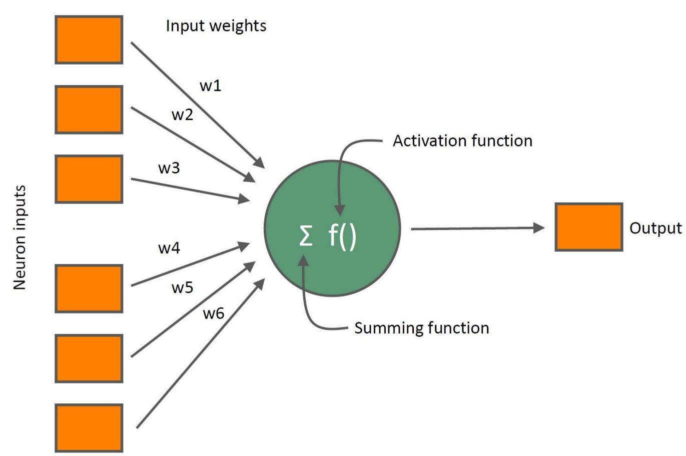
Neuralne mreže u mozgu → veštačke neuralne mreže





pojednostavljeni model!



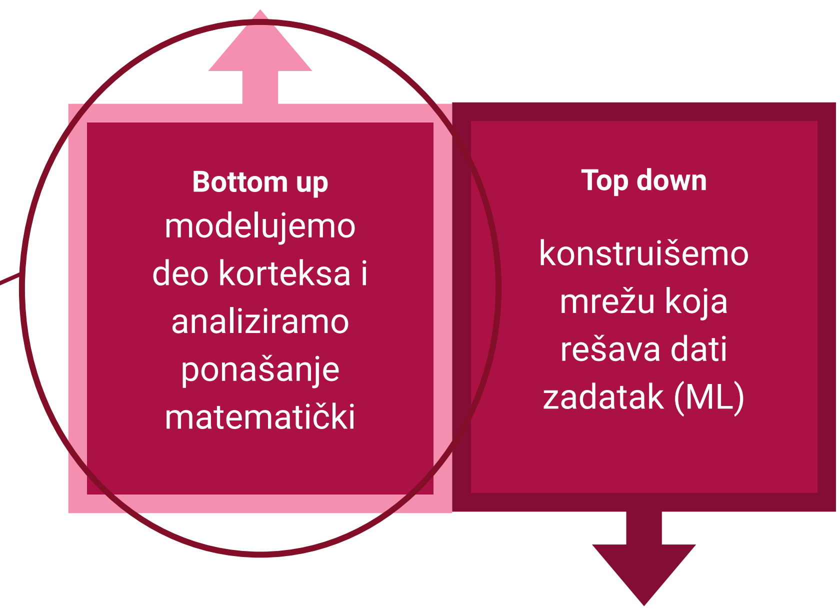


Računarska neuronauka

- puno jednostavnih (?) dinamičkih elemenata u mozgu: neurona
- veoma rekurentna arhitektura

- Kako jednostavni dinamički elementi proizvode kompleksne fenomene/računanja?

1. **Snimci neuralne aktivnosti** u mozgu
2. **Model** koji će oponašati neuralnu aktivnost u mozgu
3. Da li možemo nešto da zaključimo o **funkcionalnosti** mozga?



Agenda

1. Uvod u računarsku neuronauku
2. Autoasocijativno sećanje
3. Hopfildova mreža

Autoasocijativno sećanje

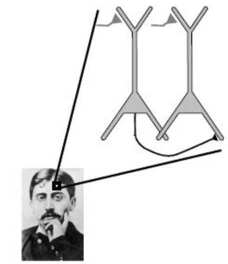
I raised to my lips a spoonful of the tea in which I had soaked a morsel of the cake. … And suddenly the memory returns. The taste was that of the little crumb of madeleine which on Sunday mornings at Combray, when I went to say good day to her in her bedroom, my aunt Léonie used to give me, dipping it first in her own cup of real or of lime-flower tea.

Marcel Proust: À la recherche du temps perdu

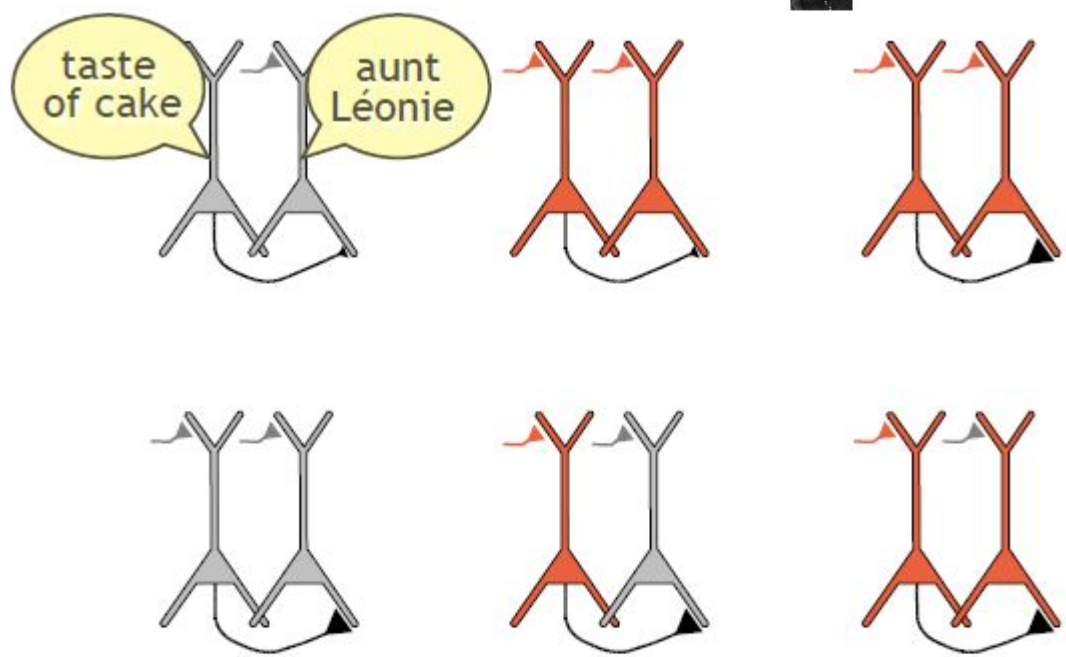
Kako se ovo desilo?

Obrada sećanja u neuralnim mrežama

- ukoliko bi postojala jedna neuralna ćelija za tetku i jedna neuralna ćelija za kolač:



čuvanje



podsećanje

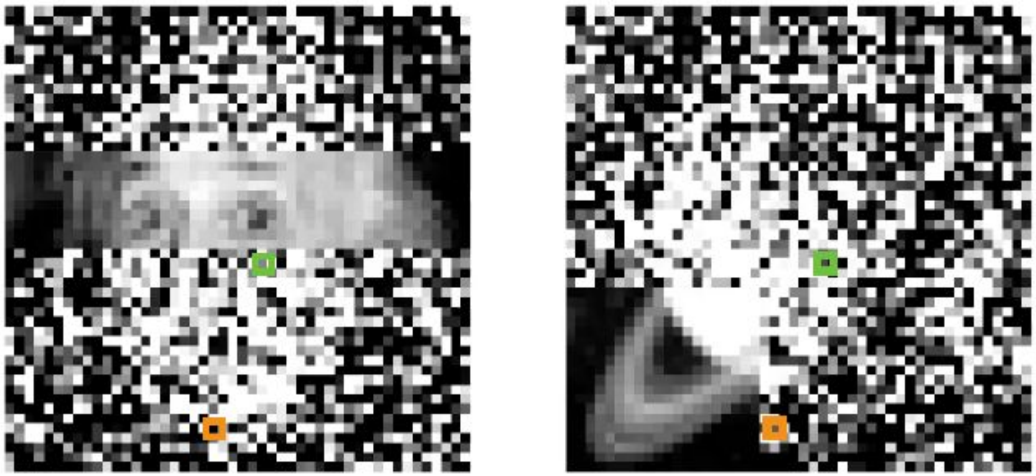
- **respodeljena** reprezentacija, bez pretpostavke ćelija koje se specijalizuju

Raspodeljena reprezentacija

Pamćenje slike

1. **piksel = neuron**
2. **vrednost piksela = aktivnost neurona**

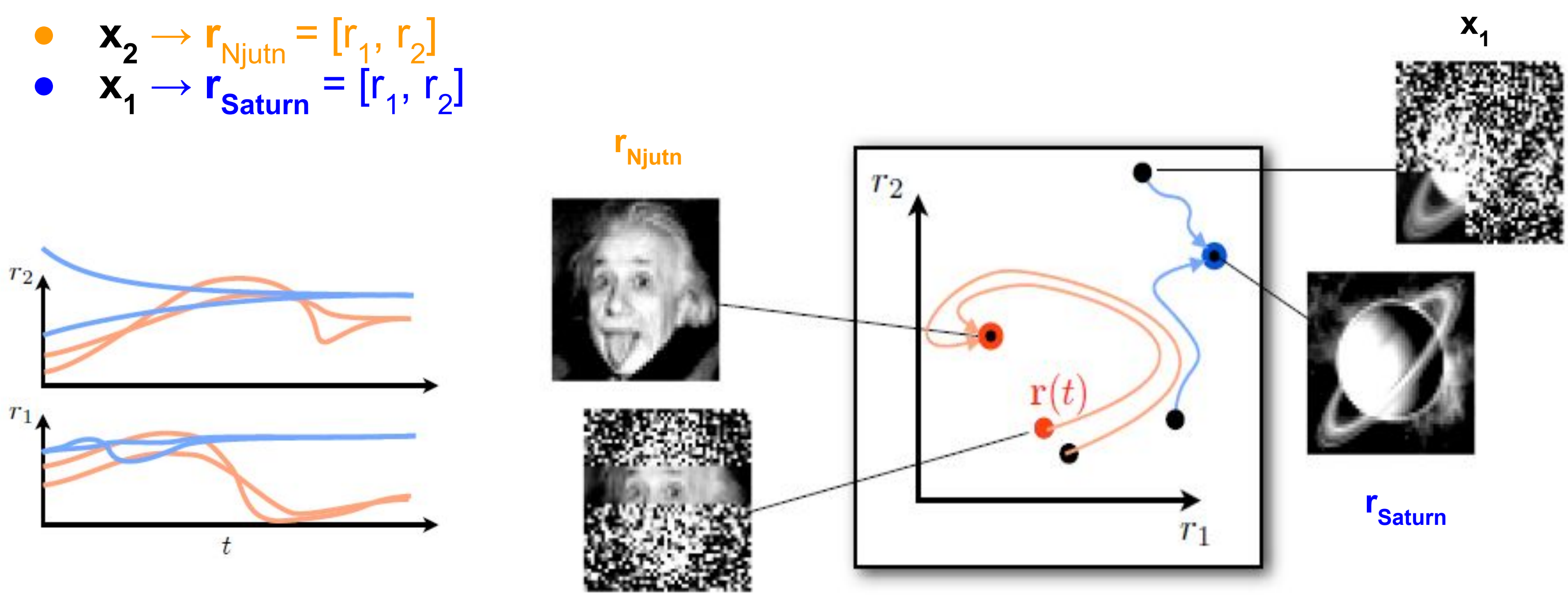
Svaki neuron (piksel) je delimično odgovoran za formiranje sećanja



Fazna ravan

- $\mathbf{r} = [r_1, r_2, \ldots, r_N]$ vektor neuralnih aktivnosti (N piksela na slici)
- r_1 i r_2 dva neurona (piksela) od N

- $\mathbf{x_2} \rightarrow \mathbf{r}_{Njutn} = [r_1, r_2]$
- $\mathbf{x_1} \rightarrow \mathbf{r}_{\mathbf{Saturn}} = [r_1, r_2]$



$\mathbf{x_2}$

fazni graf

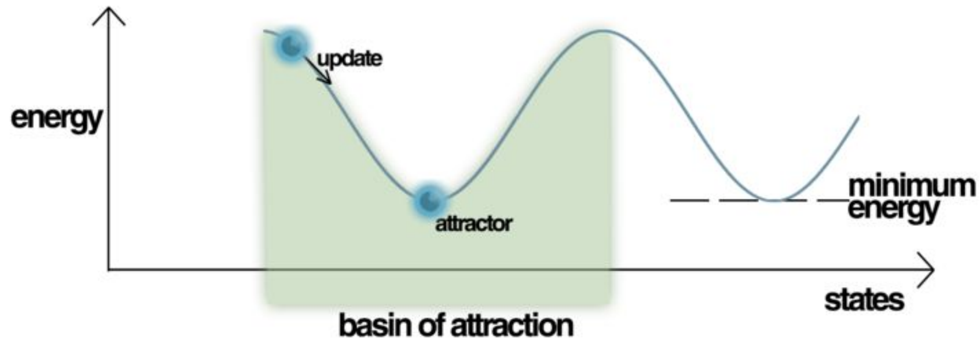
fazna ravan

Sećanja su **tačkasti atraktori** u faznoj ravni!

Agenda

1. Uvod u računarsku neuronauku
2. Autoasocijativno sećanje
3. Hopfildova mreža

Hopfildova mreža



- Dinamika mreže = zakoni promene neuralne aktivnosti do konvergencije mreže

- Da li možemo da napravimo model čija će dinamika oslikavati dinamiku prizivanja asocijativnog sećanja?

1. Definicija neurona

2. Hebijanska sinaptička plastičnost

3. Funkcija energije
   - ima stabilne fiksne tačke (tačkasti atraktori)
   - stabilne tačke = zapamćena sećanja

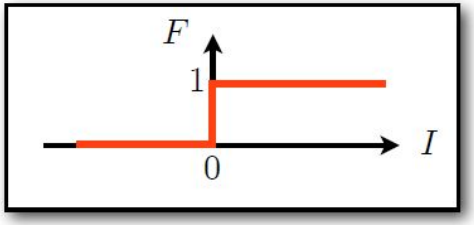
Definicija neurona i Hebijanska sinaptička plastičnost

- **Binarni neuroni**
  - svaki neuron prima input od svih ostalih N neurona

$$r_i(t) \in \{0, 1\} \quad \forall i$$

$$r_i(t + \Delta t) = F\left(\sum_{j=1}^{N} W_{ij}\, r_j(t)\right)$$

$$F(I) = \begin{cases} 1 & \text{if } I \geq 0 \\ 0 & \text{otherwise} \end{cases}$$



funkcija aktivacije

- **Hebijanska sinaptička plastičnost**
  - nekorelisana balansirana sećanja

$$r_i^{(m)} \perp r_{j \neq i}^{(m)} \perp r_j^{(m' \neq m)}$$

$$\mathrm{P}\left(r_i^{(m)} = 0\right) = \mathrm{P}\left(r_i^{(m)} = 1\right) = \frac{1}{2}$$

Sećanja su prisilno ubačena u mrežu:
zanima nas prisećanje memorija, ne učenje!

- Pravilo kovarijanse

$$W_{ij} = \sum_{m=1}^{M} \left(r_i^{(m)} - \frac{1}{2}\right)\left(r_j^{(m)} - \frac{1}{2}\right)$$

$$W_{ii} = 0$$

Funkcija energije

1. Funkcija energije koja ima stabilne fiksne tačke

- možemo da se prisetimo jedne stvari bez da nam misli odlutaju

$$\boxed{E(\mathbf{r}) = -\frac{1}{2}\sum_i \sum_{j \neq i} W_{ij}\, r_i\, r_j} \longrightarrow E(\mathbf{r}(t+\Delta t)) \leq E(\mathbf{r}(t))$$

možemo **matematički** da dokažemo

… i ima donje ograničenje

$$E(\mathbf{r}(t)) = -\frac{1}{2}\sum_i \sum_{j \neq i} W_{ij}\, r_i(t)\; r_j(t) = -r_k(t) \sum_{j \neq k} W_{kj}\, r_j(t) - \frac{1}{2}\sum_{i \neq k} \sum_{j \notin \{i,k\}} W_{ij}\, r_i(t)\; r_j(t)$$

after updating r_k: $E(\mathbf{r}(t+\Delta t)) = -r_k(t+\Delta t) \sum_{j \neq k} W_{kj}\, r_j(t) - \frac{1}{2}\sum_{i \neq k} \sum_{j \notin \{i,k\}} W_{ij}\, r_i(t)\; r_j(t)$

$$E(\mathbf{r}(t+\Delta t)) - E(\mathbf{r}(t)) = -\left[r_k(t+\Delta t) - r_k(t)\right] \underbrace{\sum_{j \neq k} W_{kj}\, r_j(t)}_{\text{local field: } H_k(t)}$$

$$= \begin{cases} \text{if } r_k(t) = 0 \text{ and } H_k(t) \geq 0 & \rightarrow \quad -(1-0)H_k(t) \leq 0 \\ \text{if } r_k(t) = 1 \text{ and } H_k(t) \geq 0 & \rightarrow \quad -(1-1)H_k(t) = 0 \\ \text{if } r_k(t) = 0 \text{ and } H_k(t) < 0 & \rightarrow \quad -(0-0)H_k(t) = 0 \\ \text{if } r_k(t) = 1 \text{ and } H_k(t) < 0 & \rightarrow \quad -(0-1)H_k(t) < 0 \end{cases}$$

$$\boxed{E(\mathbf{r}(t+\Delta t)) - E(\mathbf{r}(t)) \leq 0}$$

Funkcija energije

2. Sačuvna sećanja su fiksne (stabilne?) tačke
- ako smo bili u sećanju na detinjstvo, dinamika mreže nas neće izvući iz detinjstva

$$\boxed{\begin{aligned} r_k(t) = r_k^{(\mu)} = 1 \quad &\rightarrow \quad \langle H_k(t)\rangle > 0 \rightarrow r_k(t+\Delta t) \approx 1 = r_k^{(\mu)} \\ r_k(t) = r_k^{(\mu)} = 0 \quad &\rightarrow \quad \langle H_k(t)\rangle < 0 \rightarrow r_k(t+\Delta t) \approx 0 = r_k^{(\mu)} \end{aligned}}$$

$$\begin{aligned} H_k(t) &= \sum_{j\neq k} W_{kj}\, r_j(t) = \sum_{j\neq k} r_j(t) \sum_m \left(r_k^{(m)} - \frac{1}{2}\right)\left(r_j^{(m)} - \frac{1}{2}\right) \\ &= \sum_m \left(r_k^{(m)} - \frac{1}{2}\right) \sum_{j\neq k} r_j(t) \left(r_j^{(m)} - \frac{1}{2}\right) \end{aligned}$$

when in memory state μ: $r_j(t) = r_j^{(\mu)}$

$$H_k(t) = \underbrace{\left(r_k^{(\mu)} - \frac{1}{2}\right) \sum_{j\neq k} r_j^{(\mu)} \left(r_j^{(\mu)} - \frac{1}{2}\right)}_{\text{signal}} + \underbrace{\sum_{j\neq k} r_j^{(\mu)} \sum_{m\neq\mu} \left(r_j^{(m)} - \frac{1}{2}\right)\left(r_k^{(m)} - \frac{1}{2}\right)}_{\text{noise}}$$

averaging

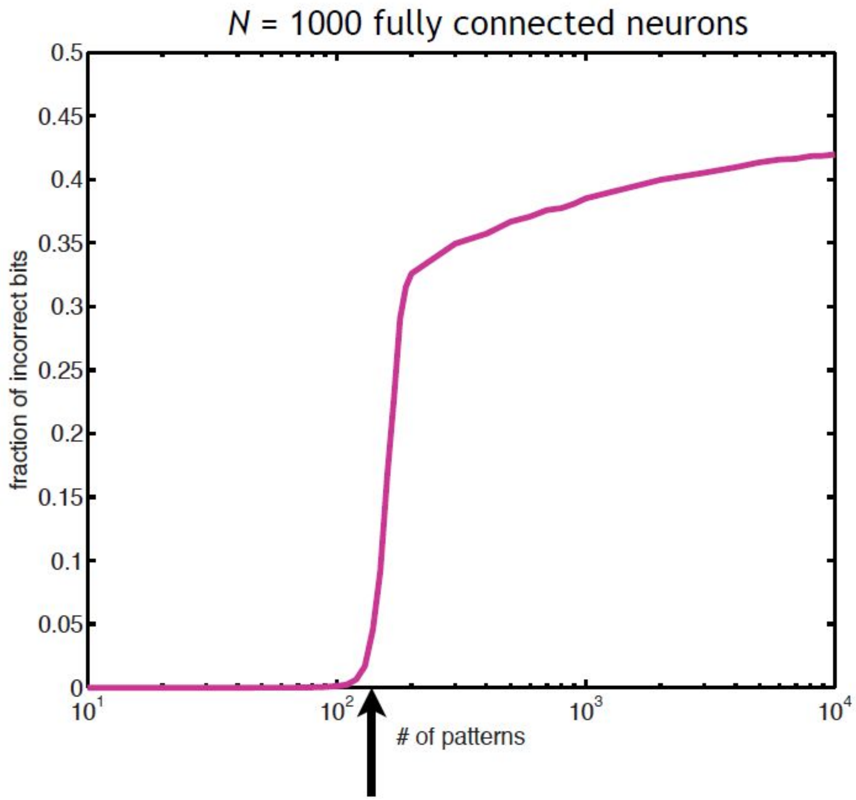
$$\langle H_k(t)\rangle_{\mathbf{r}^{(m)}} = \left(r_k^{(\mu)} - \frac{1}{2}\right) \underbrace{\sum_{j\neq k} r_j^{(\mu)} \left(r_j^{(\mu)} - \frac{1}{2}\right)}_{K_+ \geq 0} + \sum_{j\neq k} r_j^{(\mu)} (M-1) \underbrace{\left\langle \left(r_j^{(m)} - \frac{1}{2}\right)\right\rangle}_{=0} \underbrace{\left\langle \left(r_k^{(m)} - \frac{1}{2}\right)\right\rangle}_{=0}$$

$$\langle H_k(t)\rangle_{\mathbf{r}^{(m)}} = \left(r_k^{(\mu)} - \frac{1}{2}\right) K_+$$

$$\boxed{\begin{aligned} r_k(t) = r_k^{(\mu)} = 1 \quad &\rightarrow \quad \langle H_k(t)\rangle > 0 \rightarrow r_k(t+\Delta t) \approx 1 = r_k^{(\mu)} \\ r_k(t) = r_k^{(\mu)} = 0 \quad &\rightarrow \quad \langle H_k(t)\rangle < 0 \rightarrow r_k(t+\Delta t) \approx 0 = r_k^{(\mu)} \end{aligned}}$$

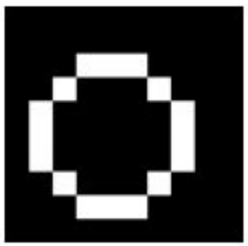
Kako se desi da ujutru pred kontrolni sve zaboravite?

- … a “bubali” ste celu prethodnu noć…
- više sačvanih sećanja → teže asocirati parcijalnu memoriju sa pravom memorijom koja treba biti prizvana
- osnovno ograničenje modela ukoliko se učenje ne dešava, već samo prisećanje

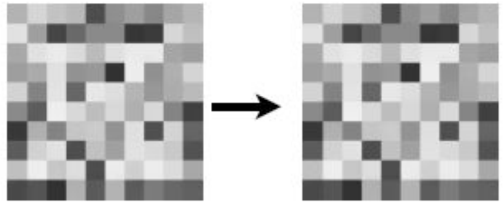


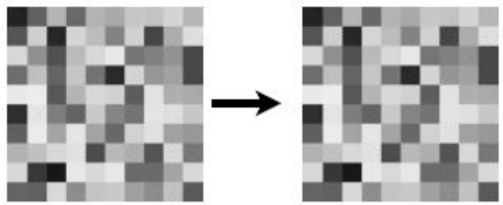
Hopfildova mreža u akciji

sačuvana sećanja u mreži

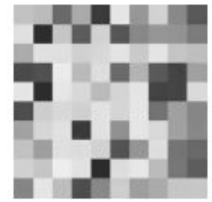


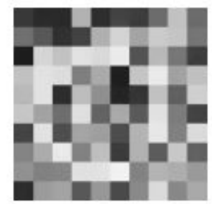
ulaz sa šumom



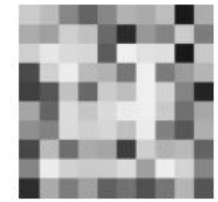


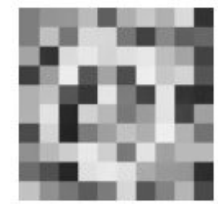
\+ 20 sačuvanih memorija



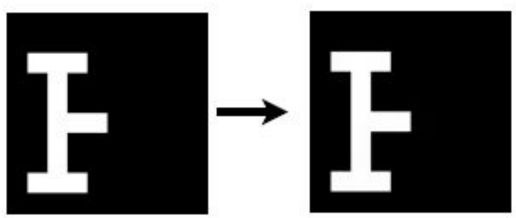


10% povezanosti





parcijalni ulaz



Problemi sa Hopfildovom mrežom:

1. Nisu samo sećanja fiksne tačke funkcije energije

2. Sećanja su nestabilne tačke

3. Sećanja su samo u proseku fiksne tačke

Šta smo (nadam se) naučili?

1. Zašto je **računarska neuronauka** važna?

2. Kako možemo da modelujemo **autoasocijativna sećanja**?
   - i sada znamo kako da napravimo model u kome će ukus torte (delimično sećanje) otkriti potpuno sećanje trenutka iz detinjstva!

3. Koja su **ograničenja modela** ukoliko se učenje ne dešava, već pokušavamo što više stvari napamet da naučimo

Hvala na pažnji!

Pitanja?