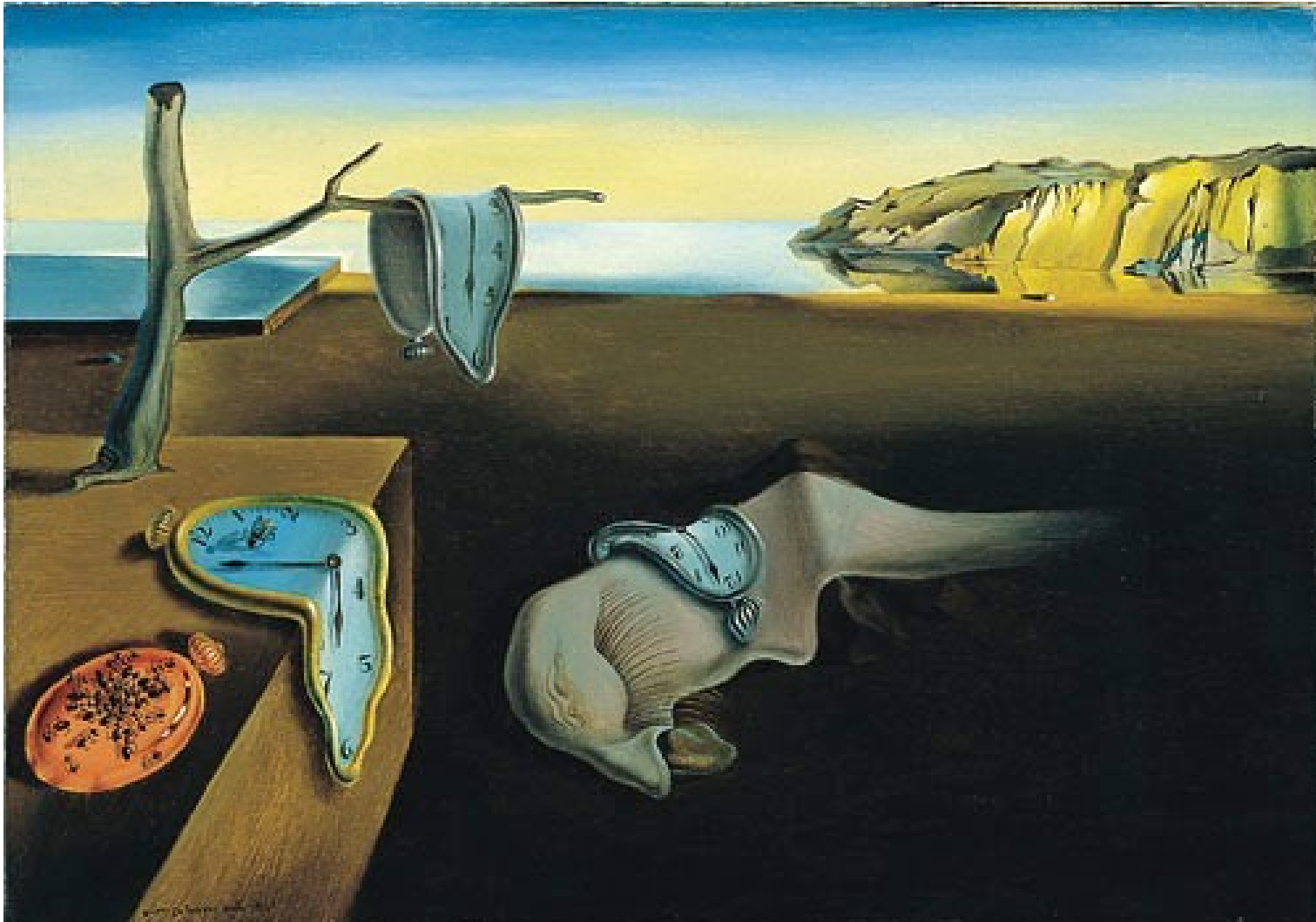


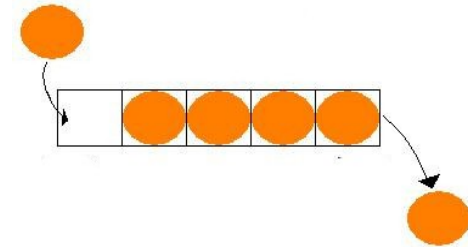
Az idegrendszeri memória modelljei



A memória típusai

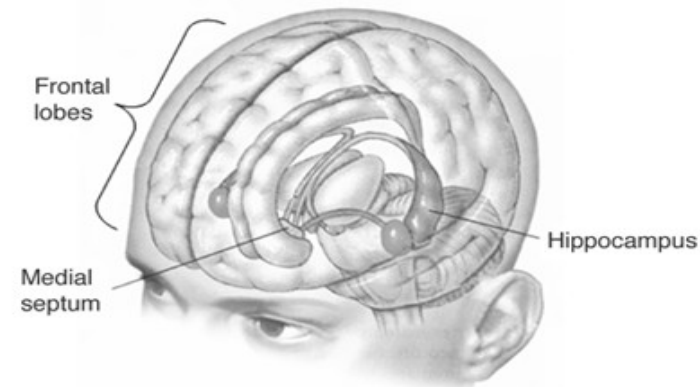
- Rövidtávú

- Working memory - az aktuális feladat
- Vizuális, auditórikus,...
- Prefrontális cortex, szenzorikus területek
- Kapacitás: 7 ± 2 minta

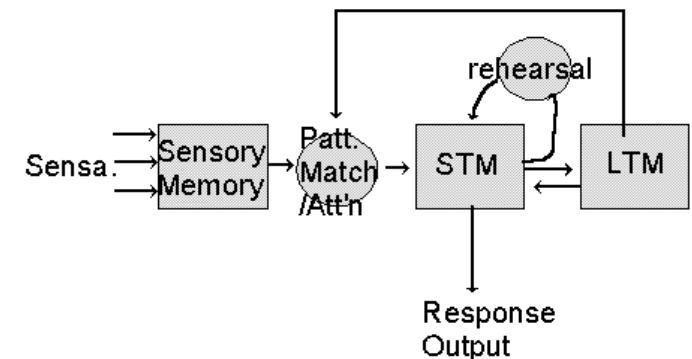


- Hosszútávú

- Epizodikus, szemantikus
- Technikailag: asszociatív
- Temporális lebeny, hippokampusz
- Interakció a rendszerek között

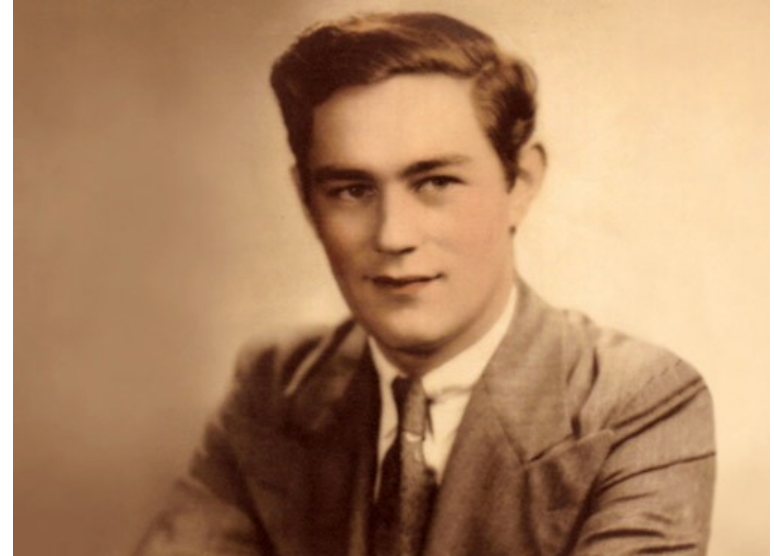


The Atkinson & Shiffrin Model of Memory



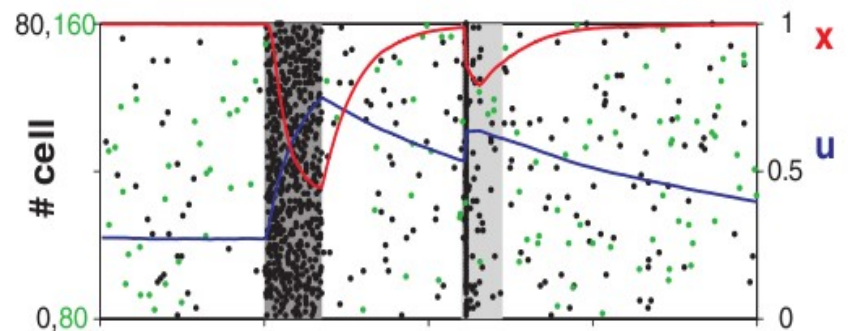
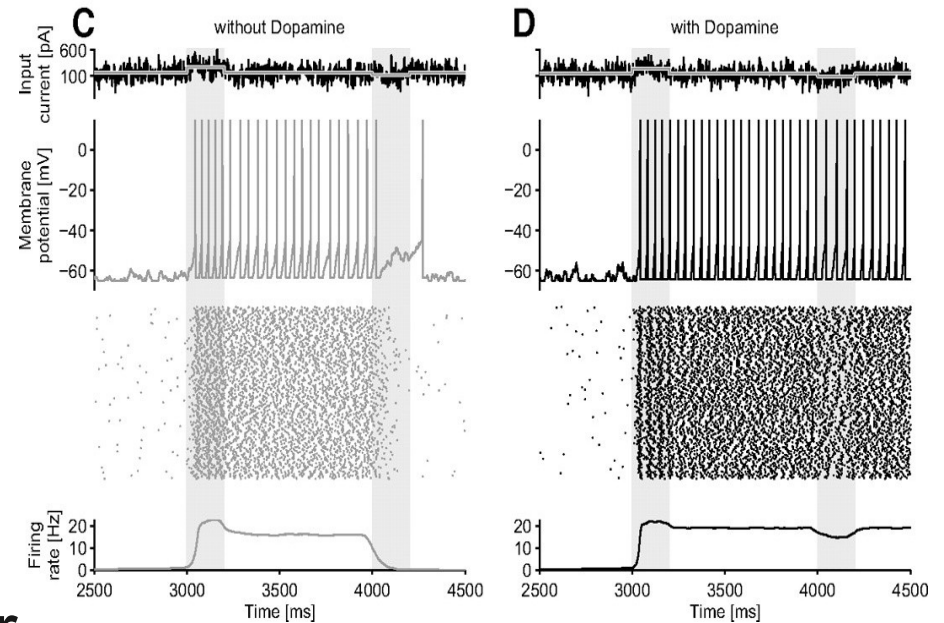
H. M.

- Súlyos epilepsziája volt, amit a hippokampusz egy részének eltávolításával orvosoltak 1953-ban.
- Ettől kezdve elvesztette az epizodikus memóriaformáció képességét – a korábbi emlékei megmaradtak.
- A rövidtávú memóriája ép maradt, valamint a motoros tanulási képessége is. Megtanult pl. tükörben rajzolni.
- A térbeli memóriája erősen sérült.
- Bizonyítékot szolgáltatott a különböző memóriarendszerek létezésére.

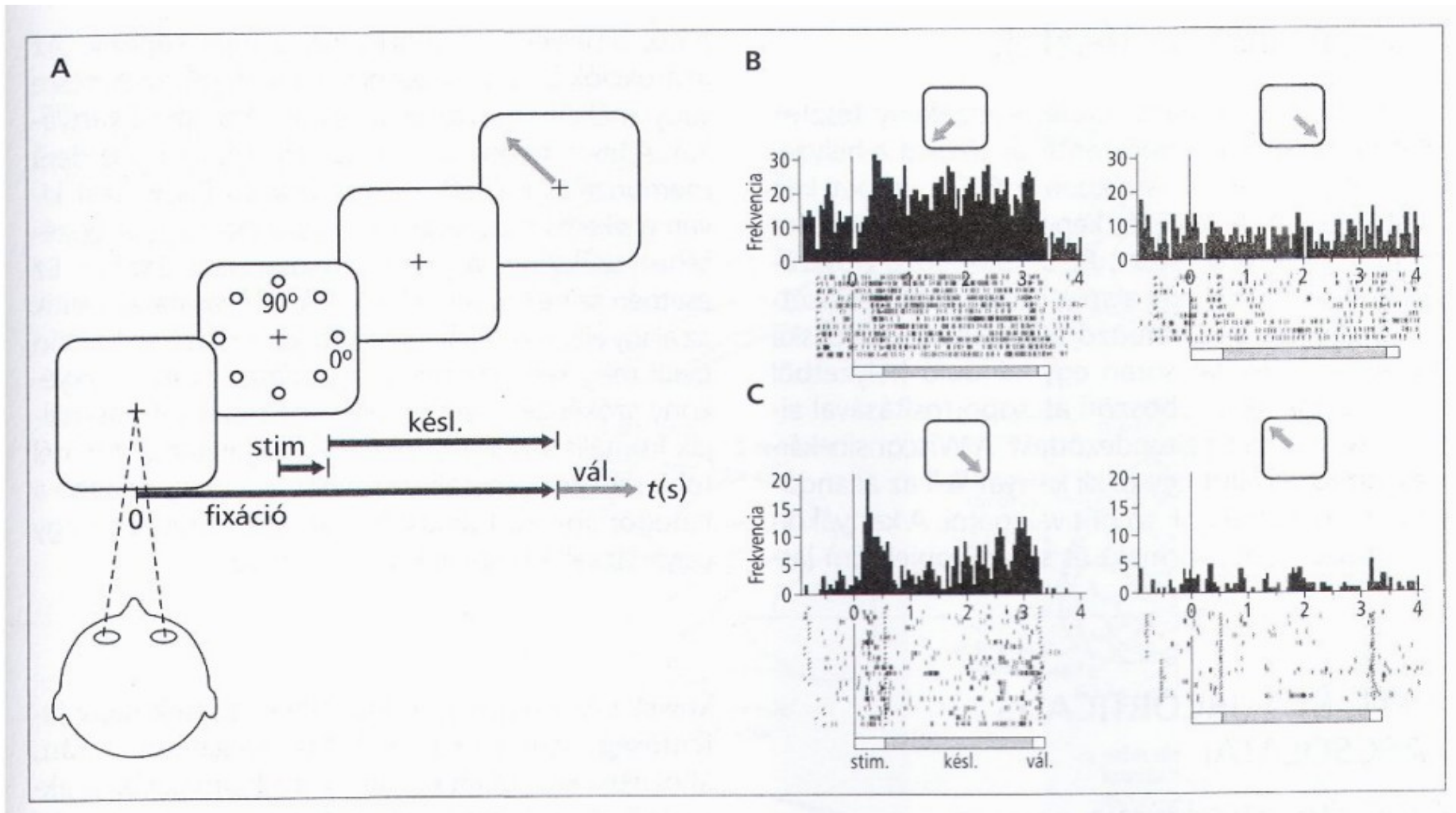


Munkamemória modelljei

- Rekurrens hálózati modellek, serkentő-gátló populációkkal
- Perzisztens aktivitás
A kódoló populáció a jel beírása után magasabb rátával tüzel
- Előfeszített állapot
A sejtek a beírásakor facilitált állapotba kerülnek, a kiolvasáskor szinkron tüzelés valósul meg
- Oszcillációs modell (később)
- Disztrakció: kis zavaró jelet ignorálni szeretnénk, nagyra viszont elromlik a memória



Perzisztens aktivitás



A majom prefrontális kérgében egyes sejtek megnövekedett aktivitást mutatnak bizonyos stimulusok után a késleltetési szakaszban, ami meghatározza az adott választ is.

Szinaptikus modell

- Szinaptikus facilitáció és depresszió dinamikája integrate and fire neuronokban

$$\dot{u}_j(t) = \frac{U - u_j(t)}{\tau_F} + U [1 - u_j(t)] \sum_k \delta(t - t_k^{(j)})$$

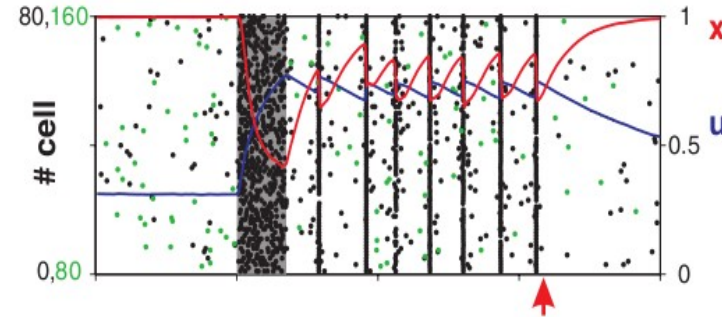
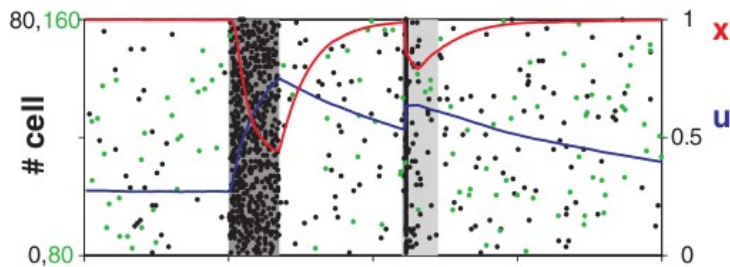
$$\dot{x}_j(t) = \frac{1 - x_j(t)}{\tau_D} + u_j(t) x_j(t) \sum_k \delta(t - t_k^{(j)})$$

$$\tau_m \dot{V}_i = -V_i + I_i^{(rec)}(t) + I_i^{(ext)}(t)$$

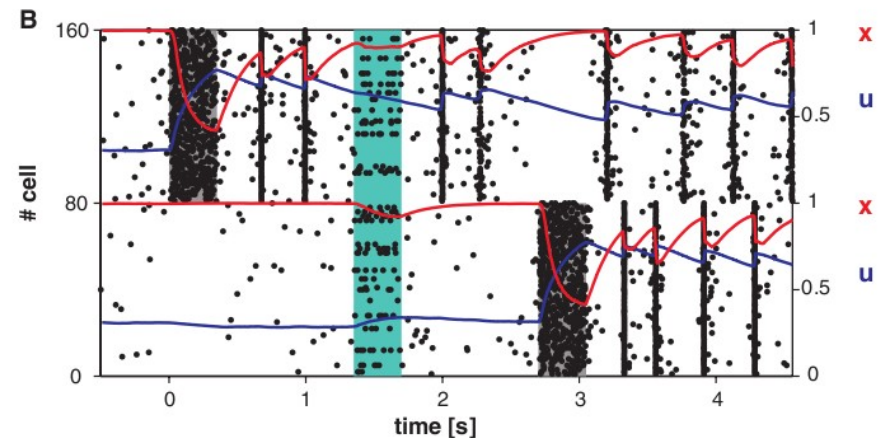
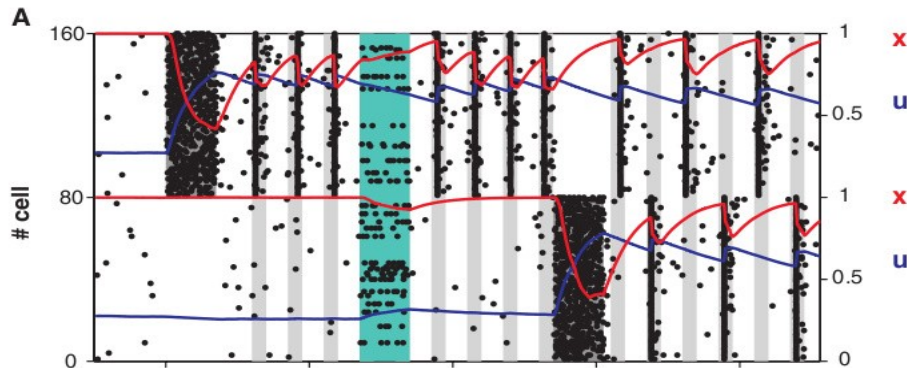
$$I_i^{(rec)}(t) = \sum_j \hat{J}_{ij}(t) \sum_k \delta(t - t_k^{(j)} - D_{ij})$$

$$\hat{J}_{ij}(t) = J_{ij} \cdot u_j(t - D_{ij}) \cdot x_j(t - D_{ij})$$

- Fixpont vagy oszcillációs dinamika

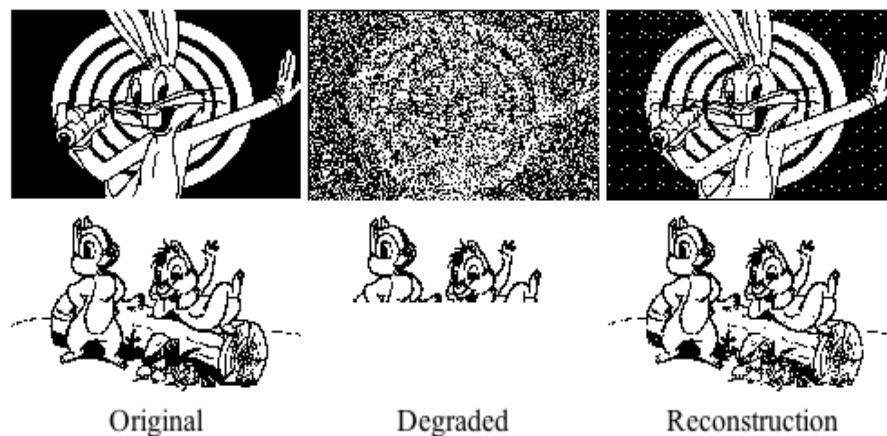
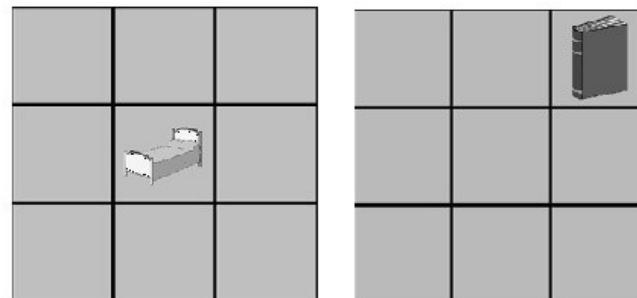


- Több elem tárolása



Asszociatív memória

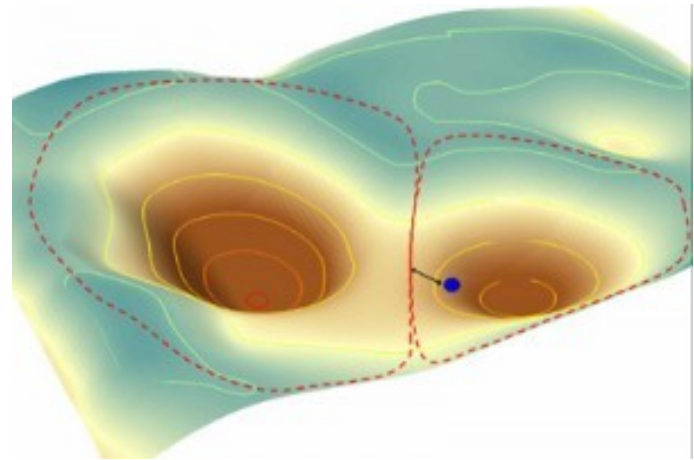
- Heteroasszociatív
 - pl. hely-objektum
- Autoasszociatív
 - Töredékes jelből az eredetit
- Különbség a számítógép memóriája és az AM között: címzés módja



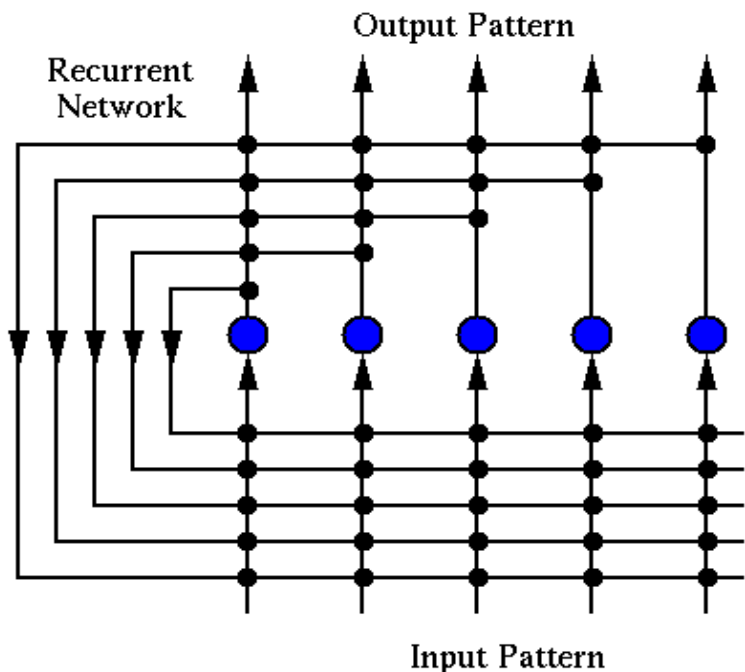
- Kapacitás: hány mintát tudunk eltárolni úgy, hogy azok visszahívhatók legyenek (többféle definíció)
- Stabilitás: minden mintára a legközelebbi tárolt mintát szeretnénk visszakapni

Attraktorhálózatok

- Attraktorok típusai
 - Pont
 - Periodikus
 - Kaotikus
- Vonzási tartományok
- Realizáció: rekurrens neurális hálózatok
- Attraktorok tárolása: szinaptikus súlyokon
 - Offline tanulás
 - Online tanulás
 - One-shot learning
- Előhívás: konvergencia tetszőleges pontból egy fix pontba



Hopfield-hálózat



- Asszociatív memória
- Bináris MCP-neuronok
- Minták tárolása: bináris vektorok
- Szimmetrikus súlymátrix
- Dale's law: egy sejt nem lehet egyszerre serkentő és gátló – ezt most megsértjük
- Rekurrens (dominánsan) hálózatok az agyban: hipokampusz CA3 régió, ...

- Offline learning
tanulandó minták:

$$\{s^1 \dots s^N\}$$

$$W_{ij} = \frac{1}{N} \sum_n s_i^n s_j^n \quad \leftarrow \text{Hebbi szabály}$$

- Léptetési szabályok: szinkron és szekvenciális

$$\mathbf{x}^{t+1} = \text{sgn}(\mathbf{W} \mathbf{x}^t - \boldsymbol{\theta}) \quad x_k^{t+1} = \text{sgn}\left(\sum_i^K W_{ik} x_i^t - \theta_k\right)$$

A HN dinamikája

- Nemlineáris rendszerek stabilitás-analízise: Lyapunov-függvény segítségével definiáljuk az állapotokhoz rendelhető energiát.

Ha a függvény:

- Korlátos
- Belátható, hogy a léptetési dinamika mindig csökkenti (növeli)

Akkor a rendszer minden bemenetre stabil fix pontba konvergál.

- Hopfield-hálózat Lyapunov-függvénye:

$$E = -\frac{1}{2} \mathbf{x}^T \mathbf{W} \mathbf{x} - \boldsymbol{\theta} \mathbf{x}$$

- Attraktorok az eltárolt mintáknál, de más helyeken is
- A HN használható kvadratikus alakra hozható problémák optimalizációjára is

A HN kapacitása

- Információelméleti kapacitás

- A tárolandó mintákat tekintjük Bernoulli-eloszlású változók halmazának

$$P(s_i^n = 1) = P(s_i^n = 0) = 0.5$$

- Követeljük meg az egy valószínűségű konvergenciát

$$\lim_{n \rightarrow \infty} P(s^a = \text{sgn}(\mathbf{W}\mathbf{s}^a)) = 1 \quad \forall a = 1 \dots M$$

- Ekkor (sok közelítő lépéssel) megmutatható, hogy

$$M \approx \frac{N}{2 \log_2 N}$$

- Összehasonlítás a CA3-mal

- Kb. 200000 sejt, kb. 6000 minta tárolható

- Más becslések

- figyelembevéve a minták ritkaságát

$$P(s_i^n = 1) = \alpha$$

$$M \approx N \frac{1}{\alpha \log_2 \frac{1}{\alpha}}$$

Reprezentációs tanulás

- Valószínűségi leírás
- 3féle dolgot tanulhatunk: csak predikció, kimenetek valószínűsége, underlying rejtett változók/dinamika
- Explicit rejtett változós modellek
- Implicit rejtett változós modellek
- Modellösszehasonlítás
- Becslési algoritmusok: EM

Boltzmann-gép

- Eloszlások reprezentációja – mennyiségek közti statisztikai összefüggések
- Sztochasztikus állapotátmenet

$$\mathbf{I} = \mathbf{W}\mathbf{u} + \mathbf{M}\mathbf{v}$$

$$P(v_a^{t+1} = 1) = \frac{1}{1 + e^{-I_a}}$$

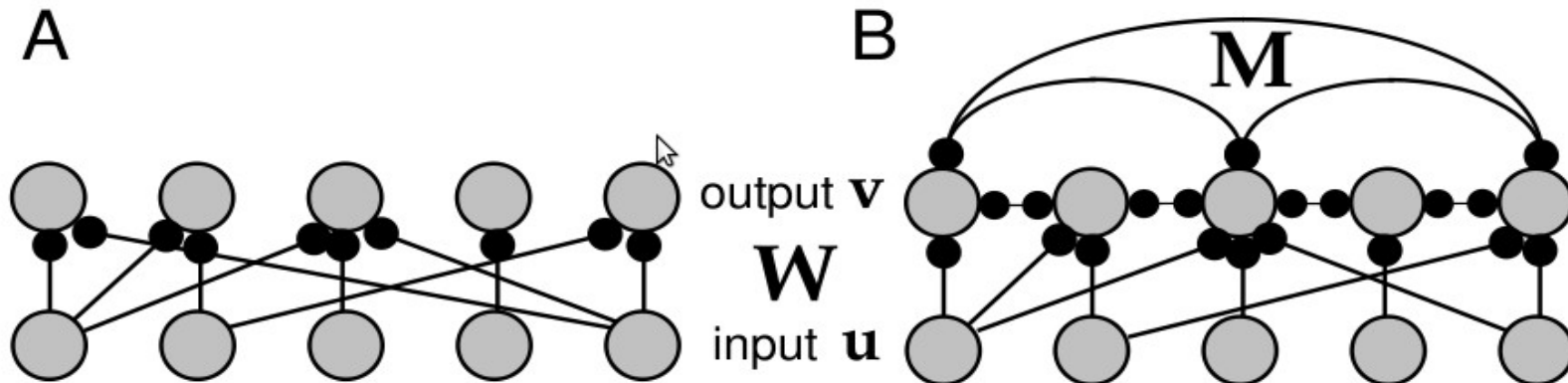
- A hálózat határeloszlása

Energia:

$$E(\mathbf{v}) = -\mathbf{v}^T \mathbf{W}\mathbf{u} + \frac{1}{2} \mathbf{v}^T \mathbf{M}\mathbf{v}$$

Boltzmann-eloszlás:

$$P(\mathbf{v}) = \frac{e^{-E(\mathbf{v})}}{\sum_{\mathbf{v}} e^{-E(\mathbf{v})}}$$



Tanulás Boltzmann-géppel

- Felügyelt tanulás, csak W -re, M analóg
- Hiba: Kullback-Leibler-divergencia a közelítendő és a megvalósított eloszlás között

$$D_{KL}[P(\mathbf{v}|\mathbf{u}), P(\mathbf{v}|\mathbf{u}, \mathbf{W})] = \sum_{\mathbf{v}} P(\mathbf{v}|\mathbf{u}) \ln \frac{P(\mathbf{v}|\mathbf{u})}{P(\mathbf{v}|\mathbf{u}, \mathbf{W})}$$

nem függ W -től

Tanulás Boltzmann-géppel

- Felügyelt tanulás, csak W -re, M analóg
- Hiba: Kullback-Leibler-divergencia a közelítendő és a megvalósított eloszlás között

$$D_{KL}[P(\mathbf{v}|\mathbf{u}), P(\mathbf{v}|\mathbf{u}, \mathbf{W})] = \sum_{\mathbf{v}} P(\mathbf{v}|\mathbf{u}) \ln \frac{P(\mathbf{v}|\mathbf{u})}{P(\mathbf{v}|\mathbf{u}, \mathbf{W})}$$

nem függ W -től

a $P(\mathbf{v}|\mathbf{u})$ -val súlyozott kimeneti összegzés helyett bemeneteke vett átlag:

$$\langle D_{KL} \rangle = -\frac{1}{N_s} \sum \ln P(\mathbf{v}^m | \mathbf{u}^m, \mathbf{W}) - K$$

Tanulás Boltzmann-géppel

- Felügyelt tanulás, csak W -re, M analóg
- Hiba: Kullback-Leibler-divergencia a közelítendő és a megvalósított eloszlás között

$$D_{KL}[P(\mathbf{v}|\mathbf{u}), P(\mathbf{v}|\mathbf{u}, \mathbf{W})] = \sum_{\mathbf{v}} P(\mathbf{v}|\mathbf{u}) \ln \frac{P(\mathbf{v}|\mathbf{u})}{P(\mathbf{v}|\mathbf{u}, \mathbf{W})} \quad \text{nem függ } W\text{-től}$$

a $P(\mathbf{v}|\mathbf{u})$ -val súlyozott kimeneti összegzés helyett bemeneteke vett átlag:

$$\langle D_{KL} \rangle = -\frac{1}{N_s} \sum \ln P(\mathbf{v}^m|\mathbf{u}^m, \mathbf{W}) - K$$

- Gradient descent – egyetlen bemenetre

$$\frac{\partial \ln P(\mathbf{v}^m|\mathbf{u}^m, \mathbf{W})}{\partial W_{ij}} = v_i^m u_j^m - \sum_{\mathbf{v}} P(\mathbf{v}|\mathbf{u}^m, \mathbf{W}) v_i u_j \quad \text{a Boltzmann-eloszlásból}$$

Tanulás Boltzmann-géppel

- Felügyelt tanulás, csak W -re, M analóg
- Hiba: Kullback-Leibler-divergencia a közelítendő és a megvalósított eloszlás között

$$D_{KL}[P(\mathbf{v}|\mathbf{u}), P(\mathbf{v}|\mathbf{u}, \mathbf{W})] = \sum_{\mathbf{v}} P(\mathbf{v}|\mathbf{u}) \ln \frac{P(\mathbf{v}|\mathbf{u})}{P(\mathbf{v}|\mathbf{u}, \mathbf{W})}$$

nem függ W -től

a $P(\mathbf{v}|\mathbf{u})$ -val súlyozott kimeneti összegzés helyett bemeneteke vett átlag:

$$\langle D_{KL} \rangle = -\frac{1}{N_s} \sum \ln P(\mathbf{v}^m | \mathbf{u}^m, \mathbf{W}) - K$$

- Gradient descent – egyetlen bemenetre

$$\frac{\partial \ln P(\mathbf{v}^m | \mathbf{u}^m, \mathbf{W})}{\partial W_{ij}} = v_i^m u_j^m - \sum_{\mathbf{v}} P(\mathbf{v} | \mathbf{u}^m, \mathbf{W}) v_i u_j^m$$

a Boltzmann-eloszlásból

- Delta-szabály – az összes lehetséges kimenetre való átlagot az aktuális értékkel közelítjük

$$W_{ij} \rightarrow W_{ij} + \epsilon_w (v_i^m u_j^m - v_i(\mathbf{u}^m) u_j^m)$$

Két fázis: hebbi anti-hebbi

Tanulás Boltzmann-géppel

- Felügyelt tanulás, csak W -re, M analóg
- Hiba: Kullback-Leibler-divergencia a közelítendő és a megvalósított eloszlás között

$$D_{KL}[P(\mathbf{v}|\mathbf{u}), P(\mathbf{v}|\mathbf{u}, \mathbf{W})] = \sum_{\mathbf{v}} P(\mathbf{v}|\mathbf{u}) \ln \frac{P(\mathbf{v}|\mathbf{u})}{P(\mathbf{v}|\mathbf{u}, \mathbf{W})}$$

nem függ W -től

a $P(\mathbf{v}|\mathbf{u})$ -val súlyozott kimeneti összegzés helyett bemeneteke vett átlag:

$$\langle D_{KL} \rangle = -\frac{1}{N_s} \sum \ln P(\mathbf{v}^m|\mathbf{u}^m, \mathbf{W}) - K$$

- Gradient descent – egyetlen bemenetre

$$\frac{\partial \ln P(\mathbf{v}^m|\mathbf{u}^m, \mathbf{W})}{\partial W_{ij}} = v_i^m u_j^m - \sum_{\mathbf{v}} P(\mathbf{v}|\mathbf{u}^m, \mathbf{W}) v_i u_j$$

a Boltzmann-eloszlásból

- Delta-szabály – az összes lehetséges kimenetre való átlagot az aktuális értékkel közelítjük

$$W_{ij} \rightarrow W_{ij} + \epsilon_w (v_i^m u_j^m - v_i(\mathbf{u}^m) u_j^m)$$

Két fázis: hebbi anti-hebbi

- Nem felügyelt

$$D_{KL}[P(\mathbf{u}), P(\mathbf{u}, \mathbf{W})]$$

Rejtett változós modellek

$$\dot{x} = f(x, u, \theta_u) + \epsilon \quad \epsilon = P(0, \Sigma_\epsilon)$$

$$y = g(x, \theta_x) + v \quad v = P(0, \Sigma_v)$$

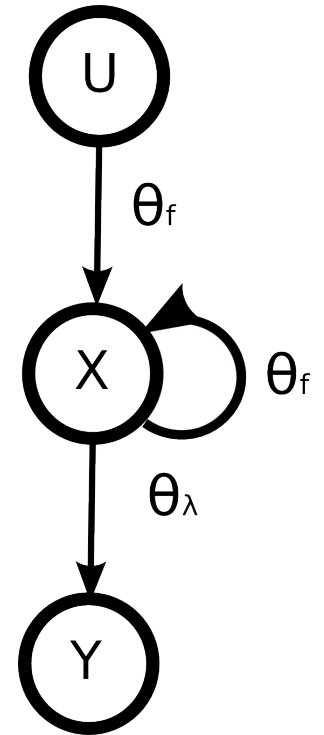
posterior

likelihood

prior

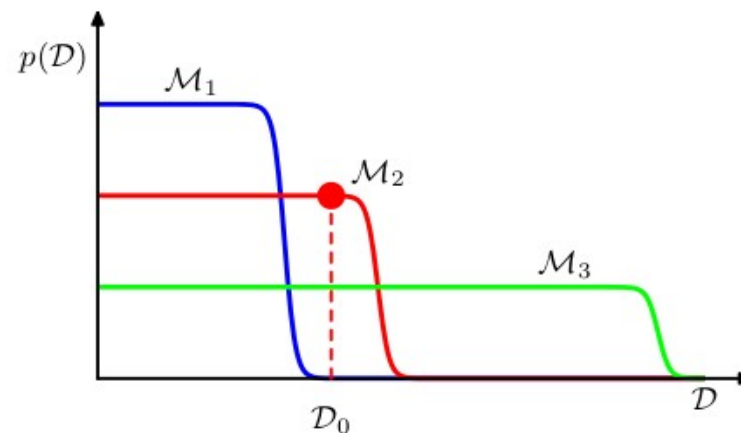
$$p(\theta|y, M) = \frac{p(y|\theta, M) p(\theta|M)}{p(y|M)}$$

Evidence
(marginal likelihood)

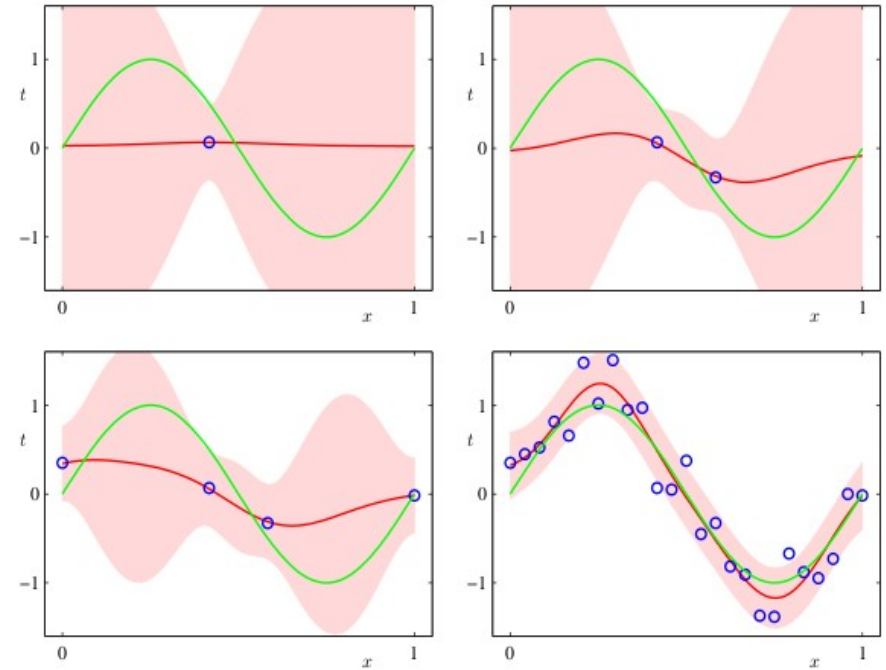
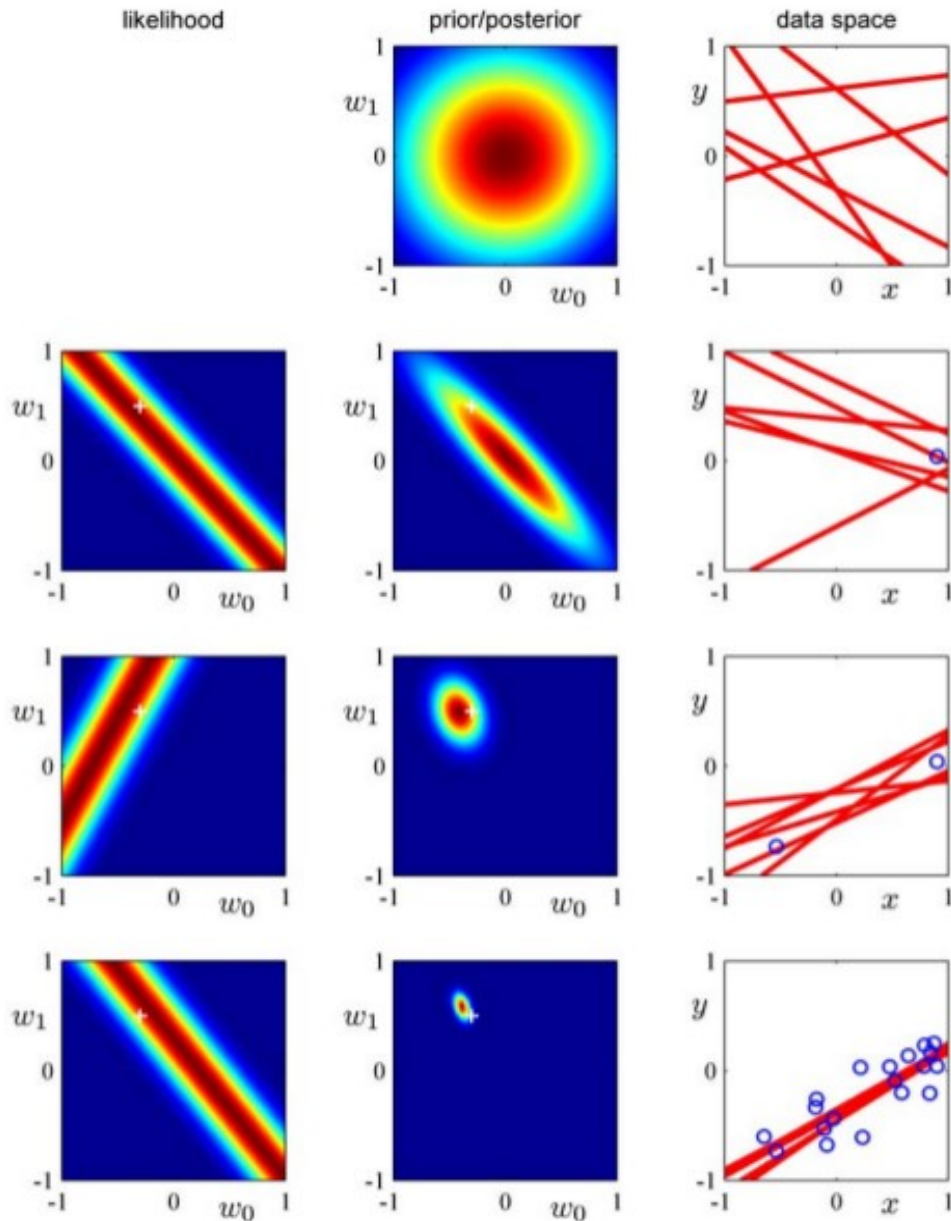


Predictive distribution:

$$p(y'|\theta, y, M)$$



Iteratív tanulás



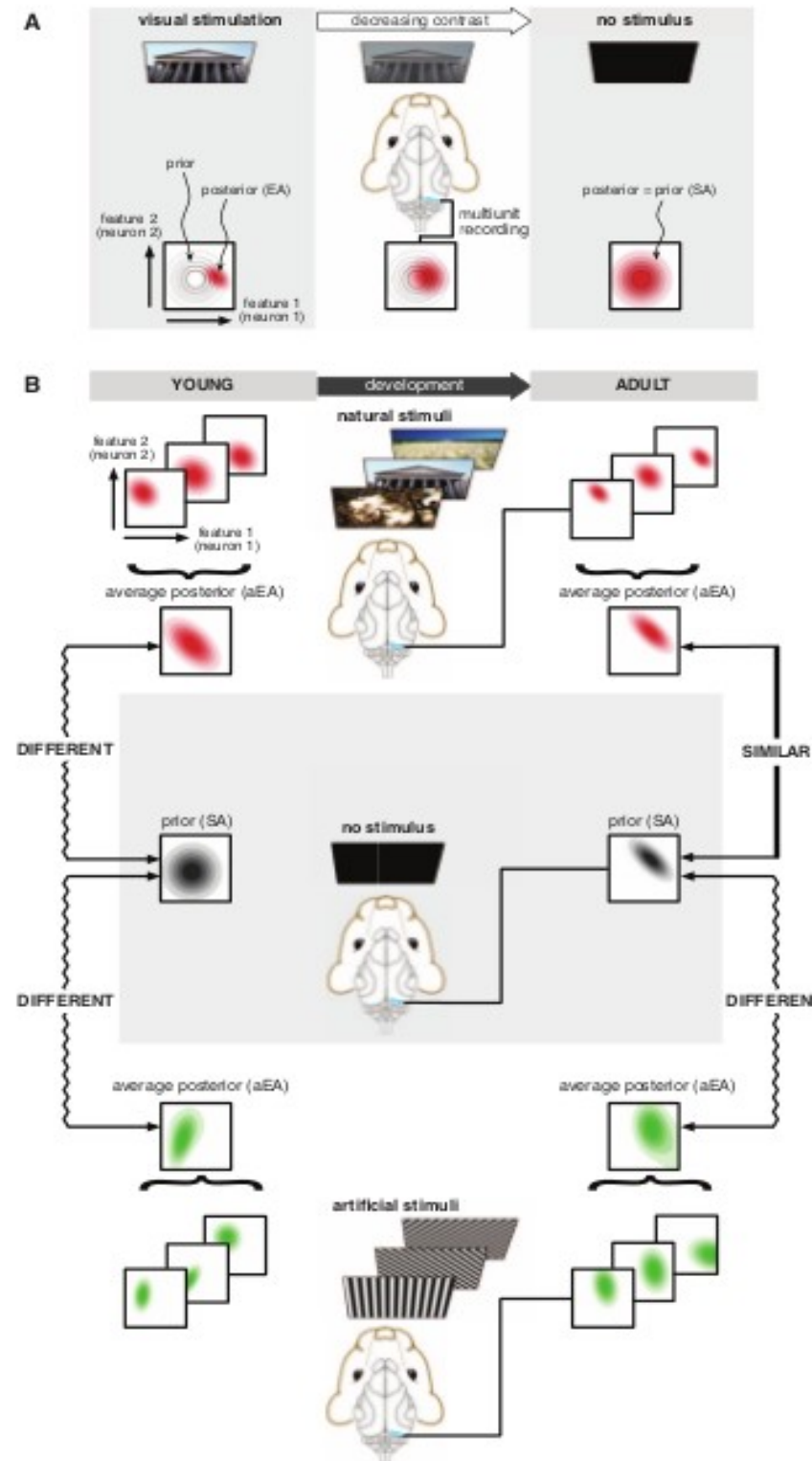
További olvasnivaló:

C. M. Bishop: Pattern Recognition and Machine Learning

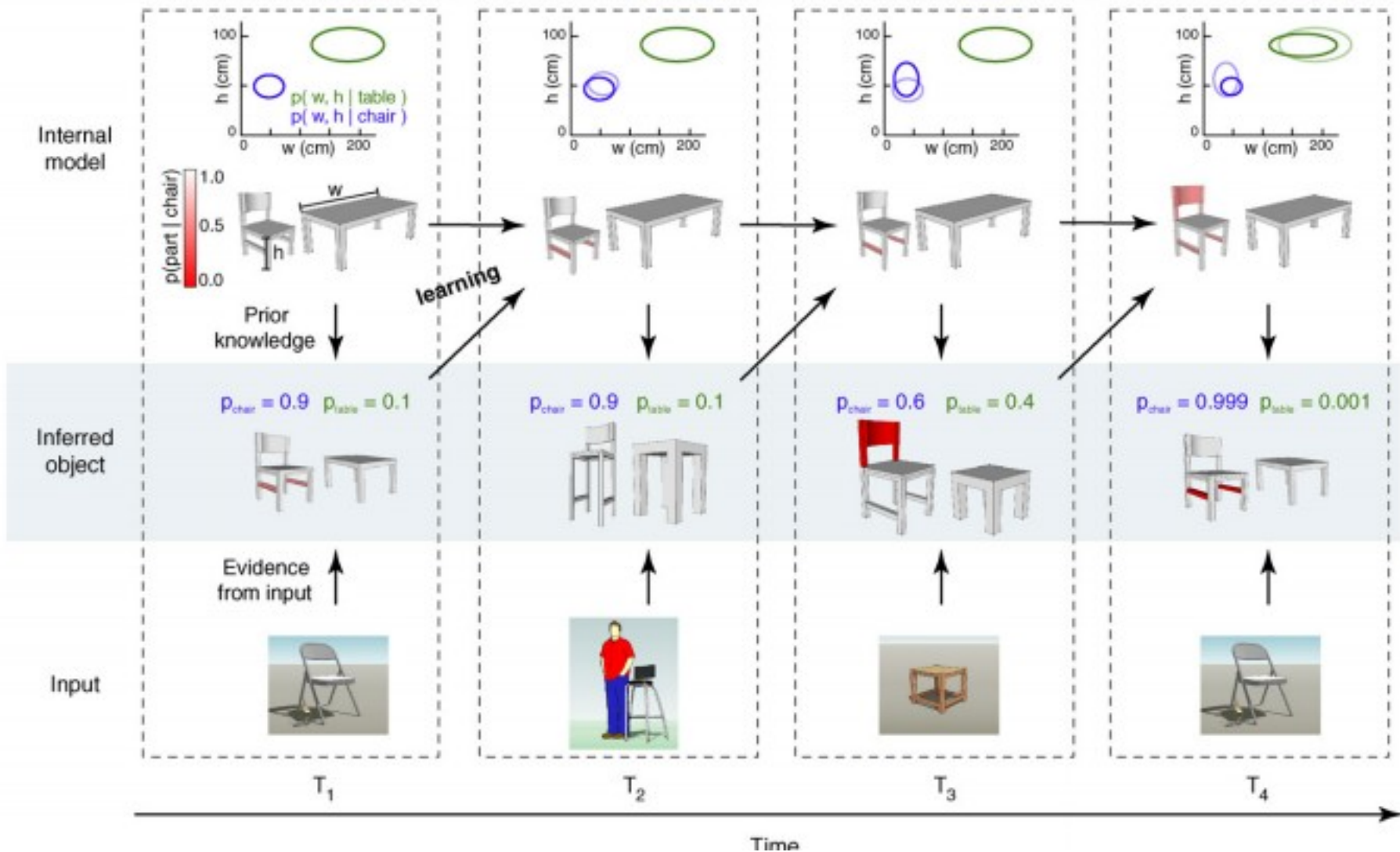
Zoubin Ghahramani: Unsupervised Learning

Bayesian Brain

- Az agy megvalósítja a Bayesiánus posterior becslését
- A prior az előző megfigyelésekből képzett posterior
- A posterior minden megfigyelés után priorrá válik
- Ha nincs input, a priort mintavételezzük
- Mérésekkel igazolható állatok különböző fejlődési szakaszaiban



Bayesian brain



Máté Lengyel, Gergő Orbán, József Fiser, Pietro Berkes