A MEMÓRIA IDEGRENDSZERI MODELLJEI

MEMÓRIA: információ tárolás esetleges előhívás céljából
→ érzékelésből és annak feldolgozásából származik

TANULÁS, MEMÓRIA/EMLÉKEZET, FELEJTÉS, ELŐHÍVÁS nem elválaszthatók

Az emlékezés folyamata információelméleti szemszögből:

→ KÓDOLÁS: információ befogadása, feldolgozása és kombinálása
→ TÁROLÁS: kódolt információ tartós megőrzése
→ ELŐHÍVÁS: felidézés valamely inger hatására

/egymásra hatással vannak/

IDEGRENDSZERI MODELL mit ír le?

Az a kérdés, hogy milyen tulajdonságokkal rendelkezik a memória

Előzmény: EMLÉKEZETKUTATÁS: filozófia → kognitív pszichológia

KORAI EMLÉKEZETKUTATÁS 3 FŐ ISKOLÁJA:

→ ÉSZAK-AMERIKA:
Hermann Ebbinghaus (1850–1909) német filozófus követői
↳ elsőként végzett kísérleteket (saját magán) → feljegyzési görbe (exponenciális)
verbális tanulási megközelítés szólistákkal
behaviorizmus: inger–válasz pszichológia (introspektív módszer elutasítása)

→ NÉMETORSZÁG:
Gestalt-pszichológusok (alaklélektan)
belső reprezentációk hangsúlyozása
észlelékutatáshoz hasonló megközelítés → aktív folyamat

→ NAGY-BRITANNIA:
Frederic Charles Bartlett "Az emlékezés" (1932)
életszerű komplex szituációk vizsgálata (idegen népmese)
hibázások vizsgálata
sémák: strukturálják a tudást és befolyásolják az előhívást

EMLÉKEZETI RENDSZEREK

néhány 100 msec

SZENZOROS EMLÉKEZET

- ikonikus
- echoikus

/észlelési folyamat részének is tekinthető/

másodperc – perc

MUNKAMEMÓRIA /RTM/

Baddeley & Hitch (1975):
- fonológiai hurok
- téri-vizuális vázlattömb
- epizódikus puffer (2000)

+ központi végrehajtó

/Ericsson & Kintsh (1995)
Cowan (2005)/

akár élethosszig

HOSSZÚ TÁVÚ MEMÓRIA

EXPLICIT (DEKLARATÍV)

SZEMANTIKUS (TÉNYEK)

→ frontalis és temporális lebeny

EPIZÓDIKUS (ESEMÉNYEK)

/tér és idő/

→ hippokampusz mediotemporális régió

IMPLICIT

PROCEDURÁLIS (KÉSZSÉGEK)

→ striatum, kisagy, motoros kéreg

PRIMING (előfeszítés)

/pl: alma → körte/

→ agykéreg

KONDICIONÁLÁS (klasszikus és operáns)

ÉRZELMI VÁLASZOK

/pl: fóbiák/

→ amigdala

VÁZIZOMZAT

→ kisagy

NEM-ASSZOCIATÍV: HABITUÁCIÓ, SZENZITIZÁCIÓ

→ genucreláris: reflexív

retrográd amnézia:
régi emlékek nem hívhatók elő

anterográd amnézia:
nem tud újakat elraktározni

H.M. (1957)
→ epilepszia miatt: temporális lebenyirtás
→ RTM ép, motoros tanulási képesség van, de nincs tudatában
→ intellektusa ép, de szemantikus memóriája drámaian leromlott + retrográd amnézia is!

- 50 első randi } filmek
- Memento

SZENZOROS EMLÉKEZET : inger meghosszabbítása feldolgozás céljából

Vizsgálata : rövid ideig bemutatott ingerekkel

IKONIKUS EMLÉKEZET :

→ 1740 Segner : izzó cigarettavég (fennmaradás x 100 msec)

→ 1960 Sperling : 3 x 4 betü 50 msec-ig → 4-5 betü visszaidézés

Ennyit láttak v. elfelejtették a beszámoló alatt ?

Ha egy sorból kérdeznek vissza, akkor 3-4 betü

⇒ hálrasnyuló emlékmyom !

<u>további effektusok</u> :

→ sötét mező előtte-utána ⇒ tovább fennmarad

→ villanás töröl (rendelkezésre álló időben lineáris)

→ maszkolási hatás retinális (nem mindegy melyik szem)

→ de létezik mintázatmaszkolás, amelynél mindegy

⇓

AZ ÉSZLELÉS FOLYAMATA TÖBB SZINTEN IS TÁROLÁST IGÉNYEL!

ECHOIKUS EMLÉKEZET :

→ 1970 Efron : szubjektiv élmény szerint egy 30 v. 100 msec-es hang egyformán 130 msec-ig szól, de ez lerövidül, ha egy második hang követi

→ Békésy György : a jelenség fontos hangversenytermek akusztikájánál (visszhang)

<u>megjegyzés</u> : Az auditoros rendszernél egyértelmü a szenzoros emlékezet szükségessége, hiszen a hang időbeli rezgés :

→ tiszta zenei hang esetében is van hangnyomás változás :
a változás felismeréséhez pedig összehasonlítás, ázat memória szükségeltetik

RÖVID TÁVÚ EMLÉKEZET:

modalitások integrációját teszi lehetővé
⇒ efölött nem érdemes modalitás szerint osztályozni

RTM ÉS A TUDAT:

„Úgy tűnik, létezik elménknek egy királyi fogadóterme, ahol a tudat ülésezik, és ahova egyszerre két – három idea jamul kihallgatásra." (Galton)

RÖVIDTÁVÚ FELEJTÉS

→ Például: megtanart bemutatkozás

Elméletek: → automatikus nyomelhalványulás } mindkét faktor fontos!
→ interferencia (hasonló jobban zavar)

proaktív
új tanulást zavarja a régi szokás

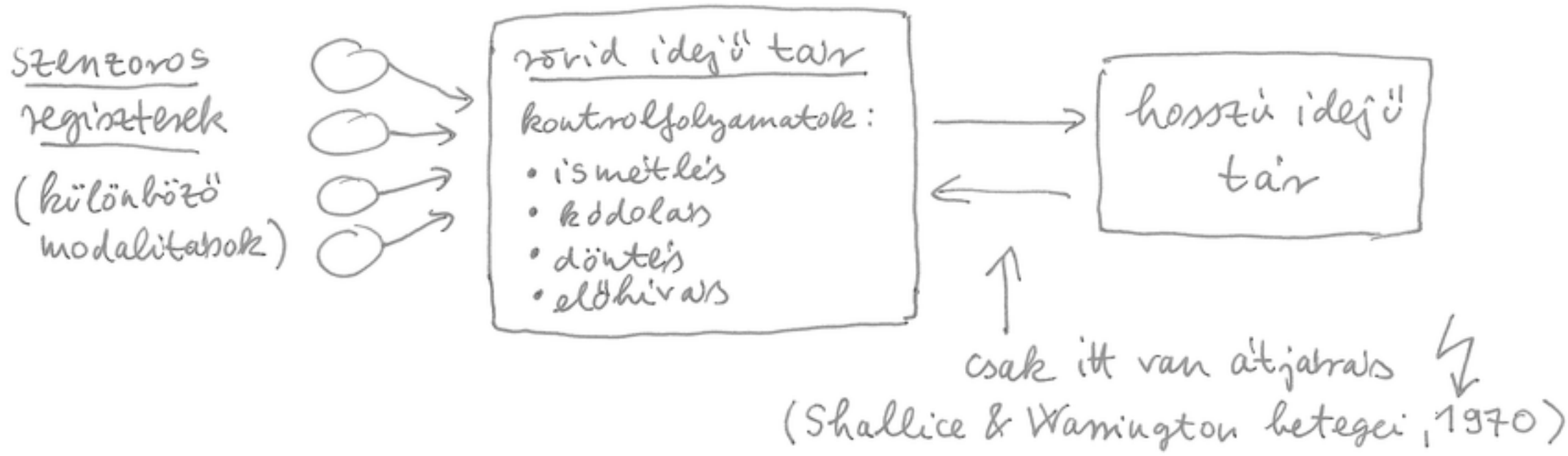
retroaktív
új tanulás zavarja a régi szokást

/pl. új PIN a régit/

Különálló (rövid/hosszú távú) rendszerek vagy sem?

→ Feldolgozási szintek elmélete:
Minél mélyebben dolgozzuk fel, annál jobban emlékezünk rá (átmenet)

→ Atkinson & Shiffrin modális modellje



INTERAKCIÓ A RENDSZEREK KÖZÖTT:

pl: deklaratív tudás lerontható a procedurálist, ha rásegítünk

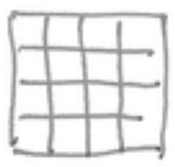
MUNKAMEMÓRIA

Több komponens? → párhuzamosan végzett feladatokkal vizsgálható

FONOLÓGIAI HUROK: tár + artikulációs kontrollfolyamat

→ nem a szótagszám, hanem a kiejtési idő" számít
→ gyorsbeszédűek előnyben
→ bizonyos nyelvek előnyben-hátrányban (számterjedelem)
→ van olyan beteg, akinél csak a kontrollfolyamat sérült
→ hasonló hangzásra a felidézés romlik
→ olvasás tanulásához és nyelvi megértéshez is köze van!

Vizuális emlékezet: milyen mértékben hasonló a percepciójához?

→ „lelki szemeink előtt" szubjektív értelben nagy a variancia
→ Shepard & Metzler 1971: mentális rotáció
→ lineáris és nem függ a komplexitástól (térben is)
→ de ha azt kell eldönteni, hogy része-e, akkor függ
→ pásztázás lelki szemeinkkel:
→ tovább tart választani, ha a képelemek távolabb esnek korábbi képen
→ hasonló eredmény egyetemi város képzeletbeli térképén (nincs mutatás)
→ ha képzeletben közelednénk: jobban kitöltjük a látóteret az objektum
→ de ha egy kockát testátló mentén ujjainkközé fogunk, akkor nem „látjuk" jól, hogy a köztes csúcsok nem egy síkba esnek!
→ az abakusz mesterek téri-vizuális mnemotechnikát használnak: számterjedelmük oda-vissza nagy (de betűkre nem) és vizuális feladat zavarja őket
→ Brooks-feladat:  mondatfelidézések:

1-es feladat (vizuális): „A következő négyzetbe jobbra tegyen 2-est"
2-es feladat (verbális): „A következő rossz négyzetbe tegyen 2-est"
(vizuális átlag: 8, verbális átlag: 6)
→ vizuális feladat (pontkövetés pálcával) ront az elsőn, de nem ront a verbálison

Téri vagy vizuális?
→ bekötött szemmel csak hangalapján pontkövetés ⇒ téri zavarás fontosabb!

Hol és Mi?
→ agykárosult beteg: tudja mi, de rossz irányba nyúl

MUNKAMEMÓRIA IDEGRENDSZERI MODELL

Korábban ún. perzisztens aktivitás modell:
a kódoló populáció a jel beírása után magasabb rátával tüzel
(az információ spike-okban tárolása metabolikusan költséges)

↓ helyette előfeszített állapot: facilitáció → kiolvasáskor szinkron tüzelés

Mongillo, Barak, Tsodyks (Science 2008):
SYNAPTIC THEORY OF WORKING MEMORY

→ prefrontális kortex: serkentő szinapszisok, facilitáció jellemző
(ellentétben a szenzoros területekkel)
→ Ca^+ szint kódol a preszinaptikus oldalon (lecsengés ~1s spike nélkül)

SZIMULÁCIÓ:

→ rekurrens integrate-and-fire neuronhálózat
→ input serkentő populáció (random)
→ Hebb prior: ugyanazon input serkentő neuronjai között erősebb szinapszisok
→ Tsodyks-Markram modell dimenziótlan változói

forrás: $x \in [0,1]$
kihasználás: u
} spike után $x \to x - ux$

→ alapszintek $x=1$, $u=U$
→ időállandók: τ_D (depresszió), τ_F (facilitáció)
x lecsengése $\tau_D \ll \tau_F$
→ előhívás: alacsony jel az egész hálózaton
→ szinkronizált válasz: 20 ms-en belül tüzel az adott populáció
→ egyúttal frissül az előfeszítés (τ_F lecsengésül)
→ növelve a hátteret: fixpont → oszcilláció
→ az ábrán 10.000 neuronból csak egy része van ábrázolva
→ az ábrán x és u átlagok vannak
→ több elem tárolása, lásd ábrán

ASSZOCIATÍV MEMÓRIÁK

A rövid távú emlékezetre tekinthetünk úgy, mint egy átmeneti tár.
A hosszú távú memóriába azonban tudást szeretnénk rögzíteni és később használni.
Minden zajos, ezért a tudással kapcsolatos következtetéseink asszociációk lesznek

FAJTÁI : → autoasszociatív (töredékes jelből az eredeti) (oda-vissza).
→ heteroasszociatív

↔ ellentéte közvetlen címzés: RAM, telefonkönyv

TÉNYEZŐK : kapacitás, stabilitás, vonzási tartomány (medence)

MEGVALÓSÍTÁS ATTRAKTORHÁLÓZATOKKAL :

rekurrens neuronális hálózat attraktorai :
→ stabil fixpont
→ periodikus oszcilláció (határciklus)
→ kaotikus

TANULÁS (szinaptikus súlyok) :
→ offline (adatbázis a kezünkben)
→ online (folyamatosan jön az adat)
→ one-shot learning (adatbázis megvan, de egyszer nyúlhatunk hozzá)

ELŐHÍVÁSRA / ASSZOCIÁCIÓRA VONATKOZÓ MEGFIGYELÉSEK:

→ sokszor a régi ismerős meg kell szólaljon, hogy megismerjük
→ bűnügyek : azonosítás hang alapján
→ elérhetőség és rendelkezésre állás között különbség :
38% felidézés mellett 96% felismerés (100 szó)
→ (B)emutatás, (F)elidézés → BFBFBF... } ugyanolyan hatékonyak
→ BFFFBFFF...
⇒ elem felidézése megnöveli az előhívási valószínűséget
→ dallam felismerés : dallamkontúr, hangmagasság intervallumok megőrződnek
Született tehetség szerepe? Nagymesteri szint?
pl: Mozart : Vatikáni kórus saját tulajdonú műve (tilos másolni)
Toscanini: karmester
→ kontextusfüggés : mélytengeri búvárok
→ gyakorisági paradoxon: gyakori szavakat jobban idézzük fel, de
rosszabbul ismerjük fel

HOPFIELD-HÁLÓZAT

(John Hopfield amerikai biológus 1982) → folytonos: '84

Mit várunk a hálózattól?

→ tárolt információ zajos, torzított, esetleg hiányos változatából (bemenet) az eredetire asszociáljon (stabil állapot)

Alapmodell: egyetlen teljesen összekötött réteg kétértékű neuronokkal:

$x_i \in \{-1, +1\}$ → $\{0,1\}$-re való áttérés: $x_i = 2y_i - 1$

DINAMIKA: $$\boxed{x_i = \operatorname{sgn}\left(\sum_j w_{ij} x_j - \Theta_i\right)}$$

→ szinkron frissítés: oszcillálhat (2-ciklus)

→ aszinkron frissítés tipikusabb (szekvenciális): minden távoli ugrás az állapottérben

Egy minta tárolása ($\underline{\Theta} = \underline{0}$ eset)

$\underline{a}$ minta stabil, ha $\forall i$: $$\boxed{\operatorname{sgn}\left(\sum_j w_{ij} a_j\right) = a_i}$$

Milyen w_{ij} súlyokat válasszunk, hogy az $\underline{a}$ minta stabil legyen?

$$\boxed{w_{ij} = a_i a_j} \Rightarrow \forall i: \operatorname{sgn}\left(\sum_j a_i a_j^2\right) = \operatorname{sgn}\left(a_i \sum_j a_j^2\right) = \operatorname{sgn}(a_i) = a_i$$

Van-e más stabil pont? Vegyük észre, hogy $-\underline{a}$ is megoldás:

$$\operatorname{sgn}\left(\sum_j a_i a_j (-a_j)\right) = \operatorname{sgn}\left(-a_i \sum_j a_j^2\right) = -a_i$$

Nem csodálkoztunk, hogy 2 megoldás $\exists$, hiszen ± 1-re szimmetrikus a rendszer! De van-e más $\underline{b} \neq \pm \underline{a}$ megoldás?

→ ha lenne ilyen stabil állapot, akkor $\exists l,k$: $b_l = a_l$, $b_k \neq a_k$

miközben $\forall i$: $\operatorname{sgn}\left(\sum_j a_i a_j b_j\right) = \operatorname{sgn}\left(a_i \sum_j a_j b_j\right) = a_i \cdot \operatorname{sgn}\left(\sum_j a_j b_j\right) = b_i$

skalár szorzat előjele attól függ, hogy egyezés vagy különbözőség a több, de mindkettő egyszerre nem lehet!

Mekkora a vonzási tartomány?

Egy adott i neuron akkor vált a kívánt a_i értékre, ha

$$\operatorname{sgn}\left(\sum_j a_i a_j b_j\right) = a_i \cdot \underbrace{\operatorname{sgn}\left(\sum_j a_j b_j\right)}_{+1 \text{ vagyis több az egyezés}} = a_i$$

Több minta ($\underline{a}^{(\alpha)} : \alpha = 1, 2, \dots, M$) tárolása :

$$W_{ij} := \sum_{\alpha=1}^{M} a_i^{(\alpha)} a_j^{(\alpha)}$$

/ tetszőleges konstans szorzó beledefiniálható : irodalomban $1/N$ normalizáció szokásos /

$\underline{a}^{(\beta)}$ stabilitási feltétele : $\forall i \quad a_i^{(\beta)} =$

$$= \operatorname{sgn}\left(\sum_{j=1}^{N} W_{ij} a_j^{(\beta)} \right) = \operatorname{sgn}\left(\sum_{j=1}^{N} \sum_{\alpha=1}^{M} a_i^{(\alpha)} a_j^{(\alpha)} a_j^{(\beta)} \right) =$$

$$= \operatorname{sgn}\left(\sum_{j=1}^{N} a_i^{(\beta)} \underbrace{a_j^{(\beta)} a_j^{(\beta)}}_{+1} + \sum_{j=1}^{N} \sum_{\alpha \neq \beta} a_i^{(\alpha)} a_j^{(\alpha)} a_j^{(\beta)} \right) =$$

$$= \operatorname{sgn}\left(N \cdot a_i^{(\beta)} + \underbrace{\sum_{\alpha \neq \beta}^{M} a_i^{(\alpha)} \sum_{j=1}^{N} a_j^{(\alpha)} a_j^{(\beta)}}_{} \right)$$

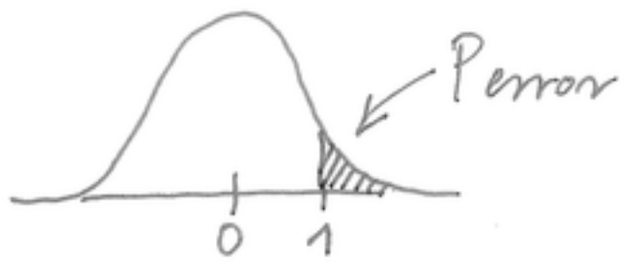
Az a kérdés, hogy ennek a tagnak $a_i^{(\beta)}$ - val megegyező az előjele, vagy N-nél kisebb az abszolút értéke?

A stabilitás és a vonzási tartomány is függ a minták számától és a relatív viszonyaiktól, ezért általánosságban csak valószínűségeket számolhatunk:

→ feltesszük, hogy a fenti tag mindhárom komponense egymástól független és a minták komponensei egyforma (0.5) valószínűséggel ± 1 - ek

N-el leosztva : az a kérdés, hogy $N(M-1)$ db véletlenszerű előjelű $1/N$ abszolút értékű tagból álló összeg hogyan viszonyul az egységhez :

→ összeg centrális Gauss $\approx \sqrt{\frac{M}{N}}$ szórással



$$P_{error} = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{1}^{+\infty} e^{-\frac{x^2}{2\sigma^2}} dx$$

(0.999)

↳ $N = 1000$ esetén $M \approx 105$

Állítás : ha $M \approx \frac{N}{2 \log_2 N}$, akkor $\lim_{n \to \infty} P_{error} = 0$

Összehasonlítás : dominánsan rekurrens CA3 régió a hippokampuszban :
≈ 200.000 sejt $\Rightarrow \approx 6000$ tárolható minta

Ritka mintákra többet is lehet : $M \approx \frac{N}{\alpha \log_2 (1/\alpha)}$

↳ $\forall m, i : P(S_i^{(m)} = 1) = \alpha$

HOPFIELD – HÁLÓZAT ENERGIAFÜGGVÉNYE

Ha a súlyok szimmetrikusak ⇒ ∃ energiafüggvény, ami monoton csökken

emlékeztető: $\underline{x}(t+1) = \mathrm{sgn}\left(\underline{\underline{W}}\,\underline{x}(t) - \underline{\Theta}\right)$ (sgn komponensenként)

$$E = -\frac{1}{2}\underline{x}^T\underline{\underline{W}}\,\underline{x} + \underline{\Theta}^T\underline{x}$$ ← korlátos

(sor: $\underline{x}^T$, oszlop (vektor): $\underline{x}$)

Állítás: energia csökken: $\Delta E = (-\underline{x}^T\underline{\underline{W}} + \underline{\Theta}^T)\Delta\underline{x}$

Tfh a k. neuront frissítjük:

$$\Delta E(x_k\downarrow) = \left[-\sum_i W_{ki}x_i(-1) + \Theta_k(-1)\right] - \left[-\sum_i W_{ki}x_i(+1) + \Theta_k(+1)\right] = 2\left(\sum_i W_{ki}x_i - \Theta_k\right)$$

$$\Delta E(x_k\uparrow) = \left[-\sum_i W_{ki}x_i(+1) + \Theta_k(+1)\right] - \left[-\sum_i W_{ki}x_i(-1) + \Theta_k(-1)\right] = -2\left(\sum_i W_{ki}x_i - \Theta_k\right)$$

$x_k\downarrow$ esetén a zárójeles rész negatív
$x_k\uparrow$ esetén a zárójeles rész pozitív

| a léptetési szabályból ⇒ $\Delta E < 0$ mindkét esetben

ENERGIAMINIMUMOK BEÁLLÍTÁSA:

→ 1 minta esetén: $E \sim \left(\sum x_i a_i\right)^2$

→ több mintára: $E \sim -\sum_\alpha\left(\sum_i x_i a_i^{(\alpha)}\right)\left(\sum_j x_j a_j^{(\alpha)}\right) = -\sum_i\sum_j\left(\sum_\alpha a_i^{(\alpha)}a_j^{(\alpha)}\right)x_i x_j$

NEM KÍVÁNT MINIMUMOK:

→ $-\underline{a}^{(\alpha)}$ /inverzek/

→ páratlan lineáris kombinációk, pl: $a_i^{(\text{hibrid})} = \mathrm{sgn}\left(\pm a_i^{(\alpha_1)} \pm a_i^{(\alpha_2)} \pm a_i^{(\alpha_3)}\right)$

Általában ezeknek nagyobb az energiájuk és kisebb a vonzási körzetük

ÖTLETEK:
→ ismételt kereséssel mélyebb energiájú állapotot találhatunk
→ szinkron frissítéssel nagyobb teret tudunk bejárni
→ sztochasztikus neuronok: szimulált lehűtés: korlátozott arányban $\Delta E > 0$ lehet

ELŐNYÖK:
→ lokális és inkrementális tanulás (lépésről-lépésre többet)
→ robosztusság (kivehetünk neuronokat)
→ heteroasszociációt is megoldja:
elválasztunk egy részt és futtatásnál random inputot kap
→ rejtett változók: reprezentáció tanulás
→ kvadratikus alakra hozható optimalizációs feladatot megold

HÁTRÁNYOK:
→ megköveteltük a szimmetriát!
→ megsértettük a Dale-törvényszerűséget:
egy sejt nem lehet egyszerre gátló és serkentő is
(vannak kivételek ez alól is)

Az emlékek nem csak a múltról szólnak, meghatározzák a jövőnket. (Az emlékek őre c. film)

Igen vacak memória az, amelyik csak hátrafelé működik. (Lewis Carroll)

A memória az a képességünk, amellyel felejtünk... (Vavyan Fable)

Úgy tűnik, létezik elmémnek egy királyi fogadóterme, ahol a tudat ülésezik, és ahova egyszerre két-három idea járul kihallgatásra. (Francis Galton)

Az emlékezet bizonytalan. Alkalmazkodik ahhoz, ahogy a világot értjük, megváltozik, hogy illeszkedjen az előítéleteinkhez. (Holly Black)

A múltad átírható. Amit most megváltoztatsz, az a múltadat is megváltoztatja. Nem magát a megtörtént eseményt, hanem a lelkedbe írt hatását. (Müller Péter)

Ha nincs, aki megerősíthet az emlékeidben, egy idő után elveszíted őket. (Colleen Hoover)

Annyi valóság van, ahányan vagyunk, az analízisben is azt mondják, nem az számít, hogy mi történt valójában, hanem az, hogyan élte át az ember, és miként emlékszik rá. (Lángh Júlia)
Azt mondják, csak olyan eseményre emlékszik az ember, ami akkor történt meg vele, amikor már ismerte az annak leírásához szükséges szavakat. (Mary E. Pearson)

Ha az ember csak vissza tud emlékezni mindarra, amit látott, akkor már sosem henyél, igazában sosem magányos, és nincs többé egyedül. (Vincent Van Gogh)

Mindannyian arra vágyunk, hogy valaki megőrizzen minket az emlékezetében. (Diego Marani)

Addig élünk, amíg módunkban áll visszafelé tekinteni. (Sütő András)

EMLÉKEZTETŐ

Hopfield-hálózat a legegyszerűbb <u>energia alapú</u> neuronhálózat.
Van egy <u>korlátos skalár</u>, <u>ami monoton csökken</u> a dinamikai fejlődés során.
<u>Rekurrens</u> hálózatok dinamikája általában nagyon bonyolult lehet, de a
$w_{ij} = w_{ji}$ szimmetria miatt ∃ energia: $\boxed{E = -\frac{1}{2}\underline{x}^T \underline{\underline{W}}\, \underline{x} + \underline{\Theta}^T \underline{x}}$

Mi az energia? Párkonelaíók súlyozott összege. (sor, oszlop (vektor))

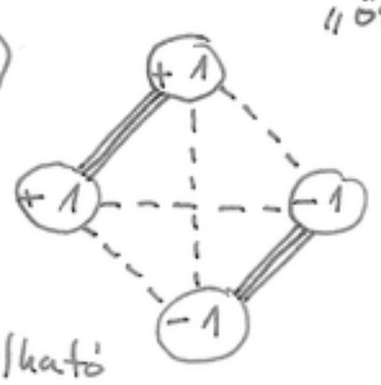
<u>Hasonlóság</u>: spinüvegek Ising-modellje felfele és lefele álló spinekkel
(annyi különbséggel, hogy a spinüveg rácsszerű véletlen súlyokkal)
szimmetrikus, kvadratikus energia + léptetési szabály $\Rightarrow \boxed{\Delta E = (-\underline{x}^T \underline{\underline{W}} + \underline{\Theta}^T)\Delta \underline{x} < 0}$

<u>állapottér</u>: N-dimenziós hiperkocka csúcsai
kiterjedt vonzási tartomány ⇒ kapacitás korlát:

TANULÁSI SZABÁLYOK

→ <u>Hebb-féle</u> tanulási szabály (<u>lokális, inkrementális</u>)

$$\underline{\underline{W}} \to \underline{\underline{W}} + \varepsilon(\underline{a}\,\underline{a}^T - \underline{\underline{I}})$$



<u>kapacitás aszimptotikusan</u>:
$M \approx \frac{N}{2\ln(N)}$, ha majdnem ∀ minta pontosan rekonstruálható
$M \approx \frac{N}{4\ln(N)}$, ha ∀ minta pontosan rekonstruálható
<u>megjegyzés</u>: ennél kisebb a kapacitás, ha korrelálatlak a minták!

→ <u>pszeudo-inverz</u> tanulási szabály (nem lokális, nem inkrementális)
kapacitás ~ N, de mátrix invertálást igényel!

→ <u>Storkey</u> (1997): lokális mezőt figyelembe vevő <u>1-rendű</u> képlet
$w_{ij} \to w_{ij} + a_i a_j - a_i h_{ji} - a_j h_{ij}$, $h_{ij} = \sum_{\substack{k=1 \\ k \neq j}}^{N} w_{ik} a_k$ ← ($w_{ii} = 0$ miatt $k \neq i$)
(i. neuron inputja j nélkül)
kapacitás: $\frac{N}{\sqrt{2\ln(N)}}$
megjegyzés: körülzéséből lokálisan kiemeli a tanult mintát: „kontrasztot"

→ <u>Unlearning</u> (Hopfield, 1983): procedúra: 1. véletlen állapot
2. stabil fixpontba fejlődés
3. $w_{ij} \to w_{ij} - \varepsilon x_i x_j$
(ismétlés)
elgondolás:
A nagyobb vonzási medencével rendelkező valódi minták jobban túlélik!
megjegyzés: vigyázni kell a megfelelő mértékre, mert egy idő után ∀ törlődik, és elkezd új hamis mintákat generálni a rendszer

→ módosítás: <u>ciklusosság</u>: learning ⇄ unlearning ciklusok
megjegyzés: Nem csak arra alkalmas, hogy a hamis memóriákat csökkentsük, hanem arra is, hogy a régieket elfelejtsük újak javára kapacitást felszabadítva!

<u>CRICK-MITCHISON hipotézis</u> (1983) neurológiai adatok nélkül: álom → felejtés
Álom: → véletlenszerű fúrcsaságok
→ sokszor negatív élmény
→ nem emlékszünk rá (legfeljebb az utolsóra RTM segítségével)

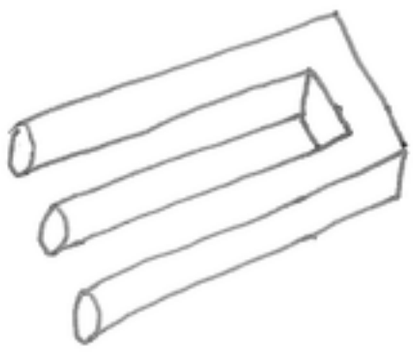
SZTOCHASZTICITÁS ELŐNYEI:

→ zaj hatására ki lehet szabadulni a lokális minimumokból

→ a feladat általában probabilisztikus:

→ zaj

→ ambivalencia: → Necker-cube
→ hollow-face
(Bajcsy-Zsilinszky Endre szobra a Deák-téren)
→ stb.

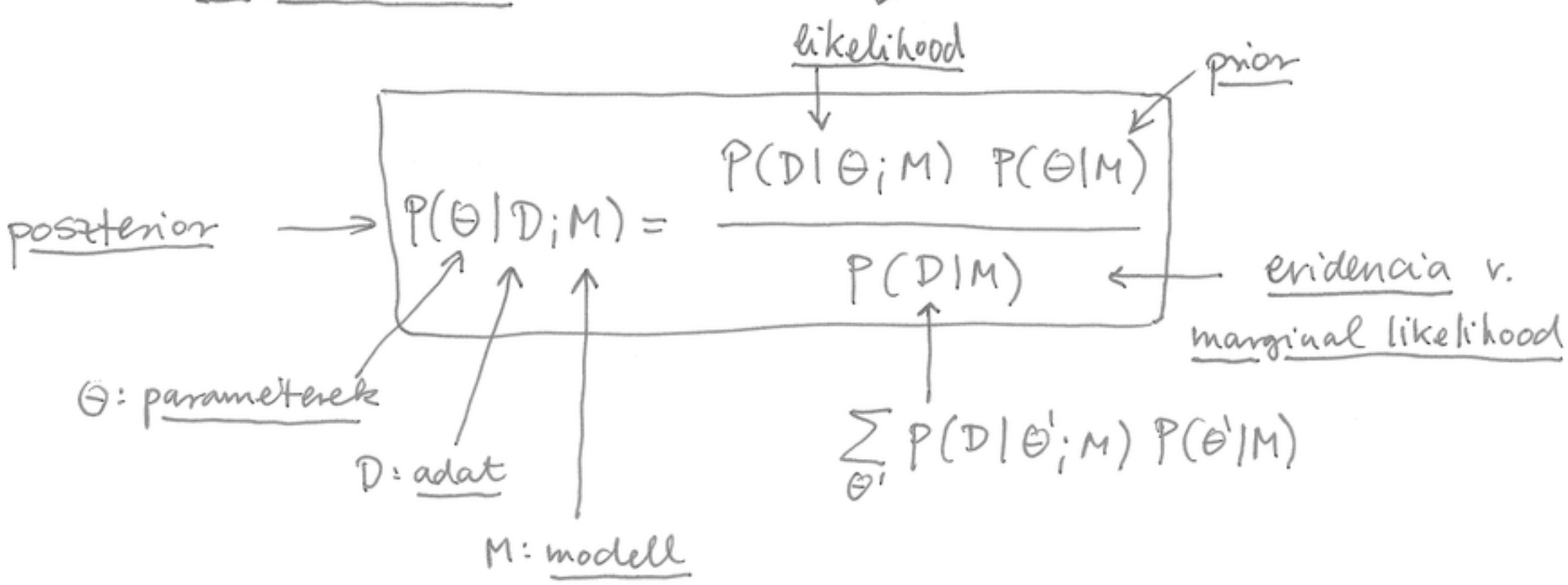


→ talán éppen emiatt az idegrendszer is sztochasztikus: szinaptikus „zaj."

Az optimális megoldás probabilisztikus megoldást kíván:
a tanulandó struktúrák valójában sokdimenziós valószínűség-eloszlások

(Hasonlóan a tudományhoz, az idegrendszer is a világot leíró generatív modellt próbálja kitalálni és jóslásokhoz felhasználni)

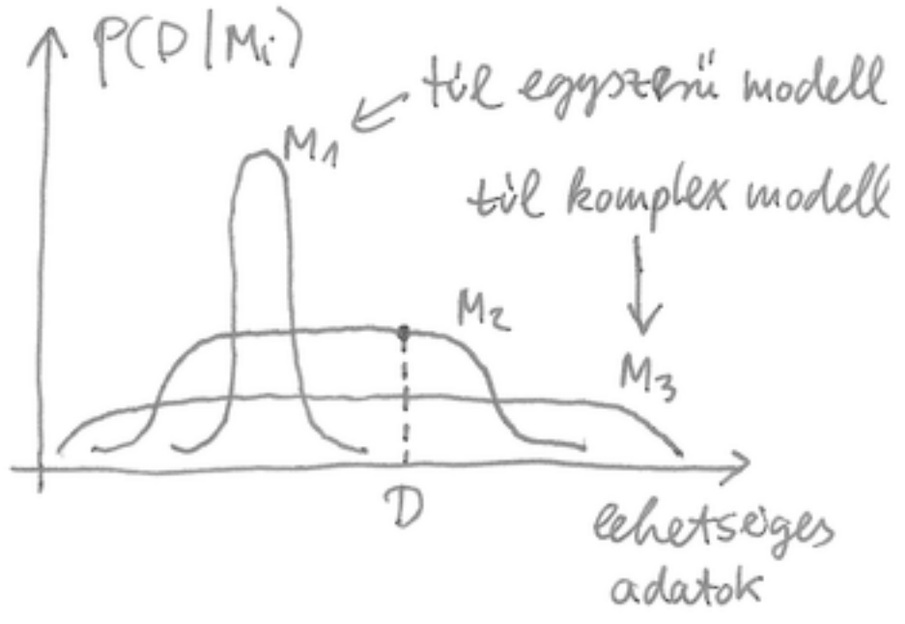
Általános Bayes-tanulási séma:



prediktív eloszlás (a jósláshoz): $P(D_{n+1}|\theta, D_{[0,n]}, M)$

prior illúziók: → checker-shadow illusion
→ Margaret Thatcher illusion
→ stb.

modell szelekció: →



reprezentációs tanulás:

→ megfigyelt változók
→ rejtett változók (paraméterek)

Például a Boltzmann-gép neuronjait két csoportba osztjuk: $(\underline{v}, \underline{h})$

megjegyzés: Ha nem lenne a megfigyelt környezetben struktúra, akkor 2^v darab valószínűségre lenne szükség a leíráshoz, amikhez a w_{ij} súlyok nem lennének elegendőek!

BOLTZMANN-GÉP

(Geoffrey Hinton & Terry Sejnowski, 1985)

→ bináris állapotok : $s_i \in \{0, 1\}$

→ aszinkron léptetési szabály :

→ s_i valószínűséggel kiválasztjuk az i. neuront

→ p_i valószínűséggel $s_i := 1$, egyébként $s_i := 0$

$$p_i = \frac{1}{1+e^{-\Delta E_i}}, \text{ ahol } \Delta E_i = \sum_j w_{ji} s_j - \Theta_i$$

$$w_{ij} = w_{ji}, \quad w_{ii} = 0$$

↳ $q_i = 1 - p_i = 1/(1+e^{\Delta E_i})$

Állítás : A rendszer (termikus) egyensúlyi-eloszlása Boltzmann-eloszlás :

$$\pi(\underline{s}) = \frac{e^{-E(\underline{s})}}{\sum_{\underline{s}'} e^{-E(\underline{s}')}} \leftarrow Z \text{ ún. állapotösszeg (normalizáció)}$$

ahol $$E(\underline{s}) = -\sum_{i<j} w_{ij} s_i s_j + \sum_i \Theta_i s_i$$

↳ ΔE_i = energia megváltozása, ha csak az i. neuron vált: $1 \to 0$

Emlékeztető : Markov-lánc (jelen esetben véletlen bolyongás az állapottérben) részletes egyensúlya : ha egy $\pi(\underline{s})$ valószínűségi mérték $\forall \underline{s}, \underline{s}'$ állapotra eleget tesz az alábbinak :

$$\pi(\underline{s})\, p(\underline{s} \to \underline{s}') = \pi(\underline{s}')\, p(\underline{s}' \to \underline{s})$$

Állítás : Ekkor viszont $\pi(\underline{s})$ stacionárius eloszlás, azaz :

$$\pi(\underline{s}) = \sum_{\underline{s}'} \pi(\underline{s}')\, p(\underline{s}' \to \underline{s})$$

Bizonyítás : $/\sum_{\underline{s}'} \Rightarrow \sum_{\underline{s}'} \pi(\underline{s}) p(\underline{s} \to \underline{s}') = \pi(\underline{s}) \underbrace{\sum_{\underline{s}'} p(\underline{s} \to \underline{s}')}_{1} = \pi(\underline{s})$

↳ Bizonyítás : Ha $\underline{s} = \underline{s}'$, akkor triviális a részletes-egyensúly

Ha $\underline{s} \neq \underline{s}'$, akkor az aszinkron léptetés miatt egyetlen neuronnál különbözik a két állapot (i)

$$\pi(\underline{s}) p(\underline{s} \to \underline{s}') = \frac{e^{-E(\underline{s})}}{Z} \cdot \frac{s_i}{1+e^{-\Delta E_i}} \quad / \cdot \frac{e^{\Delta E_i}}{e^{\Delta E_i}} = 1$$

($s_i = 0$, $s_i' = 1$)

$$\Rightarrow = \frac{e^{-E(\underline{s})+\Delta E_i}}{Z} \frac{s_i}{e^{\Delta E_i}+1} = \frac{e^{-E(\underline{s}')}}{Z} \cdot s_i \cdot q_i = \pi(\underline{s}')\, p(\underline{s}' \to \underline{s})$$

Q.E.D.

BOLTZMANN-GÉP TANÍTÁSA

A valódi megfigyelhető változók $p(\underline{v})$ eloszlása és a belső modell által szabadon generált $\pi(\underline{v})$ egyensúlyi eloszlás közti

Kullback-Leibler divergencia: $G = \sum_{\underline{v}} p(\underline{v}) \log\left(\frac{p(\underline{v})}{\pi(\underline{v})}\right) = \sum_{\underline{v}} p(\underline{v}) \left[\log(p(\underline{v})) - \log(\pi(\underline{v}))\right]$

rejtett változókra marginalizált egyensúlyi eloszlás:

$$\pi(\underline{v}) = \sum_{\underline{h}} \pi(\underline{v},\underline{h}) = \frac{1}{Z} \cdot \sum_{\underline{h}} e^{-E(\underline{v},\underline{h})}$$, ahol $Z = \sum_{\underline{u},\underline{h}} e^{-E(\underline{u},\underline{h})}$ ún. állapotösszeg

↖ joint egyensúlyi eloszlás

$$E(\underline{v},\underline{h}) = -\sum_{i<j} w_{ij} s_i^{\underline{v}\underline{h}} s_j^{\underline{v}\underline{h}} + \sum \theta_i s_i^{\underline{v}\underline{h}}$$

segéd formula: $\frac{\partial}{\partial w_{ij}} e^{-E(\underline{v},\underline{h})} = \frac{\partial}{\partial w_{ij}} e^{\sum_{i<j} w_{ij} s_i^{\underline{v}\underline{h}} s_j^{\underline{v}\underline{h}}} = s_i^{\underline{v}\underline{h}} s_j^{\underline{v}\underline{h}} e^{-E(\underline{v},\underline{h})}$

$p(\underline{v})$ w_{ij}-től független $\Rightarrow \frac{\partial G}{\partial w_{ij}} = -\sum_{\underline{v}} p(\underline{v}) \frac{\partial}{\partial w_{ij}} \log(\pi(\underline{v}))$

$$\frac{\partial}{\partial w_{ij}} \log(\pi(\underline{v})) = \frac{\partial}{\partial w_{ij}} \log\left(\sum_{\underline{h}} e^{-E(\underline{v},\underline{h})}\right) - \frac{\partial}{\partial w_{ij}} \log(Z) =$$

$$= \frac{\frac{\partial}{\partial w_{ij}} \left(\sum_{\underline{h}} e^{-E(\underline{v},\underline{h})}\right)}{\sum_{\underline{h}} e^{-E(\underline{v},\underline{h})}} - \frac{\frac{\partial}{\partial w_{ij}} Z}{Z} = \frac{\sum_{\underline{h}} s_i^{\underline{v}\underline{h}} s_j^{\underline{v}\underline{h}} e^{-E(\underline{v},\underline{h})}/Z}{\sum_{\underline{h}} e^{-E(\underline{v},\underline{h})}/Z} - \frac{\sum_{\underline{u},\underline{h}} s_i^{\underline{u}\underline{h}} s_j^{\underline{u}\underline{h}} e^{-E(\underline{u},\underline{h})}}{Z} =$$

$$= \frac{\sum_{\underline{h}} s_i^{\underline{v}\underline{h}} s_j^{\underline{v}\underline{h}} \pi(\underline{v},\underline{h})}{\underbrace{\sum_{\underline{h}} \pi(\underline{v},\underline{h})}_{\pi(\underline{v})}} - \sum_{\underline{u},\underline{h}} s_i^{\underline{u}\underline{h}} s_j^{\underline{u}\underline{h}} \pi(\underline{u},\underline{h})$$

$$\Rightarrow \frac{\partial G}{\partial w_{ij}} = -\sum_{\underline{v}} p(\underline{v}) \left[\sum_{\underline{h}} s_i^{\underline{v}\underline{h}} s_j^{\underline{v}\underline{h}} \underbrace{\pi(\underline{h}|\underline{v})}_{p(\underline{h}|\underline{v})} - \sum_{\underline{u},\underline{h}} s_i^{\underline{u}\underline{h}} s_j^{\underline{u}\underline{h}} \pi(\underline{u},\underline{h})\right]$$

kondicionálisok egyenlősége definíció szerint ⟶ $p(\underline{h}|\underline{v})$

(mert mindegy, hogy $\underline{v}$ eleve fix, vagy a szabad fejlődés során jutott ilyen egyensúlyba)

Felhasználva, hogy $p(\underline{v},\underline{h}) = p(\underline{h}|\underline{v}) \cdot p(\underline{v})$, továbbá $\sum_{\underline{v}} p(\underline{v}) = 1$ a második tagnál

$\Rightarrow \frac{\partial G}{\partial w_{ij}} = p_{ij}^- - p_{ij}^+$, ahol $p_{ij}^+ = \sum_{\underline{v},\underline{h}} p(\underline{v},\underline{h}) s_i^{\underline{v}\underline{h}} s_j^{\underline{v}\underline{h}}$ ← együttes aktivitások valószínűségei a kett fázisban

Hasonlóan: $\frac{\partial G}{\partial \theta_i} = p_i^- - p_i^+$ $\quad p_{ij}^- = \sum_{\underline{v},\underline{h}} \pi(\underline{v},\underline{h}) s_i^{\underline{v}\underline{h}} s_j^{\underline{v}\underline{h}}$

GRADIENS-MÓDSZER (lokálisan számolható tanulás): $\Delta w_{ij} = \varepsilon(p_{ij}^+ - p_{ij}^-)$

↑ tanulási ráta $R = 1/\varepsilon$

megjegyzés: Maximum likelihood ugyanezt adja:

$$\operatorname{argmax}_W \prod_{\underline{v}\in V} p(\underline{v}|W) = \operatorname{argmax}_W \sum_{\underline{v}\in V} \log(\underline{v}|W)$$

↑ azon w_{ij}-k keresése, amikre ez maximális ← V mintahalmaz

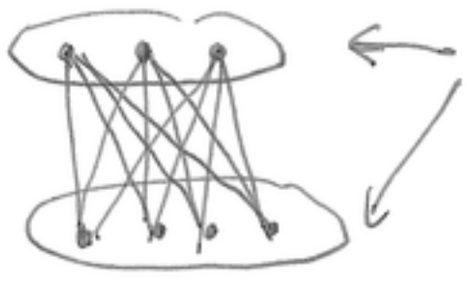
BOLTZMANN GYORSÍTÁSOK

Sajnos az alap Boltzmann-gép nagyon zajos és az egyensúly kivárása miatt nagyon lassú a tanítása.

→ trükk: tároljuk az előző egyensúly mintáját (particle), és így az egyensúly hamarabb beáll ebből indítva, ha w_{ij}-k kicsit változnak

RBM (restricted Boltzmann machine) Smolensky, 1986:

Ha a kapcsolatokat megszorítjuk, akkor praktikusan is használhatóvá válik!



← rétegeken belül nincsen kapcsolat, így rögzített $\underline{v}$-re azonnal beáll az egyensúly $\underline{h}$-ra is!

↓ ez triviális

$$\Delta w_{ij} \sim \langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_{modell}$$

CONTRASTIVE DIVERGENCE (Hinton 2002)

$$\Delta w_{ij} \sim \langle v_i h_j \rangle^0 - \langle v_i h_j \rangle^\infty \approx \langle v_i h_j \rangle^0 - \langle v_i h_j \rangle^1$$

meglepően jól működik!

↑ rekonstrukció + rekonstrukcióból generált $\underline{h}$

→ adattól távolabbi minimumokat nem látja jól, ezért a szokásos kompromisszum: CD1, CD3, CD5, ... egymás után

DBM (deep learning):

→ előre tanított RBM-ek réteges összeillesztése

például a 2. réteg feature-ök feature-jeit tanulja

analógia: absztrakciós hierarchia a látórendszerben: retina (1M) →

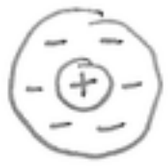
→ LGN (1M) → V1 (190M) → V2 (150M) → V4 (68M) → PIT (36M) → CIT (17M) → AIT (16M) ← neuronok száma

↑ luminancia kontraszt

↑ élek, vonalak orientáció szelektivitás

formák minták

objektumok kategóriák



illúziónkus kontúrok

Kanizsa-háromszög

felhasználás: pl. beszédfelismerés, karakter-, kép-, arc-felismerés, stb.

Boltzmann általánosítások: → magasabb rendű: $w_{ijk} s_i s_j s_k$, tanulás hasonló: $\langle s_i s_j s_k$

→ "mean field" közelítés: átlagokra vonatkozó determinisz

→ "tikus dinamika (folytonos Hopfield-al ekvivalens) $\frac{dx_i}{dt}$