



Översikt

- Beräkningsprinciper
- Bindningsproblemet
 - Rött äpple och grönt päron, eller grönt äpple och rött päron?
- Olika minnessystem
 - Rygggraden i kognitiva systemet
- Analys av nätverkets inandöme
- Parametrar i Emergent
 - Inhibering av individuella noder

Beräkningsprinciper

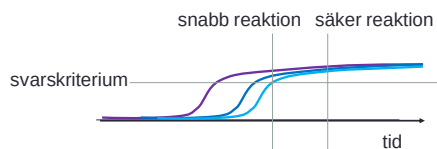
Återblick

- Kunskap om hjärnans beräkningsprinciper hjälper oss förstå hur mänskligt tänkande uppstår
- Vägleder forskning inom kognitiv neurovetenskap
 - Disengage-funktionen inom selektiv uppmärksamhet
 - Parietallobsskada
- Beräkningsprincipen i detta fall
 - Inbyggd tävlan mellan olika stimuli
 - Fokus flyttas automatiskt när nytt stimulus blir starkt aktiverat (via input eller via top-down feedback)
 - Behöver inte anta någon disengage-funktion, utan disengage-effekten uppstår ur ömsesidig tävlan (implementerat som kWTA i vårt fall)
 - Har visat sig stämma bra med empiriska data

Återblick

Hjärnan är bra på snabb beräkning (quick and dirty)

- Interaktiv aktivering
 - Aktivering sprids snabbt runtom i nätet
 - Finns möjlighet att reagera snabbt
 - Men en tidig reaktion på stimulus kan visa sig vara inkorrekt



Två beräkningsvägar

- Vad- och var-kanalen
 - Snabb beräkning i var-kanalen ger snabb motorisk reaktion
 - Långsammare, mer detaljrik beräkning i vad-kanalen ger rik bas för associativa och språkliga funktioner
- Analytiskt och intuitivt system i beslutsfattande
- När bearbetar det ena, när bearbetar det andra systemet?

Återblick

Hjärnan tillämpar (till synes) redundant beräkning

- Fallet HV (skada på vad-kanalen)
- Beräknar en sak på två olika sätt parallellt
 - T.ex. visuella mönster beräknas i både vad- och var-kanalen
- Slutlig respons baseras på en av dessa
 - Inbyggd tävlan mellan dessa två beräkningar
 - Den ström som kommer upp i aktivering först får avgöra responsen

Dator: skillnad på data och beräkning

- I en dator är det skillnad på data och bearbetning
- Man vill ha generell beräkning som kan tillämpas på olika sorters data
- Det blir då svårt att anpassa beräkningen till datas egenskaper
 - Hur äter jag denna frukt?
 - Om den är blöt och stor
 - Går lätt att dela → kniv och gaffel
 - Om ...
 - Måste uttryckligen täcka upp alla möjligheter
 - Arbetsamt, ej flexibelt
 - Vad göra när helt ny, exotisk frukt, som inte är med bland if-satserna?

Hjärnan: innehållsspecifik beräkning

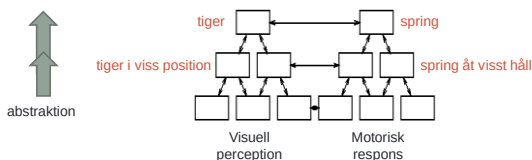
- Aktivering av data = bearbetning
 - Ingen skillnad mellan data och bearbetning
- Innehållsspecifikt beräkningssätt
 - Ex. olika särdrag bearbetas av olika hjärnceller
 - Dessa i sin tur skickar till olika andra celler, etc.
 - Olika beräkningsströmmar landar i olika associativa områden inom olika delar av hjärnan
- Ofta redundanta representationer
 - Samma eller liknande information representeras på flera ställen

Andra beräkningsprinciper

- Hierarkisk beräkning ger olika abstraktionsnivåer
 - Kan stegvis bortse ifrån detaljer
 - T.ex. mer och mer komplexa särdrag
 - Kombinerar till objekt
 - Sedan semantisk innebörd
- Olika modaliteter (t.ex. syn och motorik) kan på så sätt associeras till varandra på olika abstraktionsnivåer

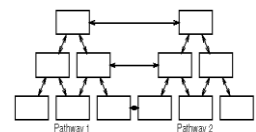
Abstraktionsnivåer

- Associationer utvecklas automatiskt på rätt nivå
 - Inlärning känslig för korrelation mellan input och rätt beteende



Hierarkiska visuella kanaler

- Olika typer av beräkning
 - Ex. färg, form – position
- Uppnår konstans (invariants) genom att stegvis abstrahera från detaljer i input
 - Ex. positionsinvariants
- Stegvis, för att undvika komplex beräkning



Storlekskonstans (invarians)



- Hjärnan kan bortse ifrån varierande storlek på näthinnan
- **Invariant mental representation**

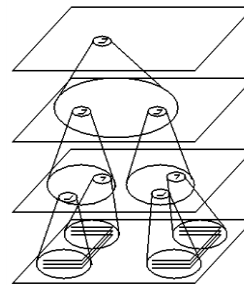
Rumslig invarians

- Hur representera objekt oberoende av position
- Vill samtidigt bevara den relativa positionen mellan objektets olika delar
- Svar:
 - Stegvis bearbetning
 - Mer och mer komplexa representationer som kodar rumslig relation mellan objektets delar
 - Mindre positionsberoende för varje steg

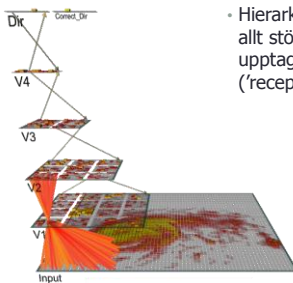
Hierarkisk beräkning

- Varför behövs hierarkisk beräkning?
 - Svårt att annars extrahera invarianser
 - Ex. parallell beräkning av...
 - Parametrar för rotation, utdragning, lutning, förstoring/förminsning
 - Matchning mellan resulterande kanonisk form och objekt lagrat i LTM
- Mycket svårt sökproblem!
 - Många möjliga transformationer
 - Många lagrade kanoniska representationer
 - Många-till-många mappning

Objektigenkänning

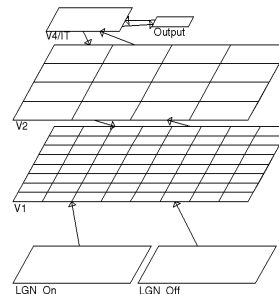


Hierarkiskt nätverk



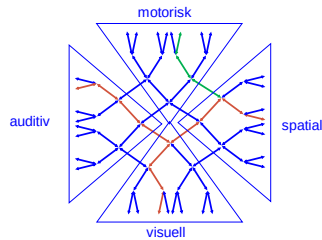
- Hierarki av neuroner med allt större upptagningsområde ('receptive field')

Objektigenkänning (kap 8 i boken)

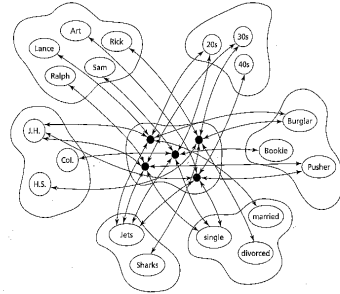


Associativa konvergenzoner

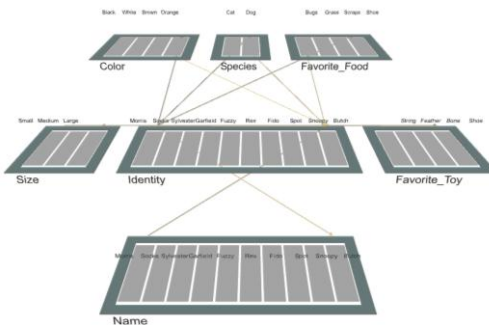
- Sammanstrålning längst upp i varje hierarki



Associativa konvergenzoner



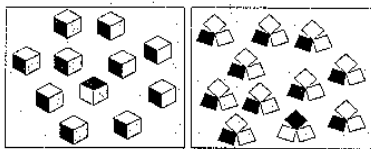
Cats and dogs



Bindningsproblemet

Kombinera färg (svart/vit) med relativ position

- Ex: 3-D landskap kan ge ögonblicklig upptäckt



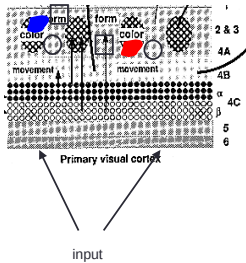
Färg kan paras ihop med
position parallellt
Inlärt

Färg kan INTE paras ihop
med position parallellt
Nytt

Bindningsproblemet

- Vilka egenskaper hör ihop med vilka?
 - Perceptuell information delas ju upp i individuella särdrag i ett tidigt bearbetningssteg
 - Hur återförenas dessa särdrag i olika objektrepresentationer?
- Problemet gäller även i större skala, t.ex. vad- och var-kanalen
 - Objektgenskaper – objektposition, rotation
 - Hur förenas dessa?

Exempel



• Hur såg input ut egentligen?

• Blå cirkel och röd fyrkant eller... röd cirkel och blå fyrkant?

Försök till lösning

- Märk representationer tillhörande samma objekt
 - Grupper av neuroner som avfyra i samma rytm antas kodar egenskaper som kommer från samma objekt
- Finns evidens för att neuroner avfyra synkroniserat i hjärnan
- Kan dock vara en sideeffekt utan egentlig funktion
 - Neuroner som aktiverar varandra tenderar att spika i samma takt

Alternativ lösning enligt boken

- Konjunktiva representationer (OM s 220)
 - En grupp av neuroner är bara aktiva för en viss *kombination* av särdrag
- Kombinationen i detta fall = viss färg + viss form
 - Räcker med att några celler kodar detta
- Stöd: Viss typ av formigenkänning förekommer även i varkanalen
- Behöver inte koda alla möjliga särdragskombinationer

Lösning på bindningsproblemet

obj1	obj2	R	G	B	F	C	T	RC/GF/B T
RF	GC	1	1	0	1	1	0	0
RC	GF	1	1	0	1	1	0	1
RF	GT	1	1	0	1	0	1	0
RT	GF	1	1	0	1	0	1	1
RF	BC	1	0	1	1	1	0	0
RC	BF	1	0	1	1	1	0	1
RF	BT	1	0	1	1	0	1	0
RT	BF	1	0	1	1	0	1	1
...	...	...	...	...	...	...	...	...

Nya kombinationer då?

- Sammanblandning av särdrag mellan olika objekt kan hindras med redundanta representationer
- Förutsätter att kombinationer av särdrag har lärts in
 - Fungerar för kombinationer av särdrag som vi ser ofta
- Hur kan vi varsebli/förena *ovanliga* kombinationer av särdrag?
 - Dessa neurongrupper har inte bildats ännu...

- Detta var utmärkt exempel när modellering och de beräkningsprinciper man upptäcker via modellering kan ge en djupare förklaring till psykologiska fenomen

Olika minnessystem

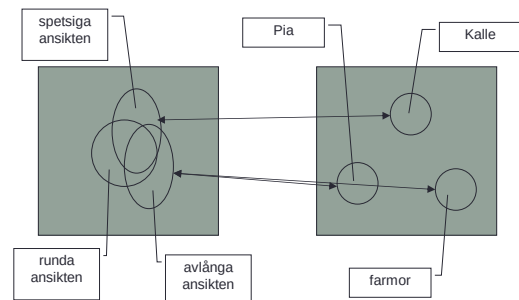
Inläring

- Integrerande
 - Ex. känna igen bokstaven A, efter att ha sett många olika a:n
 - Långsam, viktjustering i små steg
 - Vill inte rubba gammal kunskap
 - Lagras över varandra
- Separerande
 - Ex. känna igen ett ansikte
 - Snabb
 - Känna igen efter enstaka exponering
 - Kommer att lagras var för sig

Olika ställen i hjärnan

- Integrerande
 - Parietalloben
- Separerande
 - Ventrala delar, hippocampus
- Ofta samarbete mellan dessa

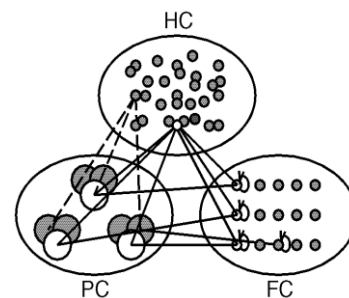
Ex. ansikten



Minnet baserat på olika mekanismer

- Långtidsminnet **viktbaserat**
 - Viktförändringar som kodar inlärd associationer
- Arbetsminne
 - **Aktiveringsbaserat** minne
 - Aktivering bevaras genom självåteraktivering (återkoppling till sig själv)

Tre minnessystem



Tre typer av minnen

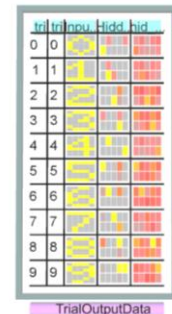
- Modalitetsspecifik, distribuerad (Parietal cortex)
 - Distribuerade, överlappande representationer
 - Likhet mellan olika särdrag
 - Kan generalisera
- Associativ, distribuerad, gles aktivering (Hippocampus)
 - En representation aktiv åt gången
 - Sammanstrålning: representerar kombinationen av ingående särdragen
- En nod för varje aktivt särdrag (Frontal cortex)
 - Arbetsminne genom själv-återaktivering av noder

Analys av nätverks inandöme

Informationsbearbetning

- Neuroner organiseras i lager
- Inputlagret antar ett visst aktiveringsmönster för varje input
 - Om lagret är **hard clamped** sätts noderas aktivering externt
 - Annars beroende på aktiveringar i resten av nätet
 - Input från input-mönstret adderas till feedbacksignaler (mer om detta på s 88 i boken)

Sifferigenkänning



Beräkning i nät

- Likheter (överlapp) mellan aktiveringsmönster avslöjar hur lagret tolkar olika input
 - Mönster som liknar varandra tyder på att lagret behandlar motsvarande input som om de tillhörde "samma kategori"
- Vad händer med dessa interna "kategorier" när lagret skickar informationen vidare till nästa lager?

Transformation mellan lager

- Likhetsförhållanden mellan aktiveringsmönster omvandlas (transformeras) mellan olika lager
- Transformation, omrepresentation
- De input som är lika i ett lager kommer kanske *inte* representeras lika i nästa lager
 - Vissa likheter kanske betonas i mottagarlagret
 - Andra likheter slätas över (tolkas som slumpmässiga)

Analys av nätet

- Analys av inandömet
- Tre sätt att ta reda på hur olika input representeras i de dolda lagren
 1. Visualisering av aktiveringar för olika input
 2. Gruppering av aktiveringar (cluster plot)
 3. Principal component analysis

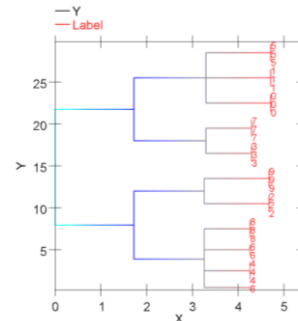
Visualisering



Klusterdiagram

- Mönstren etiketteras med de namn som angavs i input-mönstret (t.ex. "0", "1", osv)
- De aktiveringsmönster som är mest lika varandra
 - Grupperas ihop
 - Dessa grupper ligger längst ut i förgreningarna
- Aktiveringsmönster som är näst lika
 - Ligger längre inåt i förgreningarna

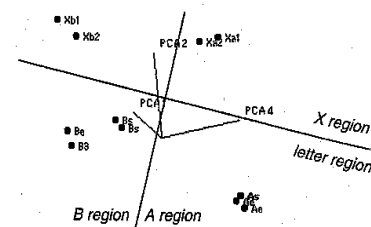
Klusterdiagram i Leabra



Användning

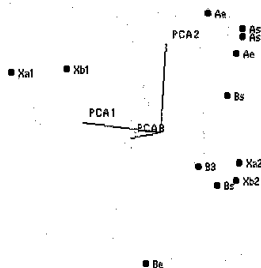
- Man kan köra klusterdiagram för olika lager i ett nät
- Visar hur nätet stegvis omvandlar input till de önskade slutkategorierna
- Lättare att förstå vilken roll ett visst lager spelar i bearbetningen av input
- Ibland vill man ha mer än klusterdiagram...

Principal Component Analysis (PCA)



Input av typen AAAXXXBBBB. Uppgift: är antal B:n = antal A:n?

Principal Component Analysis (PCA)



Input av typen AAAAAXXXBBB. Uppgift: är antal B:n = antal A:n?

Parametrar i Emergent och deras biologiska motsvarighet

Se även Appendix C i boken

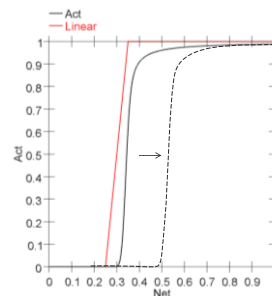
BBANN

- g_{bar}
- g
- net
- inhibering
- act

Biologi

- Cellens känslighet (e, i, l)
- Mängden insignal (e, i)
- Summan av den input som exciterande hjärnceller skickar till en mottagarcell
- (se graf på nästa sida)
- Spikfrekvens

Net $\rightarrow$ act



- Inhibering dras ifrån net
- Funktionen förskjuts i sidled
- Det krävs högre net för att få samma ut-aktivering när inhibering läggs ovanpå net

BBANN

- bias_wt
- $r_{\text{wt}}, s_{\text{wt}}$
- dav
- $\text{utaktivering} > 0.6$

Biologi

- Olika celler är olika känsliga för excitatorisk input
 - Inkommande kopplingar kan landa olika långt ifrån cellkroppen
- Synaptisk effektivitet
- Delta aktivering (mått på om nätet nått stabilt tillstånd)
- RT

BBANN

- irate
 - tillämpas på hela knippen
- $\text{act_m}, \text{act_p}$
- savg
- $\text{gain}, \text{wt_off}$

Biologi

- Inte helt klarlagt
 - Vissa forskare ser irate som dynamisk, andra som fix för varje cell
- Faser: tidsförskjutning mellan förväntan och bekräftelse
- Anpassning av Δw till förväntad aktiveringsnivå hos sändarlagret
- Kontrastförstärkning (biologin ej klarlagt)

BBANN**Biologi**

- k_{Hebb}
- distribuerad representation
 - k i KWTa
- Andel Hebb-inläring
 - Att de två inlärningsätten blandas är troligt eftersom dels långväga feedback-kopplingar, dels lokal synaptisk modifiering
- Verkar viktigt för olika typer av minne