

Bio-nätverk: Hebbiansk inlärning

729G83

Översikt

- Hebb
 - Nätet ska upptäcka korrelationer i input
- XCAL
 - Automatisk arbetsfördelning – självorganisering
- Generalisering
 - Tät vs. gles aktivering
 - Grad av överlapp mellan aktiveringsmönster

Tre former av lärande

- Övervakad inläring
 - Statistiska mönster i indata
- Övervakad inläring
 - Vilka aspekter av indata kan relateras till korrekta (önskade) utdata
- Reinforcement learning
 - Vilka händelser har lett fram till belöning
 - Tidsaspekt

Hebb

Donald Hebb's postulat (1949)

The Organization of Behavior

A NEUROPSYCHOLOGICAL THEORY

¹⁹¹³
D. O. HEBB
¹⁹²⁷
McGill University

1949

New York · JOHN WILEY & SONS, Inc.
London · CHAPMAN & HALL, Limited

A NEUROPHYSIOLOGICAL POSTULATE

Let us assume then that the persistence or repetition of a reverberatory activity (or "trace") tends to induce lasting cellular changes that add to its stability. The assumption* can be precisely stated as follows:

When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased.

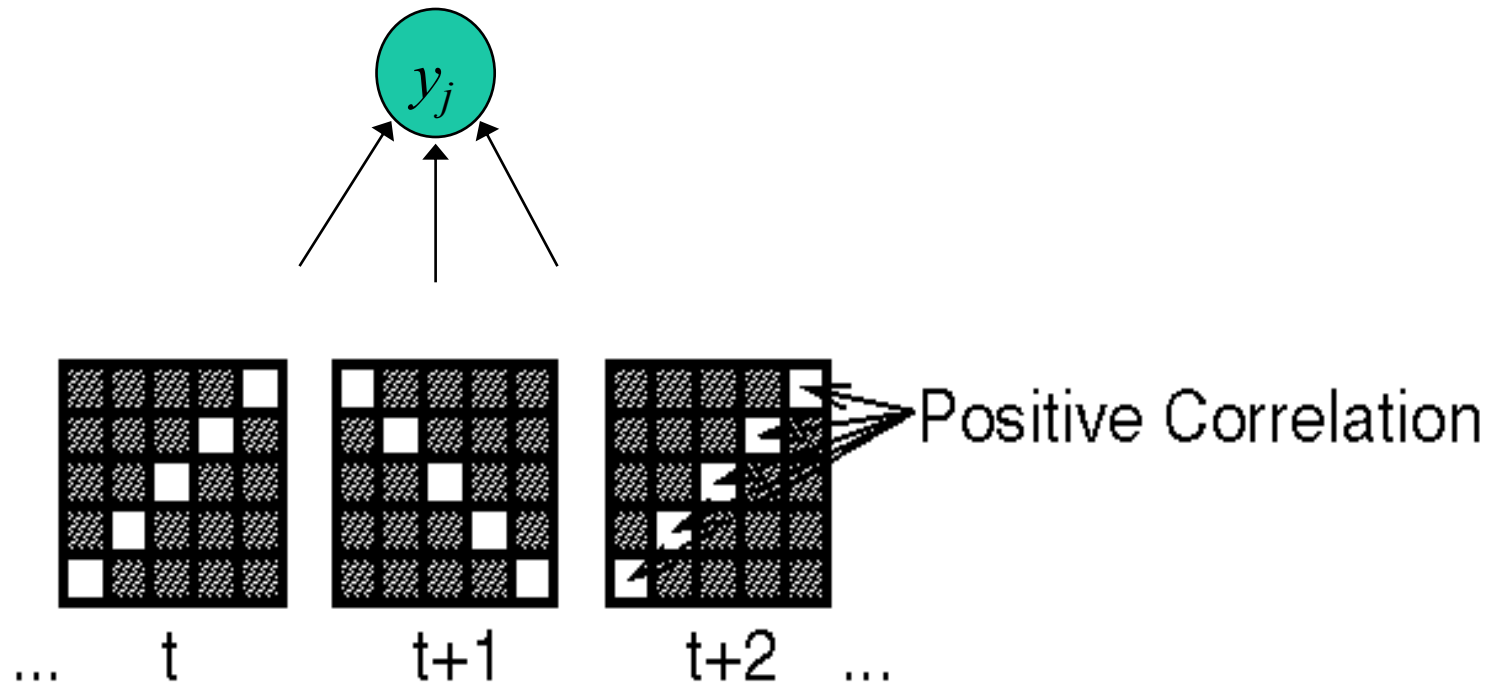
The most obvious and I believe much the most probable suggestion concerning the way in which one cell could become more capable of firing another is that synaptic knobs develop and increase the area of contact between the afferent axon and efferent soma. ("Soma" refers to dendrites and body, or all of

* See p. 229 for a further discussion of this point and an elaboration of the assumption made concerning the nature of memory.

Mål med oövervakad inlärning

- Modellinlärning = forma en intern modell av yttre verkligheten på basis av regelbundenheter i input
- Nätet gör detta utan lärare
- Extraherar återkommande mönster i indata
 - Olika delar av input som brukar förekomma tillsammans -> särdrag (features)

Korrelation

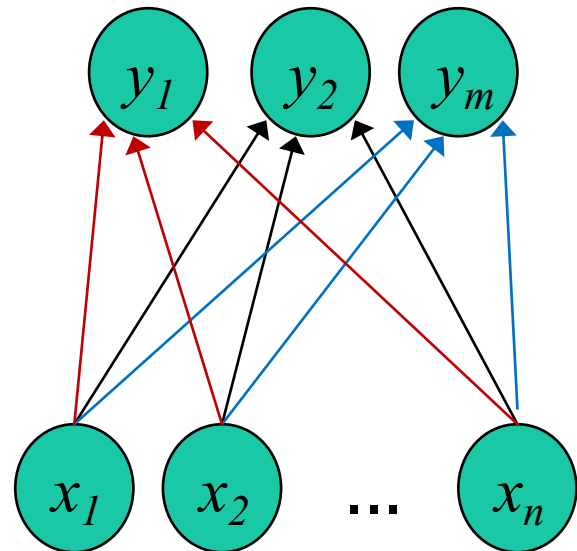


Korrelation

- Det är x_1, \dots, x_n via sina resp. vikter som tillsammans bestämmer y 's aktivering
 - Korrelationen mellan x_1, \dots, x_n lärs in av y
 - y kommer att till slut ha starka vikter till de pixlar som tillhör ena linjen

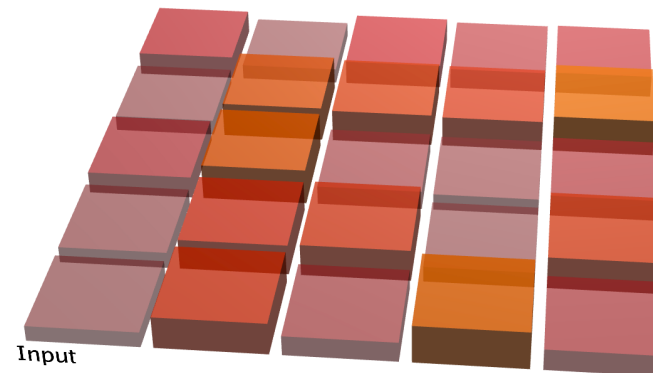
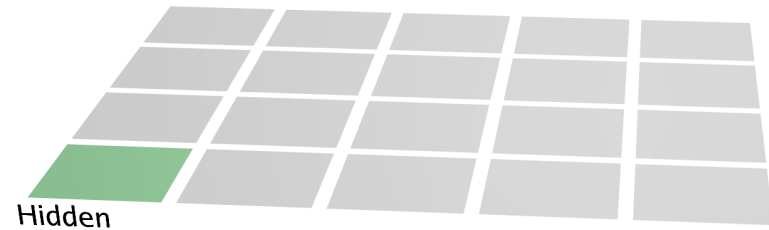
Vill ha specialiserade mottagar-enheter

- Vill att olika noder ska lära sig olika särdrag (t.ex. linjer)
- Vill/kan inte tala om vilka som ska lära sig vilka särdrag



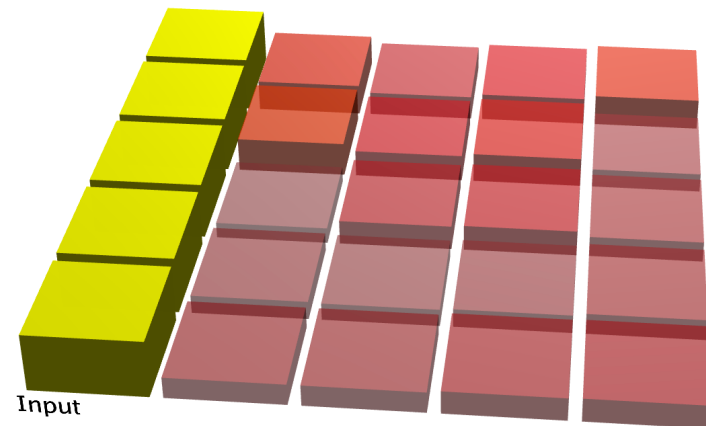
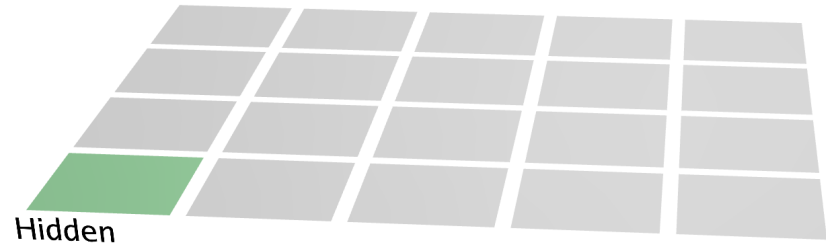
Självorganiserande nät

- Starta med små slumpvikter
- Olika enheter får då olika chans att bli aktiverade av samma input



Självorganiserande nät

- Aktiva enheter får öka sina vikter mer
- Självförstärkande effekt
 - Dessa enheter ännu lättare att aktiveras av denna input nästa gång



XCAL

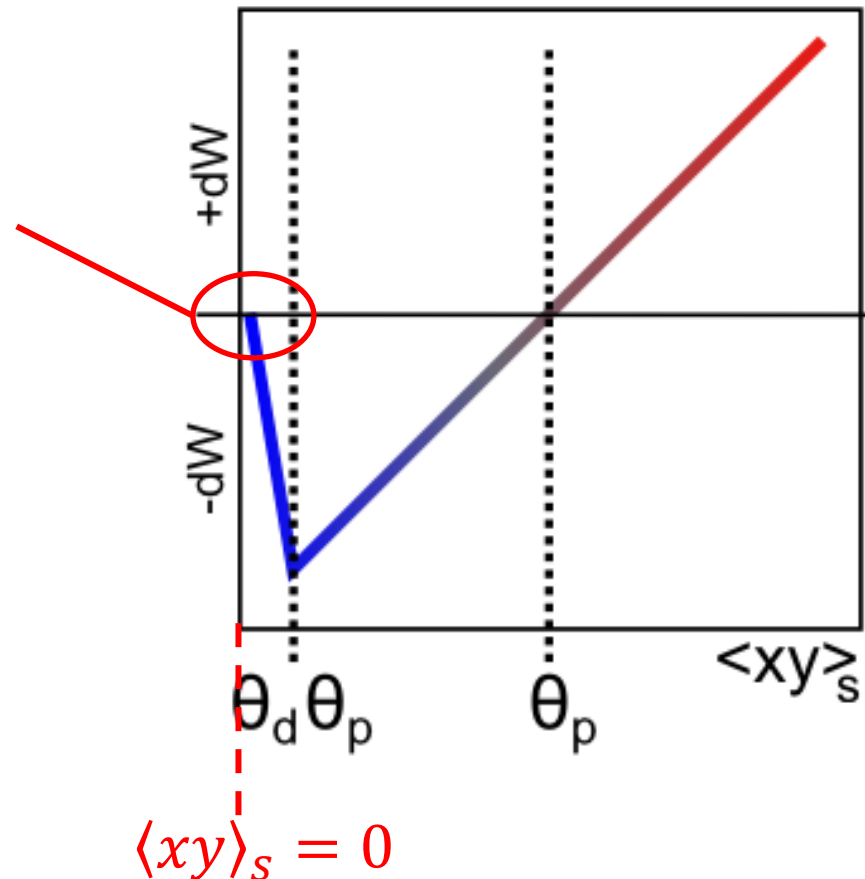
eXtended Contrastive Attractor Learning

XCAL: viktuppdatering för mottagarnoden y

- $\Delta w = f_{xcal}(xy, \theta_p) = \begin{cases} (xy - \theta_p), & xy > \theta_d \theta_p \\ -xy(1 - \theta_d)/\theta_d \end{cases}$
- Argumenten säger att viktuppdateringen är beroende av
 1. Sändarens aktivering * mottagarens aktivering
 2. Tröskeln 'täta' med index p = ('potentiation')
 - **LTP**: θ_p tröskeln säger vid vilken punkt i XCAL viktuppdateringen ska vara positiv

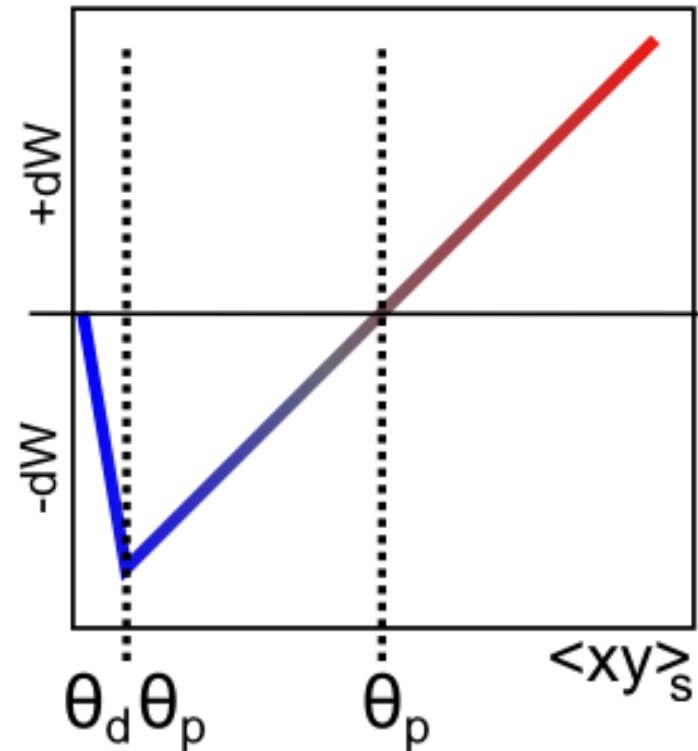
XCAL: kräver att båda noderna är aktiva

- $\langle xy \rangle_s = 0$ om x eller y är 0
 - dW blir då också 0
- Samaktivering krävs för inlärning
 - Vi är dock inte så noga med kausaliteten i emergent (jfr. STDP i hjärnan)



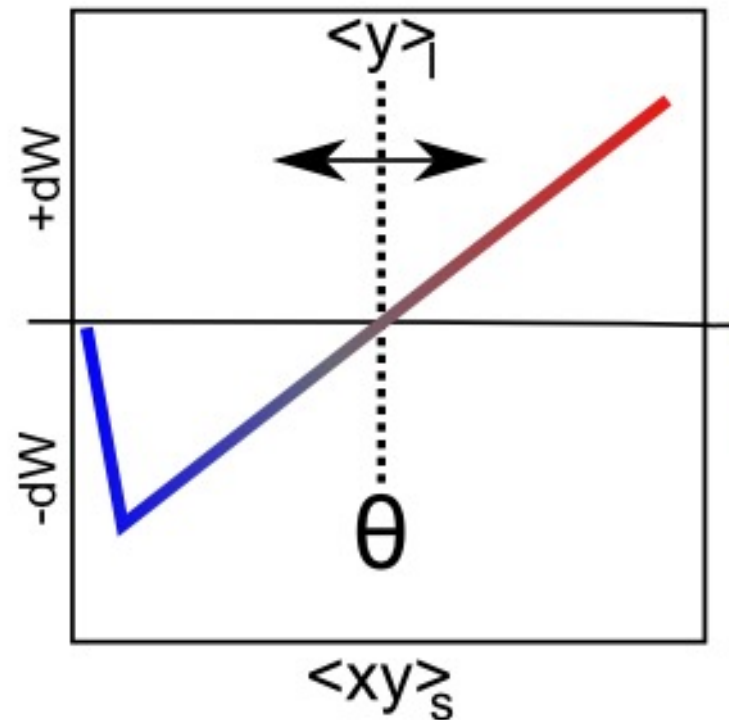
XCAL: olika trösklar

- LTP: θ_p
 - Om $\langle xy \rangle_s > \theta_p$, viktökning
 - Annars minskning
- LTD: θ_d
 - ”Knycken” i början
 - Vid θ_d -procent av θ_p



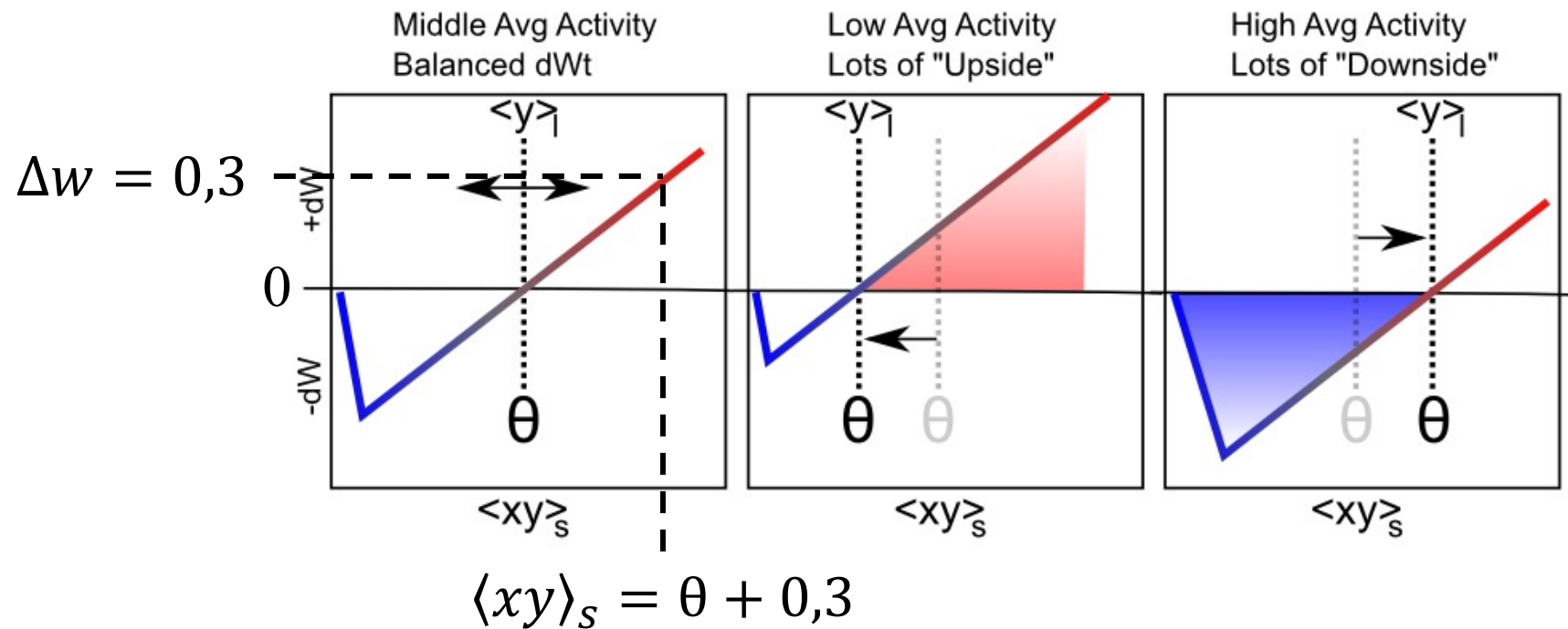
θ_p bestäms av $\langle y \rangle_l$

- Snitt aktivitet för mottagaren över längre tid
- Om $\langle xy \rangle_s > \langle y \rangle_l$, viktökning
 - Om högre aktivering än under föregående input:ar, öka vikt



Hur läsa av Δw i XCAL

- $\Delta w = \langle xy \rangle_s - \langle y \rangle_l$



Beräkning av tröskeln θ_p

- Vill hålla tröskeln inom min-max gränser $[0, 1]$
 - Tröskeln ska närma sig mjukt, dvs. mindre justering ju närmare min- eller max-gränsen

$$\bullet \theta_p = \begin{cases} y_l + \frac{1}{T_l} (\max - y_l) & , y > 0.2 \\ y_l + \frac{1}{T_l} (\min - y_l) & , y \leq 0.2 \end{cases}$$

T_l : hur brant θ_p ska närma sig golv eller tak

Running average

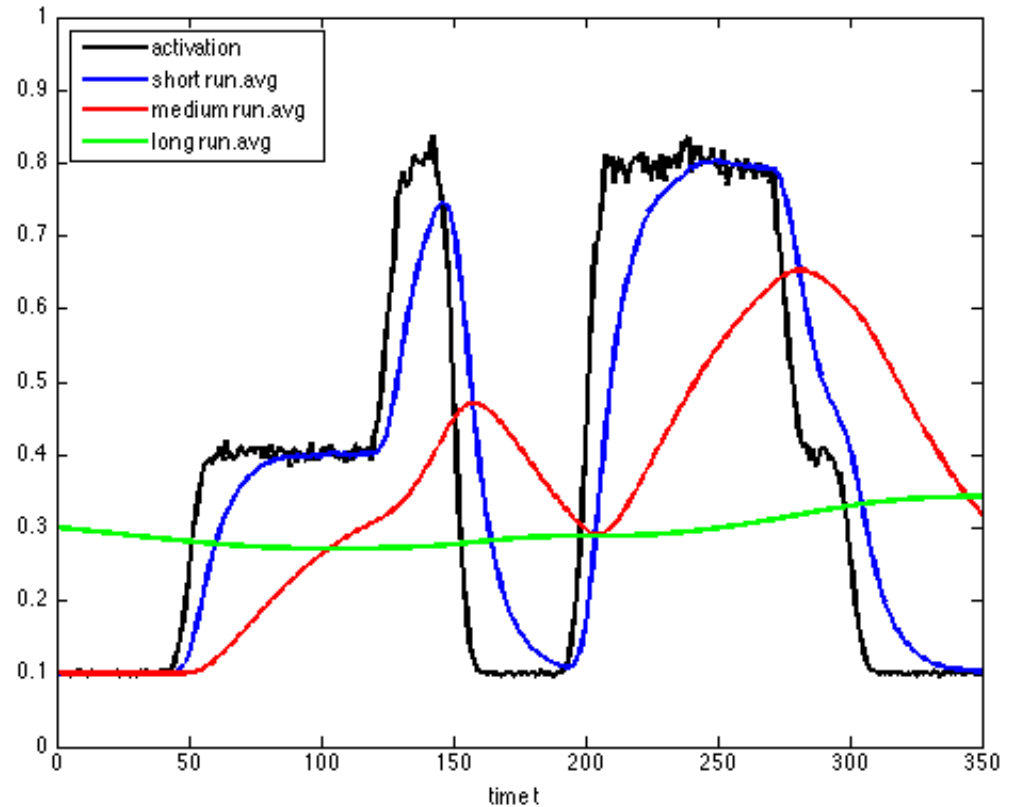
Förenklad syntax för $\langle x \rangle_{SS}$

$$x_{SS} = x_{SS} + dt_{SS}(x - x_{SS})$$

$$x_s = x_s + dt_s(x_{SS} - x_s)$$

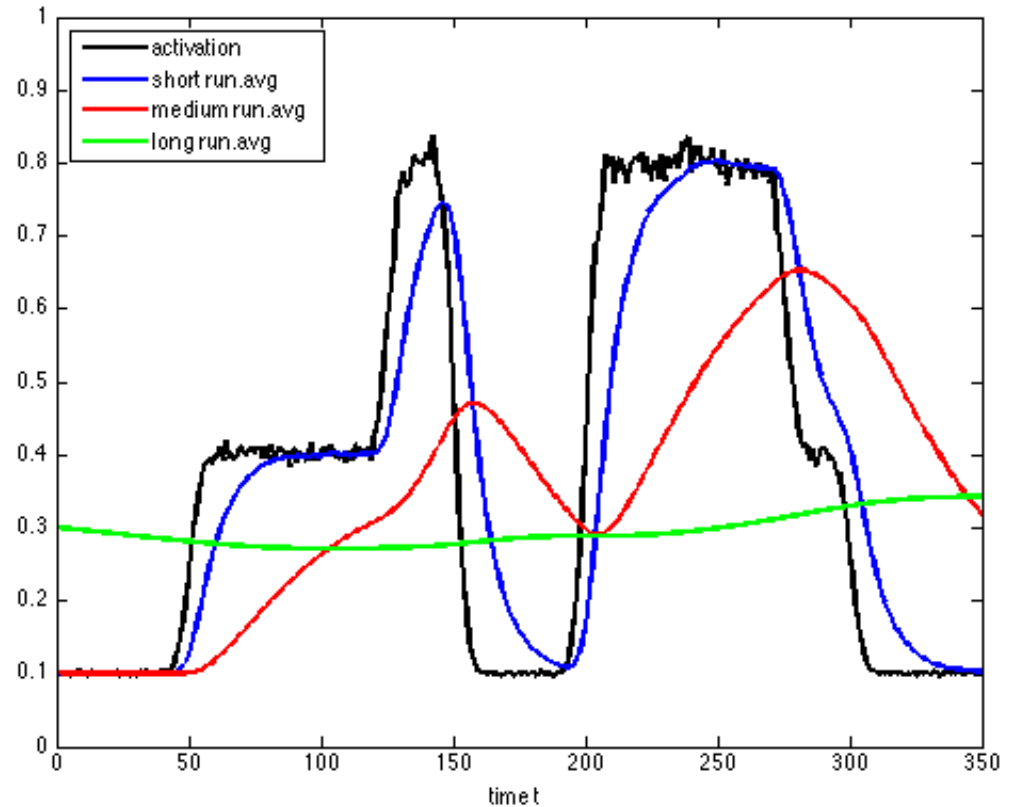
$$x_m = x_m + dt_m(x_s - x_m)$$

$$x_l = x_l + dt_l(x_m - x_l)$$



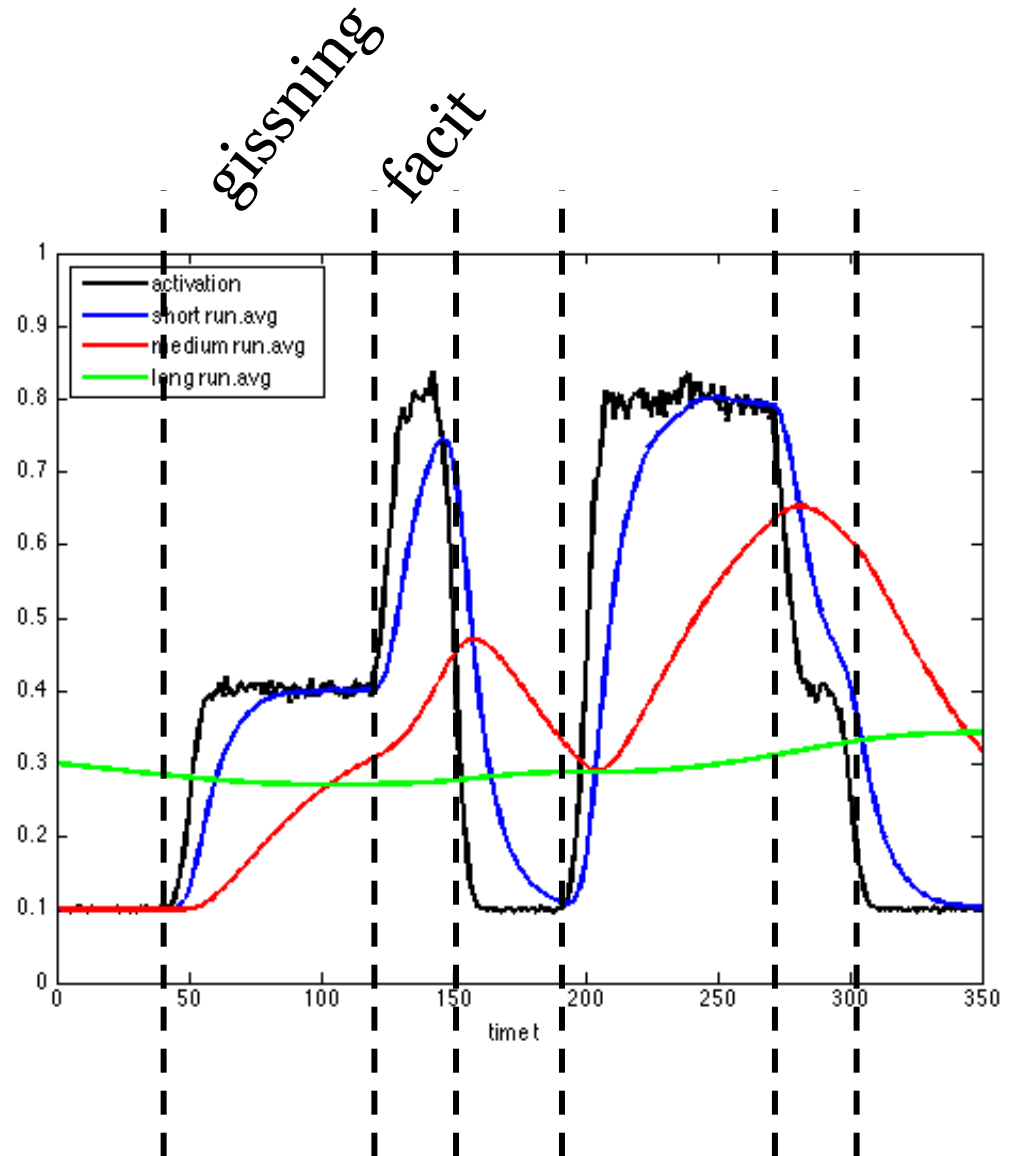
Cycle ~ 1 ms

- Ett steg simulering
 - Super-short term aktivering



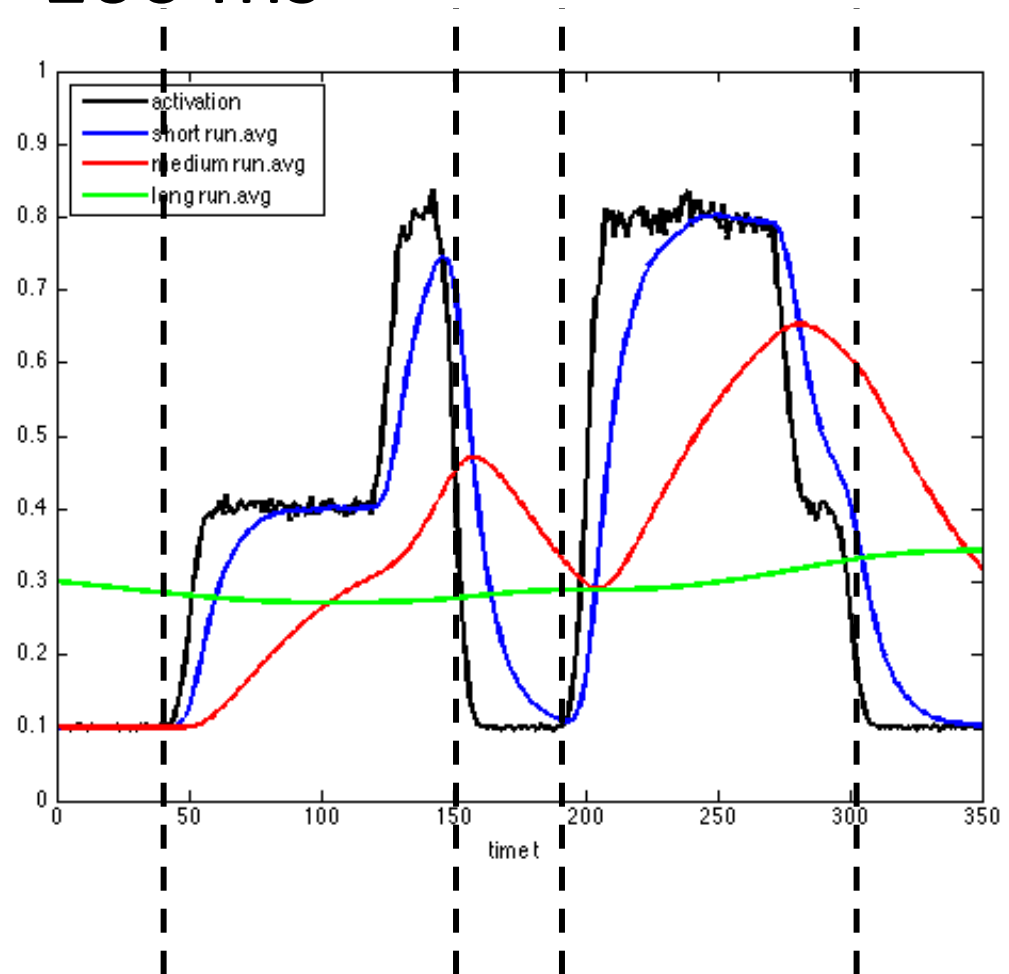
Quarter ~ 25 ms

- 3 quarters = ”gissningsfas”
 - Nätet får producera sin output
- 1 quarter = facit matas top-down, tillsammans med input



Trial = 4 quarters ~ 100 ms

- Motsvarar ett event
- En input bearbetas



Epok

- Kör igenom alla events (alla input) i data
 - Ofta i slumporder

Träning

- Kör många epoker
 - Varje epok med alla events i ny slumporder

Test

- Kör en epok med testdata
 - Ingen viktuppdatering
- Testdata är data som har lagts undan
 - Har inte använts under träning

Batch (i emergent)

- Kör flera träningar
- Ta snitt, så att man får tillförlitliga resultat

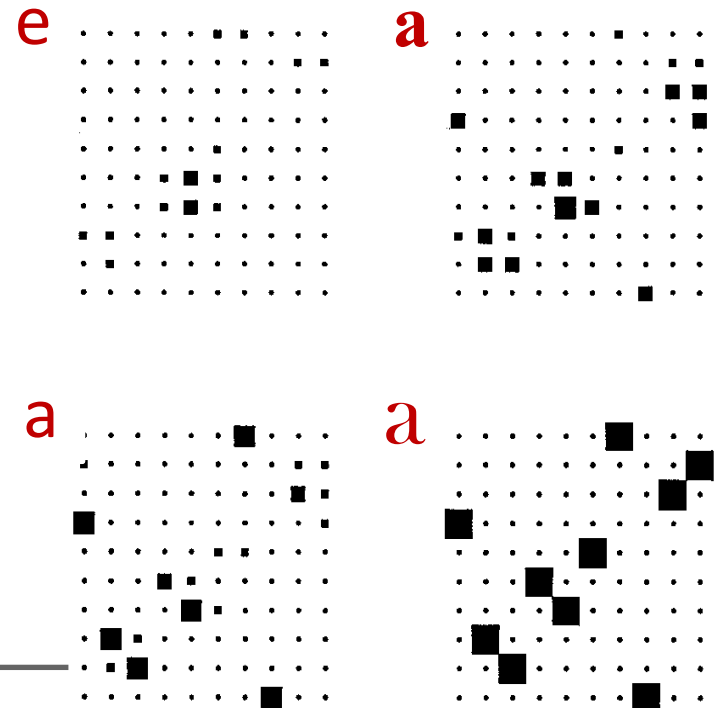
Generalisering

Generalisering

- Hela målet med träning av nät
 - Att kunna känna igen ny input
- Aldrig sett förut
 - Måste baseras på det nätet sett
 - Likhet mellan dessa och nya inputen

Distribuerad repr. för generalisering

- Överlapp mellan aktiveringsmönster repr. likhet
- Kan då reagera adekvat på ny input

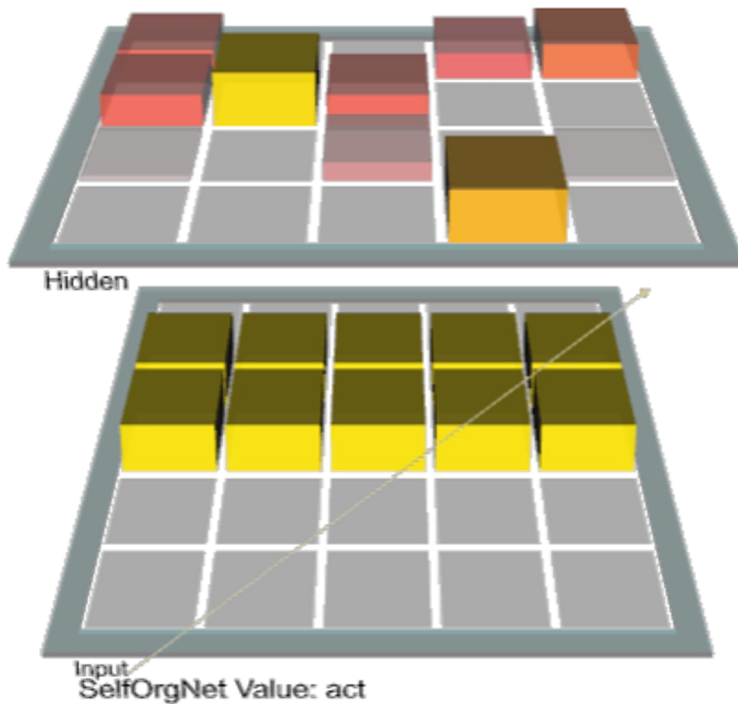


Har aldrig sett just detta 'a', men kommer att känna igen den som a

Tät (dense) – gles (sparse) aktivering

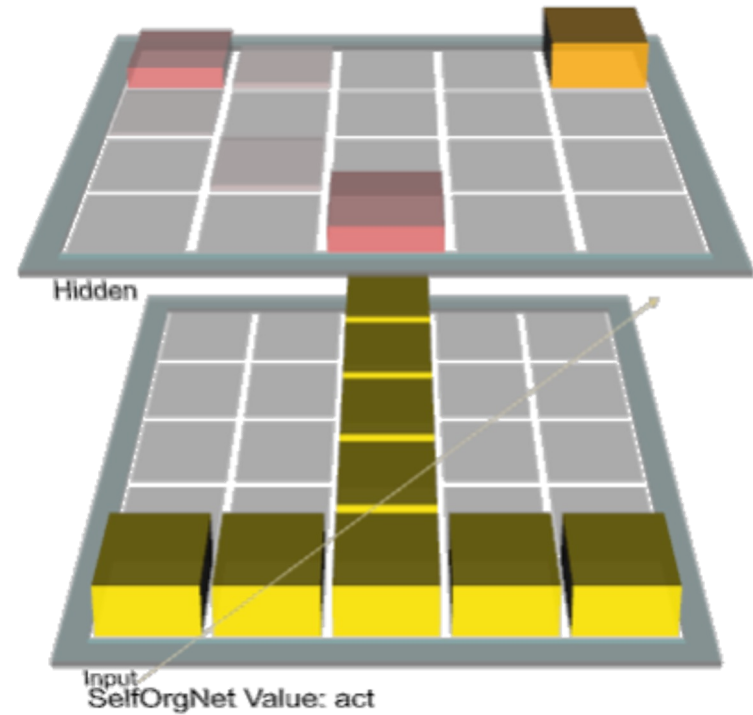
36% aktiva → 2042975 repr.

Fler möjliga representationer



4% aktiva → $25 \cdot 24 / 1 \cdot 2 = 300$ repr.

Mindre sannolikhet för överlapp



Överlapp dock inte alltid önskvärt

- Relaterar till labbfråga
 - (där efterfrågas fall när man vill ha mer eller mindre överlapp)
- Ibland vill man inte ha överlapp alls
 - Tänk t.ex. olika former av minne hos människan...?

rita.kovordanyi@liu.se

www.liu.se