Statisztikus tanulás az idegrendszerben

ORBÁN GERGŐ

www.eng.cam.ac.uk/go223

Somogyvári Zoltán

Bányai Mihály

Orbán Gergő

1. Bevezetés - G
2. Perceptron, előrecsatolt hálózatok
3. Rekurrens hálózatok, a Hopfield hálózat
4. Rejtett változós modellek
5. Reprezentációs tanulás
6. Eloszlások tanulása, a Boltzmann-gép
7. MAP paraméterbecslés, bayesi modelösszehasonlítás
8. Az EM-algoritmus, keverékmodellek
9. Az EM speciális esetei
10. PCA, ICA, divisive normalisation
11. Bayes nets, Helmholtz machine
12. DBN, kontrasztiv divergencia
13. Sampling

Unsupervised learning

Input: $x_1, x_2, \ldots, x_t$ összefoglaló néven: adat - vizuális, auditoros, szöveg

Gól: $\mathrm{P}(\mathbf{x})$

(Reinforcement learning:

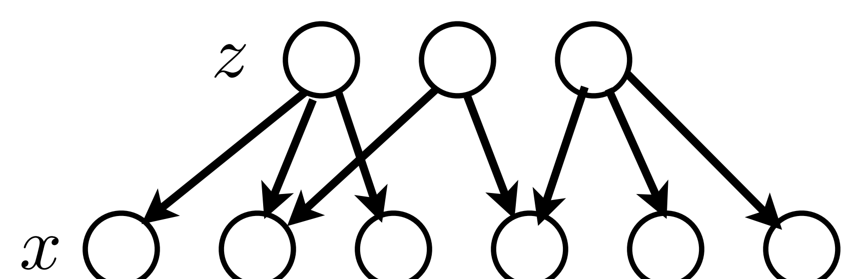
Input: $\{x_1, y_1\}, \{x_2, y_2\}, \ldots, \{x_t, y_t\}$

Gól: $\mathrm{P}(\mathbf{x} \,|\, \mathbf{y})$)

$\mathrm{P}(\mathbf{x})$ Bonyolult!

Miért is?

Egyszerűsítés: $\mathrm{P}(\mathbf{x}) = \mathrm{P}(\mathbf{x} \,|\, \mathbf{z})\, \mathrm{P}(\mathbf{z})$



- az adatot a “z”-k terében reprezentáljuk
- kategorizáció, dimenzió redukció
- általánosabban a feladat: predikció, döntéshozatal, kommunikáció

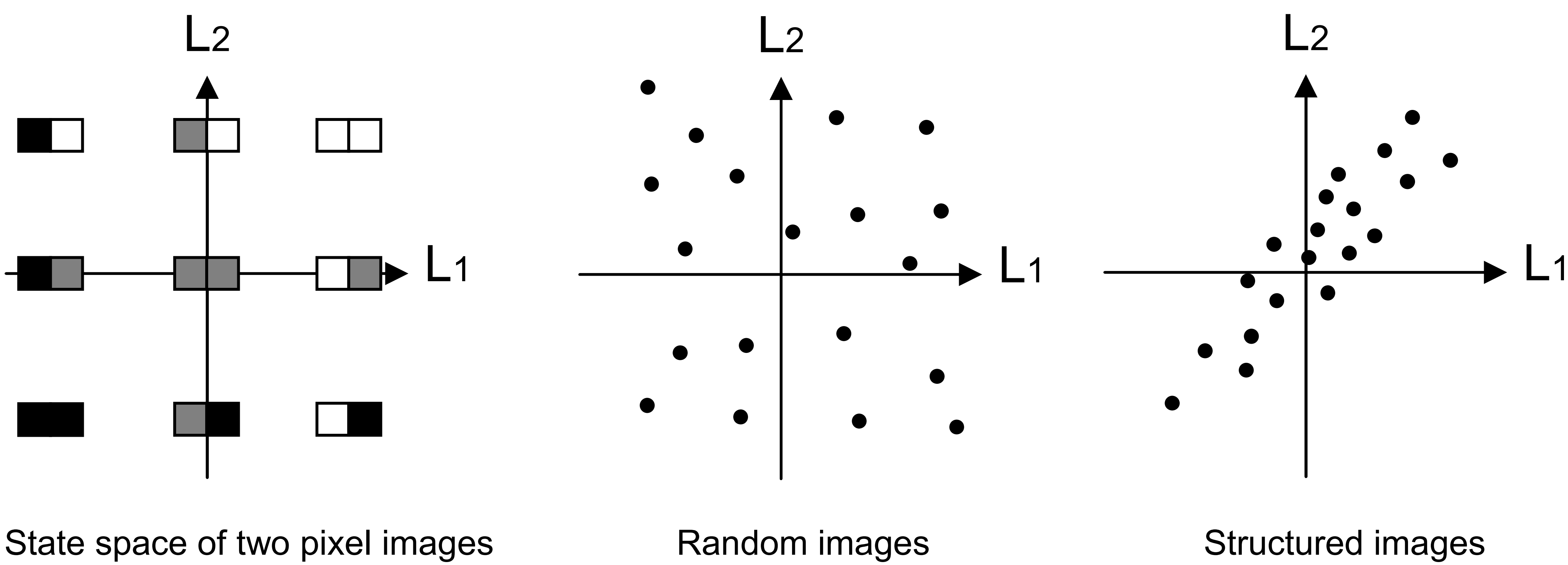
Lineáris modellek

$$\mathrm{P}\left(x \,|\, z\right) = \mathrm{Normal}\left(x; z, \theta\right) = C \exp\left(\left(x - Az\right)^{T} \Sigma^{-1} \left(x - Az\right)\right)$$

$$x = \mathbf{A} \cdot z + \epsilon$$

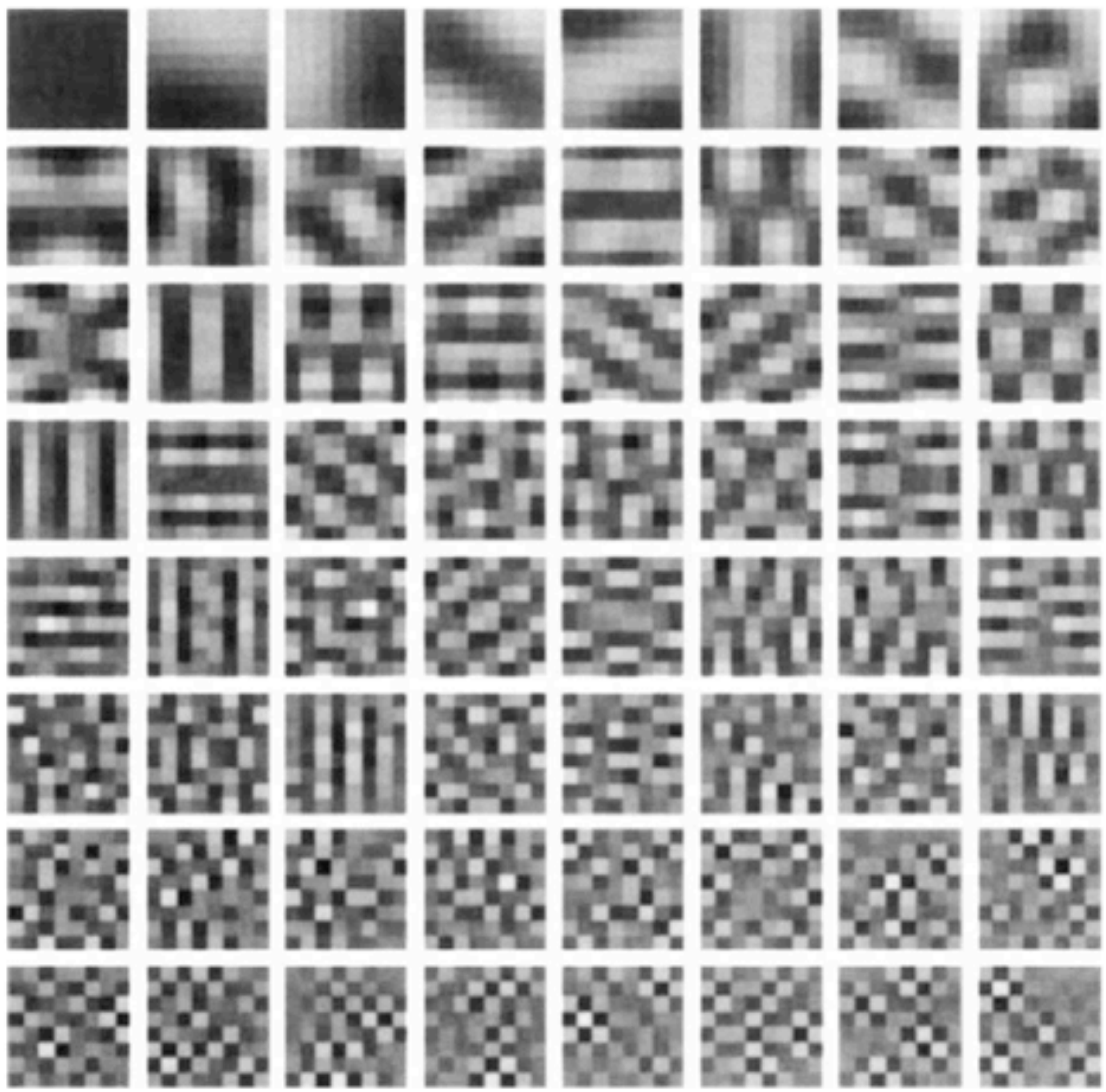
PCA

- A oszlopvektorai ortogonalisak
- D(x) = D(z)
- Izotróp zaj



PCA tulajdonságok

- Kompakt kódot eredményez
- Egy adatponért leírásáért általában a teljes hálózat felel



Sparse kódolás, ICA

$$x = \mathbf{A} \cdot z + \epsilon$$

- “z”-k függetlenek
- y priorja “ritka”(P(z))

Komputációs kritériumok:

- Hiteles rekonstrukció
  költség egy adatpontra (képre):

$$\mathrm{cost}_1 = \left(x - \sum_i A'_i \cdot z_i \right)^2$$

- Kis “energiafelhasználás (kevés szimultán aktiv neuron)
  további költség a kód “ritkasága”:

$$\mathrm{cost}_2 = -\sum_i S\left(\frac{z_i}{\sigma}\right)$$

  S a Gauss-nál nagyobb kurtózissal bíró eloszlás

- teljes költség (~energia):

$$E = -\mathrm{cost}_1 - \lambda \mathrm{cost}_2$$

Sparse kód tanulása: E-M

Algoritmus:

- Itáráció EM lépésekkel
- Random kezdeti feltételek
- Adott konnektivitási mátrixnál az aktiviások segítségével a költség minimalizálása
- Adott aktivitásokkal a költség minimalizálása a súlyok adaptálásával

Adott konnektivitási mátrix esetén a legjobb aktivitások megtalalása:

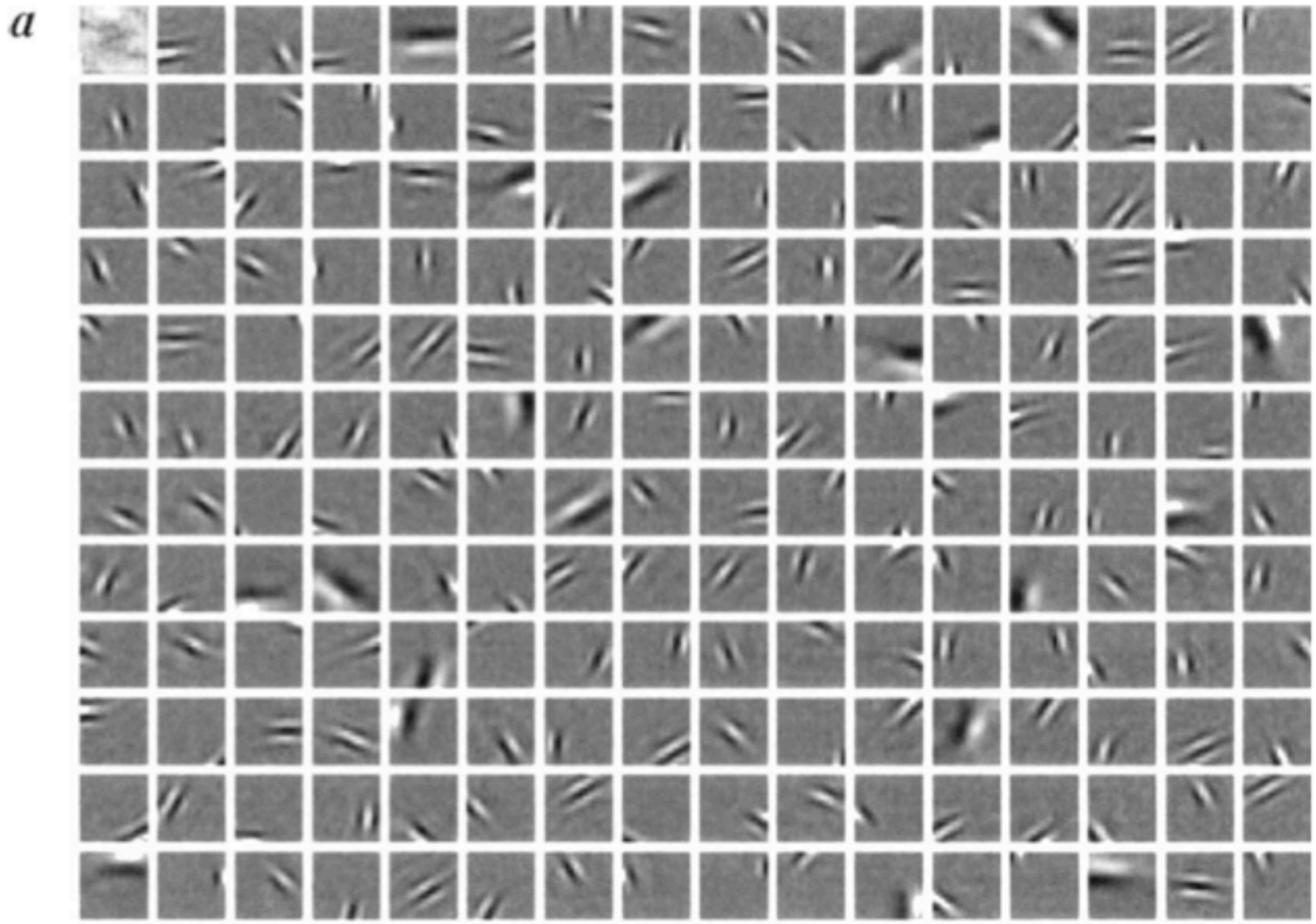
$$\dot{z}_i = \mathbf{A}_i x_t - \sum_j \mathbf{A}_i' \mathbf{A}_j z_j - \frac{\lambda}{\sigma} S' \left(\frac{z_i}{\sigma} \right)$$

Adott konnektivitási aktivációk esetén a legjobb súlyok megtalalása:

$$\Delta A_i = \eta \left\langle a_i \left[x - \hat{x} \right] \right\rangle_t$$

Sparse kódolás: eredmény

tréningezés természetes képekkel

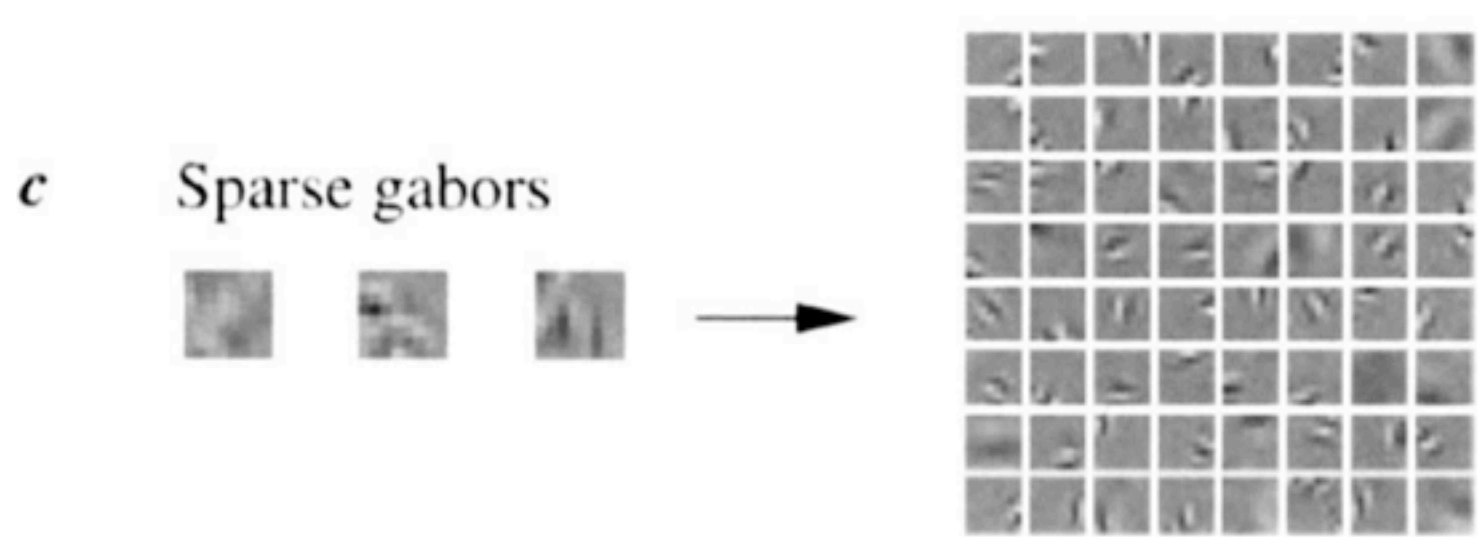


Olshausen & Field '96

A kialakult bázis:

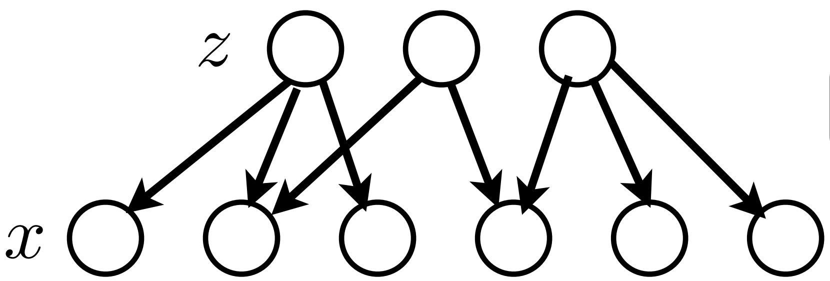
- irányított
- térbeli sávszűrést valósít meg
- lokalizált

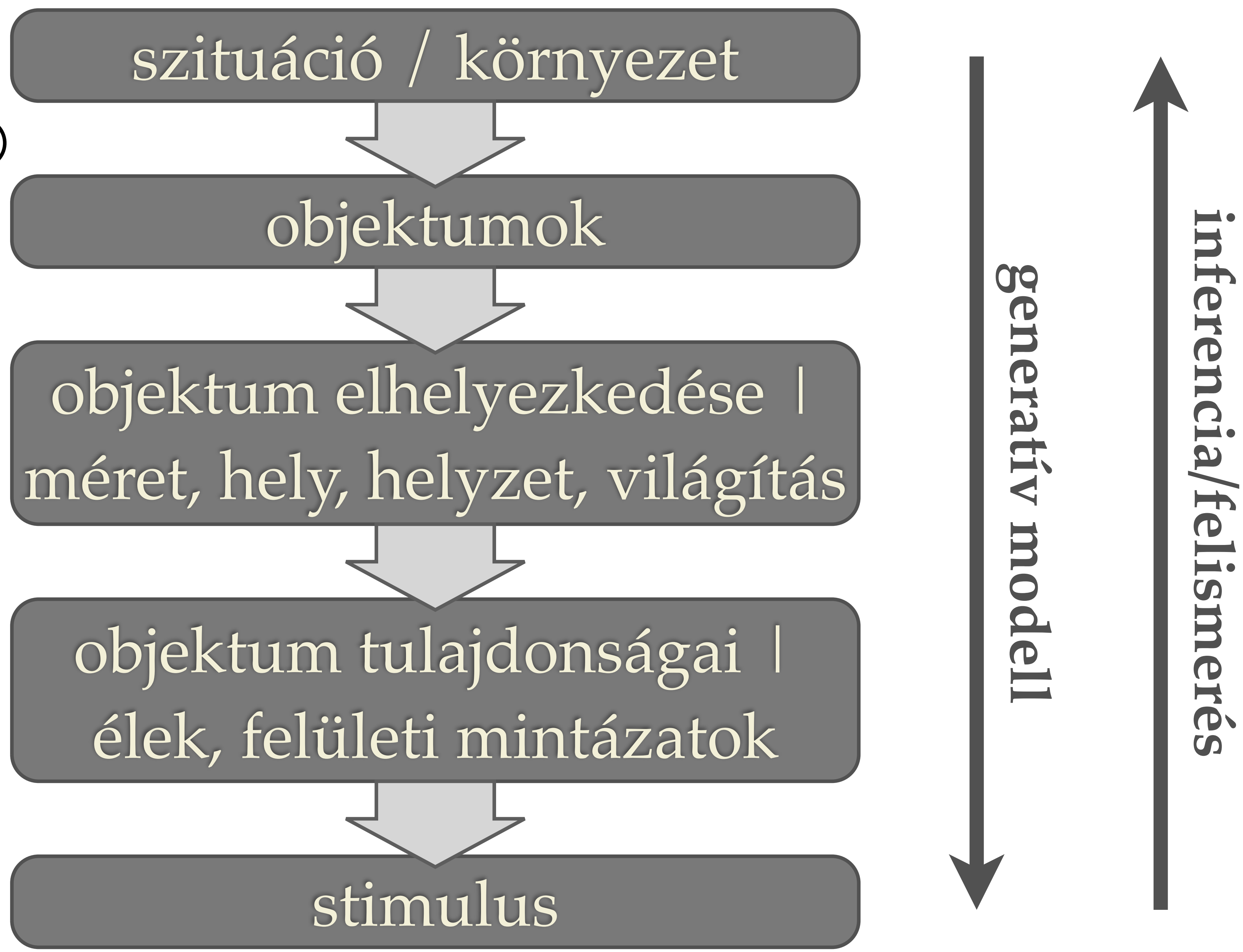
Tanulás és stimulus statisztika



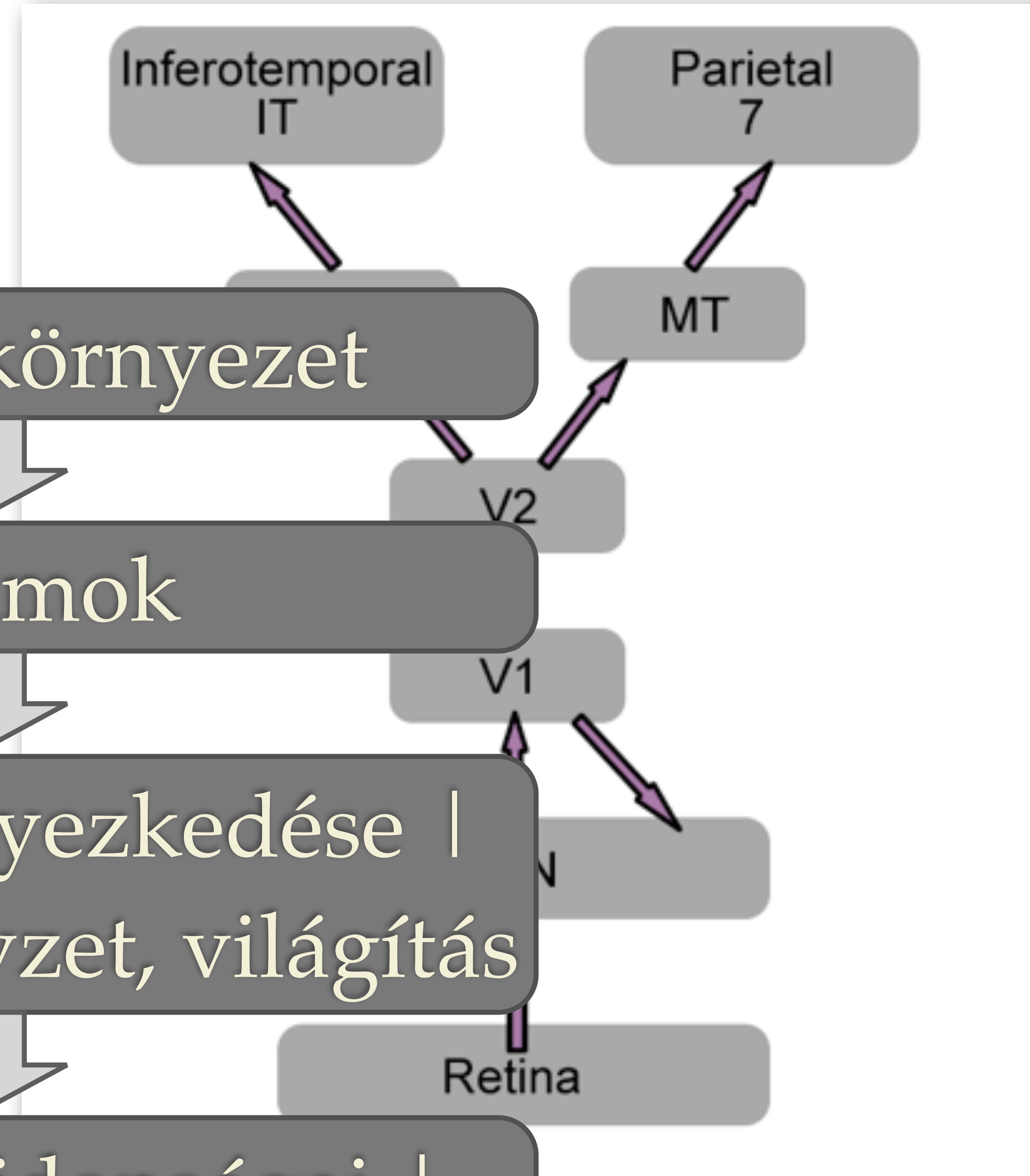
Generatív/rekogniciós modell

$$\mathrm{P}(\mathbf{x}) = \mathrm{P}(\mathbf{x} \,|\, \mathbf{z})\, \mathrm{P}(\mathbf{z})$$





Generatív/rekogniciós modell



szituáció / környezet

objektumok

objektum elhelyezkedése | méret, hely, helyzet, világítás

objektum tulajdonságai | élek, felületi mintázatok

stimulus

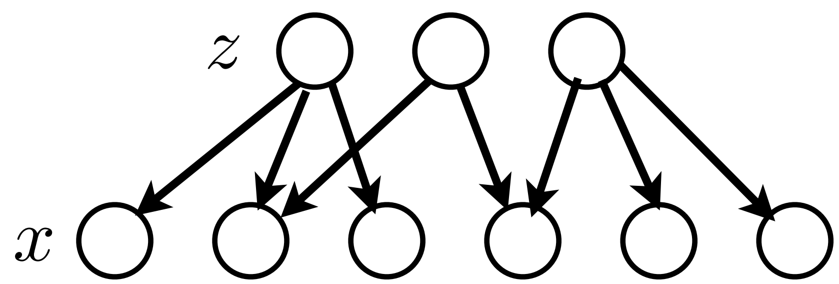
Modell definíció -> rekogníció:

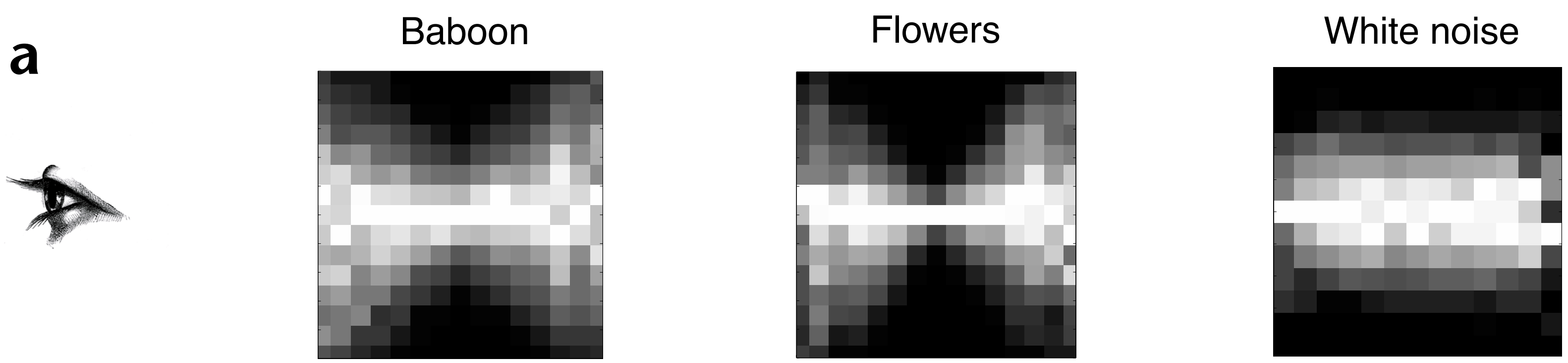
$$\mathrm{P}\left(x\,|z\right)$$

Inferencia igénye -> rekogníció:

$$\mathrm{P}\left(z\,|x\right)$$

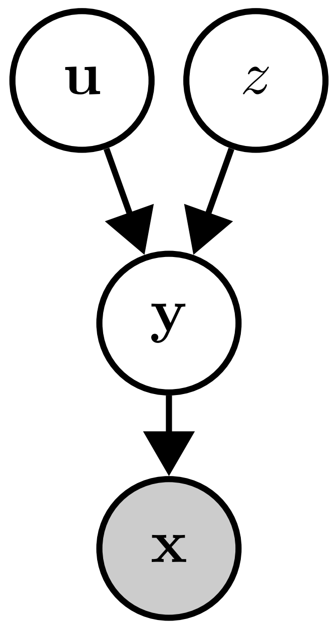
Independens komponensek



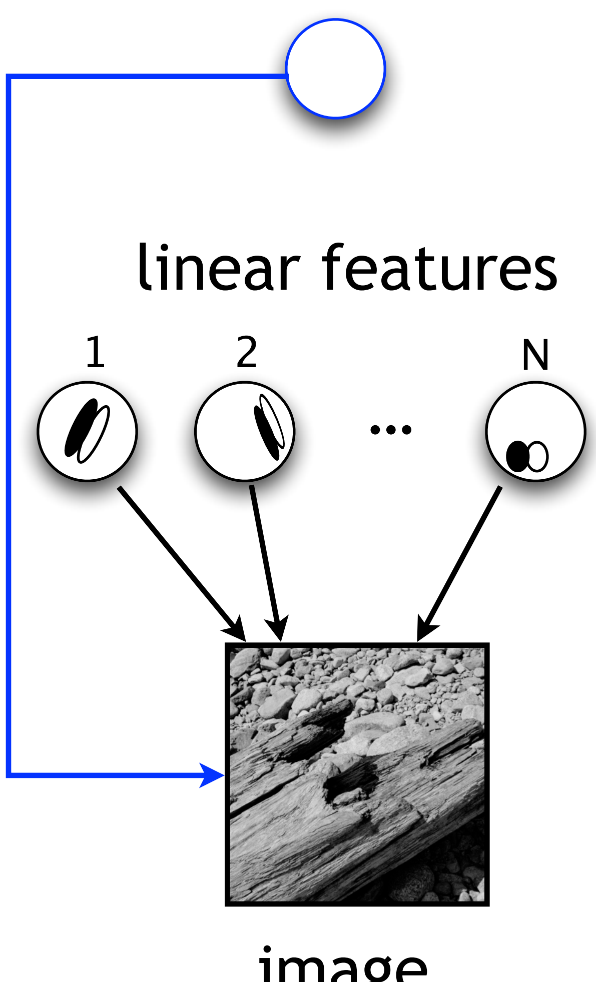


Schwartz & Simoncelli, 2001

Gaussian Scale Mixtures



$$
\begin{aligned}
\mathrm{P}(\mathbf{x}|\mathbf{y}) &= \mathcal{N}\left(\mathbf{x}; \mathbf{A}\mathbf{y}, \sigma_{\mathrm{x}}^2 \mathbf{I}\right) \\
\mathbf{y} &= z\,\mathbf{u} \\
\mathrm{P}(\mathbf{u}) &= \mathcal{N}(\mathbf{u}; \mathbf{0}, \mathbf{C}) \\
\mathrm{P}(z) &= \mathrm{Gamma}(z; k, \theta)
\end{aligned}
$$



$$\text{image} = \text{contrast} \times \left(a_1\, \text{feature}_1 + a_2\, \text{feature}_2 + \ldots + a_N\, \text{feature}_N + \text{noise} \right)$$

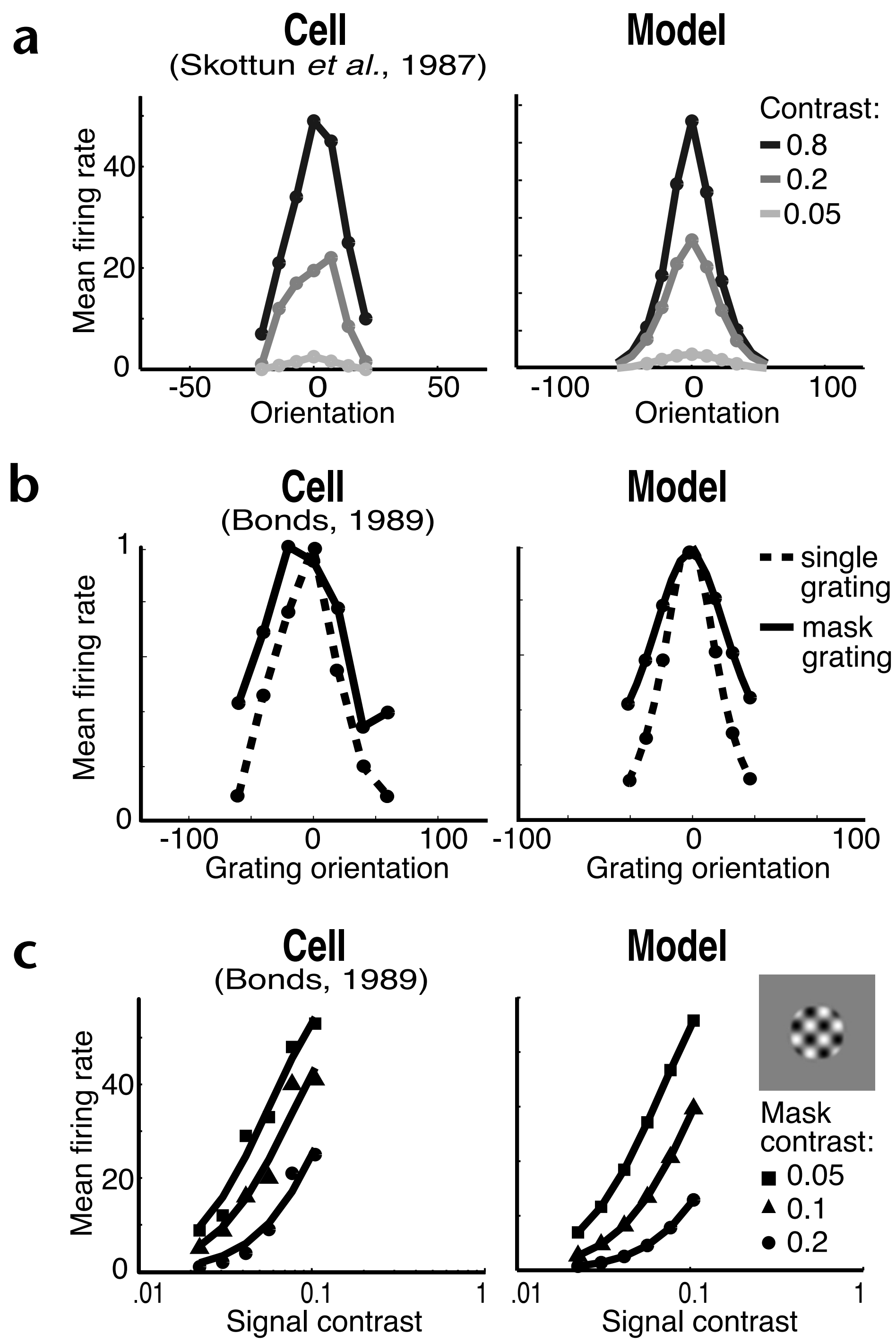
$$var\,(L_1 | L_2) = wL_2^2 + \sigma^2$$

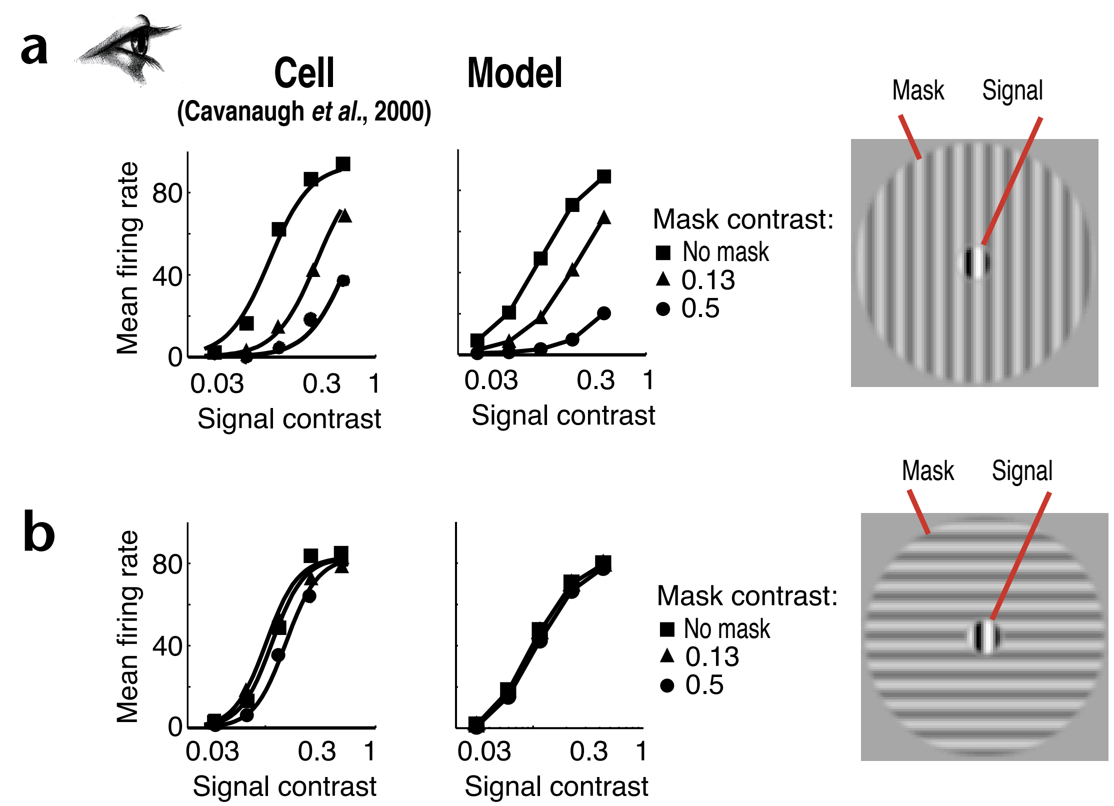
$$\mathrm{R}_1 = \frac{L_1^2}{wL_2^2 + \sigma^2}$$

$$var\,(L_i | \{L_j, j \in N_i\}) = \sum \mathrm{w}_{ji} \mathrm{L}_j^2 + \sigma^2$$

$$R_i = \frac{L_i^2}{\sum_j \mathrm{w}_{ji} \mathrm{L}_j^2 + \sigma^2}$$

Neurális adatok és GSM





Schwartz & Simoncelli, 2001