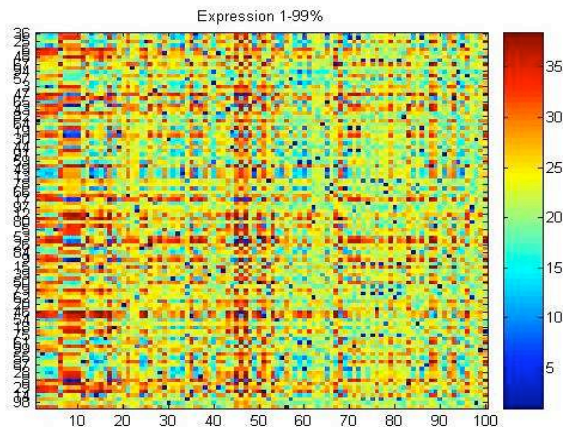


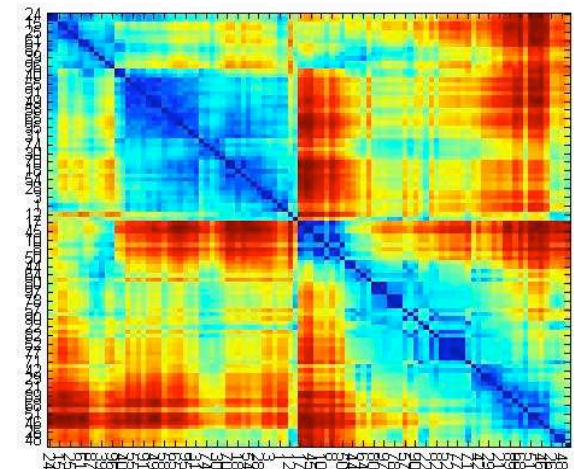
Rio de Janeiro 17/11/08

Física e Informação

Nestor Caticha



IF USP
17/11/ 2008



Sumário

Sistemas que processam informação

- Lógica Aristotélica, extensões
- Informação Completa vs Incompleta
- Qual é a estrutura matemática adequada?

Probabilidades

- racionalidade e informação incompleta
- Distribuição *a priori*

Atribuição e Atualização de probabilidades

- Teorema de Bayes
- Entropia

Aplicações

- Classificação
- Categorização

Sistemas que processam informação

- **Lógica Aristotélica**
- Informação Completa vs Incompleta
- Qual é a estrutura matemática adequada?

A, B : Asserções

Podem ser verdadeiras ou falsas

Como decidir?

Qual deve ser a nossa crença?

CRENÇA?



We all start from naive realism, i.e., the doctrine that things are what they seem. We think that grass is green, that stones are hard, and that snow is cold. But physics assures us that the greenness of grass, the hardness of stones, and the coldness of snow are not the greenness, hardness, and coldness that we know in our own experience, but something very different. The observer, when he seems to himself to be observing a stone, is really, if physics is to be believed, observing the effects of the stone upon himself.'

Einstein



'Finally, the reformed subjectivist principle must be repeated: that apart from the experiences of subjects there is nothing, nothing, nothing, bare nothingness.'

Whitehead



Sistemas que processam informação

- **Lógica Aristotélica**
- Informação Completa vs Incompleta
- Qual é a estrutura matemática adequada?

A, B : Asserções

Podem ser verdadeiras ou falsas

Como decidir?

Qual deve ser a nossa crença?

CRENÇA? Verdadeira ou falsa?

Sistemas que processam informação

- Lógica Aristotélica
- Informação Completa vs Incompleta

SILOGISMO

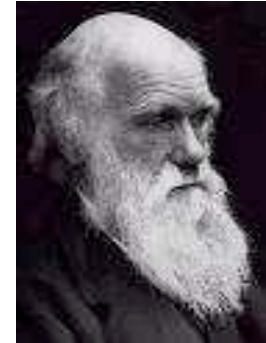
C: $A \longrightarrow B$

A implica B

Informação : C e A são VERDADEIROS
Como decidir?

CRENÇA? B deve ser verdadeiro

Realismo e Evolução Natural



evolução natural
leva a sistemas que processam informação
que acreditam que

B deve ser verdadeiro?

Sistemas que processam informação

- Lógica Aristotélica
- Informação Completa vs Incompleta

SILOGISMO

C: $A \longrightarrow B$

A implica B

Informação : C e B são VERDADEIROS
Como decidir?

Não podemos concluir nada?

Exemplo

CONTEXTO

A: 'CHOVE'

B: 'HÁ NUVENS'

C: 'A implica B'

Informação Completa

dados: C e A

B deve ser verdadeiro

Exemplo

CONTEXTO

A: 'CHOVE'

B: 'HÁ NUVENS'

C: 'A implica B'

Informação **Incompleta**

C e B

Aristoteles: nada posso concluir

Exemplo

CONTEXTO

A: 'CHOVE'

B: 'HÁ NUVENS'

C: 'A implica B'

Informação **Incompleta**

Dados C e B

Aristóteles: nada posso concluir

Mamãe: leva o guarda chuva, pois parece que vai chover.

Bertrand Russell (jovem)

“induction is either disguised deduction or a mere method of making plausible guesses”.

Harold Jeffreys

“É muito melhor trocar a ordem dos dois termos e que muito do que normalmente passa por dedução é indução disfarçada, e que até alguns dos postulados de Principia Mathematica foram adotados por motivações indutivas” (e adiciona , são falsos).

Russell mudou de posição, dobrado pela evidência e diz no fim da sua autobiografia:

“I was troubled by scepticism and **unwillingly forced to the conclusion** that most of what passes for knowledge is open to reasonable doubt”.

Sobre indução disse ainda:

“The general principles of science, such as the belief of the reign of law, and the belief that every event must have a cause, are as completely dependent on the inductive principle as are the beliefs of daily life.”

(On Induction)



Devo ter

Diferente crença sobre **A** ?

- Se não sei nada sobre **B**

- Se sei que **B** é verdade



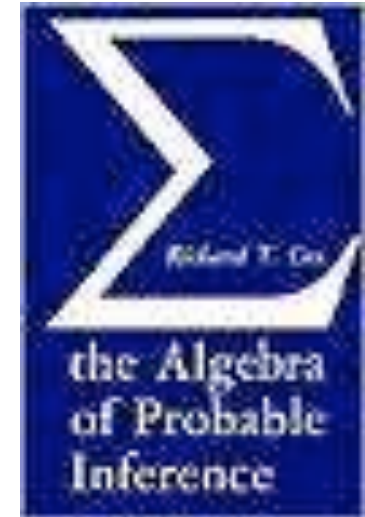
Devo ter

Diferente crença sobre **A**

- Diferente ?
- Qual é a matemática adequada?

R.T. COX 1946 Lógica Estendida

- Qual é a matemática adequada?
- Probabilidade? Não sei





R.T. COX 1946

Lógica Estendida

- Considera
Casos particulares Informação incompleta
Pessoas razoáveis concordam (?????)
 - Teoria geral existe deve funcionar em casos particulares
- Se há número muito grande, talvez sejam incompatíveis, talvez não.

R.T. COX 1946

Lógica Estendida

- Considera
Casos particulares Informação incompleta

1. Crença deve ser codificada em
Número Real

$(A|C)$ crença em A quando acredito em C
Por exemplo

$$(A|BC) \geq (A|C)$$

Que valores? **NÃO SEI.**

R.T. COX 1946

Lógica Estendida

- Considera
Casos particulares Informação incompleta

2. Atribuição de Número real às crenças
deve ser **CONSISTENTE**

Ou pelo menos não manifestamente inconsistente:

Se há mais de uma forma de atribuir um
número real a uma crença: **IGUAIS**

R.T. COX 1946

Lógica Estendida

- Considera
Casos particulares Informação incompleta

2)

Se há mais de uma forma de atribuir
um número real a uma crença

Operadores lógicos: e.g. Produto Lógico, soma Lógica

$$x = A | C,$$

$$y = B | C,$$

$$z = A | B C,$$

$$w = B | A C,$$

$$u = AB | C,$$

$$x' = \tilde{A} | C$$

R.T. COX 1946

Lógica Estendida

- Considera
Casos particulares Informação incompleta

3. Atribuição de Número real às crenças
de A e de sua negação \tilde{A} deve ser
relacionada

$$(A|C) = f(\tilde{A}|C)$$

Existe uma função f , mas não sabemos qual é

R.T. COX 1946

Lógica Estendida

- Considera
Casos particulares Informação incompleta

4. Atribuição de Número real às crenças
de asserções compostas deve ser
relacionada

$$(AB|C) = g(D)$$

Existe uma função g , mas não sabemos qual é

R.T. COX 1946

Lógica Estendida

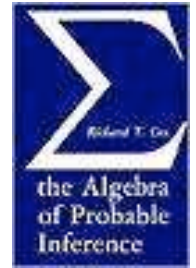
- Considera
Casos particulares Informação incompleta

4. Atribuição de Número real às crenças
de asserções compostas deve ser
relacionada

$$(AB|C) = g(D)$$

$$\begin{aligned}x &= A | C, \\y &= B | C, \\z &= A | B C \\w &= B | A C\end{aligned}$$

D: algum subconjunto
de $\{x, y, z, w\}$



- Teoremas de regradação de Cox

$P(A|C)$ número desconhecido⁴.

$P(A|C)$ função MONOTÔNICA desconhecida

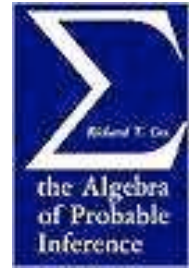
$$0 \leq P(A|C) \leq 1$$

$$P(A|C) = 1 - P(\tilde{A}|C)$$

Se $(A|C)$ é verdade $P(A|C)=1$, $P(\tilde{A}|C)=0$

$$\begin{aligned} P(AB|C) &= P(A|C) P(B|AC) \\ &= P(B|C) P(A|BC) \end{aligned}$$

“unwillingly forced to the conclusion”



- Teoremas de regradação de Cox

(A/C) número desconhecido

$P(A/C)$ função desconhecida

$$0 \leq P(A/C) \leq 1$$

$$P(A/C) = 1 - P(\tilde{A}/C)$$

Se (A/C) é verdade $P(A/C)=1$, $P(\tilde{A}/C)=0$

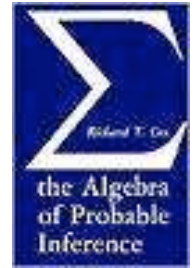
$$\begin{aligned} P(AB/C) &= P(A/C) P(B/AC) \\ &= P(B/C) P(A/BC) \end{aligned}$$

K
O
L
M
O
G
O
R
O
V

“unwillingly forced to the conclusion”

R.T. COX 1946

Lógica Estendida

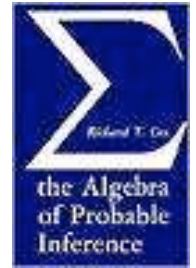


- Consistência

$$P(AB | C) = P(A | C)P(B | AC)$$

$$P(AB | C) = P(B | C)P(A | BC)$$

$$P(A | C)P(B | AC) = P(B | C)P(A | BC)$$



- Consistência

$$P(AB | C) = P(A | C)P(B | AC)$$

$$P(AB | C) = P(B | C)P(A | BC)$$

$$P(A | C) \boxed{P(B | AC)} = P(B | C) \boxed{P(A | BC)}$$

Teorema de Bayes



Bayes

$$P(B | AC) \propto P(A | BC)$$

Teorema de Bayes



Bayes

$$P(AB|C) = P(A|C)P(B|AC)$$

$$P(AB|C) = P(B|C)P(A|BC)$$

$$P(B|AC) = \frac{P(B|C)P(A|BC)}{P(A|C)}$$



Laplace



Laplace

$$P(AB|C) = P(A|C)P(B|AC)$$

$$P(AB|C) = P(B|C)P(A|BC)$$

~~Teorema de Bayes~~



Bayes

$$P(H|DC) = \frac{P(H|C)P(D|HC)}{P(D|C)}$$



Laplace

$$P(AB|C) = P(A|C)P(B|AC)$$

$$P(AB|C) = P(B|C)P(A|BC)$$

Inferência Bayesiana



Bayes

$$P(H|DC) = \frac{P(H|C)P(D|HC)}{P(D|C)}$$

H: hipótese

D: Dados

C: contexto da experiência



Laplace

$$P(AB|C) = P(A|C)P(B|AC)$$

$$P(AB|C) = P(B|C)P(A|BC)$$

Inferência Bayesiana



Bayes

$$P(H|DC) = \frac{P(H|C)P(D|HC)}{P(D|C)}$$

$P(H|C)$: Probabilidade a priori (não usa dados)

$P(D|HC)$:

$P(D|C)$:

$P(H|DC)$:



Laplace

$$P(AB|C) = P(A|C)P(B|AC)$$

$$P(AB|C) = P(B|C)P(A|BC)$$

Inferência Bayesiana



Bayes

$$P(H|DC) = \frac{P(H|C)P(D|HC)}{P(D|C)}$$

$P(H|C)$:

$P(D|HC)$: Probabilidade de que sendo a hipótese verdadeira, os dados sejam observados

$P(D|C)$:

Verossimilhança

$P(H|DC)$:



Laplace

Inferência Bayesiana



Bayes

$$P(AB|C) = P(A|C)P(B|AC)$$

$$P(AB|C) = P(B|C)P(A|BC)$$

$$P(H|DC) = \frac{P(H|C)P(D|HC)}{P(D|C)}$$

$P(H|C)$:

$P(D|HC)$:

$P(D|C)$: Evidência – Normalização

$P(H|DC)$:



Laplace

Inferência Bayesiana



Bayes

$$P(AB|C) = P(A|C)P(B|AC)$$

$$P(AB|C) = P(B|C)P(A|BC)$$

$$P(H|DC) = \frac{P(H|C)P(D|HC)}{P(D|C)}$$

$P(H|C)$:

$P(D|HC)$:

$P(D|C)$:

$P(H|DC)$:

Probabilidade posterior

Leva em conta a informação nos dados

A="Há nuvens", B="Chove"

$$P(B | AC) = \frac{P(B | C)P(A | BC)}{P(A | C)}$$

$P(A | BC) = 1$ Chove, logo há nuvens

$P(A | C) \leq 1$ Probabilidade de
"há nuvens" menor ou igual a 1

$$P(B | AC) \geq P(B | C)$$

A="Há nuvens", B="Chove"

$$P(B | AC) = \frac{P(B | C)P(A | BC)}{P(A | C)}$$

$P(A | BC) = 1$ Chove, logo há nuvens

$P(A | C) \leq 1$ Probabilidade de
"há nuvens" menor ou igual a 1

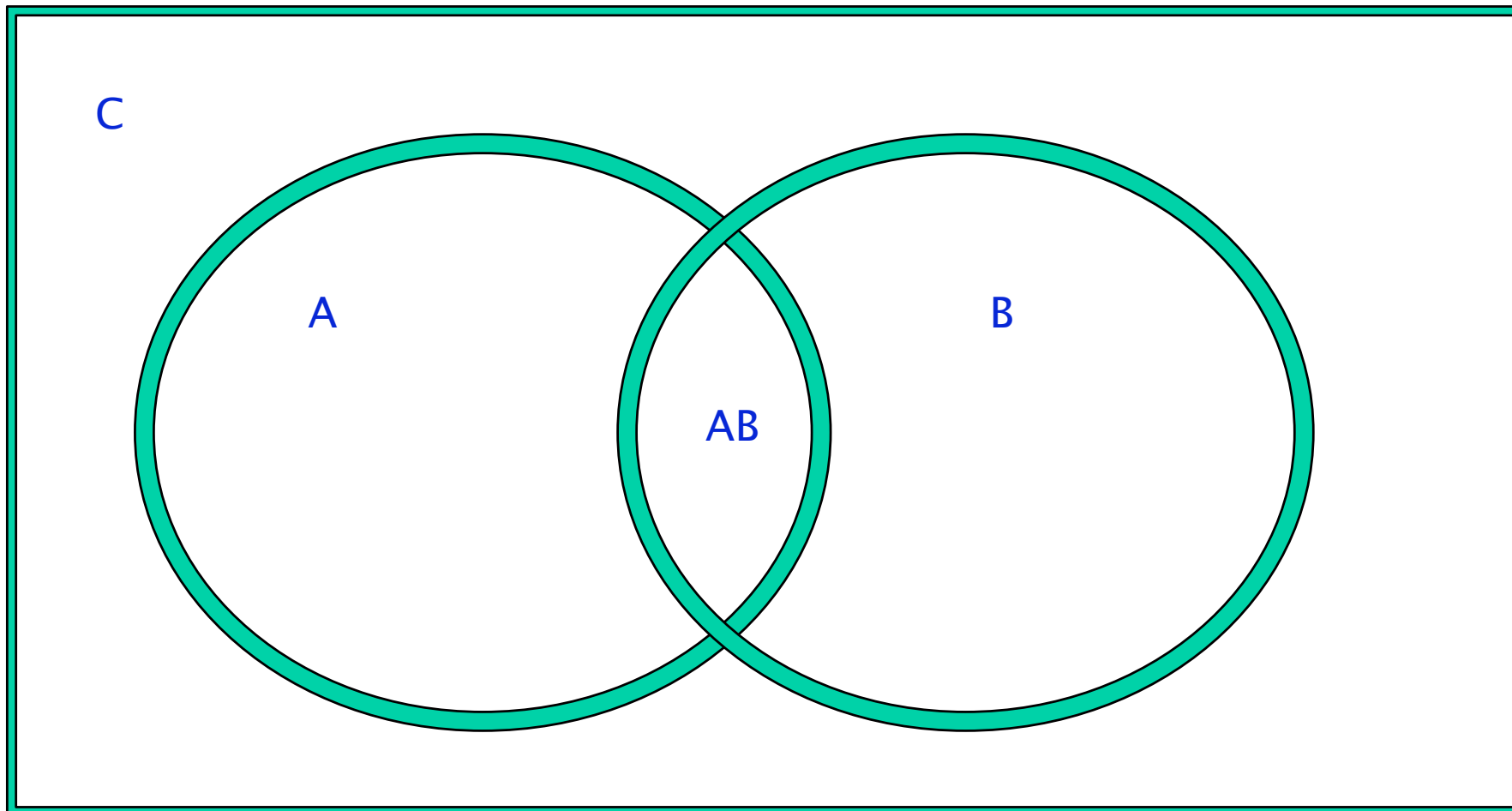
$$P(B | AC) \geq P(B | C)$$

Mamãe estava certa: "leva o guarda chuva, filinho"

probabilidade

Relação com frequência

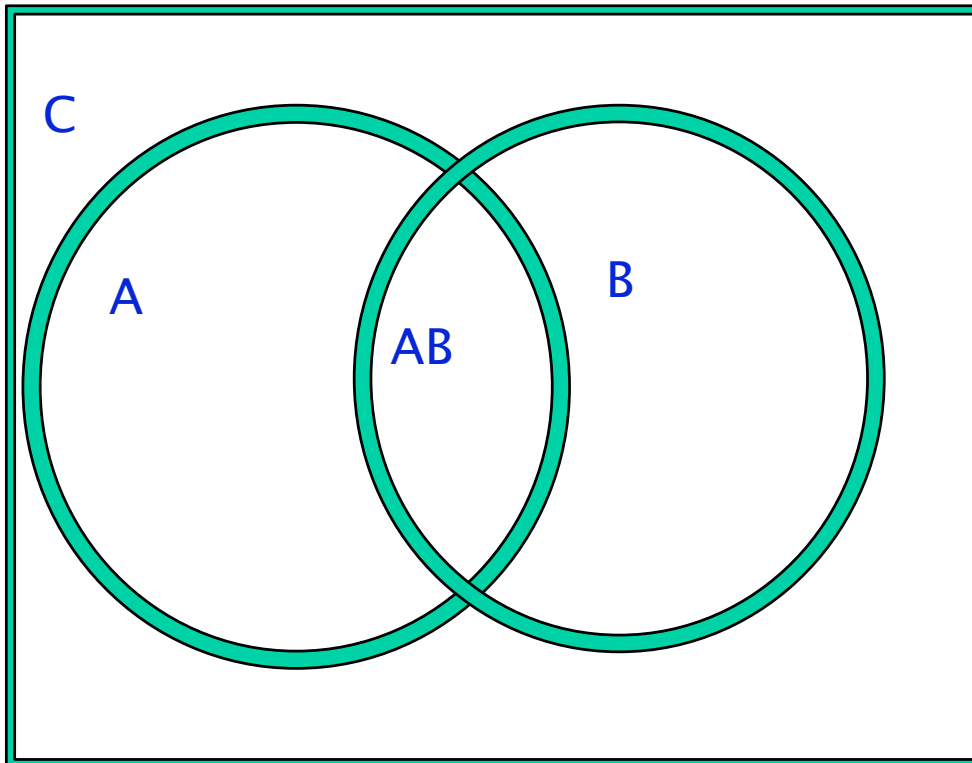
Diagrama de Venn



probabilidade

Relação com frequência

Diagrama de Venn



$$f_{AB} = \frac{n_{AB}}{n_C} = \frac{n_{AB}}{n_C}$$

$$= \frac{n_{AB}}{n_A} \frac{n_A}{n_C} = f_{B|A} f_A$$

$$= \frac{n_{AB}}{n_B} \frac{n_B}{n_C} = f_{A|B} f_B$$

probabilidade

Relação com frequência

Diagrama de Venn

$$\begin{aligned} f_{AB} &= \frac{n_{AB}}{n_C} = \frac{n_{AB}}{n_C} \frac{n_A}{n_A} \\ &= \frac{n_{AB}}{n_A} \frac{n_A}{n_C} = f_{B|A} f_A \xrightarrow{n \rightarrow \infty} P(AB) = P(B|A)P(A) \\ &= \frac{n_{AB}}{n_B} \frac{n_B}{n_C} = f_{A|B} f_B \xrightarrow{n \rightarrow \infty} P(AB) = P(A|B)P(B) \end{aligned}$$

probabilidade

Relação com frequência

Diagrama de Venn

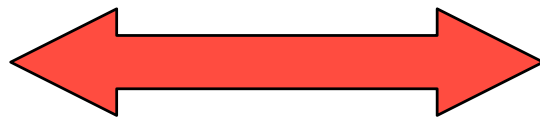
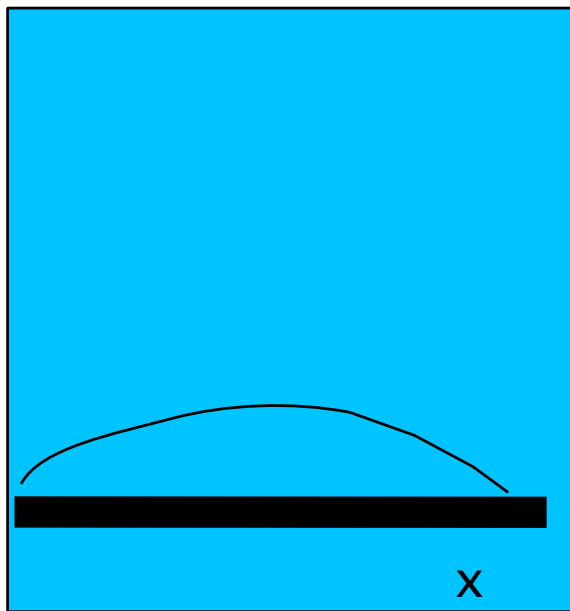
$$f_{AB} = \frac{n_{AB}}{n_C} = \frac{n_{AB}}{n_C}$$

$$= \frac{n_{AB}}{n_A} \frac{n_A}{n_C} = f_{B|A} f_A \xrightarrow{n \rightarrow \infty} P(AB) = P(B|A)P(A)$$

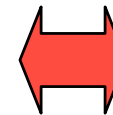
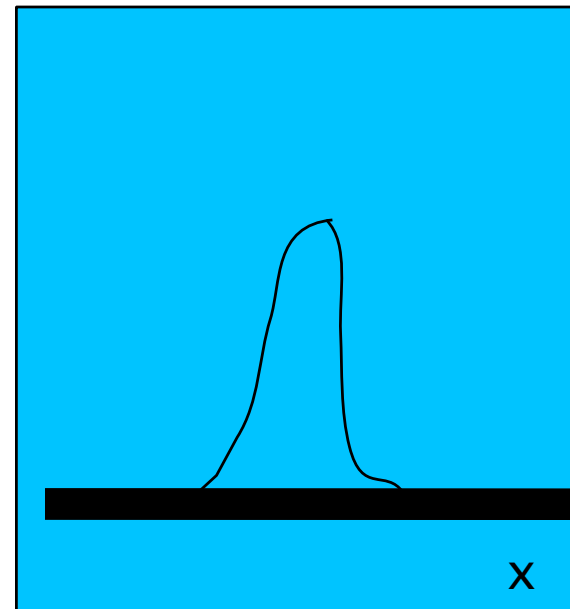
$$= \frac{n_{AB}}{n_B} \frac{n_B}{n_C} = f_{A|B} f_B \xrightarrow{n \rightarrow \infty} P(AB) = P(A|B)P(B)$$

CONVERGÊNCIA
EM PROBABILIDADE

Probabilidades e informação

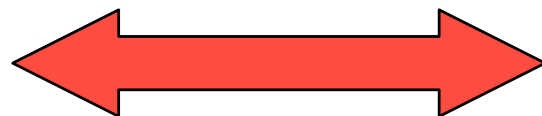
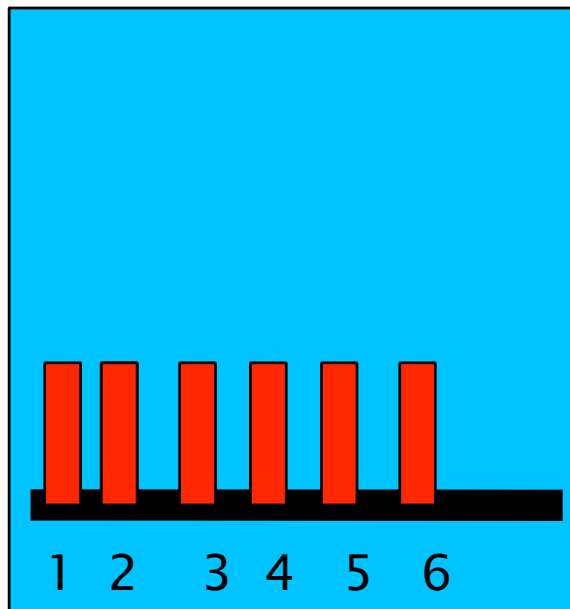


Muita incerteza
Ignorância grande

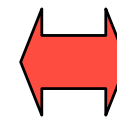
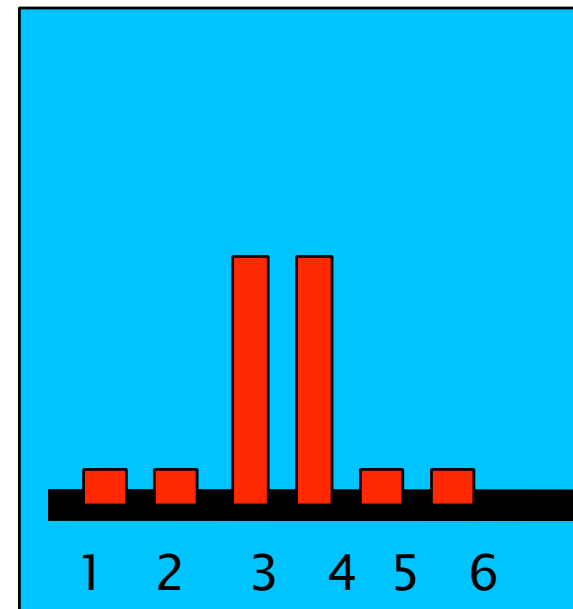


Pouca incerteza
Ignorância pequena

Probabilidades e informação



Muita incerteza
Ignorância grande

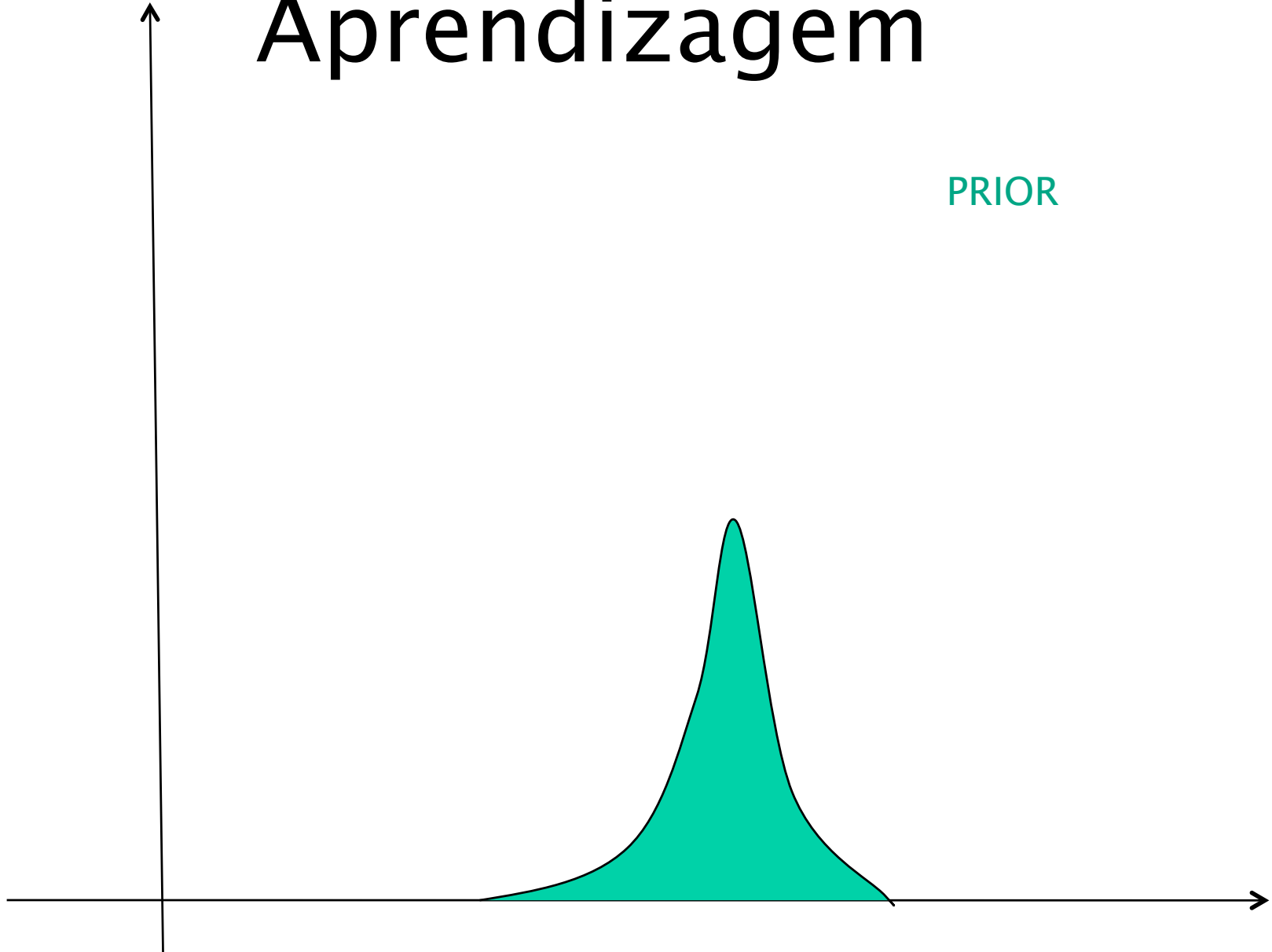


Pouca incerteza
Ignorância pequena

Aprendizagem

P

PRIOR

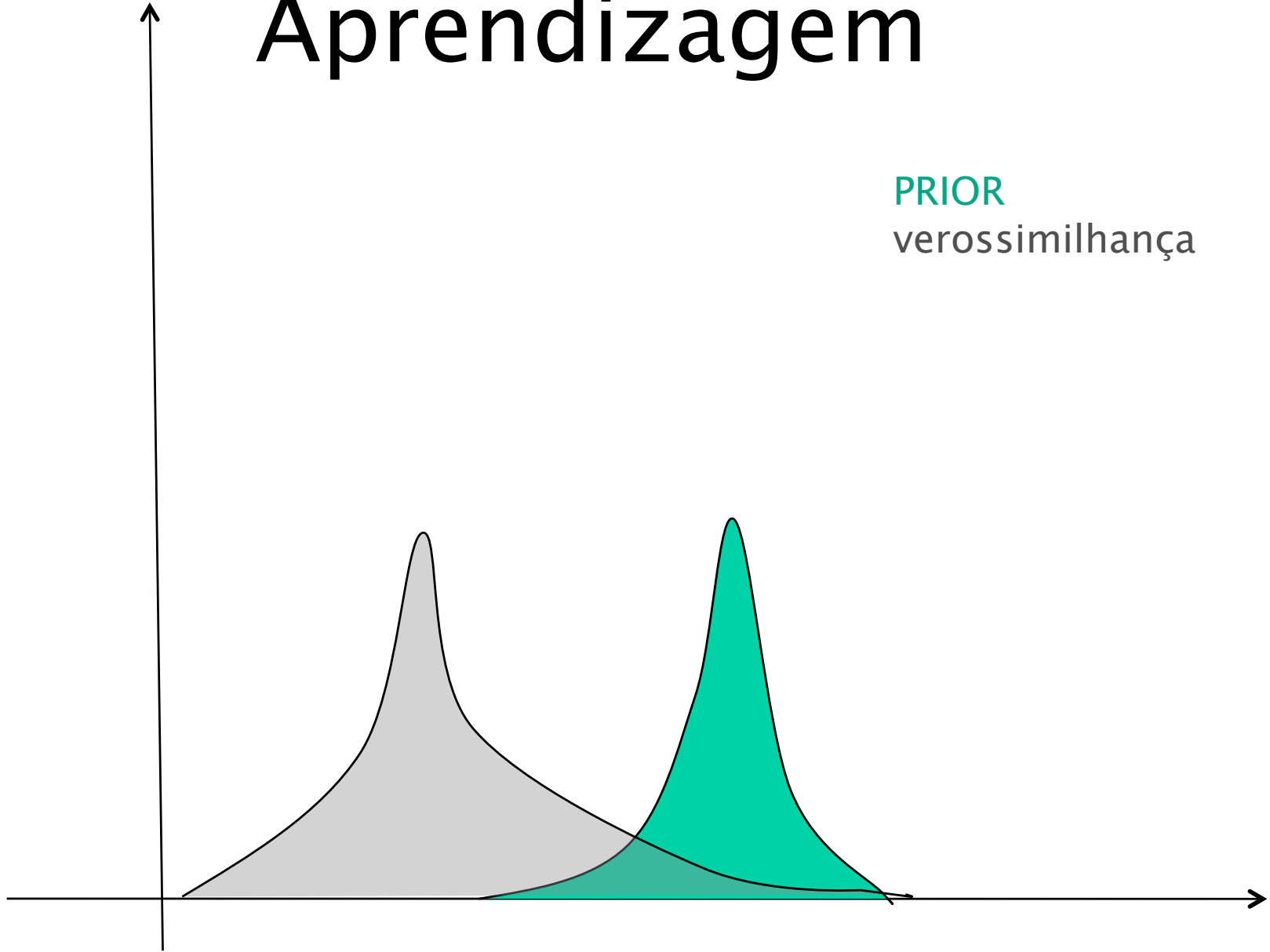


x

Aprendizagem

P

PRIOR
verossimilhança



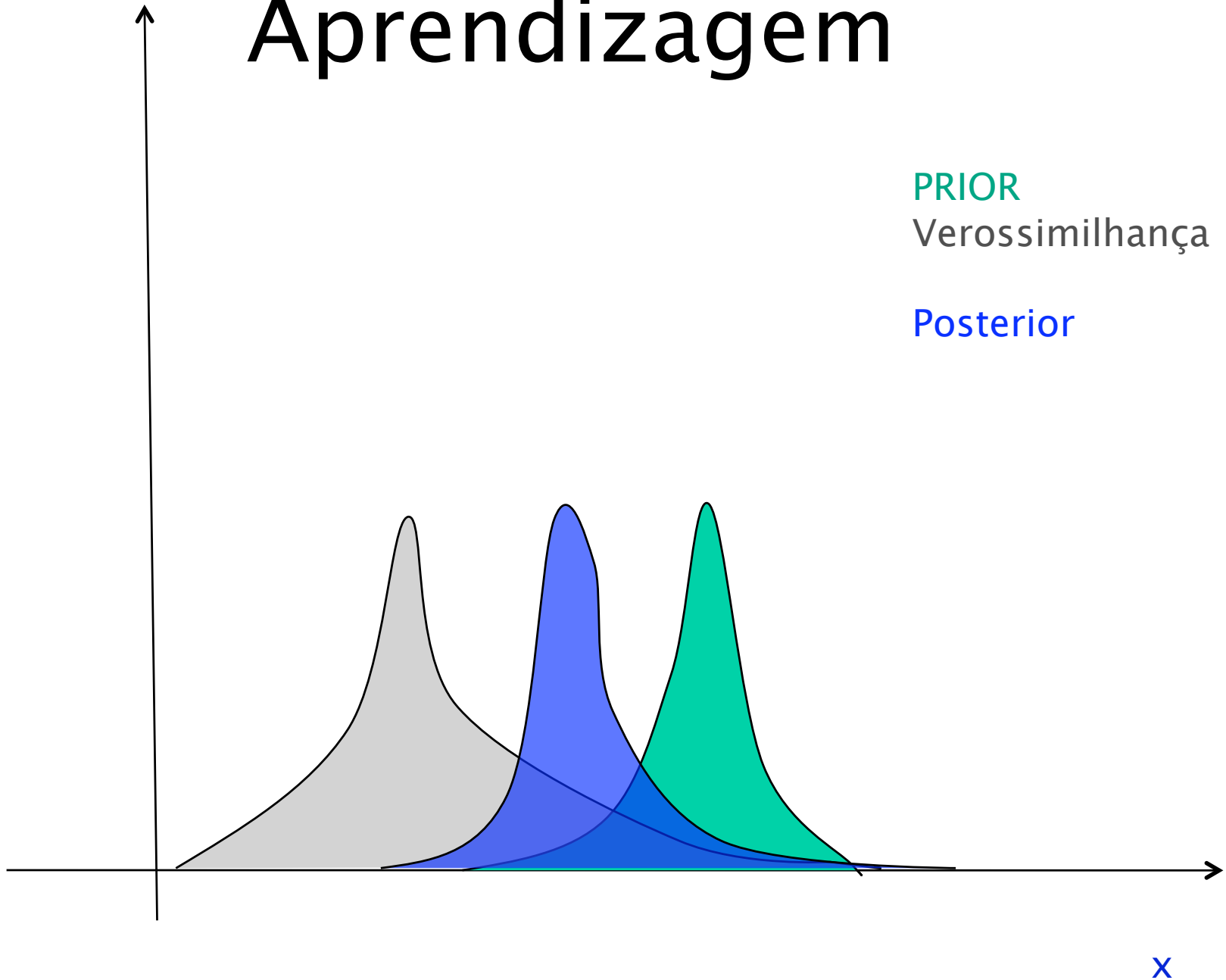
X

Aprendizagem

P

PRIOR
Verossimilhança

Posterior

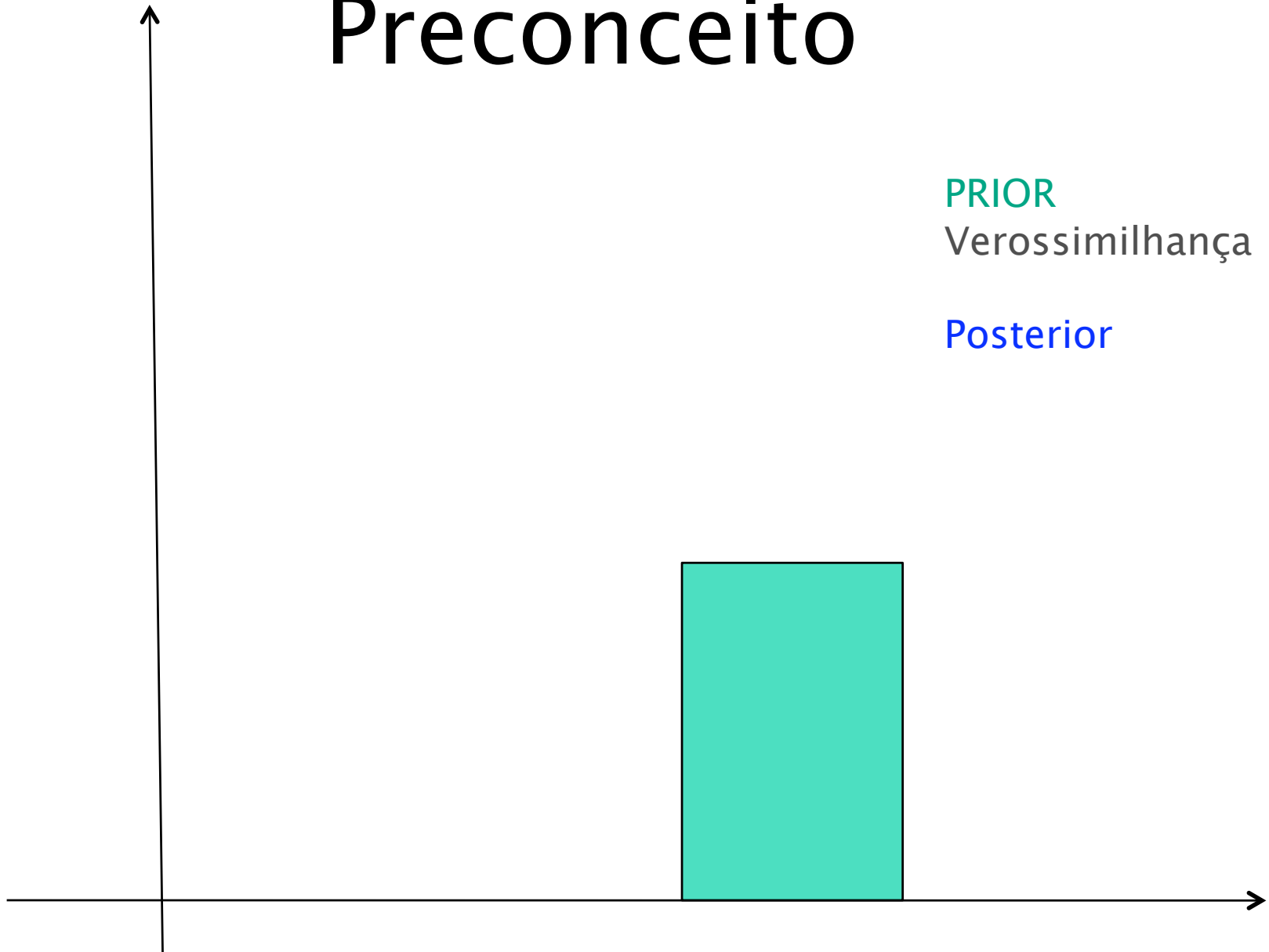


Preconceito

P

PRIOR
Verossimilhança

Posterior



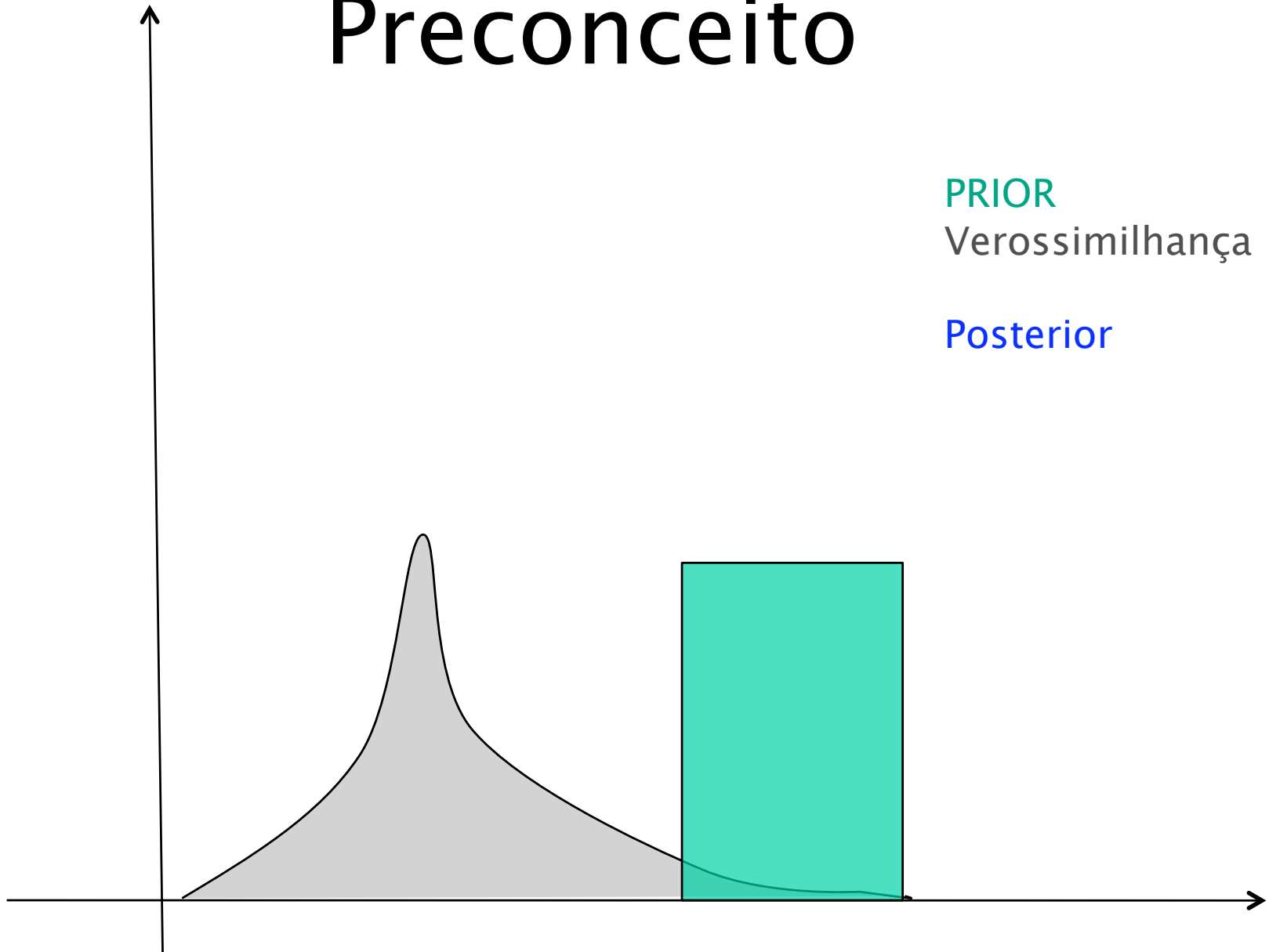
X

Preconceito

P

PRIOR
Verossimilhança

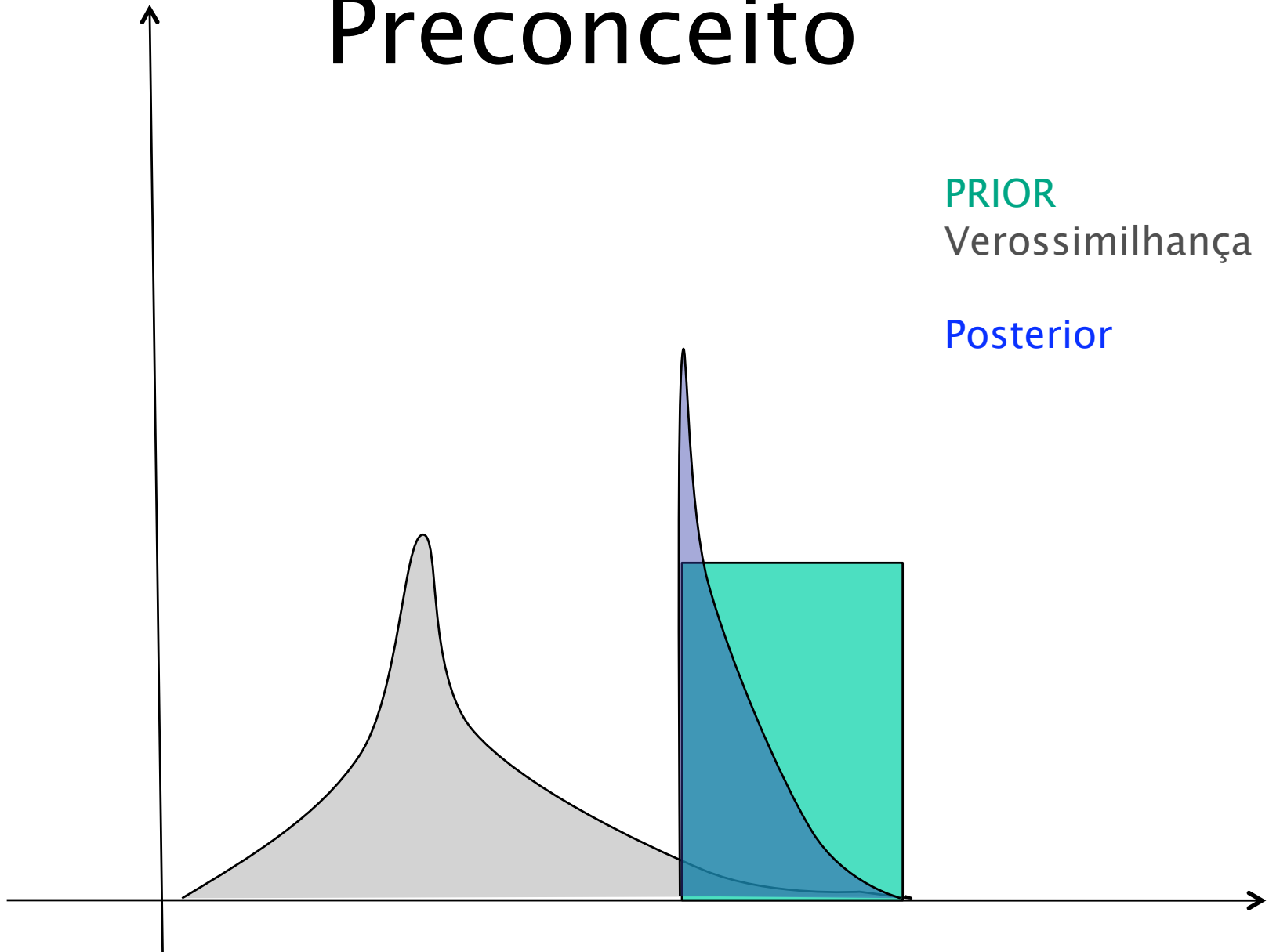
Posterior



X

Preconceito

P



PRIOR
Verossimilhança

Posterior

X

Relação com redes neurais

- Supõe forma paramétrica para classe de distribuições (gaussiana)
- Média J e covariância C
- Para uma dada arquitetura da máquina que algoritmo melhor aproxima o aprendizado Bayesiano?

Classificador $\sigma_J = T(J.S),$

$$\sigma_J = \{\pm 1\}, S \in R^N$$

Aprendizagem

P

PRIOR

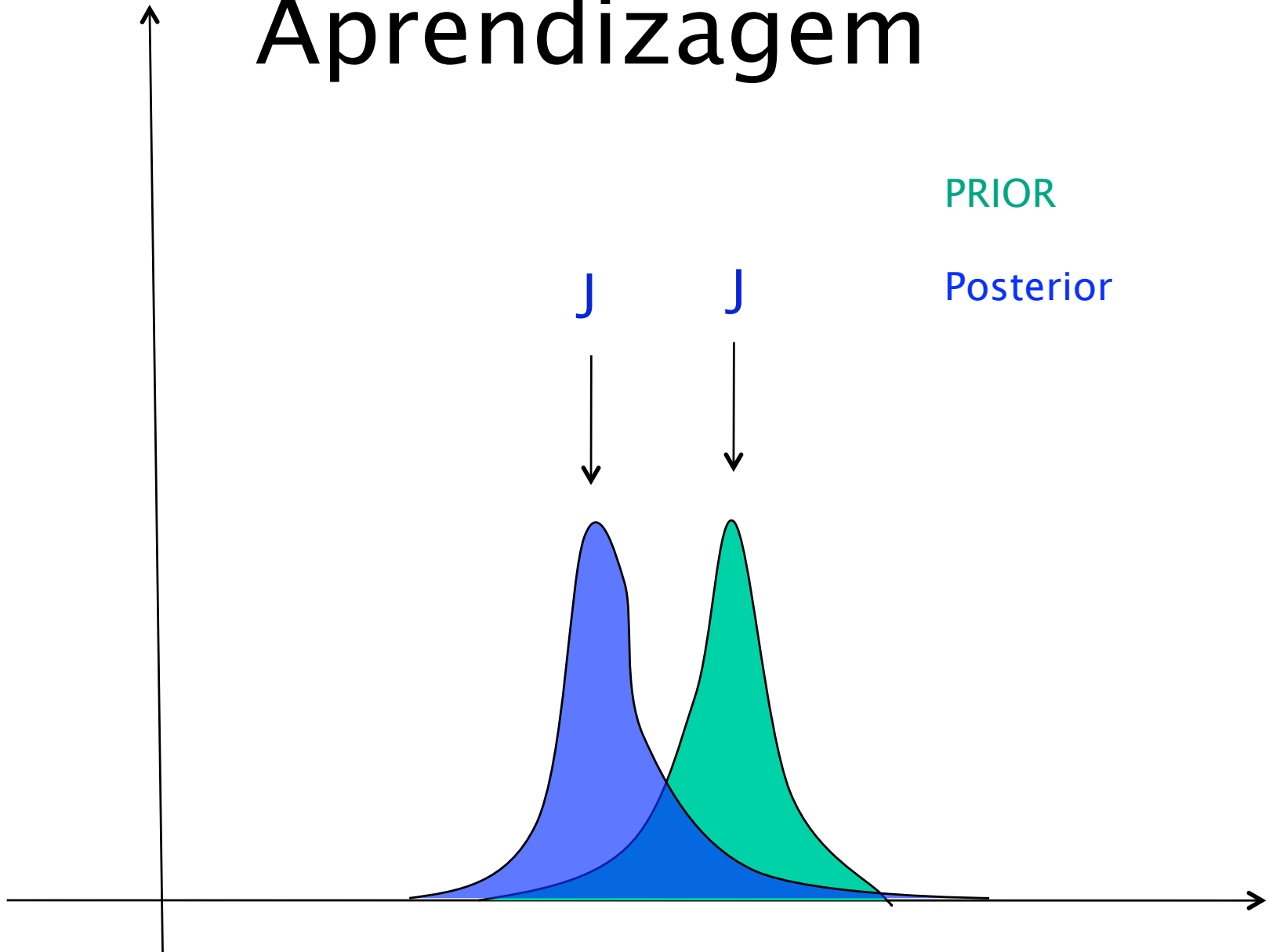
Posterior

J

J

Média J e covariância C

x



Relação com redes neurais

- Para uma dada arquitetura da máquina que

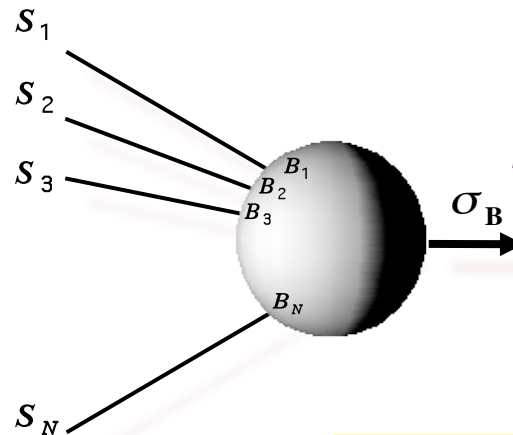
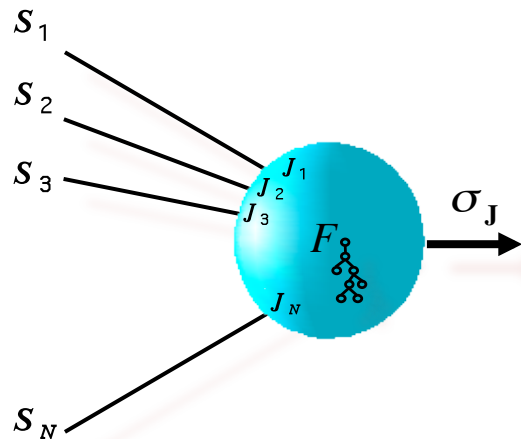
algoritmo melhor aproxima o aprendizado Bayesiano?

$$J_i(p+1) = J_i(p) + \Delta J_i(p)$$

$$C_{ij}(p+1) = C_{ij}(p) + \Delta C_{ij}(p)$$

Agents and environment (supervised learning)

Both professor (black) and student (cyan) are simple perceptrons that classify a vector \mathbf{S} with one of two possible labels: +1 or -1



The learning algorithm is represented by the modulation function

$$F(\sigma_{Bp}, \mathbf{S}_p; \mathbf{J}_p)$$

$$\mathbf{J}_{p+1} = \mathbf{J}_p + F(\sigma_{Bp}, \mathbf{S}_p; \mathbf{J}_p) \frac{\sigma_{Bp} \mathbf{S}_p}{\sqrt{N}}$$

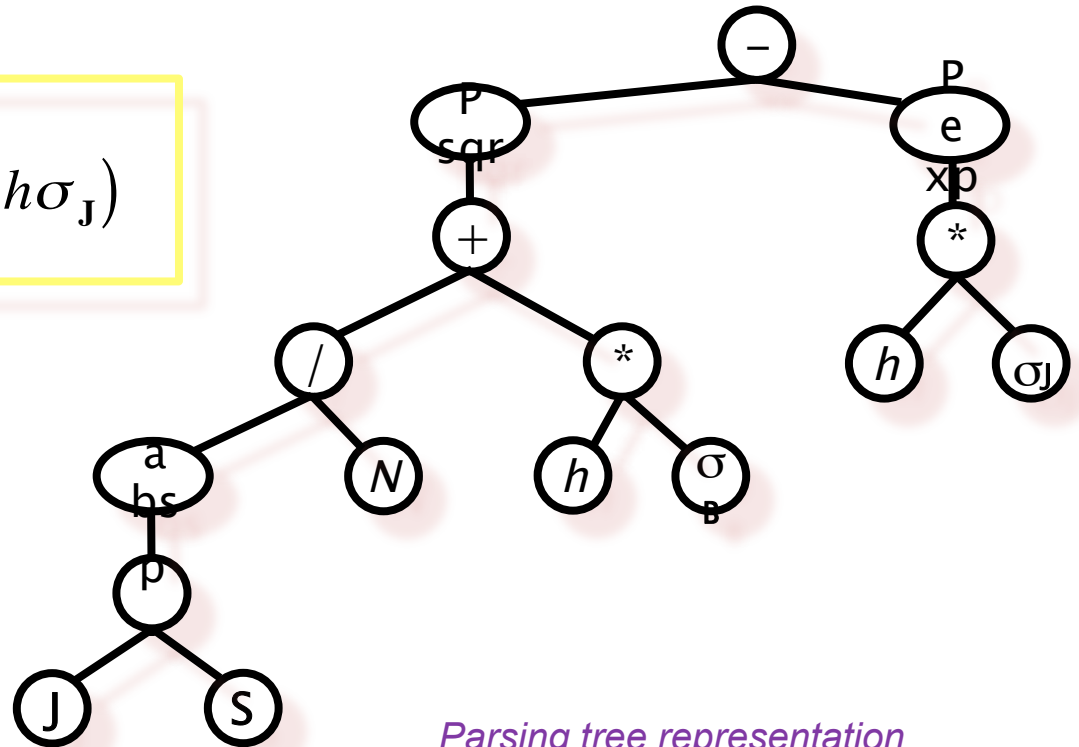
The learning algorithm is the update rule for the synaptic vector \mathbf{J} when new information is available



Example of learning algorithm as a tree

$$F = \sqrt{\frac{|\mathbf{J} \cdot \mathbf{S}|}{N}} + h\sigma_B - \exp(h\sigma_J)$$

Formula representation
of the learning algorithm



Parsing tree representation
of the learning algorithm



Genetic Programming (GP)

- Population (individuals represented by perceptrons)
- Fitness (as a function of the generalization error)

$$\vartheta = \{JSh\sigma_B\sigma_J\}$$

F = {+ - * % Psqr Pexp Plog abs p. pN. ev* vv+ vv-}

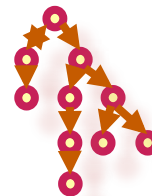
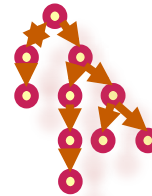


Fitness

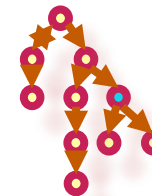
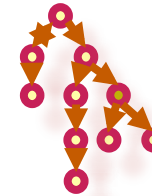
$$f = c \sum_{p=1}^P pe_g(p)$$

P is the total number of examples
c is an appropriate constant

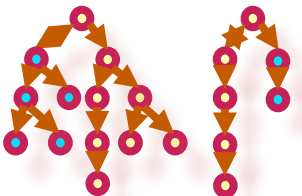
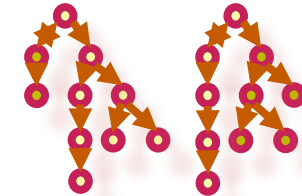
Reproduction



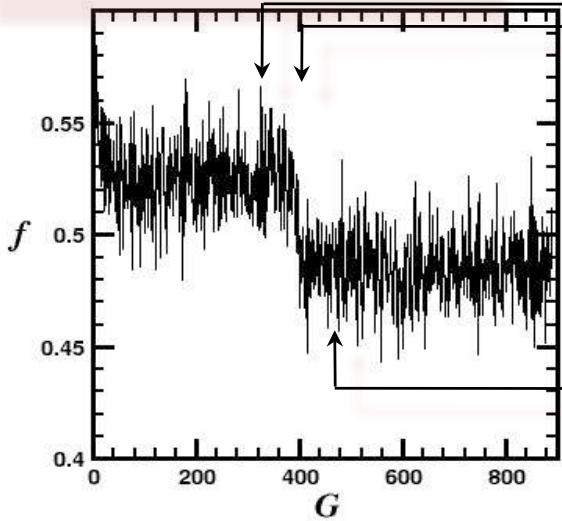
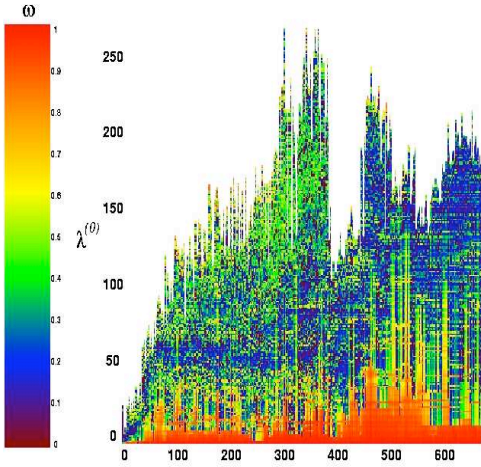
Mutation



Cross-over



GP Results J. P. Neirotti & N. C.



350

ABS	%	%	ABS	-	%	ABS	-	%	%	+	SIGB	SIGB	+	+	+	SIGB	SIGB	SIGB	SIGB	SIGB	-	H	SIGB	+					
-	%	+	SIGB	+	+	SIGB	SIGB	SIGB	+	+	+	+	+	+	+	SIGB	SIGB	SIGB	SIGB	SIGB	+	+	SIGB	-	H				
SIGB	SIGB	+	+	SIGB	SIGB	+	SIGB	SIGB	+	SIGB	SIGB	+	SIGB	SIGB	+	SIGB	SIGB	+	SIGB	SIGB	+	SIGB	SIGB	+	H				
SIGB	SIGB	-	%	ABS	%	ABS	-	%	*	-	H	SIGB	SIGJ	+	+	SIGB	SIGB	SIGB	-	H	SIGB	SIGB	SIGB	-	H				
H	SIGB	+	+	SIGB	%	%	%	%	+	SIGB	%	*	-	H	SIGB	SIGJ	-	H	SIGB	+	SIGB	+	SIGB	+	H				
SIGB	SIGB	SIGB	%	%	-	H	SIGB	%	+	-	H	SIGB	SIGJ	+	+	*	-	H	SIGB	SIGJ	SIGB	SIGB	%	%	H				
+	SIGB	SIGB	+	+	SIGB	SIGB	SIGB	SIGB	SIGB	SIGB	+	SIGB	SIGB	+	SIGB	SIGB	+	%	SIGB	+	+	%	+	SIGB	%	H			
*	-	H	SIGB	-	H	SIGB	+	+	SIGB	SIGB	SIGB	SIGB	SIGB	+	%	+	SIGB	SIGB	+	SIGB	SIGB	+	SIGB	SIGB	+	H			
+	+	+	SIGB	+	SIGB	+	SIGB	SIGB	SIGB	+	+	SIGB	SIGB	+	SIGB	SIGB	+	SIGB	SIGB	+	SIGB	SIGB	+	SIGB	SIGB	+	H		
+	+	+	+	SIGB	+	SIGB	SIGB	+	SIGB	+	+	SIGB	SIGB	+	SIGB	SIGB	+	SIGB	SIGB	+	SIGB	SIGB	+	SIGB	SIGB	+	H		
-	H	SIGB	SIGJ	+	+	SIGB	SIGB	SIGB	-	H	SIGB	+	+	SIGB	SIGB	+	SIGB	SIGB	+	SIGB	SIGB	+	SIGB	SIGB	+	H			
*	-	H	SIGB	SIGJ	-	H	SIGB	+	SIGB	SIGB	SIGB	+	H	SIGB	SIGB	SIGB	+	H	SIGB	SIGB	SIGB	%	%	%	%	+	SIGB	%	H

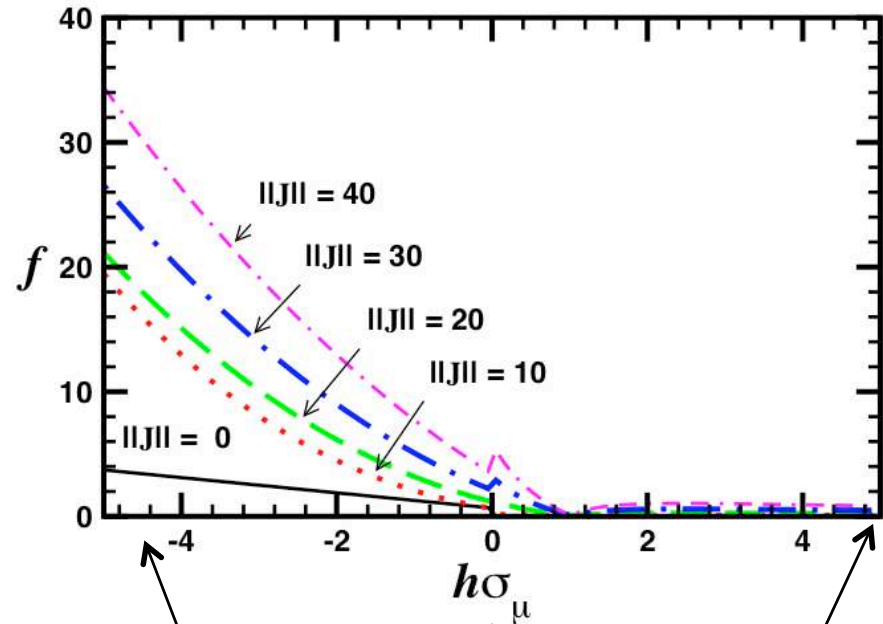
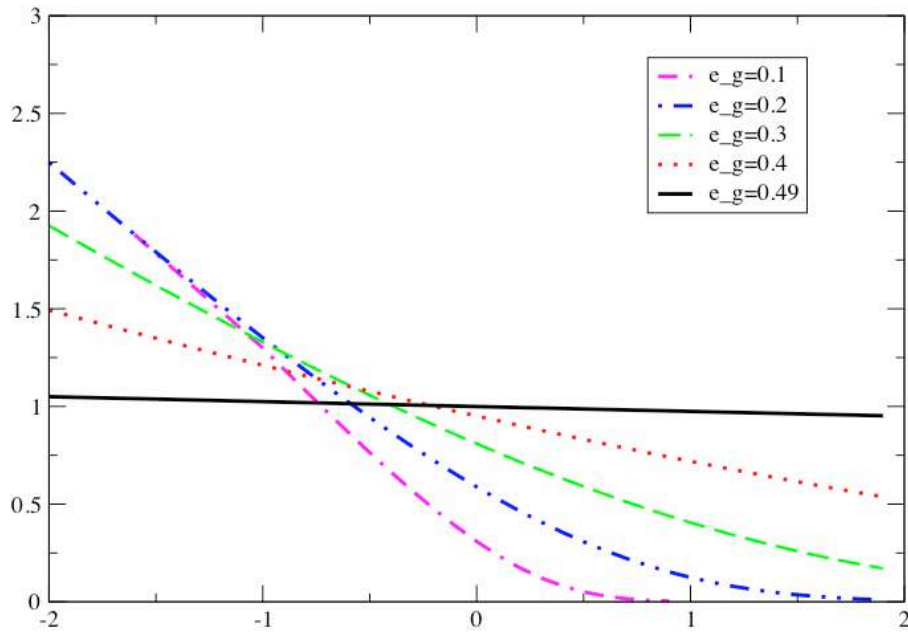
400

ABS	%	%	ABS	-	%	ABS	-	%	%	PN	J	J	SIGB	+	+	SIGB	SIGB	SIGB	SIGB	-	H	SIGB	+	PN		
%	*	-	H	SIGB	SIGJ	+	+	SIGB	SIGB	+	SIGB	SIGB	-	%	ABS	%	ABS	-	%	*	-	H	%	+	PN	
PN	J	J	SIGB	SIGJ	+	+	SIGB	%	PN	J	J	SIGB	SIGB	-	H	SIGB	SIGB	SIGB	-	H	SIGB	+	+	SIGB	+	PN
%	%	%	%	+	SIGB	%	*	-	H	SIGB	SIGJ	-	H	SIGB	+	SIGB	+	SIGB	+	-	H	SIGB	SIGB	SIGB	+	PN
%	%	SIGB	%	*	-	H	SIGB	SIGJ	+	%	%	%	PN	J	J	SIGB	+	+	SIGB	SIGB	SIGB	SIGB	SIGB	%	+	PN
%	+	SIGB	SIGB	+	SIGB	SIGB	SIGB	SIGB	+	SIGB	%	%	PN	J	J	SIGB	+	+	SIGB	SIGB	SIGB	SIGB	SIGB	%	+	PN

450

ABS	%	%	ABS	-	%	ABS	-	%	%	PN	J	J	SIGB	+	+	SIGB	SIGB	SIGB	-	H	%	%	%	PN	PN			
J	J	SIGB	+	+	SIGB	SIGB	SIGB	SIGB	+	%	*	-	H	SIGB	SIGJ	+	+	%	%	SIGB	%	%	%	PN	PN			
J	J	SIGB	+	SIGB	SIGB	SIGB	SIGB	SIGB	%	%	+	%	+	+	SIGB	SIGB	SIGB	%	PN	J	J	SIGB	SIGB	+	PN			
+	SIGB	SIGB	SIGB	SIGB	+	SIGB	SIGB	-	%	ABS	%	ABS	-	%	*	-	H	%	PN	J	J	SIGB	SIGJ	+	PN			
+	SIGB	+	SIGB	SIGB	SIGB	-	H	SIGB	SIGB	SIGB	-	H	SIGB	+	+	SIGB	%	%	%	+	SIGB	%	%	+	SIGB	%	PN	
-	H	SIGB	SIGJ	-	H	SIGB	+	%	*	-	H	SIGB	SIGJ	+	+	SIGB	%	%	SIGB	+	SIGB	SIGB	SIGB	SIGB	+	PN		
+	SIGB	SIGB	SIGB	%	%	SIGB	%	*	-	H	SIGB	SIGJ	+	%	%	%	%	PN	J	J	SIGB	+	+	SIGB	+	PN		
SIGB	SIGB	PN	J	J	SIGB	SIGB	%	%	+	SIGB	SIGB	+	SIGB	SIGB	SIGB	SIGB	+	SIGB	SIGB	+	SIGB	SIGB	+	SIGB	+	PN		
J	SIGB	+	+	SIGB	SIGB	%	ABS	-	%	%	%	%	+	PN	J	J	SIGB	PN	J	J	SIGB	+	+	PN	+	PN		
H	%	PN	J	J	SIGB	SIGJ	SIGB	%	%	PN	J	J	SIGB	+	SIGB	SIGB	SIGB	-	H	+	SIGB	+	+	SIGB	+	PN		
SIGB	SIGB	SIGB	+	%	PN	J	J	SIGB	+	%	%	SIGB	%	%	%	PN	J	J	SIGB	+	SIGB	SIGB	SIGB	+	SIGB	+	PN	
SIGB	SIGB	ABS	-	%	%	%	PN	J	J	SIGB	+	SIGB	SIGB	SIGB	-	H	%	%	%	PN	J	J	SIGB	+	SIGB	+	PN	
+	SIGB	SIGB	SIGB	SIGB	SIGB	SIGB	SIGB	+	SIGB	SIGB	SIGB	+	SIGB	SIGB	SIGB	+	SIGB	SIGB	+	SIGB	SIGB	+	SIGB	SIGB	+	PN	+	PN



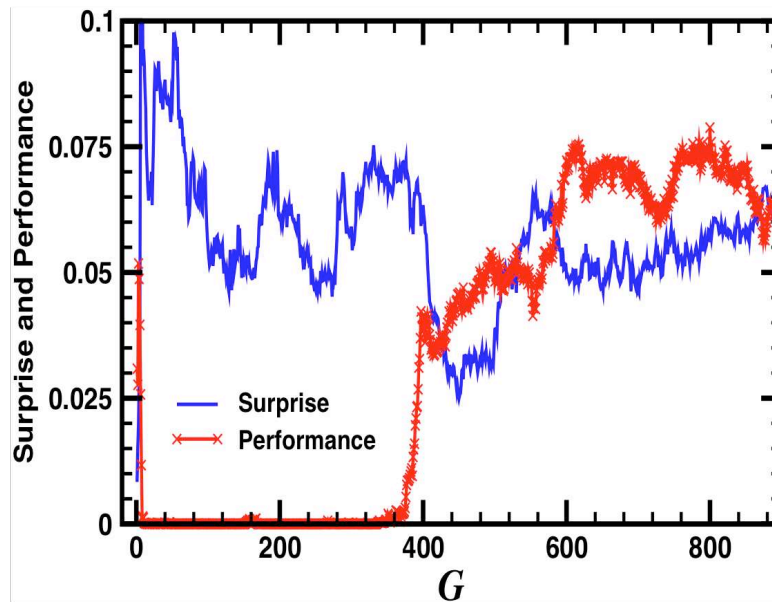
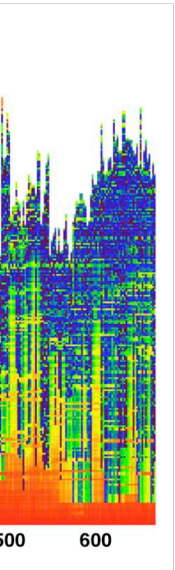


errado

difícil

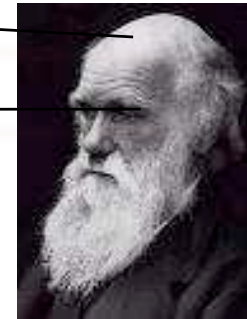
Certo

Evolução: invasão por algoritmos mais eficientes



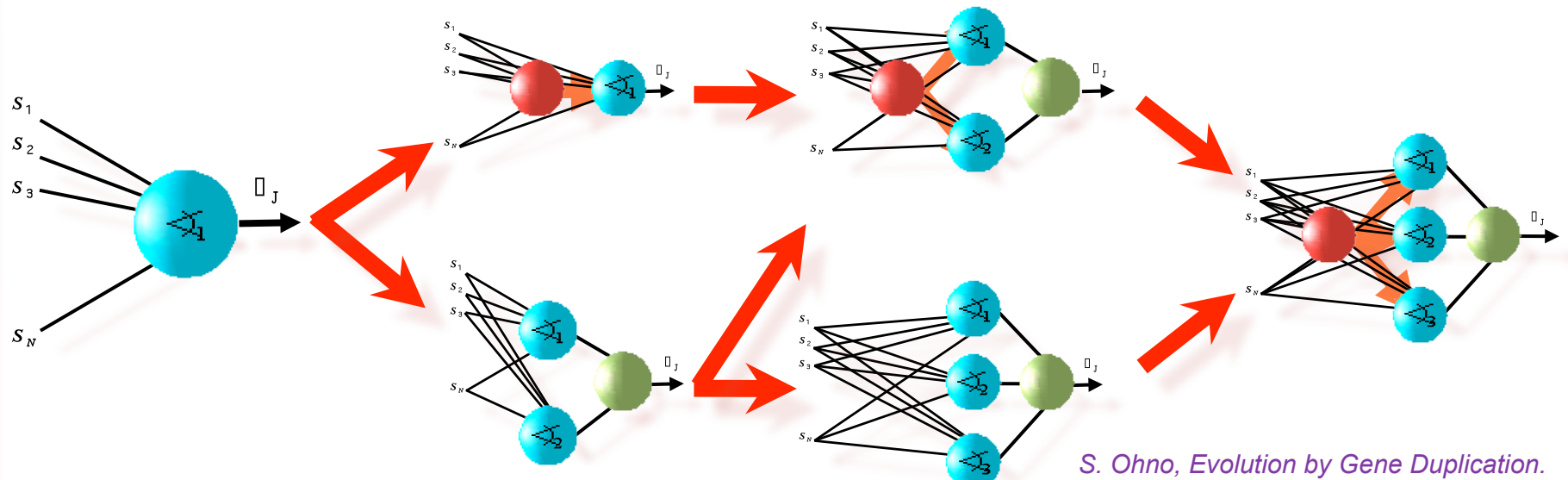
Prefrontal cortex (P)





Amygdala (S)



Evolutionary paths

- Reduced set of variables $\vartheta = \{\sigma_B S; J\}$

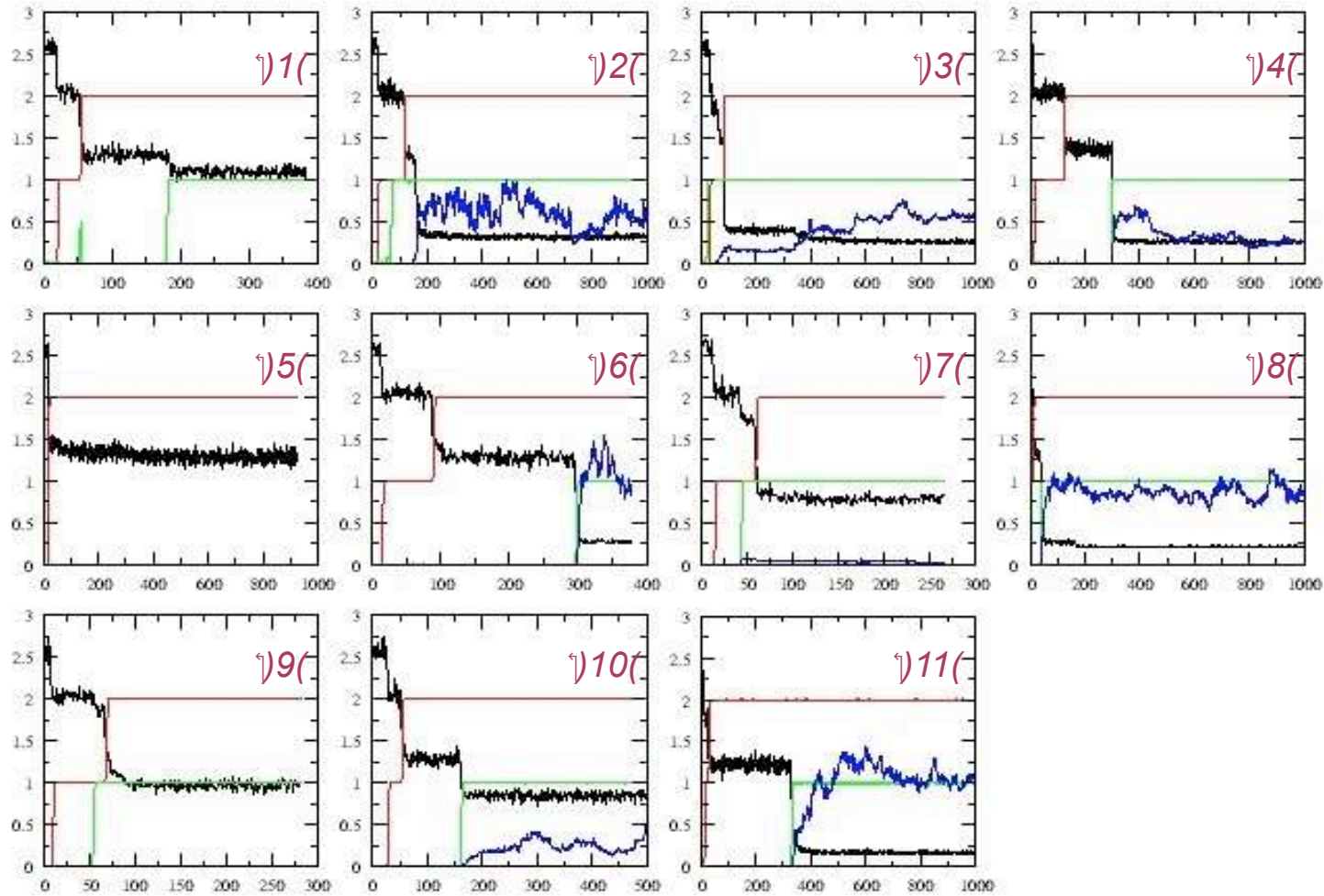


-  Module with evolving modulation function ϑ
-  Module with constant modulation function (Hebbian module). Introduces new variables to the other modules through 
-  Integrator

S. Ohno, *Evolution by Gene Duplication*. Springer, Berlin, 1970.



Preliminary results



— Fitness
— ●
— ●
— P

(
1)2(
1)3(
1)4(
1)6(
1)8(
1)11(
)



Informação: vínculos

- Muitas formas

Cada forma representa um tipo de mudança de crenças.

Cada uma representa um tipo

mudança de
distribuições de probabilidades

a partir de distribuições *a priori* arbitrárias

Parte II: Teoria de informação

- Tipicamente

Primeiro

uma estrutura matemática e depois uma interpretação

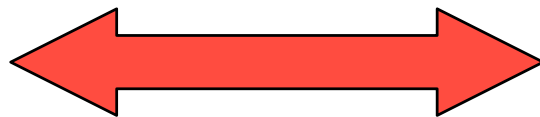
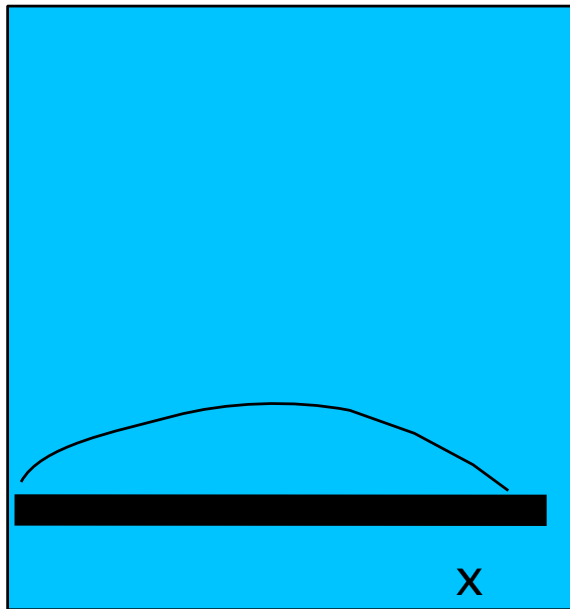
Procedimento inverso: primeiro interpretação e depois estrutura matemática



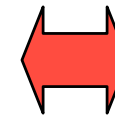
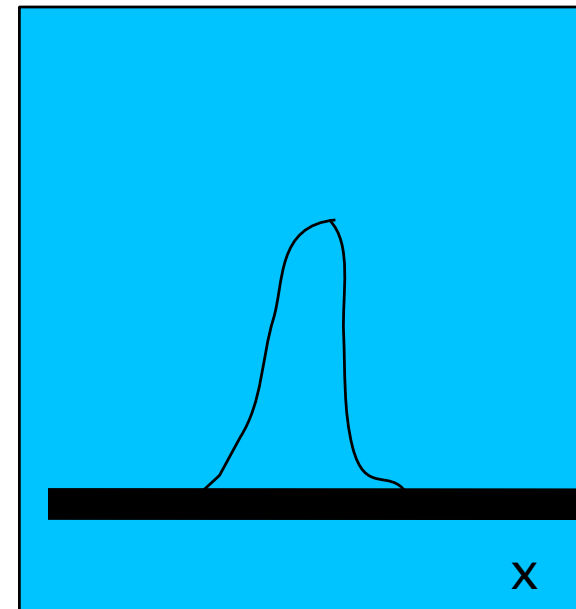
Definição de informação

- Qualquer coisa que me leve a mudar crenças.
- Qualquer coisa que me leve a mudar pdf

Probabilidades e informação não pode estar sempre certo



Muita incerteza
Ignorância grande



Pouca incerteza
Ignorância pequena

Teoria de informação

- Se existe uma teoria geral deve valer em casos particulares.
- Se existem casos particulares que não são satisfeitos por uma teoria, essa teoria é descartada.
- Se um número suficientemente gde de casos for conhecido, então a teoria geral poderá ser determinada (ou não...pode ser que não exista.)

Axiomas ou desejos que devem ser satisfeitos por uma teoria

- Como atribuir atualizar a distribuição de probabilidades quando chega nova informação?
- **(axioma 1) Ranking** : a distribuição melhor colocada no ranking , dentro daquelas que satisfazem a informação que temos, será a escolhida : atribuir um número

Axiomas ou propriedades satisfeitas por esse número?

- Axioma 2. Informação local tem efeitos locais

$$S[p, q] = \int dx F(p(x), q(x), x)$$

$$q(x) \Rightarrow p(x)$$

atualização

Axiomas ou desejos que devem ser satisfeitos por uma teoria

- (axioma 2) Invariante ante reparametrizações (mudanças de variáveis)

Uma tomada de decisão não deveria mudar se se fizer uma mudança de variáveis.

O Sistema de coordenadas é arbitrário

Covariância....

Axiomas ou desejos que devem ser satisfeitos
por uma teoria
covariância $x = \Gamma(x')$

$$\gamma(x') = \left| \frac{\partial x}{\partial x'} \right|, \dots dx = \gamma(x') dx'$$

densidade

$$m(x) = \frac{m'(x')}{\gamma(x')}, p(x) = \frac{p'(x')}{\gamma(x')}, q(x) = \frac{q'(x')}{\gamma(x')}, \quad \frac{p(x)}{m(x)} = \frac{p'(x')}{m'(x')}$$

$$S[p, q] = \int dx F(p(x), q(x), x) = \\ = \int \gamma(x') dx' F\left(\frac{p'(x')}{\gamma(x')}, \frac{q'(x')}{\gamma(x')}, \Gamma(x')\right)$$

2. Axioma ou desejos que devem ser
satisfeitos por uma teoria

covariância

$$\begin{aligned} S[p, q] &= \int dx F(p(x), q(x), x) = \\ &= \int \gamma(x') dx' F\left(\frac{p'(x')}{\gamma(x')}, \frac{q'(x')}{\gamma(x')}, \Gamma(x')\right) \\ &S[p', q'] = \int dx' F(p'(x'), q'(x'), x') \end{aligned} \left. \vphantom{\int} \right\} \text{iguais}$$

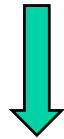
invariantes

$$S[p, q] = \int dx m(x) \Psi\left(\frac{p(x)}{m(x)}, \frac{q(x)}{m(x)}\right)$$

Axioma 1 Localidade

Se não há informação não há motivo para mudar crenças

$$p(x) = q(x)$$




$$m(x) = q(x)$$

$$S[p, q] = \int dx q(x) \Psi\left(\frac{p(x)}{q(x)}\right)$$

Consistência para sistemas independentes

- **(3) Independência :**

A priori dois objetos são independentes e a nova informação não introduz dependência, então devem continuar a ser independentes.


$$q(x,y) = q_1(x)q_2(y)$$

$$p(x,y) = p_1(x)p_2(y)$$

Duas formas de fazer:
Juntas ou separados
Deve dar o mesmo resultado

$$S[p,q] = - \int dx p(x) \log \left(\frac{p(x)}{q(x)} \right)$$

3 passos

- 1) aplica para subsistemas idênticos

$$S_{\eta}[p, q] = \frac{1}{\eta(\eta + 1)} \left(1 - \int dx p^{\eta+1} q^{\eta} \right)$$

3 passos

- 2) aplica para subsistemas diferentes.

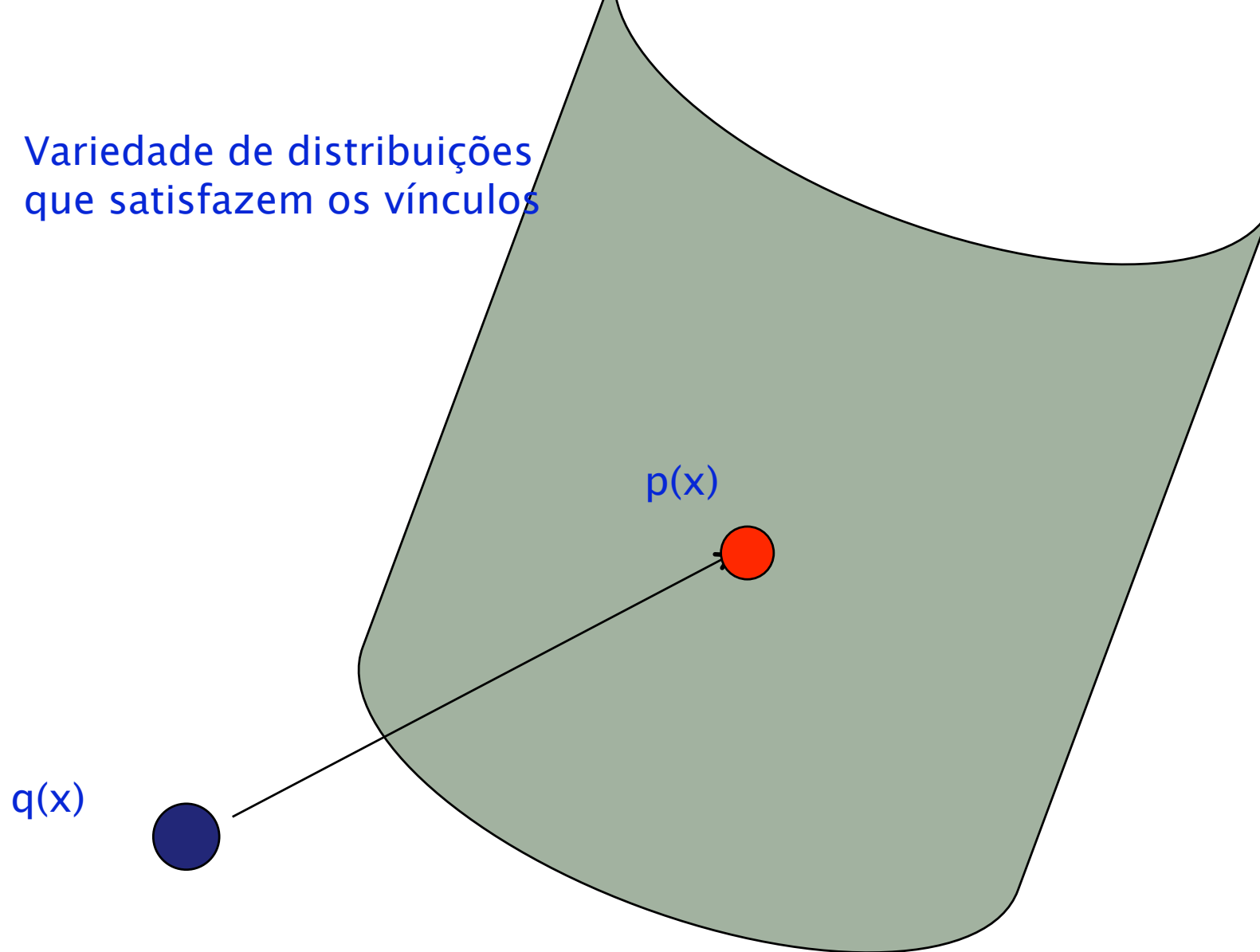
poderia ter parâmetros η_1 e η_2 diferentes

Consistência $\eta_1 = \eta_2$

- 3) aplica para N objetos, N muito grande

$$\eta_1 = \eta_2 = 0$$

Variedade de distribuições
que satisfazem os vínculos



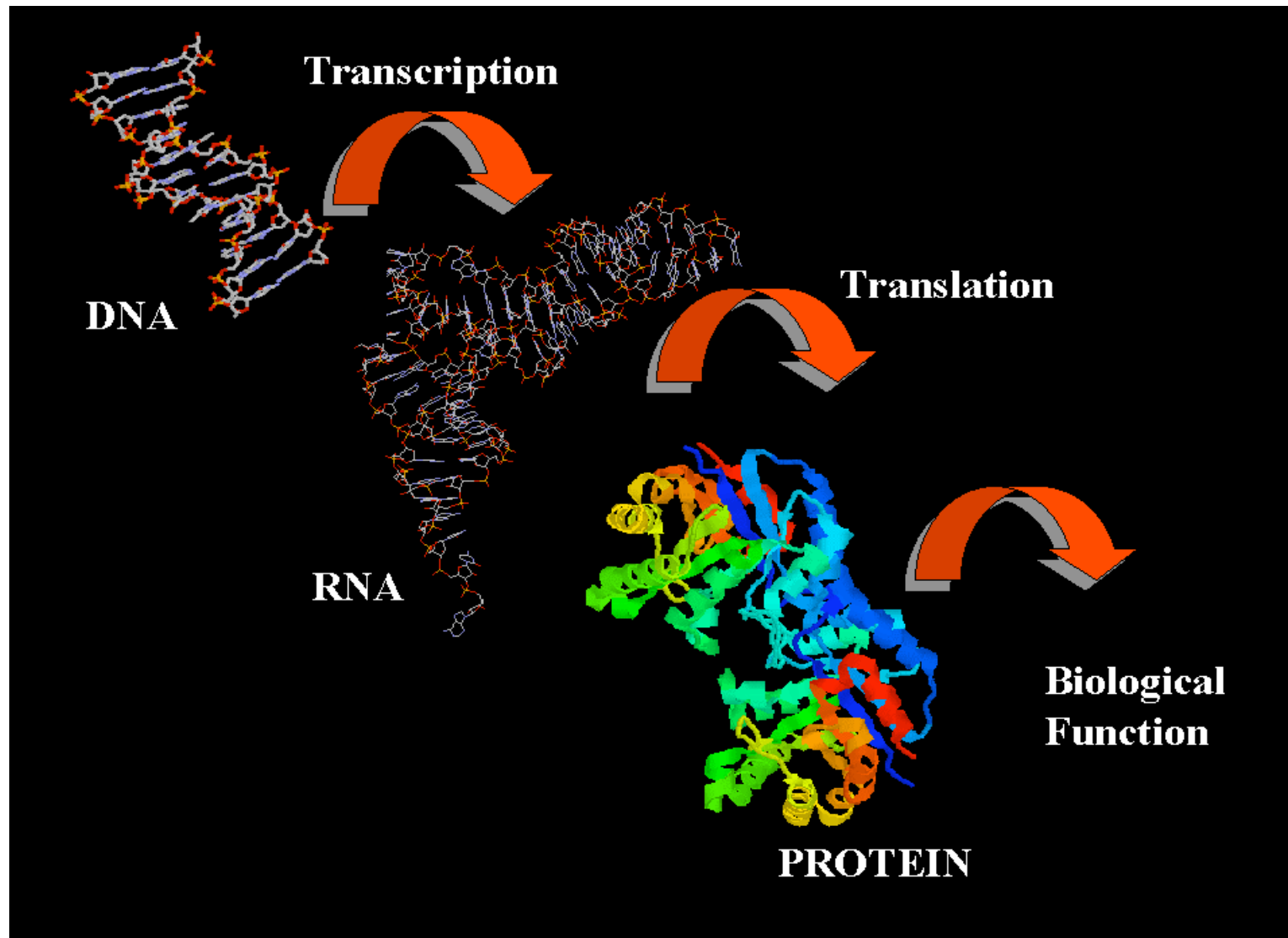
Desigualdade de Bogoliuvov



Aplicações

- Clustering e data mining em problemas de bioinformática
- Dinâmica de opiniões em sociedades de máquinas que processam informação

Gene expression



c/o R. Breitling

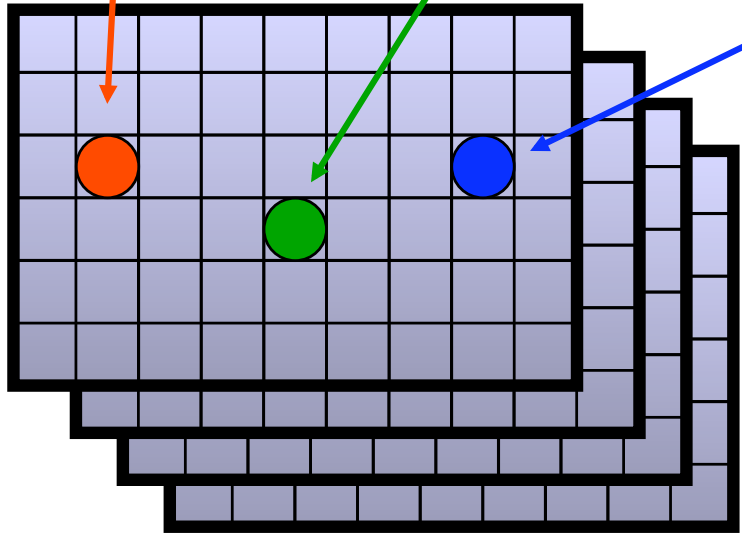
Genomic tiling array data

sequence covered by 'path' or 'tiling'

ACTTACAAGGAGTCTAGGCA ... CATTACGACT



C. elegans



microarray: transcription intensity vs. genomic position

- repeated for many samples:
- different developmental stages
 - varying external conditions
 - different strains (variants)
 - mutants ...

perfect match intensities
mismatch intensities

GAGTCTAGG (PM)
GAGTGTAGG (MM)

Categorização : supervisionada ou não?

- Cada padrão ($i=1 \dots N$) : vetor de “características” $\mu=1 \dots P$
- Identificação de prototipos : alguns exemplos classificados
- Ou nenhuma informação sobre quantas classes
- problema fundamental o que significa distância? Tensor métrico

$$D_{ij} = g_{\mu\nu} (x_i^\mu - x_j^\mu) (x_i^\nu - x_j^\nu)$$

$\{D_{ij}\}$ Distâncias

β resolução

-problema fundamental o que significa distância? Tensor métrico

$$D_{ij} = g_{\mu\nu} (x_i^\mu - x_j^\mu) (x_i^\nu - x_j^\nu)$$

$\{D_{ij}\}$ Distâncias

β 1 / resolução

$s_i = \{1, 2, \dots, q\}$ rótulo das classes

Dadas as distâncias e a resolução:

Qual é a classe de um padrão?

-problema fundamental o que significa distância? Tensor métrico

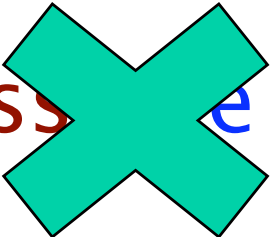
$$D_{ij} = g_{\mu\nu} (x_i^\mu - x_j^\mu) (x_i^\nu - x_j^\nu)$$

$\{D_{ij}\}$ Distâncias

β 1 / resolução

$s_i = \{1, 2, \dots, q\}$ rótulo das classes

Dadas as distâncias e a resolução:

Qual é a classe  e um padrão?

-problema fundamental o que significa distância? Tensor métrico

$$D_{ij} = g_{\mu\nu} (x_i^\mu - x_j^\mu) (x_i^\nu - x_j^\nu)$$

$\{D_{ij}\}$ Distâncias

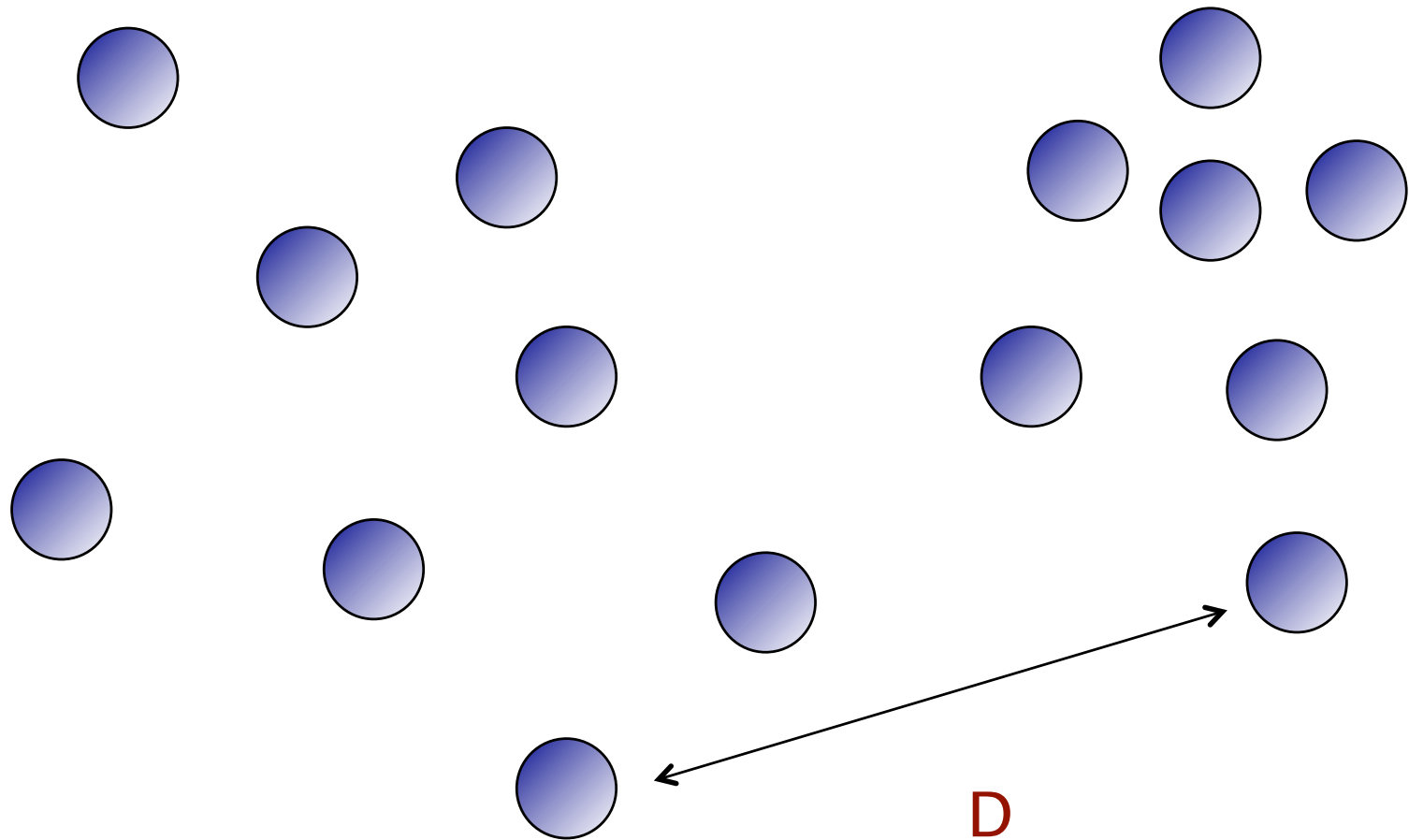
β 1 / resolução

$s_i = \{1, 2, \dots, q\}$ rótulo das classes

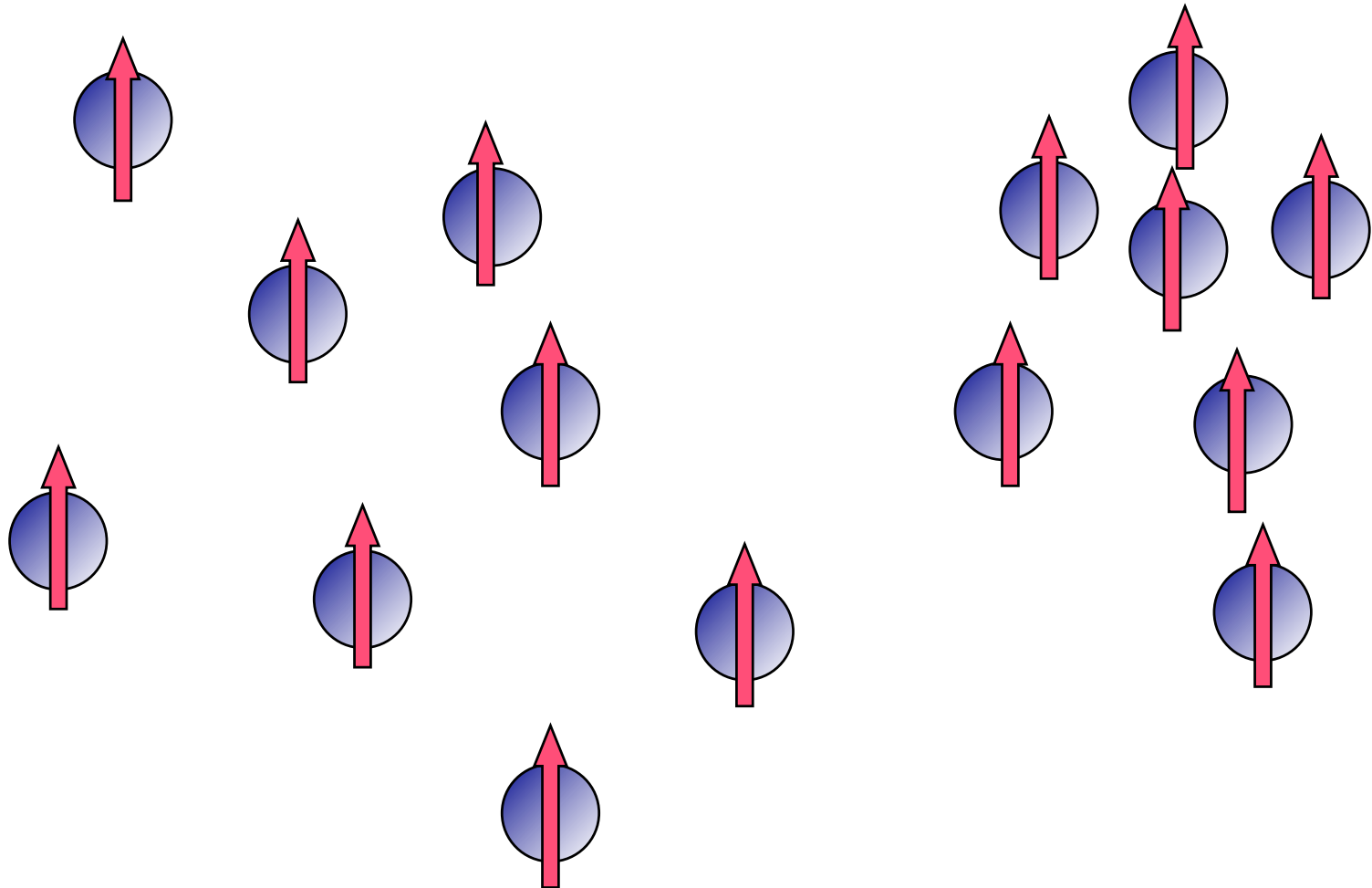
Dadas as distâncias e a resolução:

Qual é a probabilidade de que dois padrões pertençam à mesma classe?

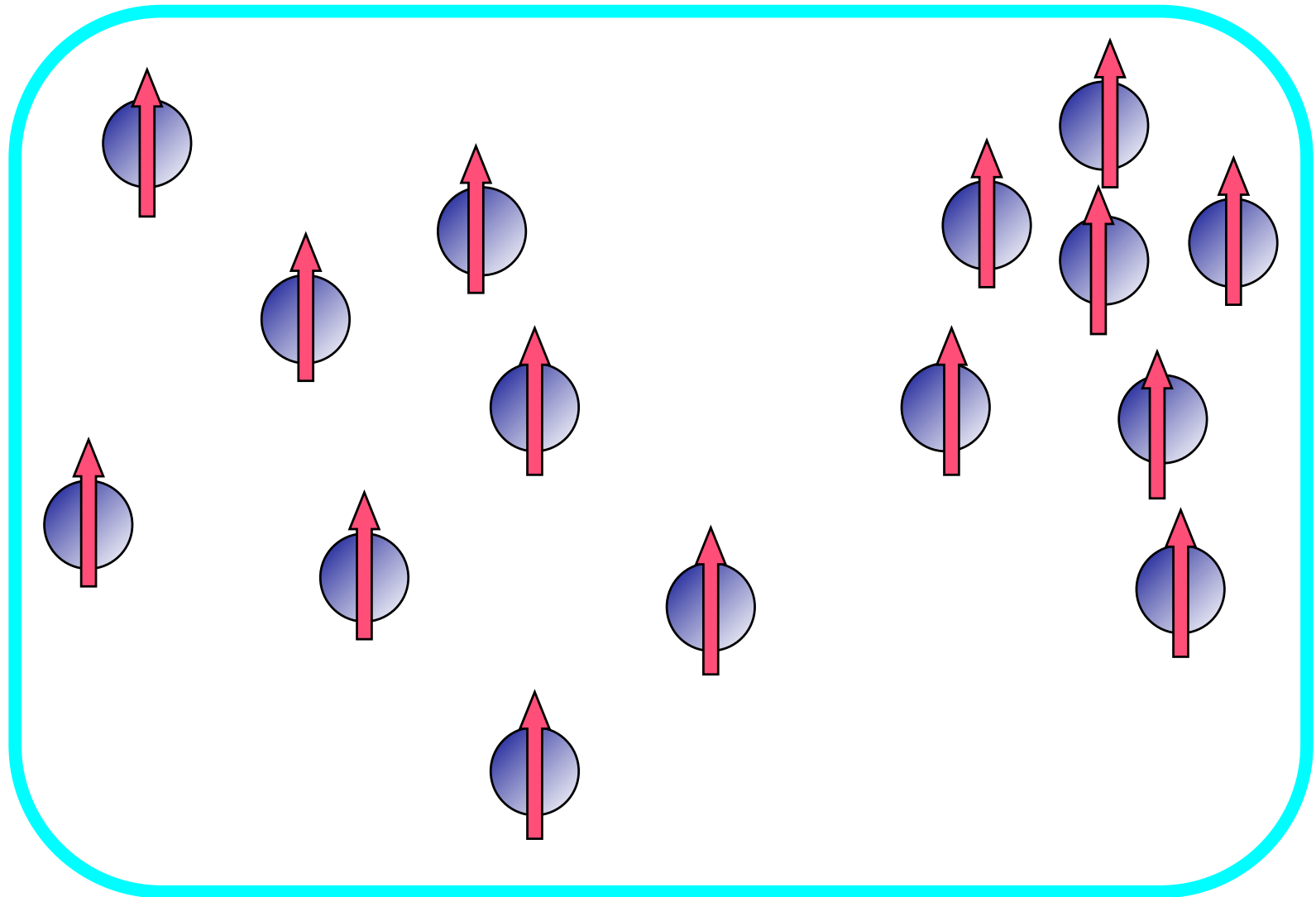
Não Supervisionado: estruturas?



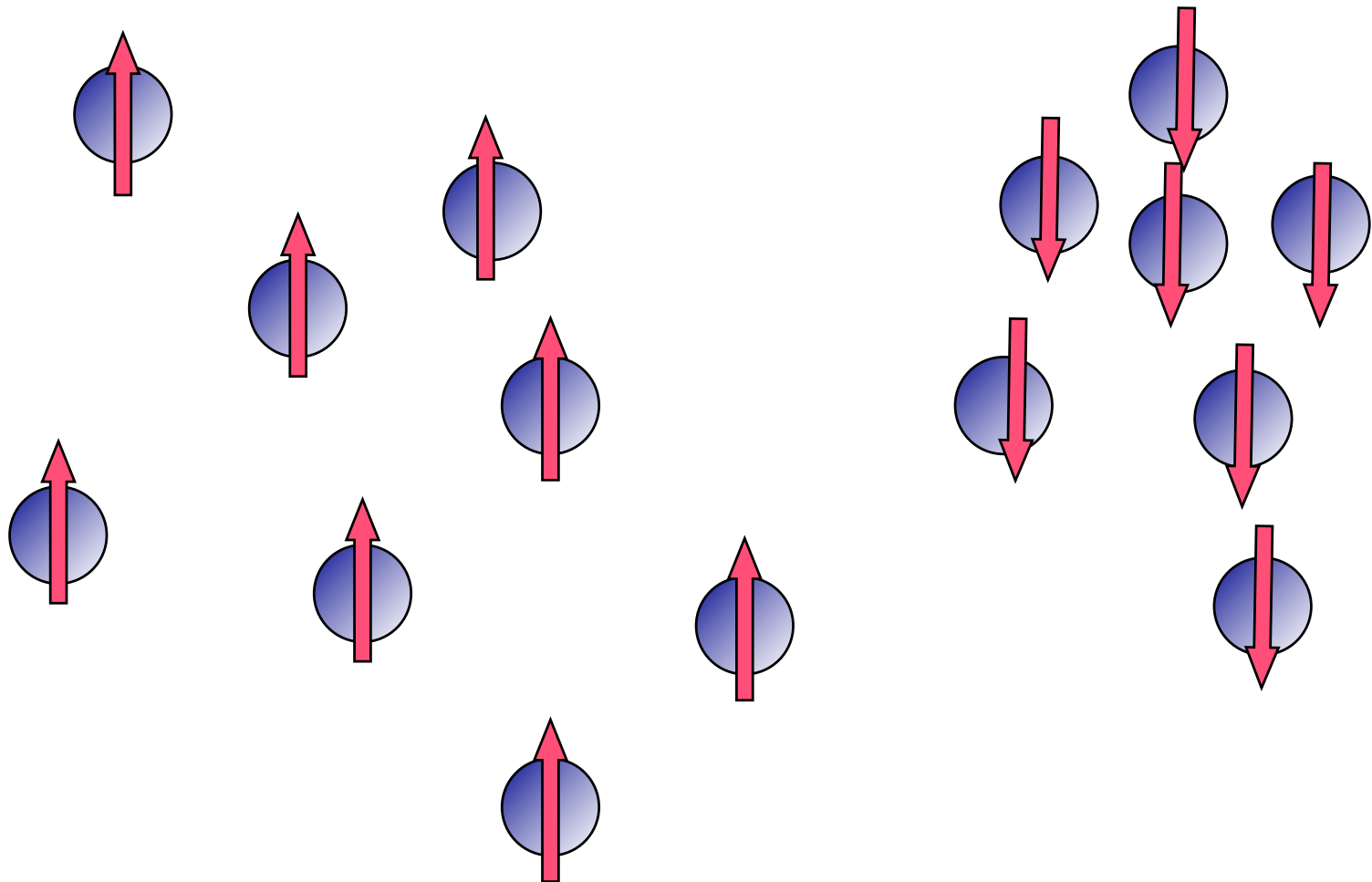
Não Supervisionado: estruturas?



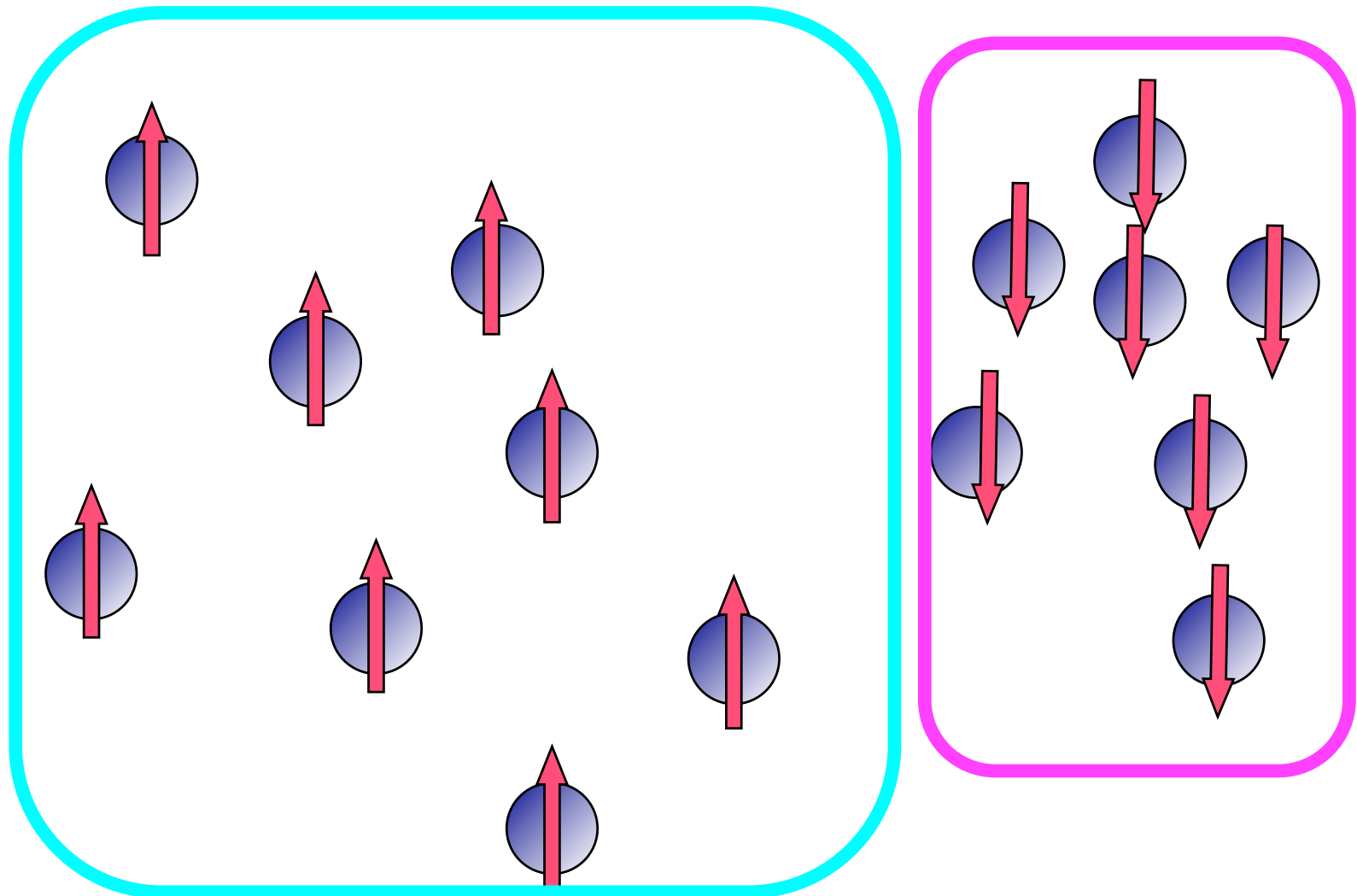
Não Supervisionado: estruturas?



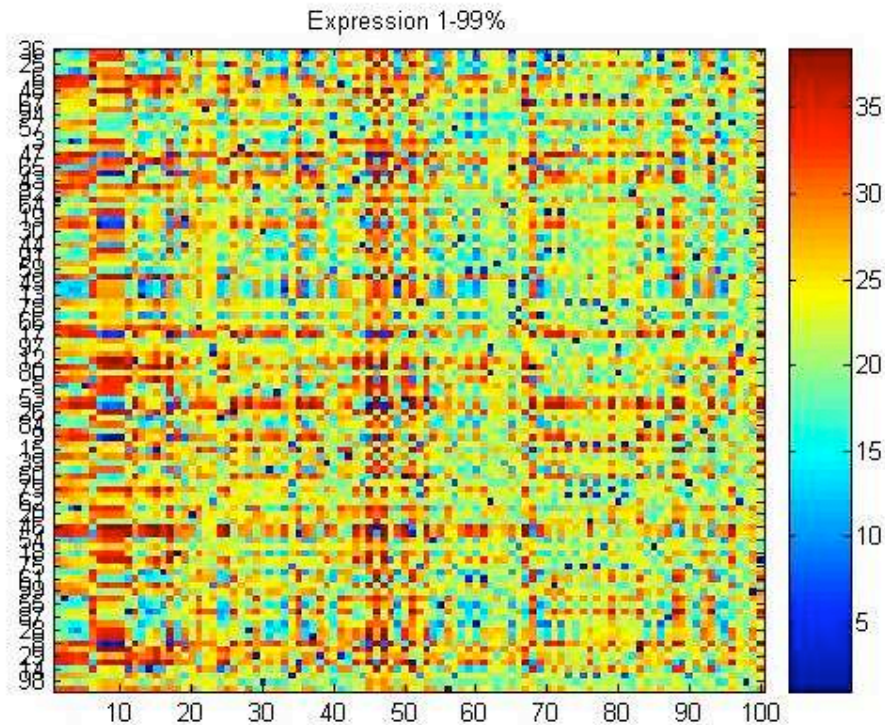
Não Supervisionado: estruturas?



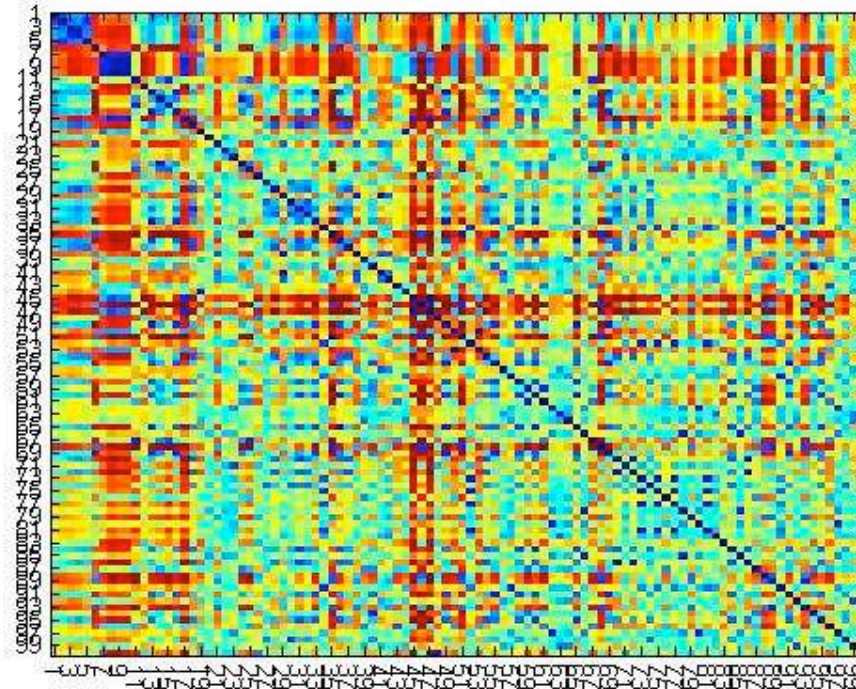
Não Supervisionado: estruturas?



Gene Expression profile

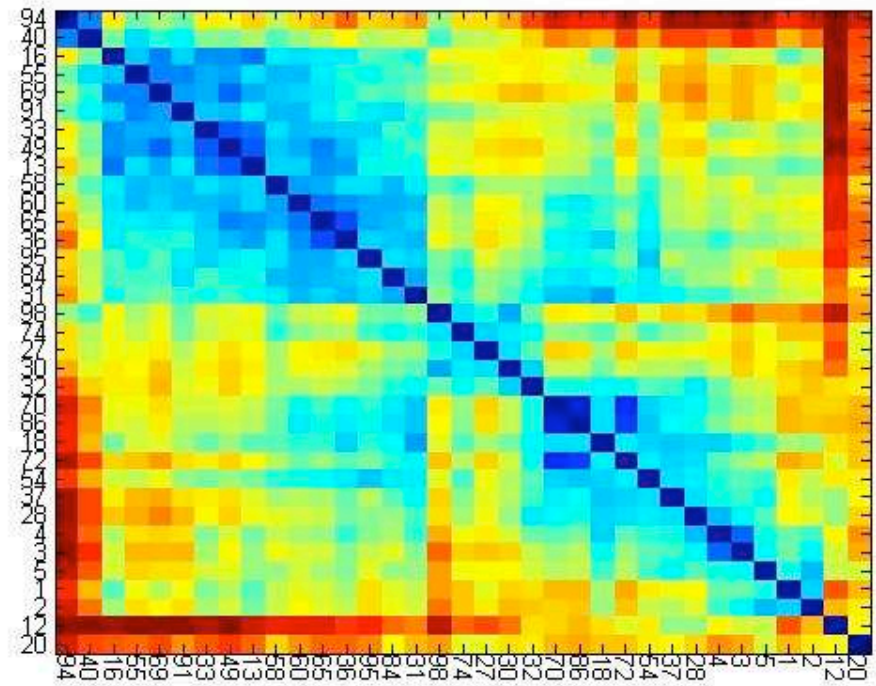
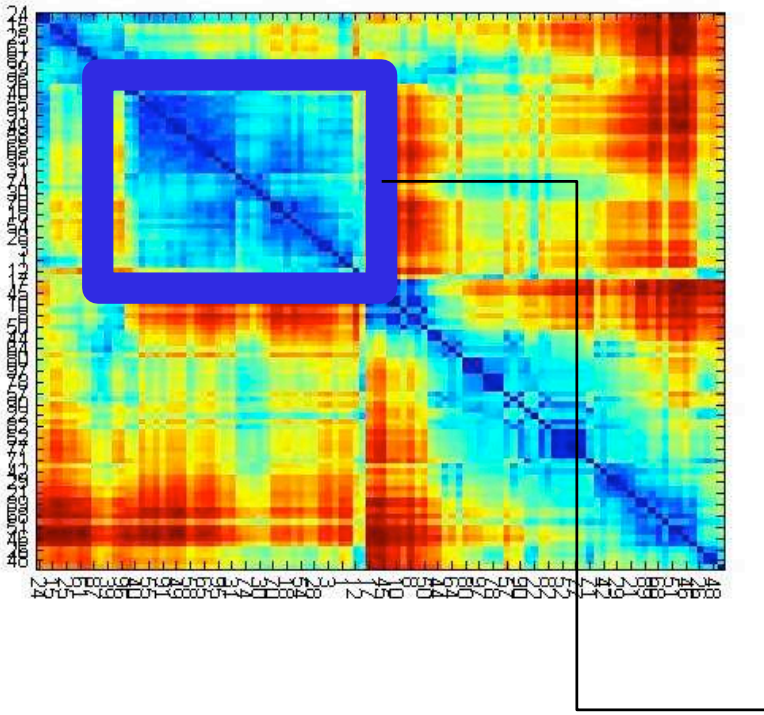


Colunas = amostras de tecidos
Linhas = expressão gênica



Gene distance

Coloca uma interação entre genes tal que a posição na matriz depende da distância no espaço génico



SPINned distance matrix zoomed

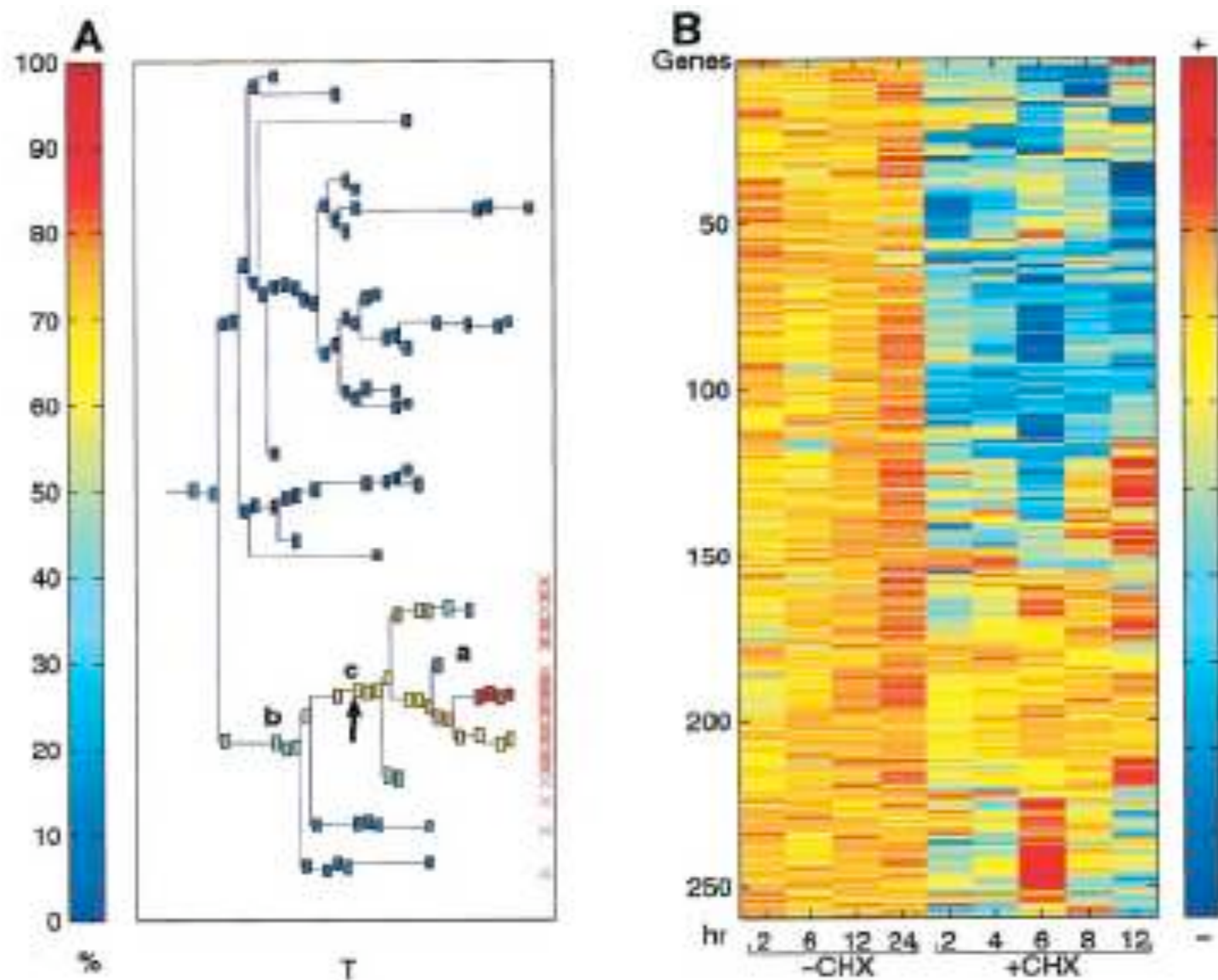
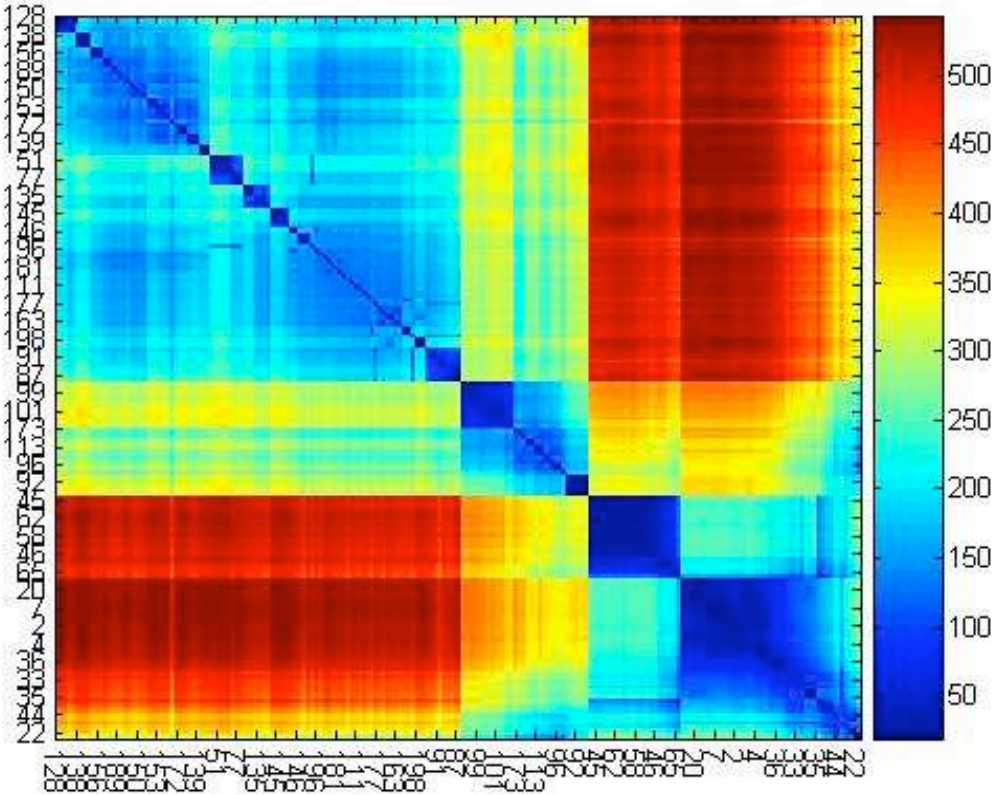


Figure 3 Clustering results using super-paramagnetic clustering (SPC) for the 259 genes that were upregulated at three time points or more upon activation of p53 in the experiment without cycloheximide (CHX). (A) The dendrogram of the genes that include clusters of size 4 and larger. Each cluster is represented by a box colored according to the per cent of primary target genes (38 genes, see Table 1) contained in the cluster. The distribution of the 38 primary target genes is marked by red crosses at the right. (B) The normalized log ratio of the nine experiments (four without CHX and five with CHX) are plotted. The genes are ordered according to the dendrogram on the left. The color represents induction (red) or repression (blue). T, a parameter of the SPC algorithm that controls the resolution at which the cluster is found. %, per cent of primary target genes in the cluster. The cluster marked by an arrow (c) contains 87% of the 38 primary genes. The cluster marked by b contains all the 38 primary genes and the cluster marked by a contains the nine genes that show different kinetics (Table 1 and Figure 4)

Spinned Homologies matrix



Learning Vector Quantization (LVQ)

- identification of prototype vectors from labelled example data
- parameterization of distance based classification (e.g. Euclidean)

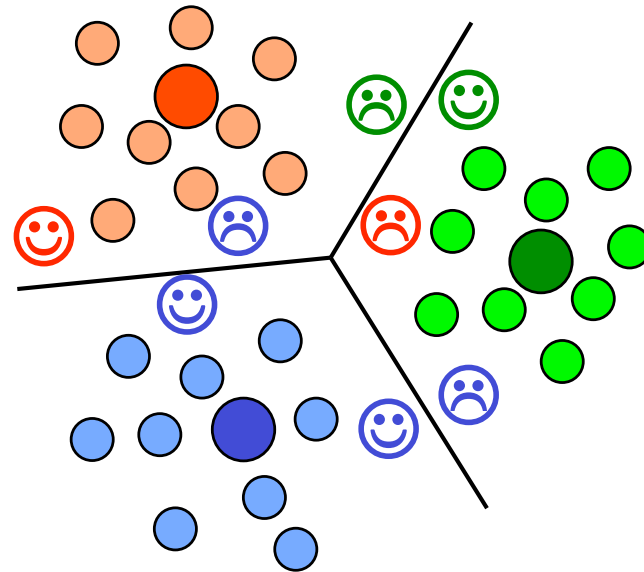
often: heuristically motivated variations of competitive learning

example: basic LVQ scheme “LVQ1” [Kohonen]

classification:

Atribuição de vetor \mathbf{x}
à classe do prototipo \mathbf{w}
mais próximo

objetivo: habilidade de generalização
Classificação de novos dados
após treinamento



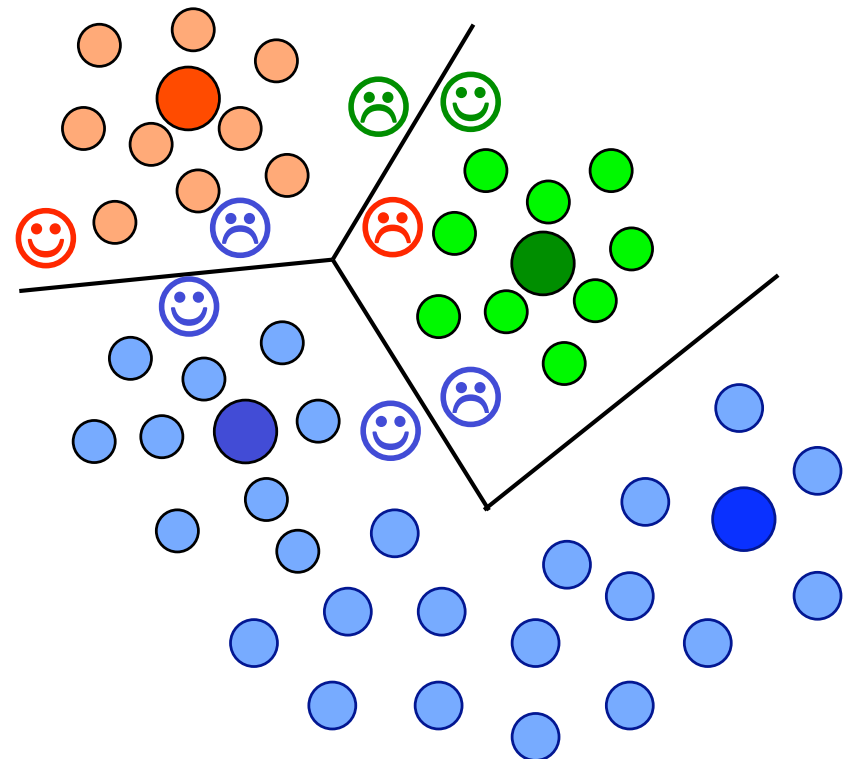
Learning Vector Quantization (LVQ) supervisionado

- Identificação de protótipo exemplo rotulado

classificação:

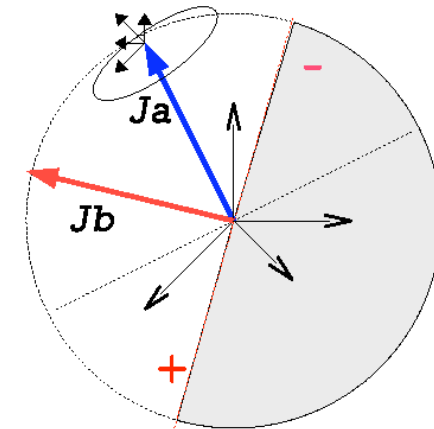
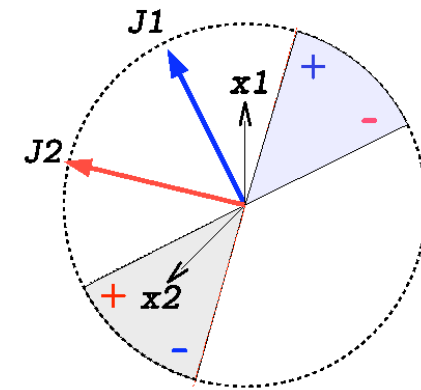
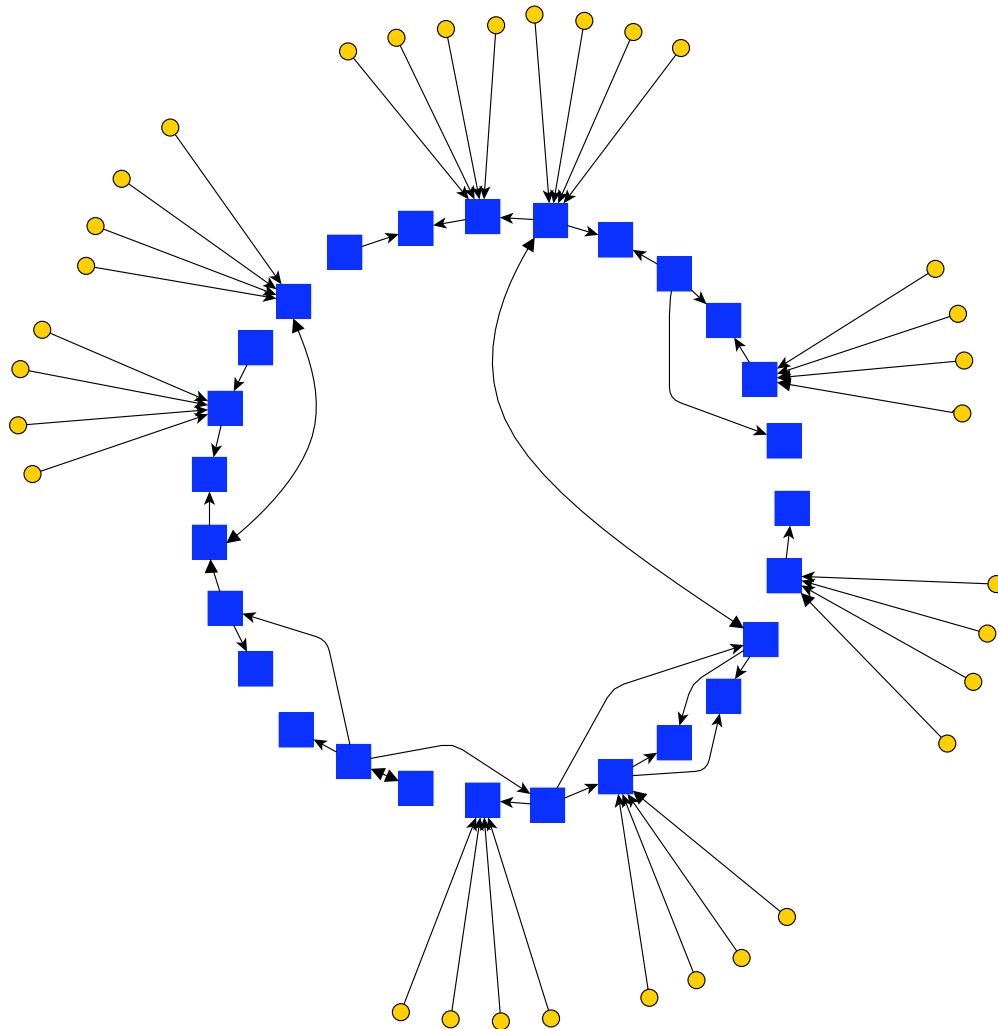
assignment of a vector ξ
to the class of the closest
prototype w

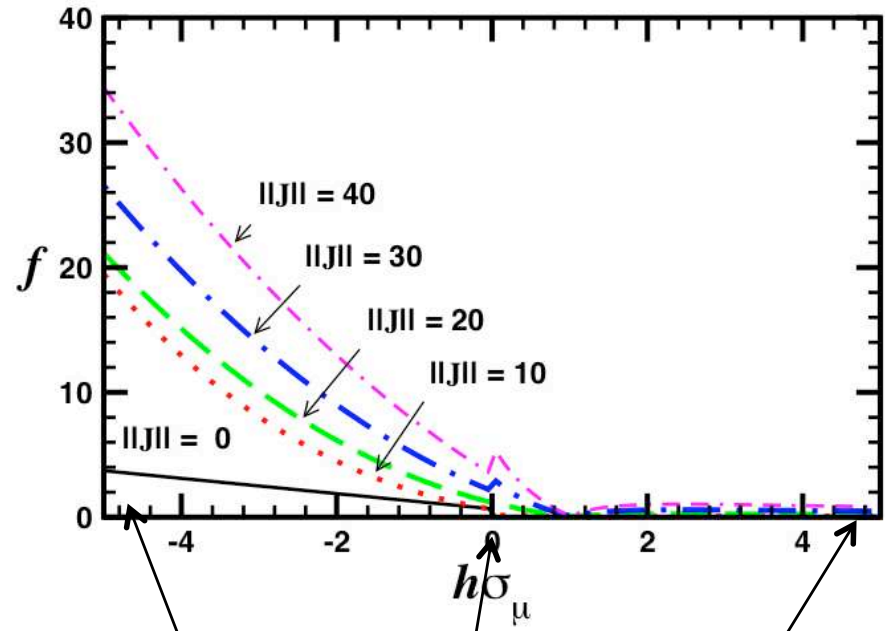
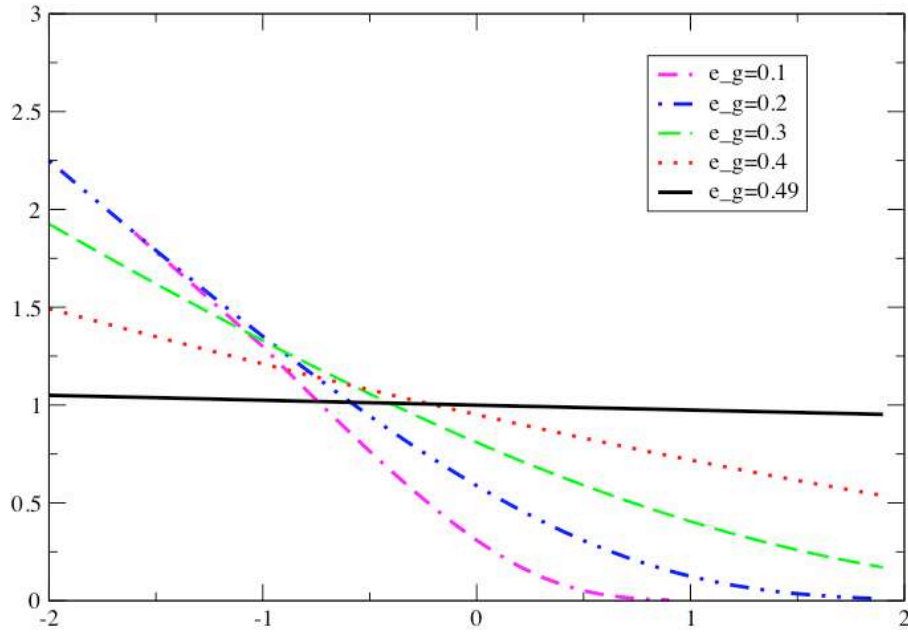
aim: **generalization ability**
classification of novel data
after learning from examples



piecewise linear decision boundaries 96

Dinâmica de Opiniões: Perceptopolis

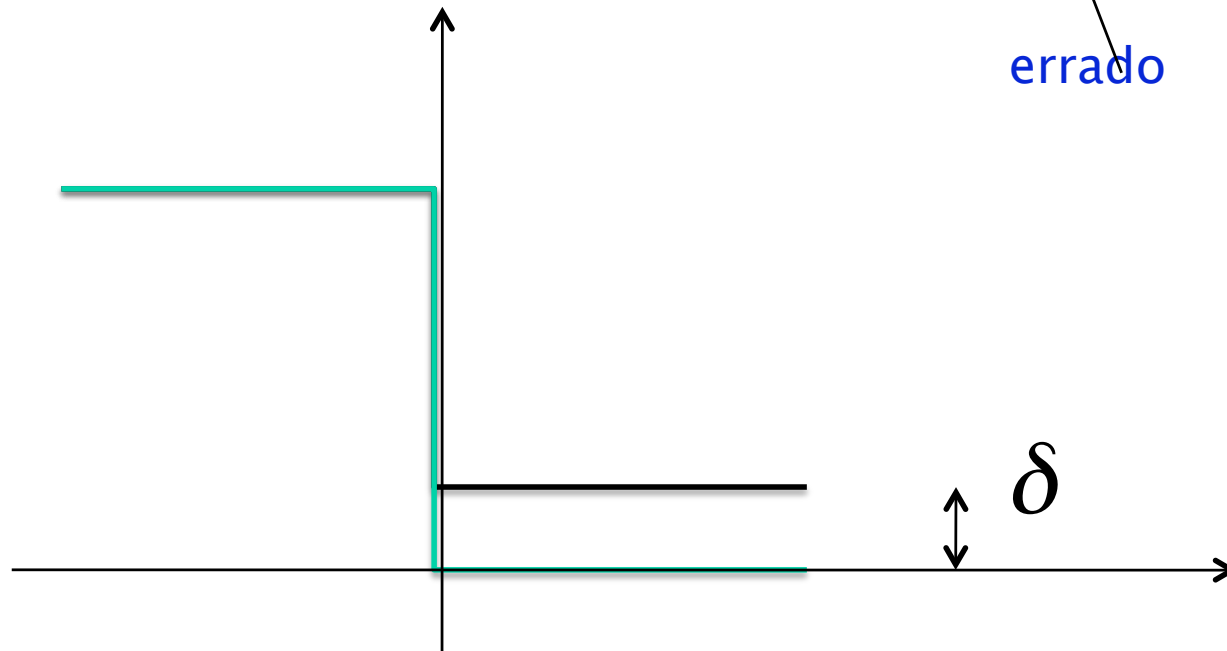




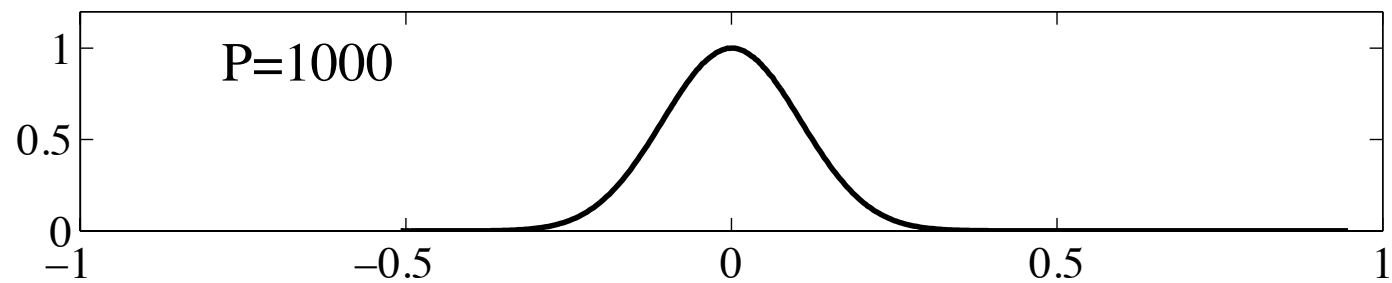
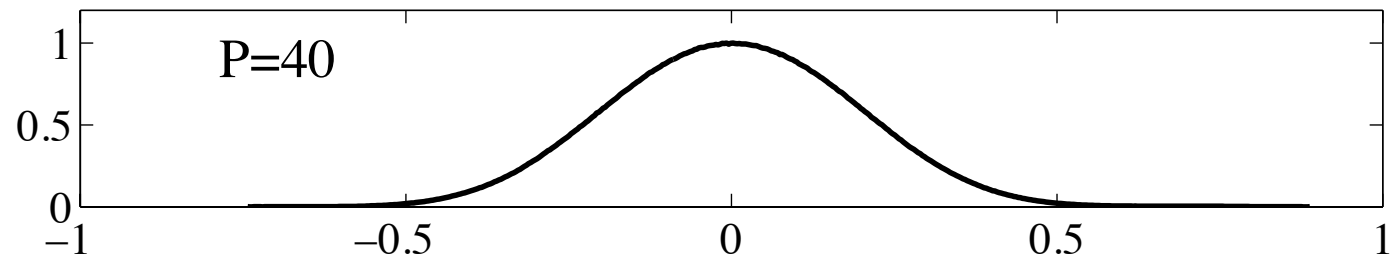
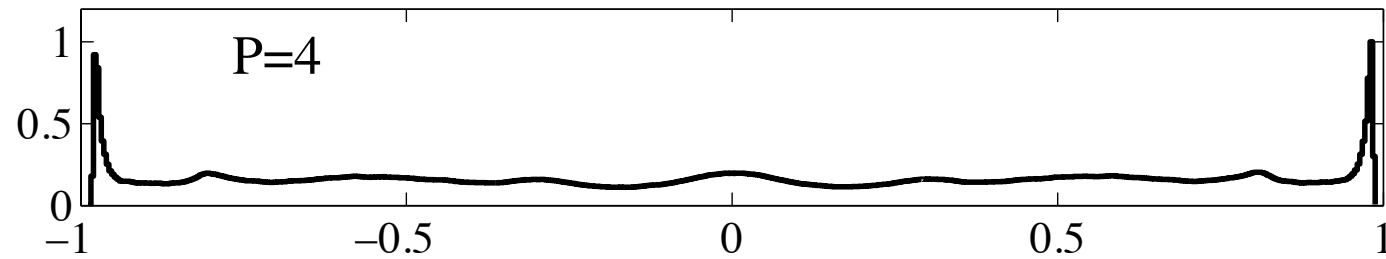
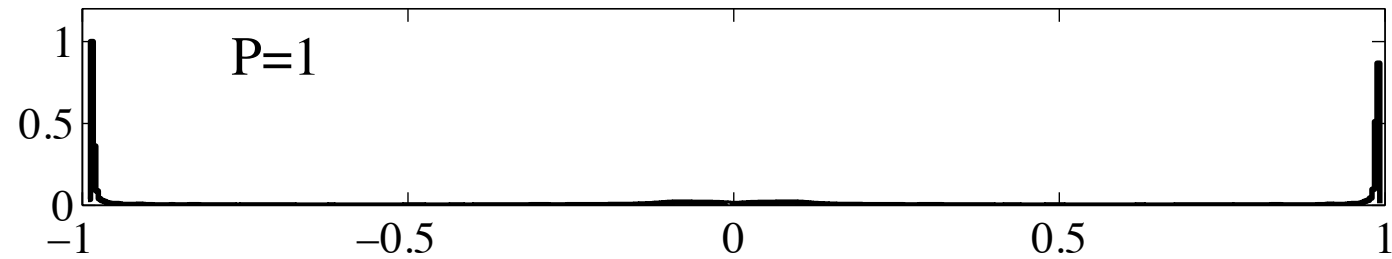
errado

difícil

Certo



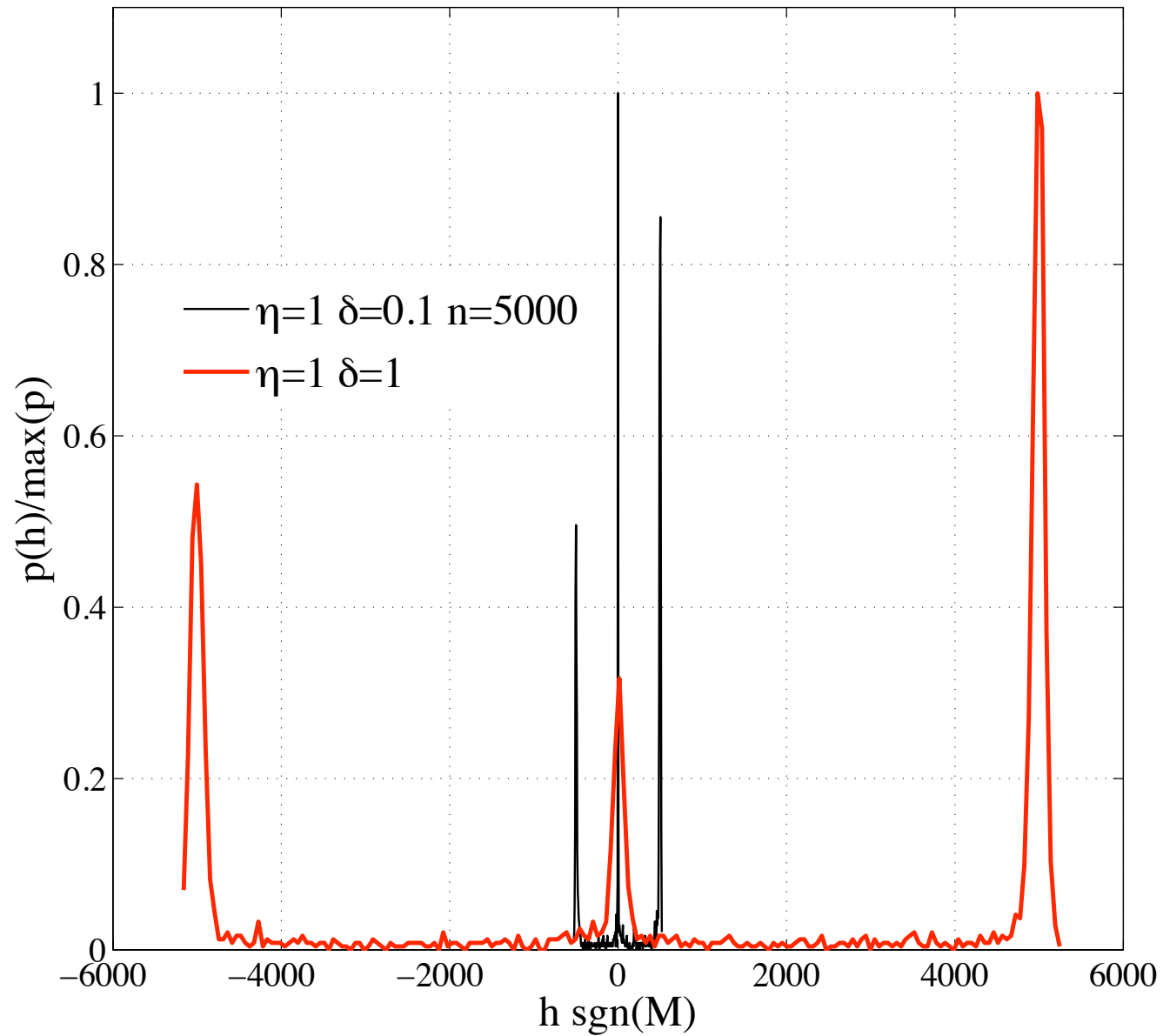
facções



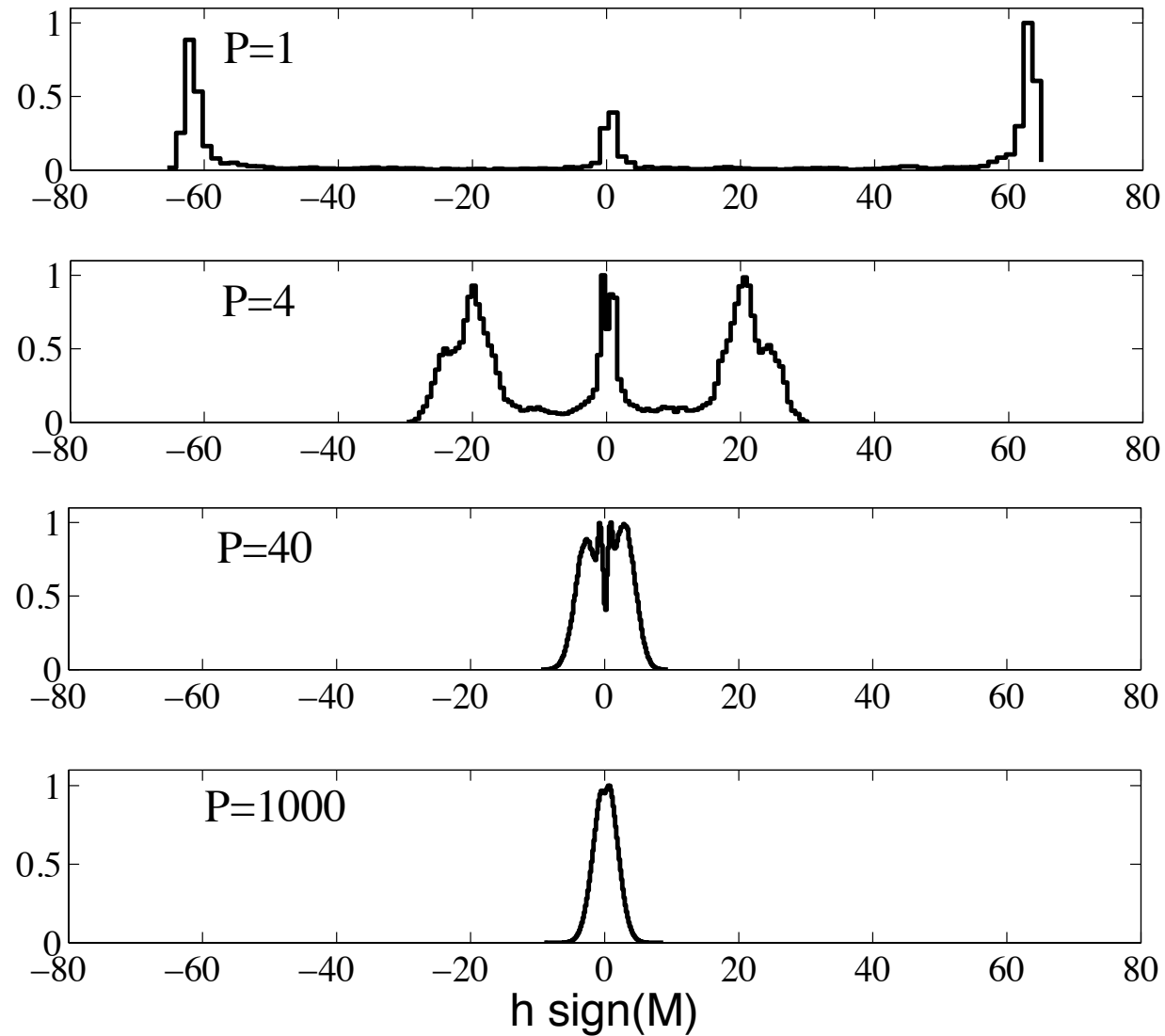
vítreo

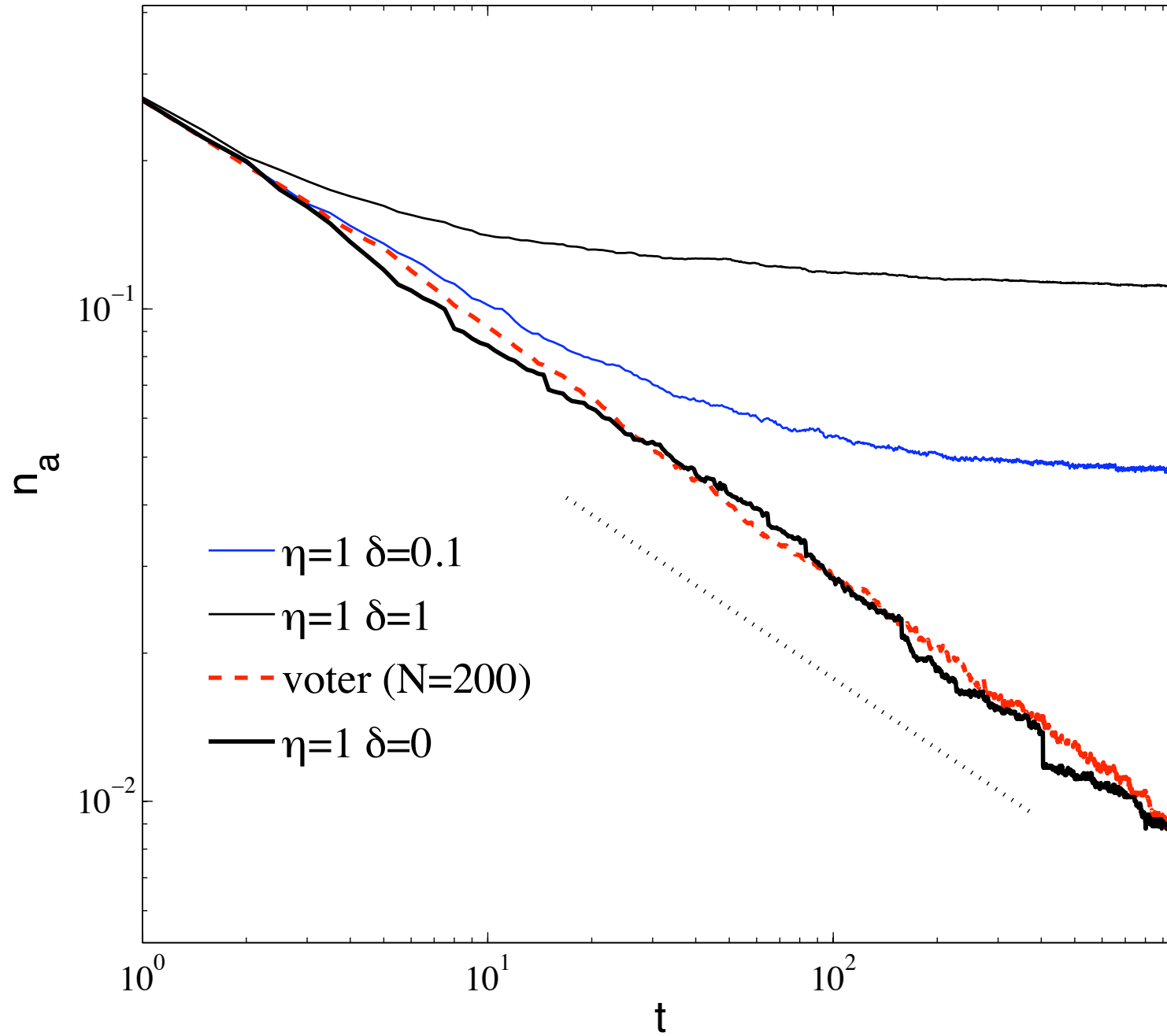
ρ

Perceptopolis

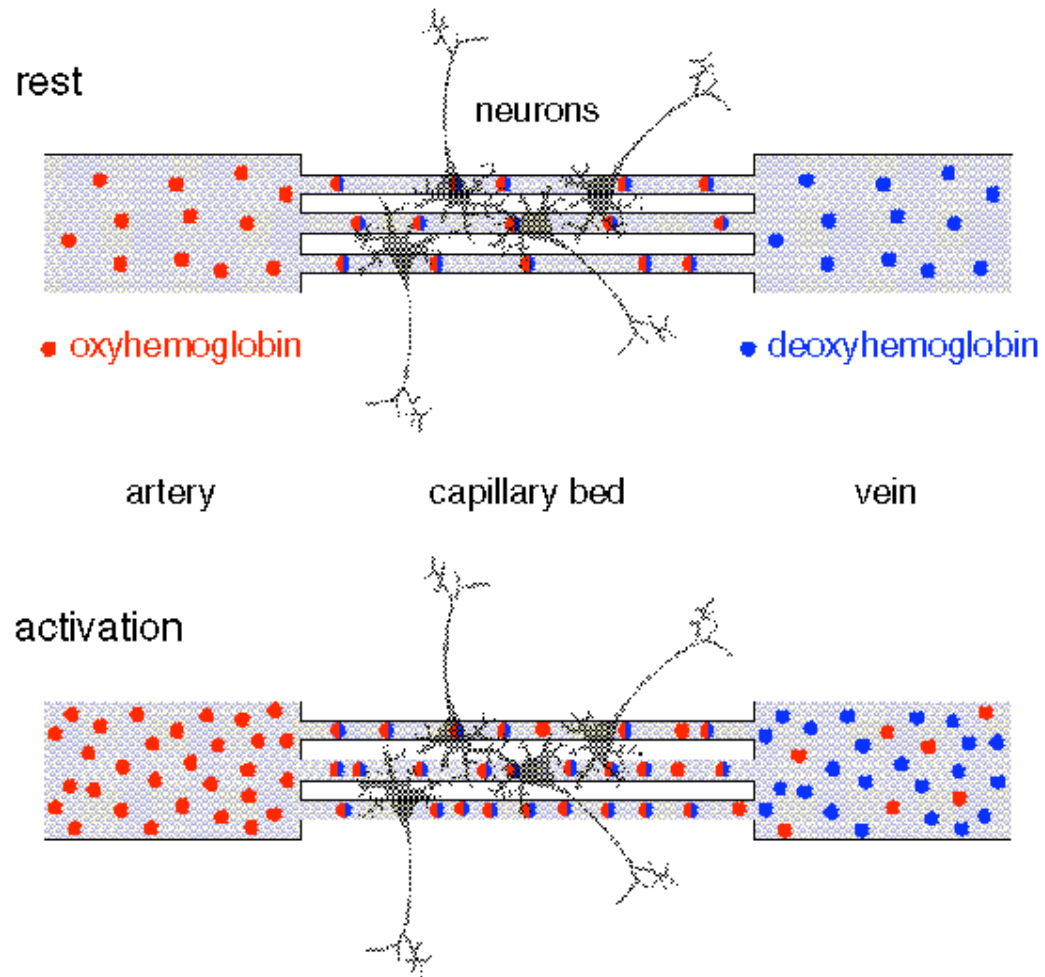


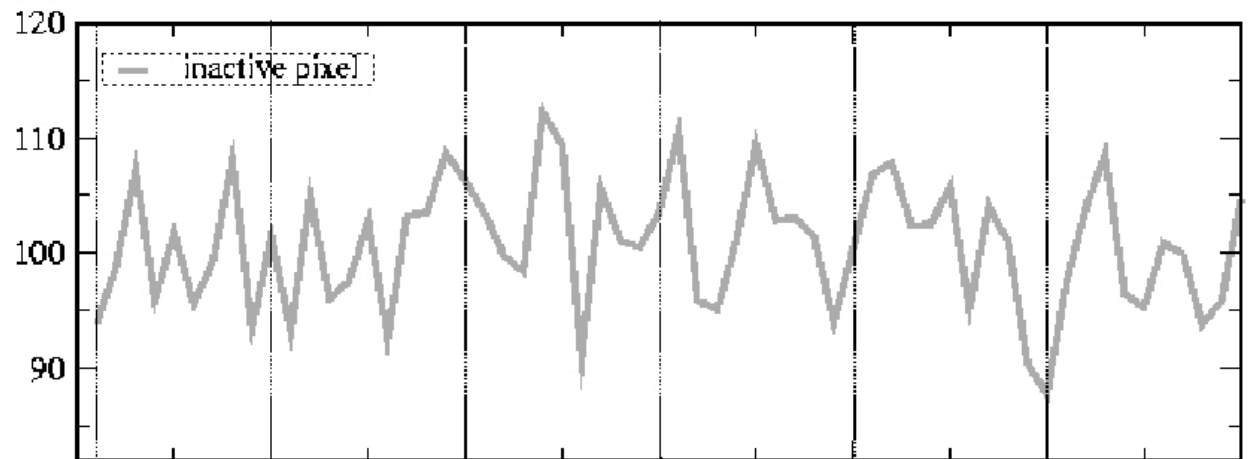
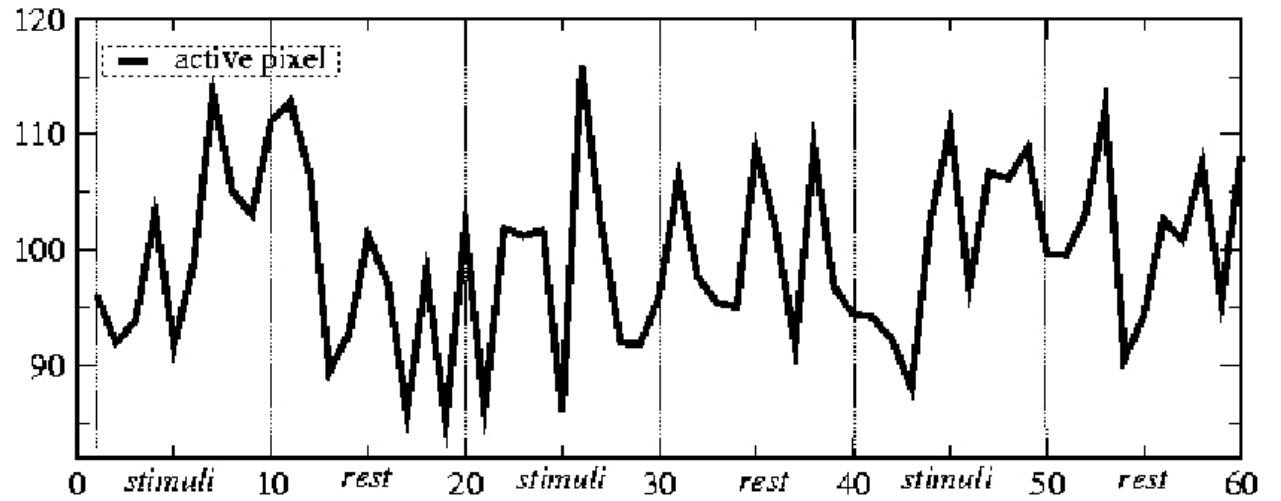
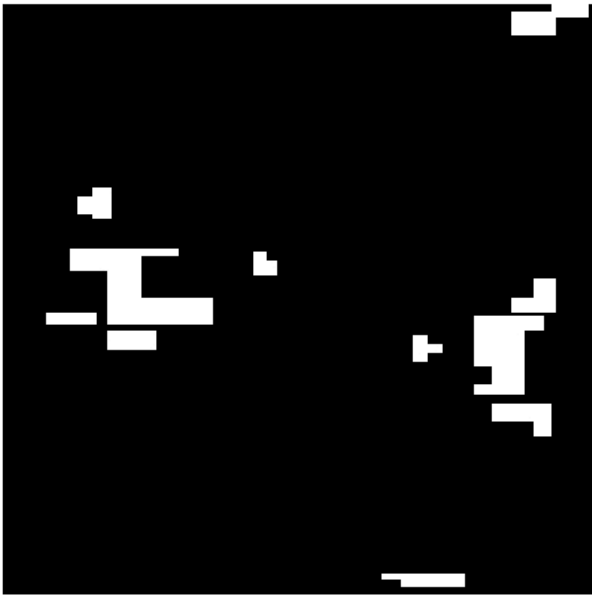
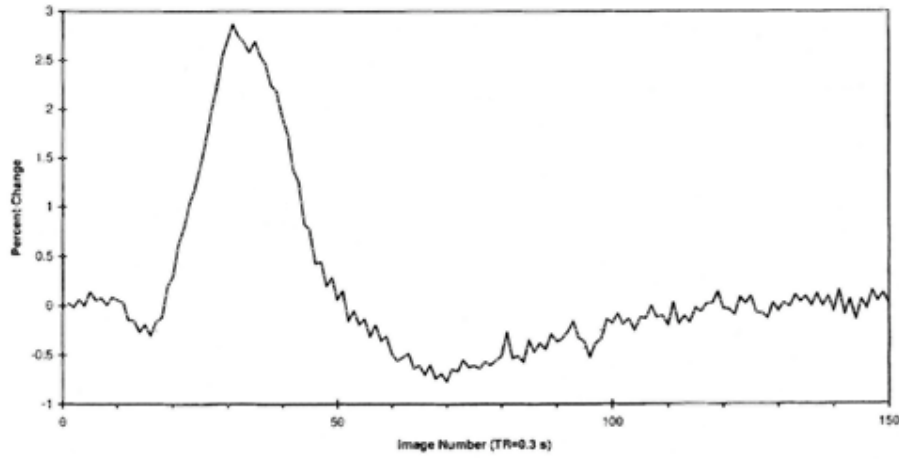
Perceptopolis

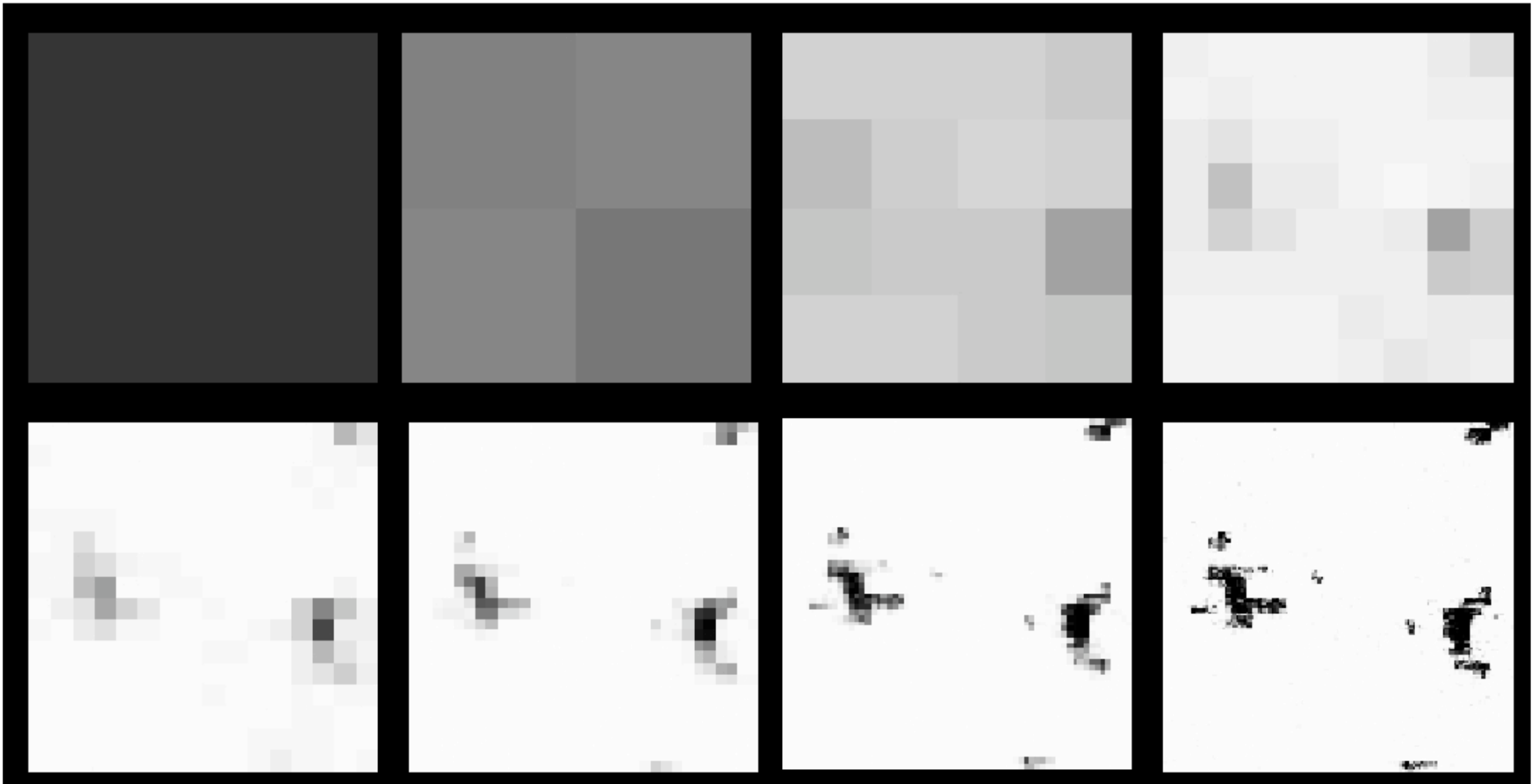


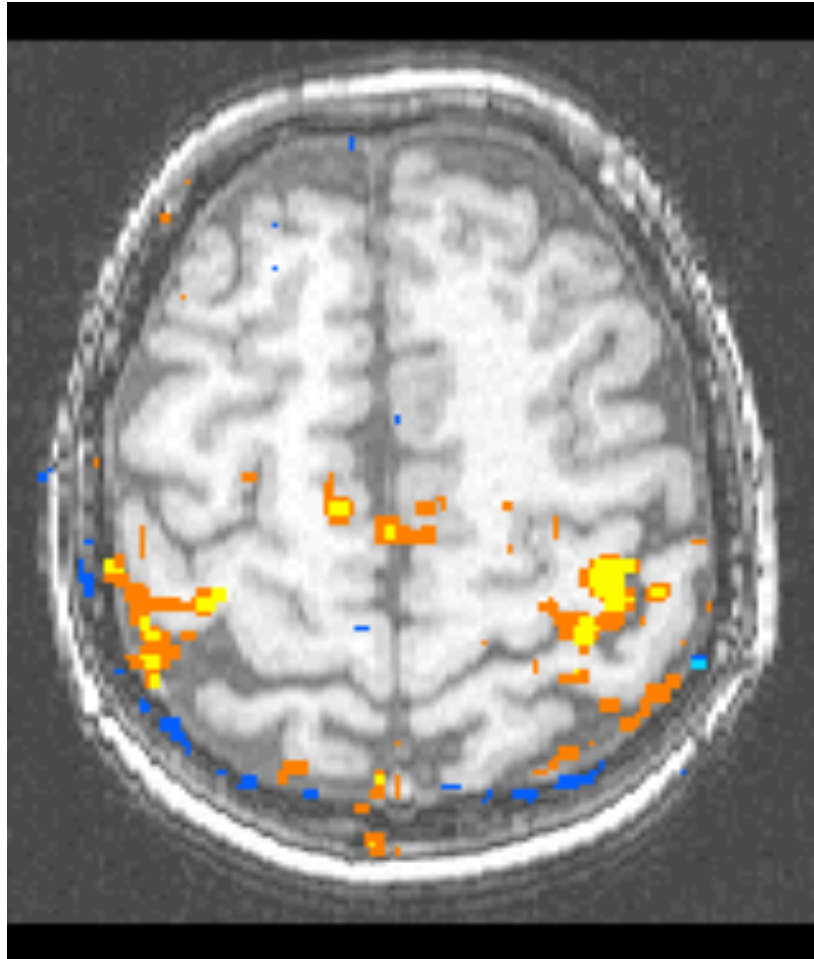


Imagens por
Ressonância
Magnética
funcional









Regiões ativas
durante movimento
de dedos

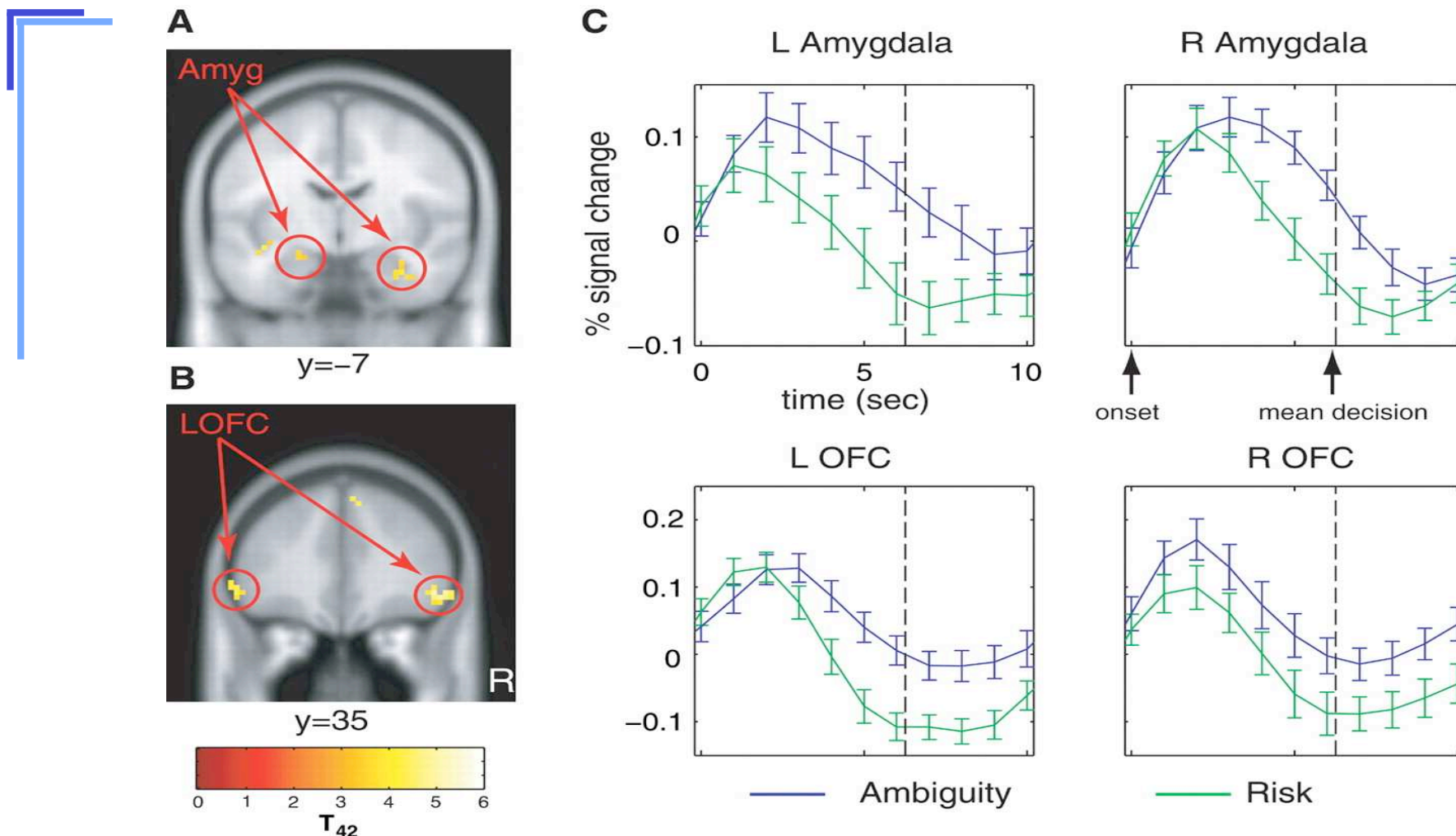


Fig. 2. Regions showing greater activation in response to **ambiguity** than in response to **risk**. Random-effects analysis of all three treatments revealed regions that are differentially activated in decision-making under ambiguity relative to risk ($P \leq 0.001$, uncorrected; cluster size $k \geq 10$ voxels). These regions include (A) left amygdala and right amygdala/parahippocampal gyrus (coronal section shown at $y = 7$ in MNI space; heat map represents t statistic with 42 degrees of freedom) and (B) bilateral OFC. (C) Mean time courses of amygdala and OFC (time synched to trial onset, dashed vertical lines are mean decision times; error bars are SEM; $n = 16$). Hsu et al Science 2005



vimos

- Teoria de informação porque devemos usar **probabilidades**,
- Como atualizar pdf porque entropia
- Limites, evolução

Aplicadas:

- Classificadores (supervisionados)
- Categorização (não supervisionada) Transições de fase
- segmentação de imagens, identificação **genes (cancer)**, **séries temporais (economia, fMRI) estudo do cérebro que faz a inferência**



Colaboradores

- Ariel Caticha
- Osame Kinouchi
- Renato Vicente
- Marcus V. Baldo
- Said Rabbani
- J.P. Neirotti
- Vitor B. P. Leite
- Roberto Alamino
- Michel Biehl
- Peter Riegler
- M. Copelli
- Selene Amaral
- André Martins
- Fabiano Ribeiro