UFES – NINFA

Introdução aos Conceitos de Classificação Baseada em Padrões

Francisco de Assis Boldt

- Professor do IFES – Coordenadoria de Informática
- Mestre em Informática pela UFES (2008)
- Candidato a aluno de Doutorado pela UFES.

Aprendizagem de máquina

- **Aprendizado Supervisionado**
  - **Reconhecimento de Padrões**
    - **Classificação**
      - **Redes Neurais**
- Aprendizado Não Supervissionado
- Aprendizado por Reforço
- Transdução
- Aprendizado Multitarefa

Histórico

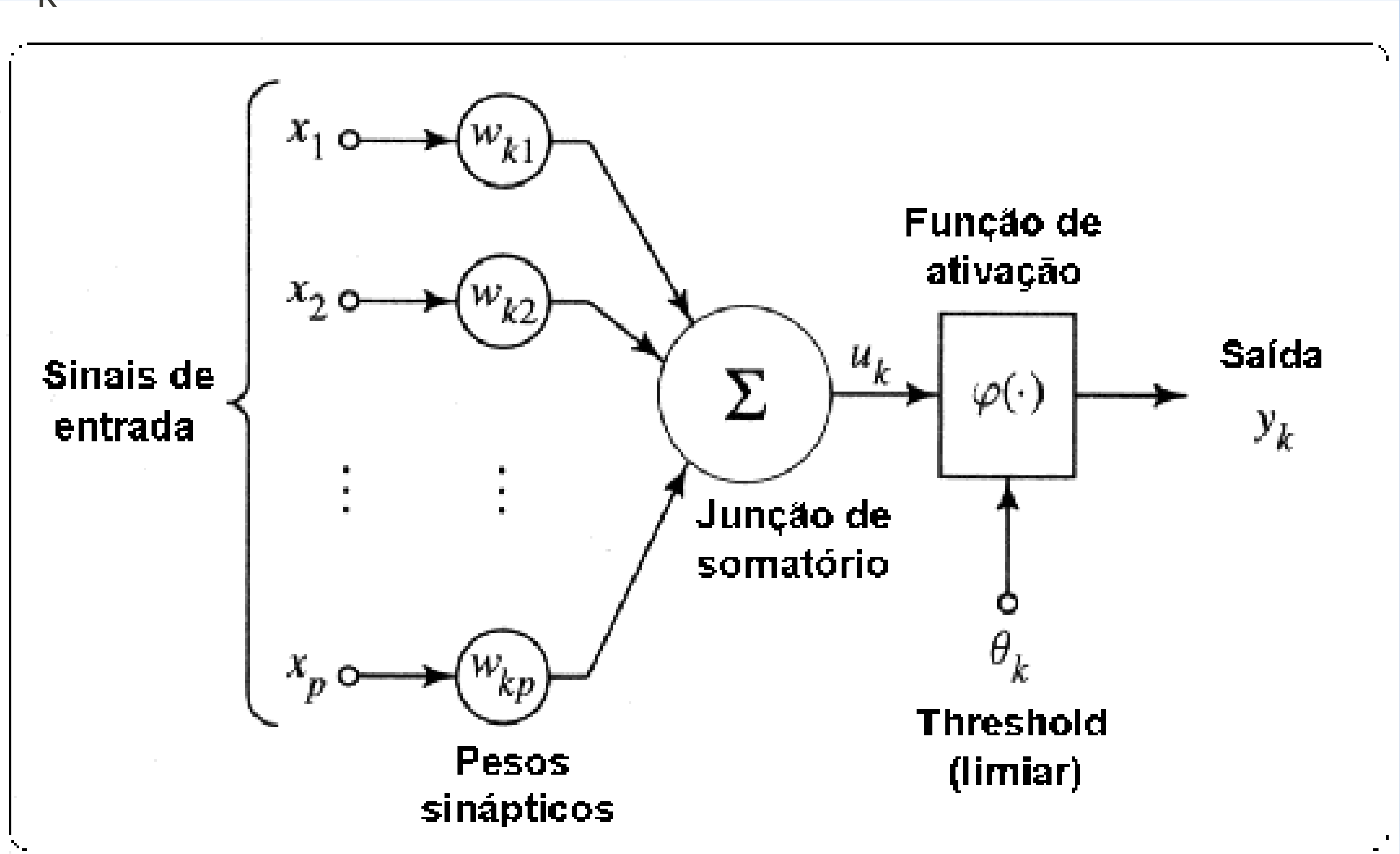
- 1943 – Warren McCulloch, Walter Pitts
  - Primeiro modelo artificial de um neurônio biológico.
- 1949 – Donald Hebb
  - Primeiro trabalho com ligação direta com aprendizado. (regra de Hebb)
- 1958 – Frank Rosenblat
  - Perceptron (retina, peso, resposta)
- 1960 – Widrow e Hoff
  - Regra Delta

Histórico (cont...)

- 1969 – Minsky e Papert
  - Limitações do Perceptron.
- 1982 – John Hopfield
  - Retomada das pesquisas na área.
- 1986 -
  - Back-propagation

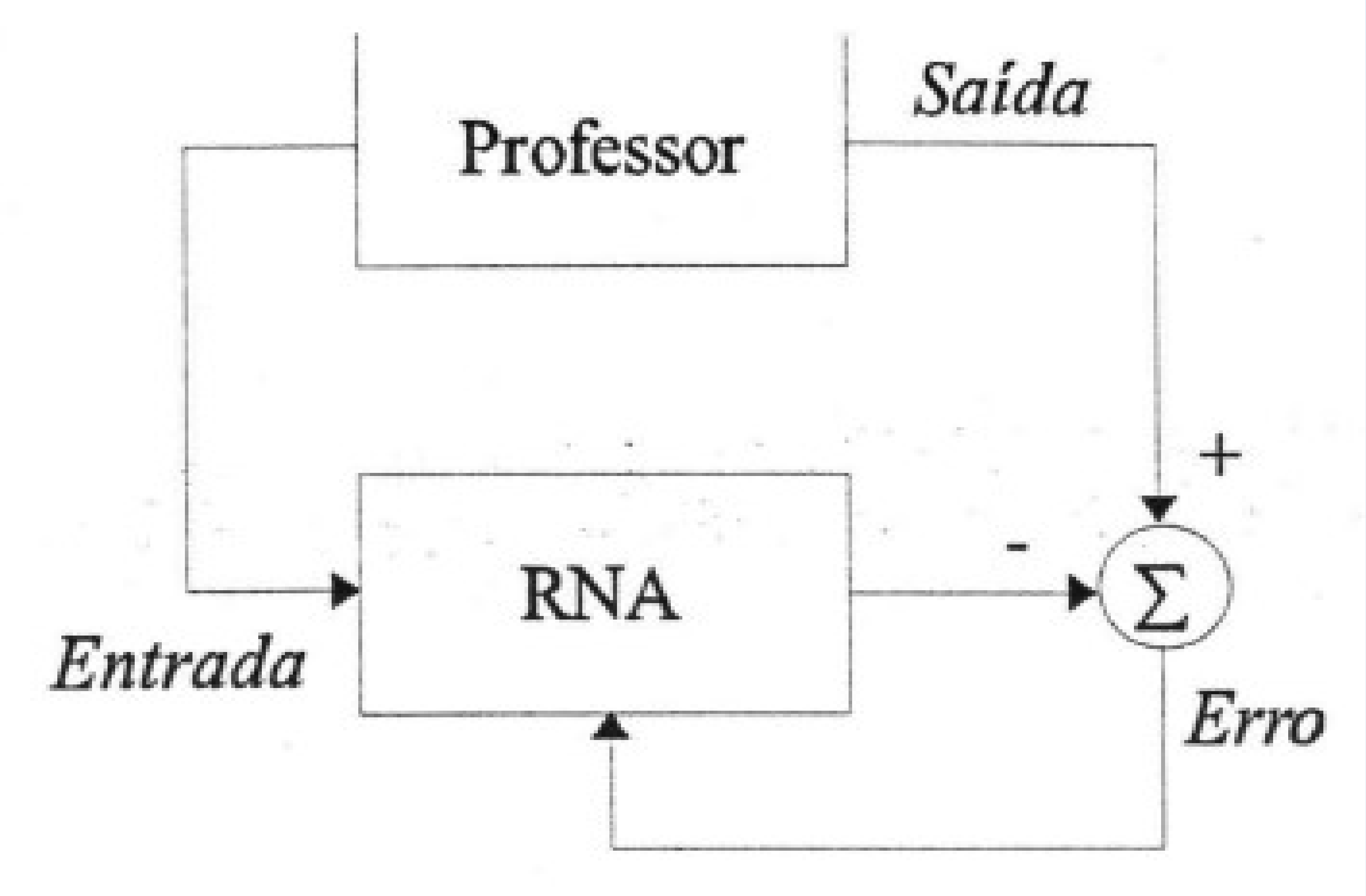
Neurônio de McCulloch e Pitts

- $y_k = f(x^t * w)$, ex.: $f(x) = x > \Theta$



Aprendizado Supervionado

- $w_i(t + 1) = w_i(t) + \eta e x_i$



Exemplo: AND

$w_i(t+1) = w_i(t) + \eta e(t) x_i$

$\Theta = 0.1$

$\eta = 0.05$

$f(x) = 1$ para $x > \Theta$, 0 c.c.

padrão	x1	x2	y (d)
1	0	0	0
2	0	1	0
3	1	0	0
4	1	1	1

t	w1	w2
0	0	0
1	0	0
2	0	0
3	0	0
4	0.05	0.05
5	0.05	0.05
6	0.05	0.05
7	0.05	0.05
8	0.1	0.1
9	0.1	0.1
10	0.1	0.1
11	0.1	0.1

padrão	Σxi*wi	f(Σxi*wi)=y(o)	Erro (d-o)
1	0	0	0
2	0	0	0
3	0	0	0
4	0	0	1
1	0	0	0
2	0.05	0	0
3	0.05	0	0
4	0.1	0	1
1	0	0	0
2	0.1	0	0
3	0.1	0	0
4	0.2	1	0

Exemplo: OR

$w_i(t + 1) = w_i(t) + \eta e(t)x_i$

$\Theta = 0.1$

$\eta = 0.05$

$f(x) = 1$ para $x > \Theta$, 0 c.c.

padrão	x1	x2	y (d)
1	0	0	0
2	0	1	1
3	1	0	1
4	1	1	1

t	w1	w2
0	0	0
1	0	0
2	0	0.05
3	0.05	0.05
4	0.1	0.1
5	0.1	0.1
6	0.1	0.15
7	0.15	0.15
8	0.15	0.15
9	0.15	0.15
10	0.15	0.15
11	0.15	0.15

padrão	Σxi*wi	f(Σxi*wi)=y(o)	Erro (d-o)
1	0	0	0
2	0	0	1
3	0	0	1
4	0.1	0	1
1	0	0	0
2	0.1	0	1
3	0.1	0	1
4	0.3	1	0
1	0	0	0
2	0.15	1	0
3	0.15	1	0
4	0.3	1	0

Exemplo: XOR

$w_i(t + 1) = w_i(t) + \eta e(t)x_i$

$\Theta = 0.1$

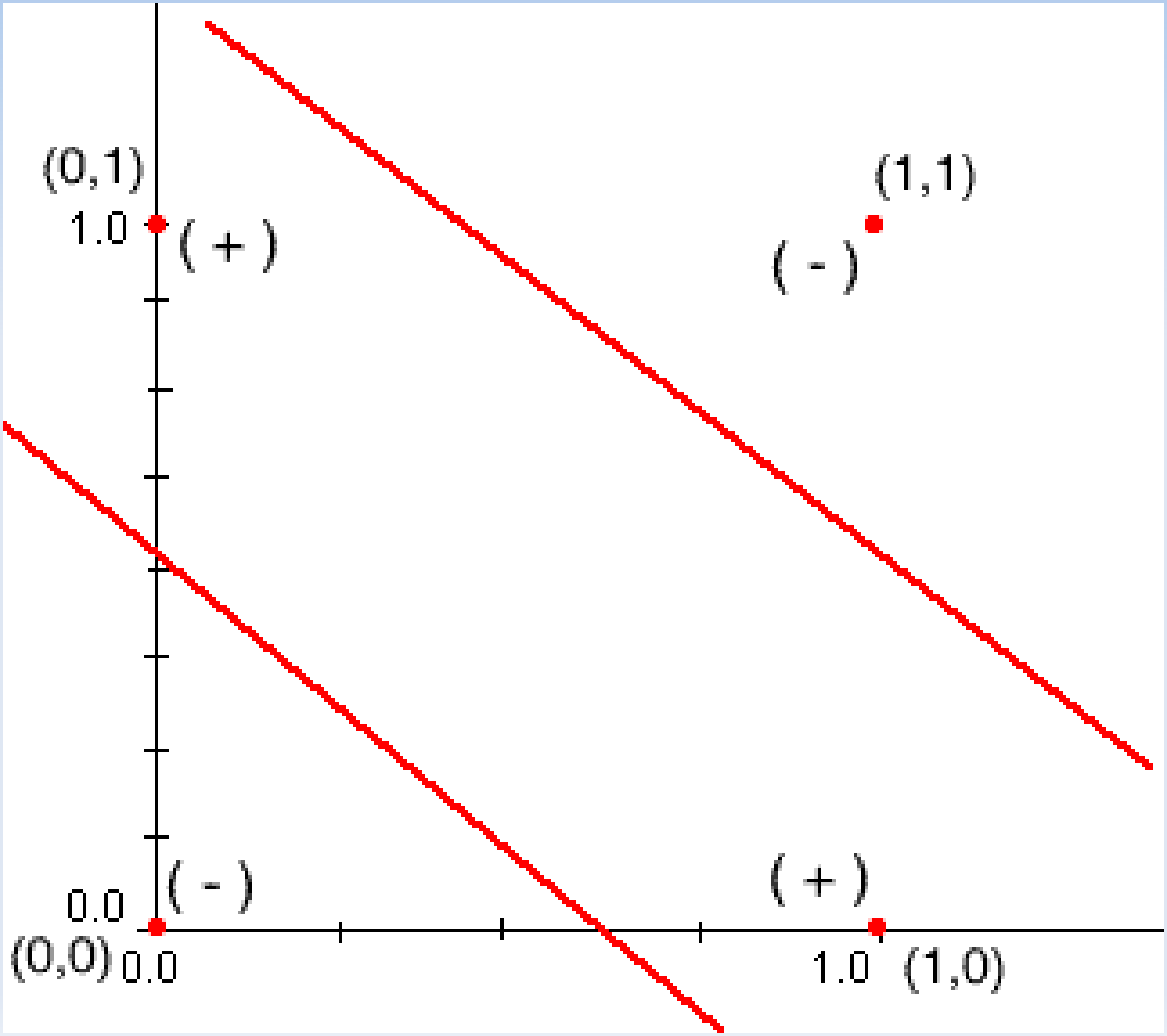
$\eta = 0.05$

$f(x) = 1$ para $x > \Theta$, 0 c.c.

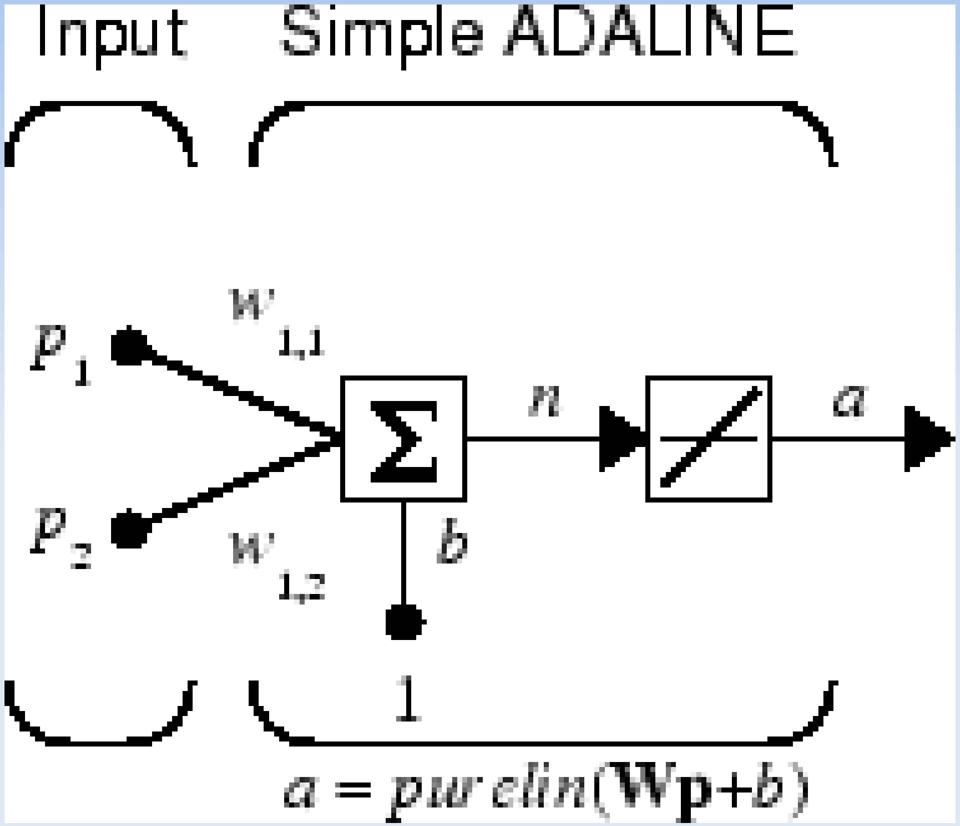
padrão	x1	x2	y (d)
1	0	0	0
2	0	1	1
3	1	0	1
4	1	1	0

t	w1	w2	padrão	Σxi*wi	f(Σxi*wi)=y(o)	Erro (d-o)
0	0	0	1	0	0	0
1	0	0	2	0	0	1
2	0	0.05	3	0	0	1
3	0.05	0.05	4	0.1	0	0
4	0.05	0.05	1	0	0	0
5	0.05	0.05	2	0.05	0	1
6	0.05	0.1	3	0.05	0	1
7	0.1	0.1	4	0.2	1	-1
8	0.05	0.05	1	0	0	0
9	0.05	0.05	2	0.05	0	1
10	0.05	0.1	3	0.05	0	1
11	0.1	0.1	4	0.2	1	-1
12	0.05	0.05	1	0	0	0

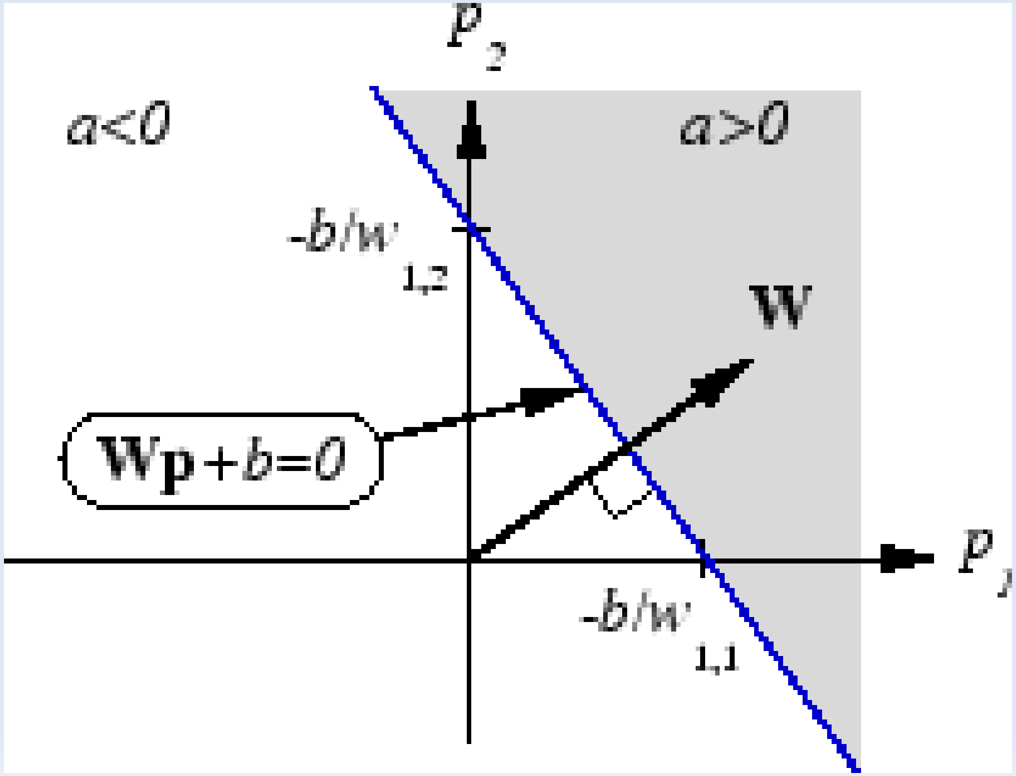
Não Linearidade do XOR



Adaline



Input
Simple ADALINE
p_1
p_2
$w_{1,1}$
$w_{1,2}$
n
a
b
1
$a = purelin(\mathbf{W}\mathbf{p}+b)$



p_2
$a<0$
$a>0$
$-b/w_{1,2}$
$\mathbf{W}$
$\mathbf{W}\mathbf{p}+b=0$
p_1
$-b/w_{1,1}$

Exemplo Adaline: AND

$w_i(t + 1) = w_i(t) + \eta e(t) x_i$

$\Theta = 0.1$

$\eta = 0.05$

$f(x) = 1$ para $x > \Theta$, 0 c.c.

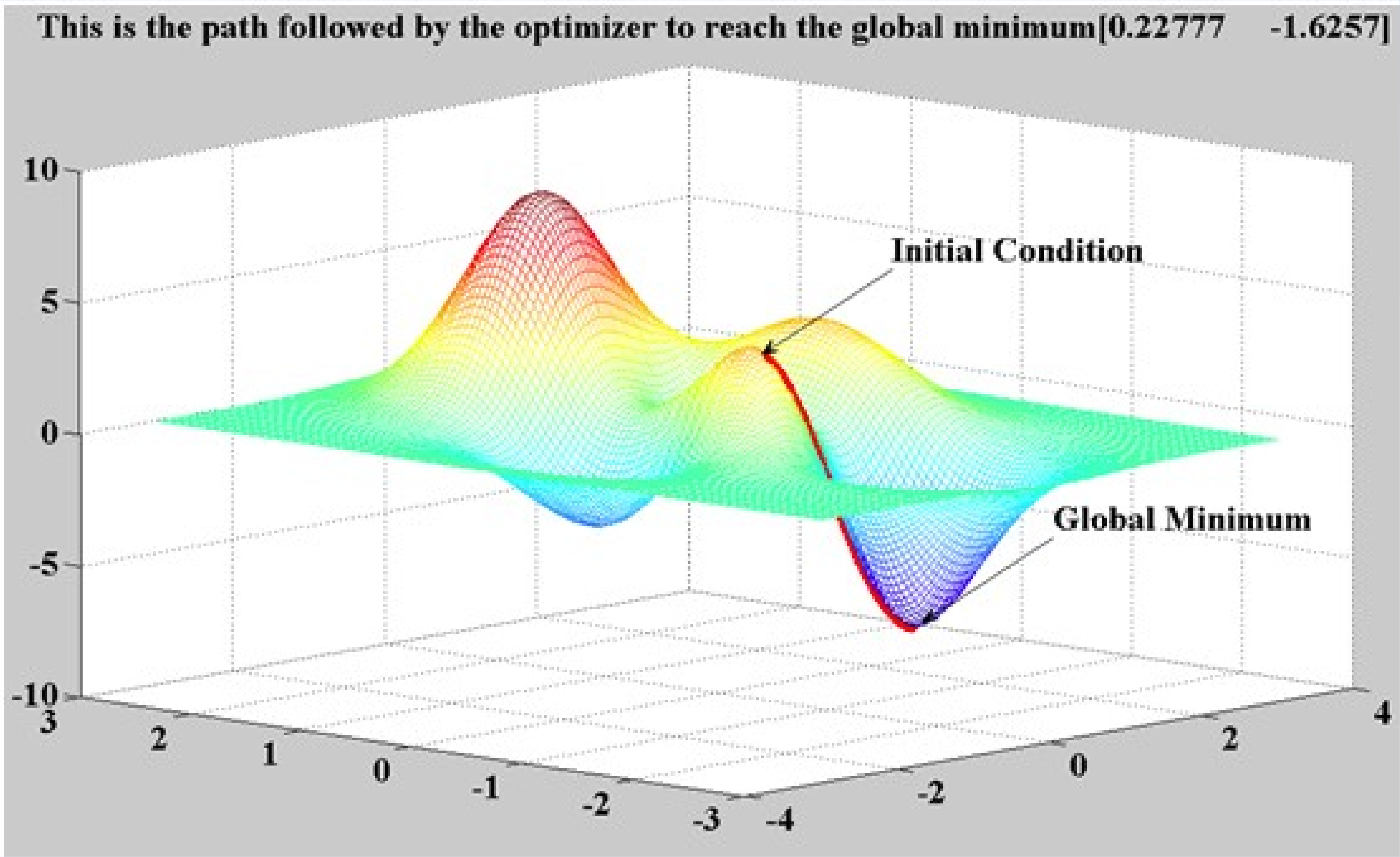
max sme = 0.1

padrão	**x1**	**x2**	**y (d)**
1	0	0	0
2	0	1	1
3	1	0	1
4	1	1	1

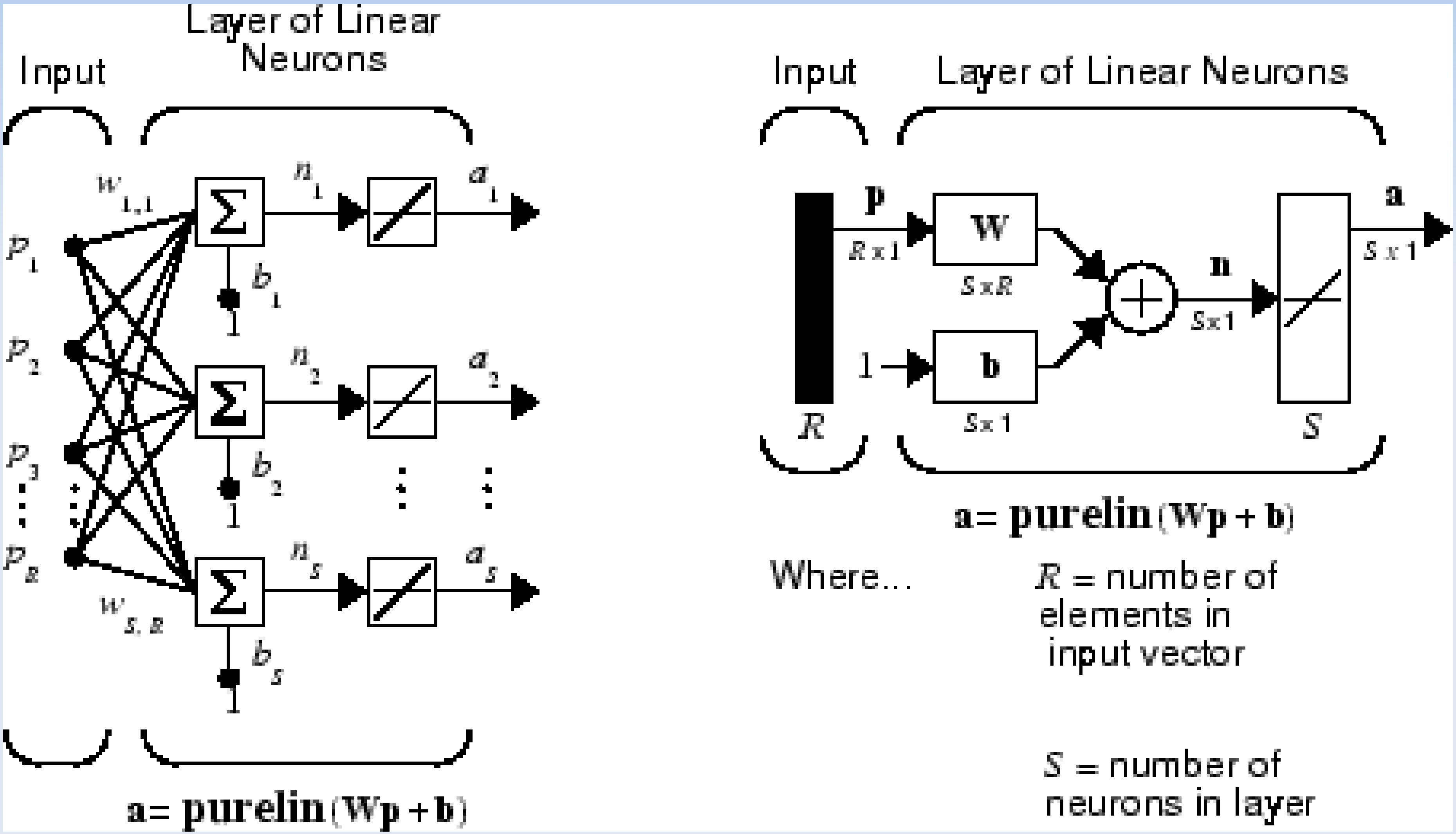
t	**w1**	**w2**	**padrão**	**Σxi*wi**	**Erro (d-o)**	
0	0	0	**1**	0	0	
1	0	0	**2**	0	1	
2	0	0.05	**3**	0	1	
3	0.05	0.05	**4**	0.1	0.9	0.7
4	0.1	0.1	**1**	0	0	
5	0.1	0.1	**2**	0.1	0.91	
6	0.1	0.14	**3**	0.1	0.91	
7	0.14	0.14	**4**	0.28	0.72	0.54
8	0.18	0.18	**1**	0	0	
9	0.18	0.18	**2**	0.18	0.82	
10	0.18	0.22	**3**	0.18	0.82	
11	0.22	0.22	**4**	0.43	0.57	0.42

60	0.59	0.59	**1**	0	0	
61	0.59	0.59	**2**	0.59	0.41	
62	0.59	0.61	**3**	0.59	0.41	
63	0.61	0.61	**4**	1.23	-0.23	0.1
64	0.6	0.6	**1**	0	0	
65	0.6	0.6	**2**	0.6	0.4	
66	0.6	0.62	**3**	0.6	0.4	
67	0.62	0.62	**4**	1.24	-0.24	0.09

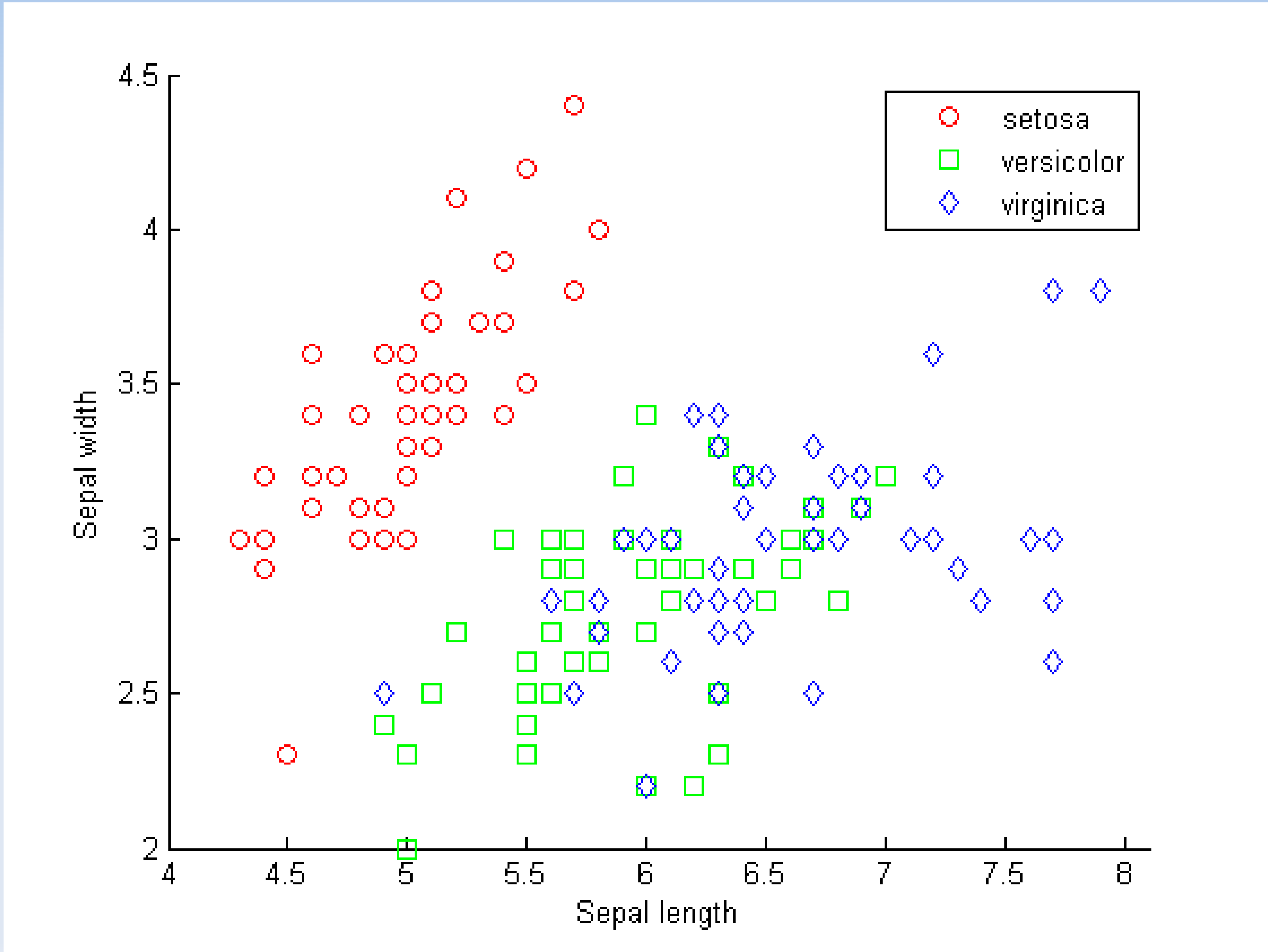
Descidade de Gradiente



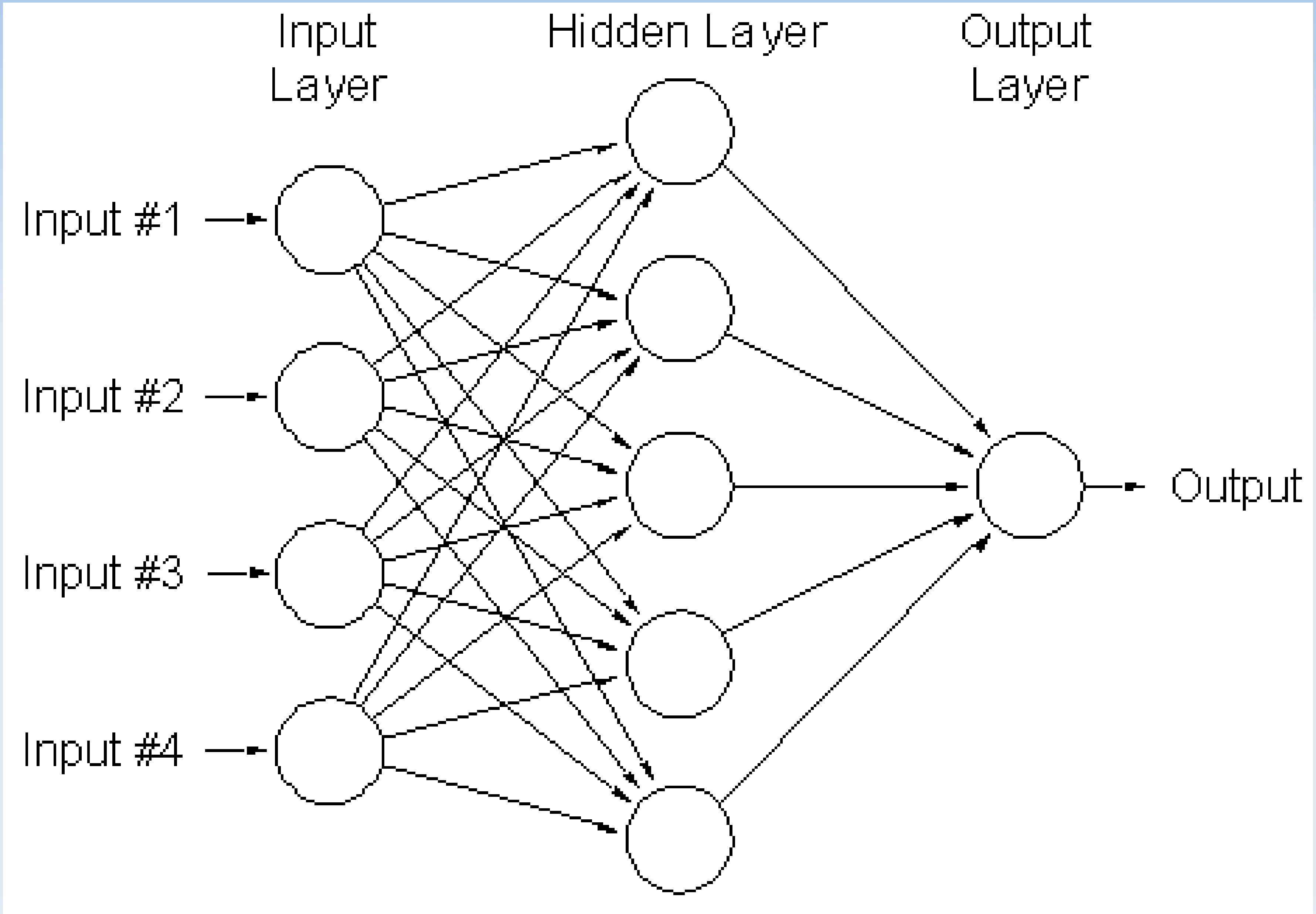
Mais de Dois Padrões



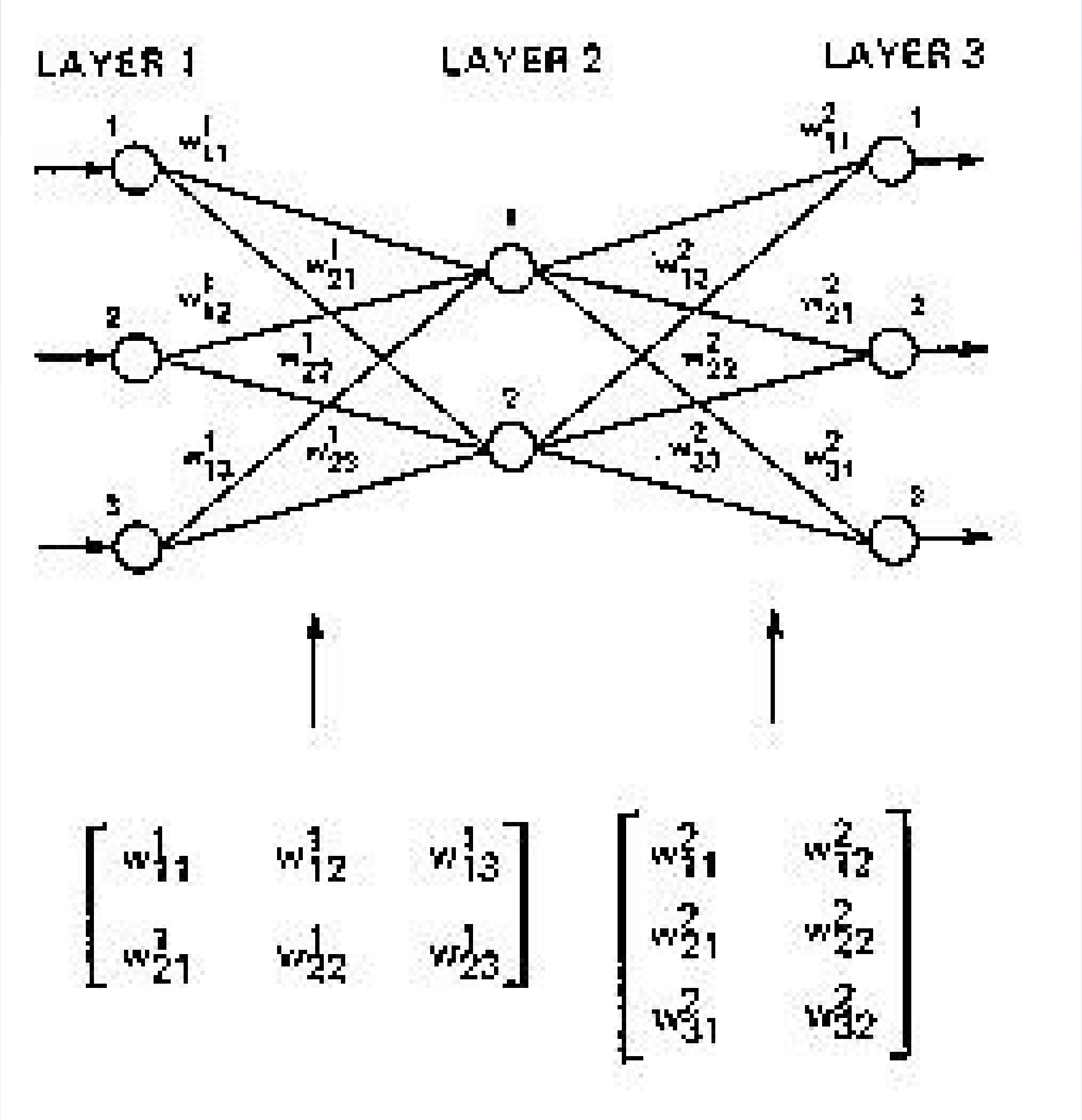
Exemplo Iris



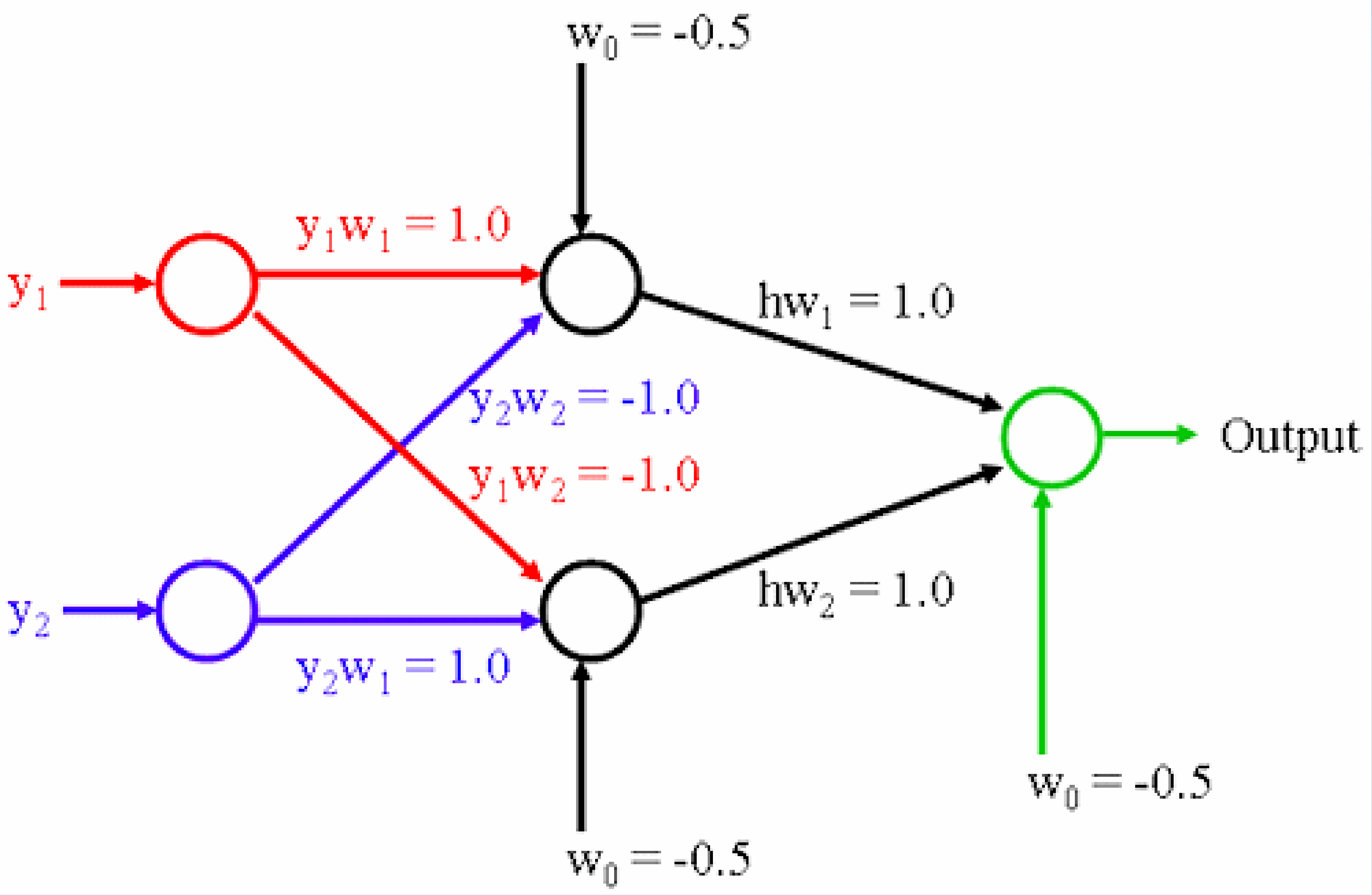
Back-Propagation



Vários padrões



Percepetron Multicamadas XOR



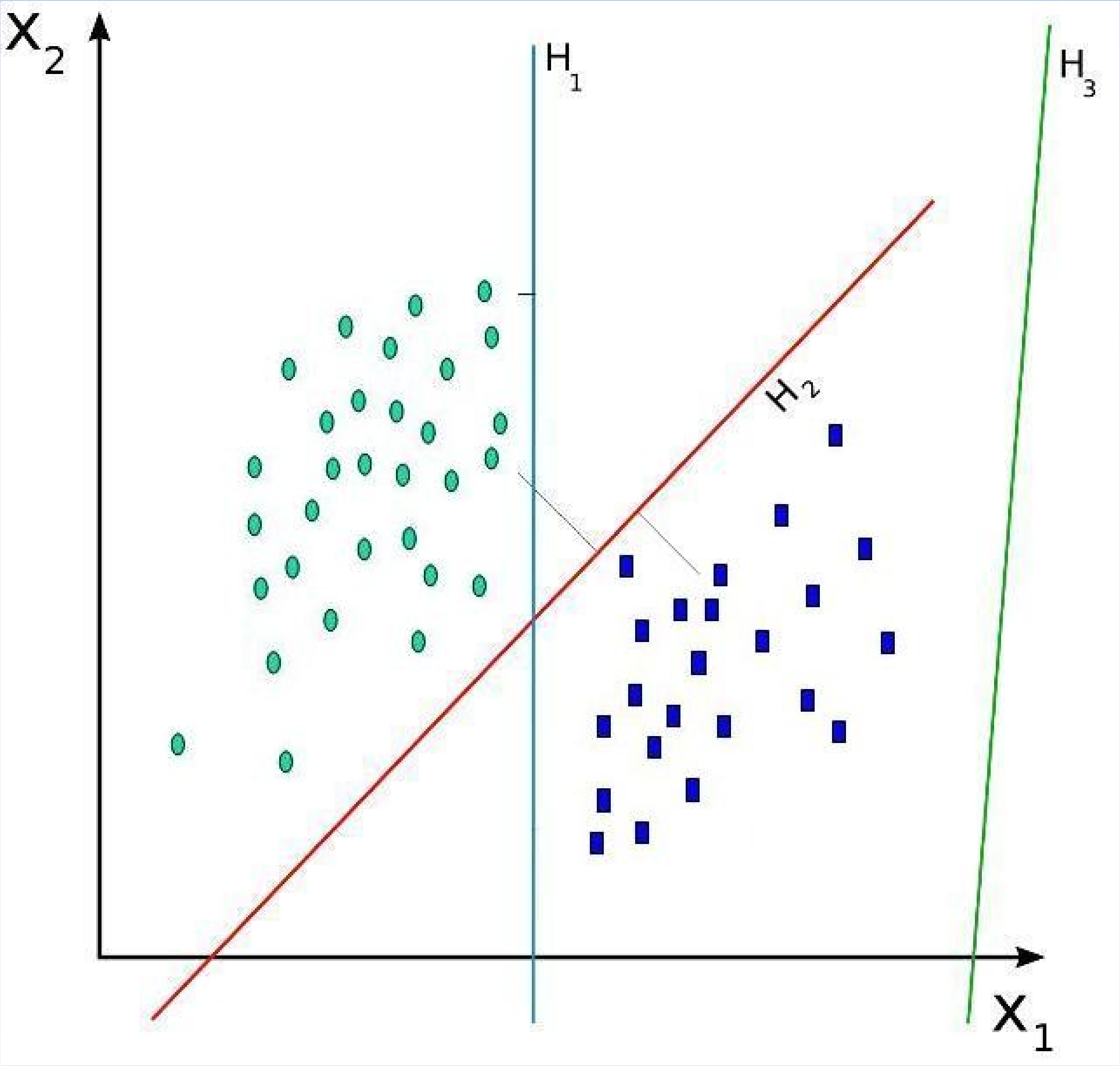
Perceptron Multicamadas XOR

f(x) = x > O

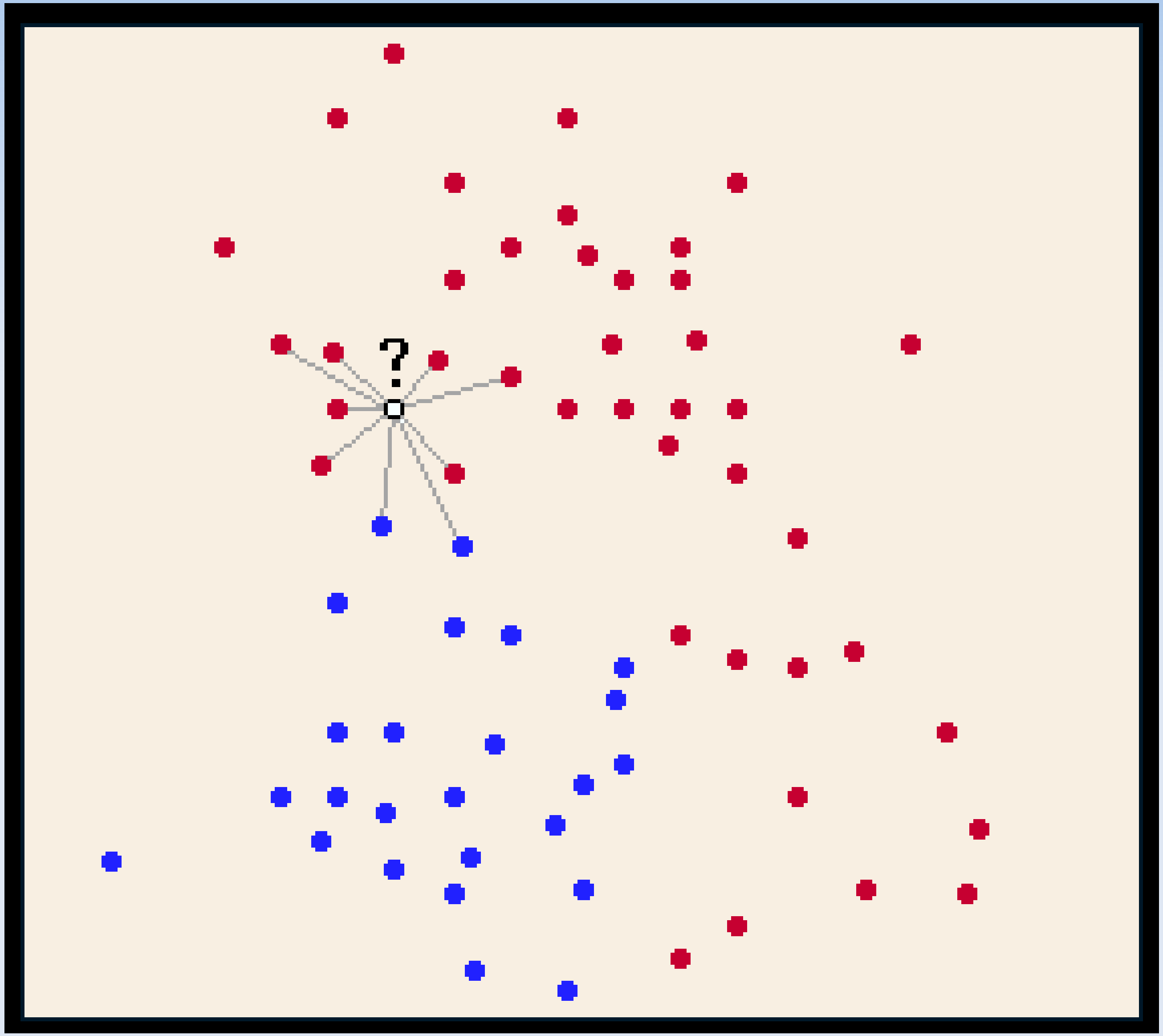
O = 0

x1	x2	h1w0	h2w0	h1w1	h2w1	h1w2	h2w2	Σ1	Σ2	**f(Σ1)**	**f(Σ2)**	ow0	ow1	ow2	Σ3	**f(Σ3)**
0	0	0	0	1	-1	-1	1	0	0	**0**	**0**	0	1	1	0	**0**
0	1	0	0	1	-1	-1	1	-1	1	**0**	**1**	0	1	1	1	**1**
1	0	0	0	1	-1	-1	1	1	-1	**1**	**0**	0	1	1	1	**1**
1	1	0	0	1	-1	-1	1	0	0	**0**	**0**	0	1	1	0	**0**

SVM



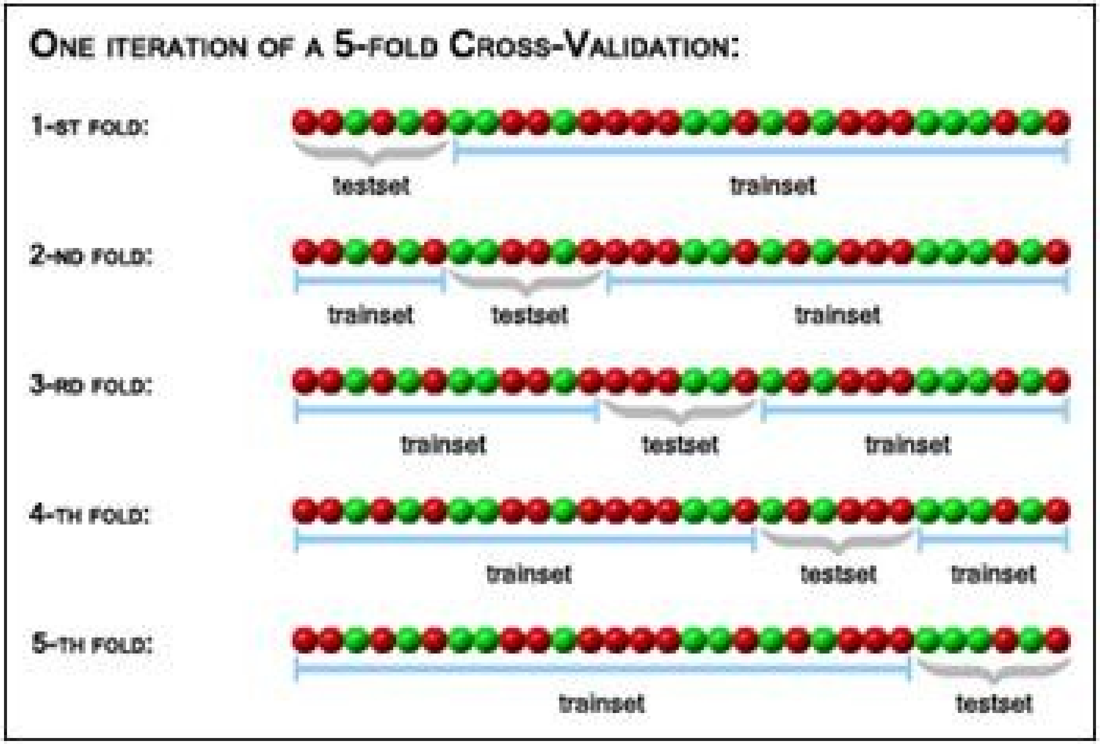
KNN

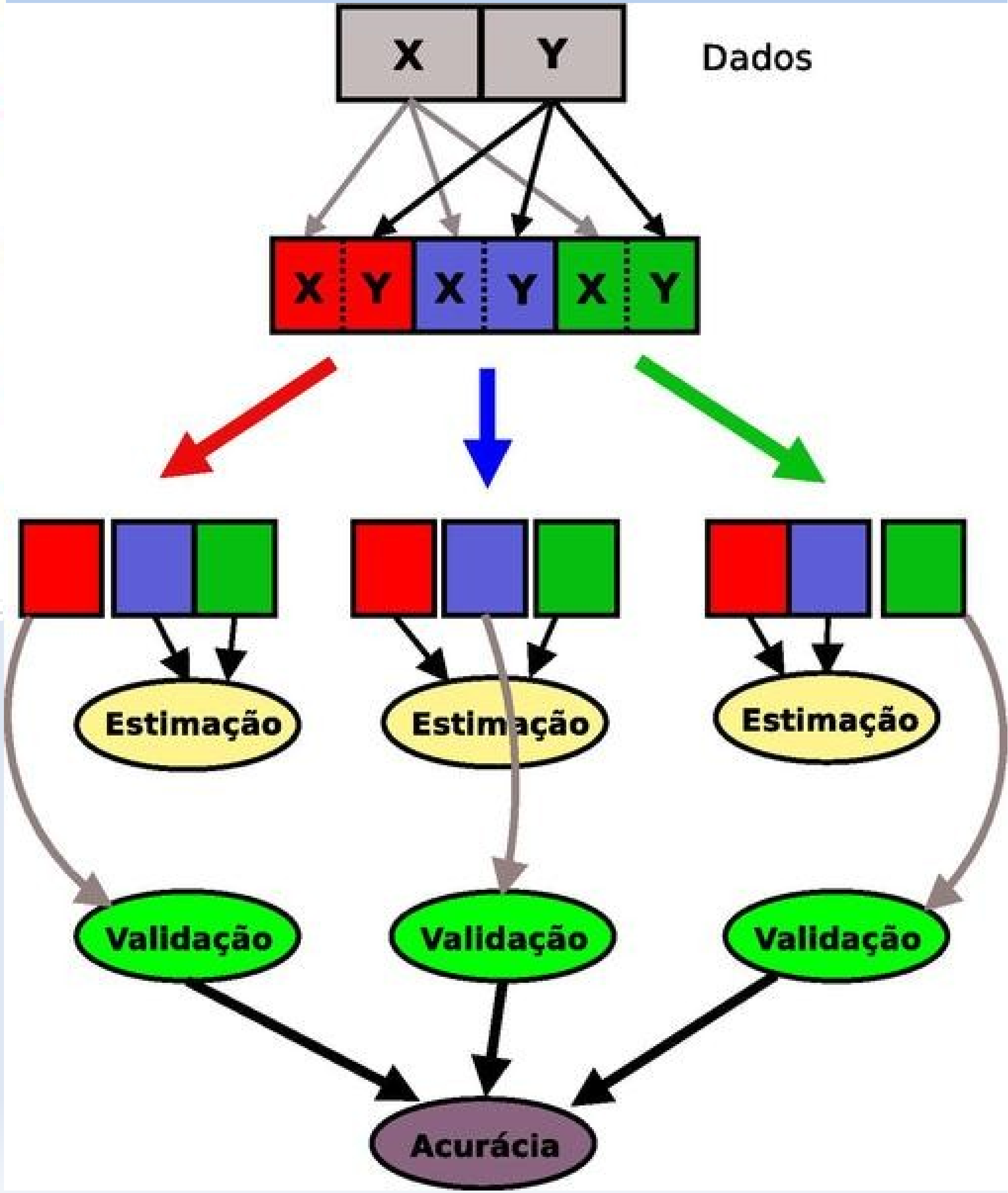


Validações e Testes

- Resubstituição
- Holdout
- K-fold cross-validation
- Leave-one-out cross-validation

K-Fold Cross Validation





Possível Trabalho

- As técnicas atuais de descida de gradiente utilizam a soma do erros quadráticos ou os erros quadráticos médios.
- Esta abordagem é muito ”democrática” porque todos os padrões influenciam igualmente no cálculo do erro e, consequentemente, na posição e inclinação do hiperplano separador.
- Se o erro for calculado em relação ao centro de massa das classes, ao invés de cada padrão, é possível que o hiperplano separador busque uma posição mais adequada.

Obrigado

Dúvidas?