Redes neuronais artificiais

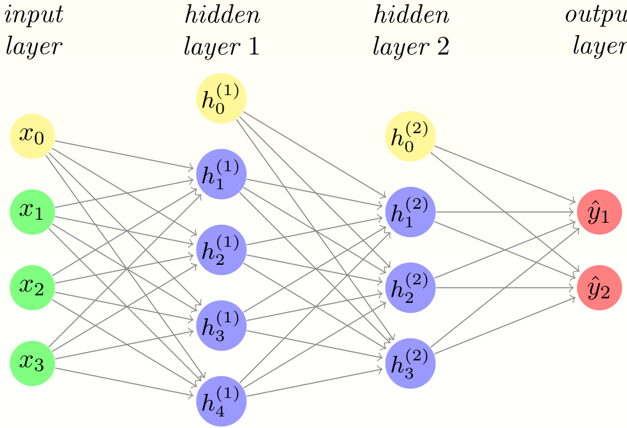
Gaspar J. Machado

Departamento de Matemática, Universidade do Minho

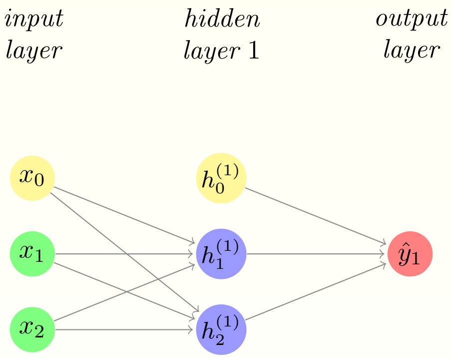
março de 2024

RNA

- Como tratar bases de dados que não são linearmente separáveis?
  - Linearização dos dados aumentando a dimensão do espaço dos atributos.
  - Redes neuronais artificiais — RNA (ou *Perceptron* multicamada — *multilayer Perceptron*, MLP, que é um nome enganador).



RNA | XOR (i)



$$\tilde{U} = \begin{bmatrix} \tilde{u}_{01} & \tilde{u}_{02} \\ \tilde{u}_{11} & \tilde{u}_{12} \\ \tilde{u}_{21} & \tilde{u}_{22} \end{bmatrix} \qquad \tilde{V} = \begin{bmatrix} \tilde{v}_0 \\ \tilde{v}_1 \\ \tilde{v}_2 \end{bmatrix}$$

$$\tilde{U} = \begin{bmatrix} \tilde{u}_{01} & \tilde{u}_{02} \\ \tilde{u}_{11} & \tilde{u}_{12} \\ \tilde{u}_{21} & \tilde{u}_{22} \end{bmatrix} = \begin{bmatrix} -\frac{1}{4} & \frac{5}{4} \\ 1 & -1 \\ 1 & -1 \end{bmatrix} \qquad \tilde{V} = \begin{bmatrix} \tilde{v}_0 \\ \tilde{v}_1 \\ \tilde{v}_2 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} \\ 1 \\ 1 \end{bmatrix}$$

$$h_1 (= h_1^{(1)}) = \operatorname{sgn}(\tilde{U}(:,1)^\top \tilde{x}) = \operatorname{sgn}\left(-\frac{1}{4} + x_1 + x_2\right)$$

$$h_2 (= h_2^{(1)}) = \operatorname{sgn}(\tilde{U}(:,2)^\top \tilde{x}) = \operatorname{sgn}\left(\frac{5}{4} - x_1 - x_2\right)$$

$$\hat{y} (= \hat{y}_1) = \operatorname{sgn}(\tilde{V}^\top \tilde{h}) = \operatorname{sgn}\left(-\frac{1}{2} + h_1 + h_2\right)$$

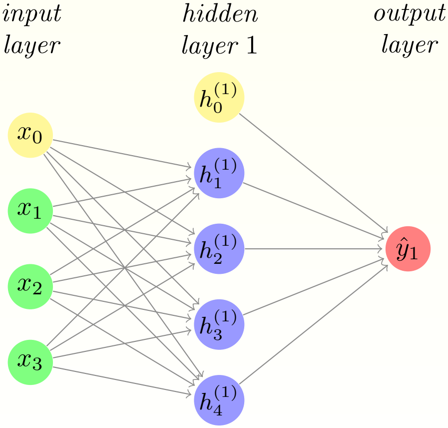
x_1	x_2	y	h_1	h_2	$\hat{y}$
0	0	-1	$\operatorname{sgn}(-\frac{1}{4}) = -1$	$\operatorname{sgn}(\frac{5}{4}) = +1$	$\operatorname{sgn}(-\frac{1}{2}) = -1$
0	1	$+1$	$\operatorname{sgn}(\frac{3}{4}) = +1$	$\operatorname{sgn}(\frac{1}{4}) = +1$	$\operatorname{sgn}(\frac{3}{2}) = +1$
1	0	$+1$	$\operatorname{sgn}(\frac{3}{4}) = +1$	$\operatorname{sgn}(\frac{1}{4}) = +1$	$\operatorname{sgn}(\frac{3}{2}) = +1$
1	1	-1	$\operatorname{sgn}(\frac{7}{4}) = +1$	$\operatorname{sgn}(-\frac{3}{4}) = -1$	$\operatorname{sgn}(-\frac{1}{2}) = -1$

RNA1 (i)

- Contexto: classificador binário com a base de dados $D = (x^n, y^n)_{n=1}^N$, $x^n \in \mathbb{R}^I$, $y^n \in \{-1, +1\}$.
- Neste curso vamos considerar redes neuronais artificiais do tipo *feed-forward* com uma camada oculta (*shallow neural network*), ou seja, com três camadas:
  - a camada de entrada (*input layer*) com I $(+1)$ nós;
  - uma camada oculta (*hidden layer*) com J $(+1)$ nós (é um hiperparâmetro da ML);
  - a camada de saída (*output layer*), com 1 nó, pois o contexto é o da classificação binária.
- Vamos denotar por **RNA1** as redes neuronais artificiais do tipo *feed-forward* com uma camada oculta.

RNA1 (ii)

- Exemplo com $I = 3$ e $J = 4$:



RNA1 (iii)

- Parâmetros de uma RNA1: matrizes

$$\tilde{U} = \begin{bmatrix} \tilde{u}_{01} & \tilde{u}_{02} & \cdots & \tilde{u}_{0J} \\ \tilde{u}_{11} & \tilde{u}_{12} & \cdots & \tilde{u}_{1J} \\ \tilde{u}_{21} & \tilde{u}_{22} & \cdots & \tilde{u}_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{u}_{I1} & \tilde{u}_{I2} & \cdots & \tilde{u}_{IJ} \end{bmatrix} \in \mathbb{R}^{(I+1)\times J} \quad \text{e} \quad \tilde{V} = \begin{bmatrix} \tilde{v}_0 \\ \tilde{v}_1 \\ \vdots \\ \tilde{v}_J \end{bmatrix} \in \mathbb{R}^{J+1}.$$

- Treinar uma RNA1: determinar as matrizes $\tilde{U}^* \in \mathbb{R}^{(I+1)\times J}$ e $\tilde{V}^* \in \mathbb{R}^{J+1}$ que minimizam uma função custo dada.

RNA1 (iv)

- Arquitetura da ML:
  - *inputs*: $x \in \mathbb{R}^I$, $\tilde{U} \in \mathbb{R}^{(I+1)\times J}$, $\tilde{V} \in \mathbb{R}^{J+1}$, $f : \mathbb{R} \to \mathbb{R}$, $g : \mathbb{R} \to \mathbb{R}$;
  - *output*: $\hat{y} \in \mathbb{R}$;
  - passos

    1. $s = \tilde{U}^\top \tilde{x}$,
    2. $h = f(s) \equiv (f(s_1), f(s_2), \ldots, f(s_J))^\top$,
    3. $r = \tilde{V}^\top \tilde{h}$,
    4. $\hat{y} = g(r)$.

- Função custo genérica:

$$E(\tilde{U}, \tilde{V}; D) = \frac{1}{N} \sum_{n=1}^{N} E_n(\hat{y}^n; y^n),$$

$$\hat{y}^n = \hat{y}(\tilde{U}, \tilde{V}; x^n).$$

Algoritmo RNA1–MGE

Input: $D = (x^n, y^n)_{n=1}^{N}$, $x^n \in \mathbb{R}^I$, $y^n \in \{0,1\}$, $\tilde{U}_{(0)} \in \mathbb{R}^{(I+1)\times J}$,
$\tilde{V}_{(0)} \in \mathbb{R}^{J+1}$, $\eta \in \mathbb{R}^+$

Output: $\tilde{U}^* \in \mathbb{R}^{(I+1)\times J}$ e $\tilde{V}^* \in \mathbb{R}^{J+1}$

$t \leftarrow 0$;

while V **do**

 selecionar $n \in \{1, \ldots, N\}$ aleatório;

 $d_{(t)}^{\tilde{v}} \leftarrow \nabla_{\tilde{v}} E_n(\hat{y}^n; y^n)$, $\tilde{V}_{(t+1)} \leftarrow \tilde{V}_{(t)} - \eta d_{(t)}^{\tilde{v}}$;

 $d_{(t)}^{\tilde{u}} \leftarrow \nabla_{\tilde{u}} E_n(\hat{y}^n; y^n)$, $\tilde{U}_{(t+1)} \leftarrow \tilde{U}_{(t)} - \eta d_{(t)}^{\tilde{u}}$;

 if $CP{=}V$ **then**

  $\tilde{U}^* \leftarrow \tilde{U}_{(t+1)}$, $\tilde{V}^* \leftarrow \tilde{V}_{(t+1)}$; **return** $(\tilde{U}^*, \tilde{V}^*)$;

 else

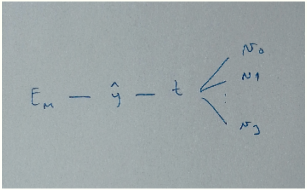
  $t \leftarrow t + 1$;

RNA1 | Cálculo das derivadas | $\tilde{V}$

- Arquitetura da ML:

$$s = \tilde{U}^\top \tilde{x}^n, h = f(s), r = \tilde{V}^\top \tilde{h}, \hat{y}^n = g(r).$$

- Seja $j \in \{0, \ldots, J\}$. Então:



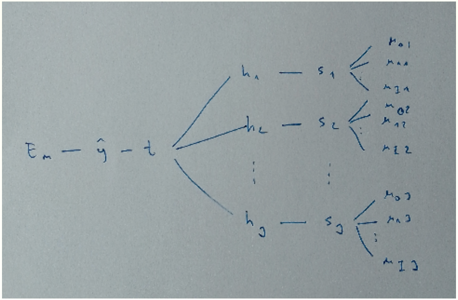
$$\frac{\partial E_n}{\partial \tilde{v}_j}(\hat{y}^n; y^n) = \frac{\mathrm{d}E_n}{\mathrm{d}\hat{y}} \frac{\mathrm{d}\hat{y}}{\mathrm{d}r} \frac{\partial r}{\partial \tilde{v}_j} = \frac{\mathrm{d}E_n}{\mathrm{d}\hat{y}} g'(r) h_j.$$

RNA1 | Cálculo das derivadas | $\tilde{U}$

- Arquitetura da ML

$$s = \tilde{U}^\top \tilde{x}^n, h = f(s), r = \tilde{V}^\top \tilde{h}, \hat{y}^n = g(r).$$

- Sejam $i \in \{0, \ldots, J\}$ e $j \in \{1, \ldots, J\}$. Então:



$$\frac{\partial E_n}{\partial \tilde{u}_{ij}}(\hat{y}^n; y^n) = \frac{\mathrm{d}E_n}{\mathrm{d}\hat{y}} \frac{\mathrm{d}\hat{y}}{\mathrm{d}r} \frac{\partial r}{\partial h_j} \frac{\mathrm{d}h_j}{\mathrm{d}s_j} \frac{\partial s_j}{\partial \tilde{u}_{ij}} = \frac{\mathrm{d}E_n}{\mathrm{d}\hat{y}} g'(r) v_j f'(s_j) x_i.$$

RNA1_1 | função logística + “cross entropy”

- Sejam:

$$\begin{aligned} f(z) &= \sigma(z), \\ g(z) &= \sigma(z), \\ E_n(\hat{p}^n; y^n) &= -y^n \ln(\hat{p}^n) - (1 - y^n) \ln(1 - \hat{p}^n). \end{aligned}$$

- Então:
  - Para $j \in \{0, \ldots, J\}$:

$$\frac{\partial E_n}{\partial \tilde{v}_j}(\hat{p}^n; y^n) = (\hat{p}^n - y^n)\, \tilde{h}_j.$$

  - Para $i \in \{0, \ldots, J\}$, $j \in \{1, \ldots, J\}$:

$$\frac{\partial E_n}{\partial \tilde{u}_{ij}}(\hat{p}^n; y^n) = (\hat{p}^n - y^n)\, v_j h_j (1 - h_j) \tilde{x}_i.$$

  - *Backpropagation* ...
- Vamos denotar por **RNA1$_1$–MGE** o algoritmo da rede neuronal artificial do tipo *feed-forward* com uma camada oculta cujas funções de ativação são a função logística e cuja função é a “cross entropy”com o MGE.

Input: $D = (x^n, y^n)_{n=1}^{N}$, $x^n \in \mathbb{R}^I$, $y^n \in \{0,1\}$, $\tilde{U}_{(0)} \in \mathbb{R}^{(I+1)\times J}$, $\tilde{V}_{(0)} \in \mathbb{R}^{J+1}$, $\eta \in \mathbb{R}^+$

Output: $\tilde{U}^* \in \mathbb{R}^{(I+1)\times J}$ e $\tilde{V}^* \in \mathbb{R}^{J+1}$

$t \leftarrow 0$;

while V **do**

  selecionar $n \in \{1, \ldots, N\}$ aleatório;

  $s \leftarrow \tilde{U}_{(t)}^{\top} \tilde{x}^n$, $h \leftarrow \sigma(s)$, $r \leftarrow \tilde{V}_{(t)}^{\top} \tilde{h}$, $\hat{p}^n \leftarrow \sigma(r)$;

  $d_{j,(t)}^{\tilde{v}} \leftarrow (\hat{p}^n - y^n)\tilde{h}_j$, $\tilde{v}_{j,(t+1)} \leftarrow \tilde{v}_{j,(t)} - \eta d_{j,(t)}^{\tilde{v}}$, $j = 0, \ldots, J$;

  $d_{ij,(t)}^{\tilde{u}} \leftarrow (\hat{p}^n - y^n)\, v_{j,(t)} h_j (1 - h_j) \tilde{x}_i$,
   $\tilde{u}_{ij,(t+1)} \leftarrow \tilde{u}_{ij,(t)} - \eta d_{ij,(t)}^{\tilde{u}}$, $i = 0, \ldots, I$, $j = 1, \ldots, J$;

  if $CP{=}V$ **then**

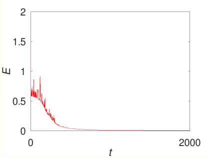
    $\tilde{U}^* \leftarrow \tilde{U}_{(t+1)}$, $\tilde{V}^* \leftarrow \tilde{V}_{(t+1)}$; **return** $(\tilde{U}^*, \tilde{V}^*)$;

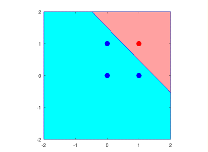
  else

    $t \leftarrow t + 1$;

ex1 (AND) | RNA1_1–MGE | $J = 2$

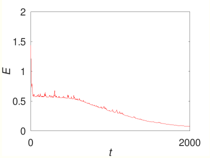
$t_{\text{MAX}} = 2000,\ \eta = 0.5$
$E = 0.0042,\ acc = 1$

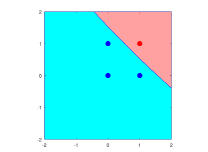




y^n	0	0	0	1
$\hat{p}^n$	0.0002	0.0041	0.0035	0.9910
$\hat{y}^n$	0	0	0	1

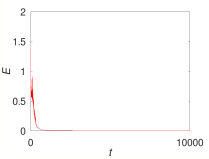
$t_{\text{MAX}} = 2000,\ \eta = 0.1$
$E = 0.0750,\ acc = 1$

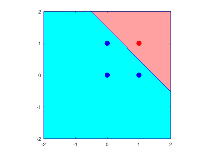




y^n	0	0	0	1
$\hat{p}^n$	0.0043	0.0800	0.0595	0.8600
$\hat{y}^n$	0	0	0	1

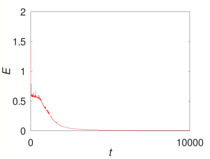
$t_{\text{MAX}} = 10000,\ \eta = 0.5$
$E = 0.0006,\ acc = 1$

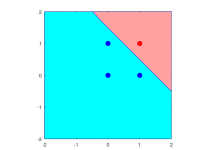




y^n	0	0	0	1
$\hat{p}^n$	0.0000	0.0005	0.0005	0.9988
$\hat{y}^n$	0	0	0	1

$t_{\text{MAX}} = 10000,\ \eta = 0.1$
$E = 0.0039,\ acc = 1$

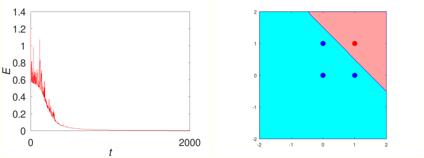




y^n	0	0	0	1
$\hat{p}^n$	0.0001	0.0037	0.0033	0.9916
$\hat{y}^n$	0	0	0	1

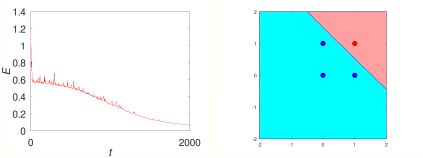
ex1 (AND) | $\mathrm{RNA1}_1$–MGE | $J = 3$

$t_{\mathrm{MAX}} = 2000,\ \eta = 0.5$
$E = 0.0039,\ acc = 1$



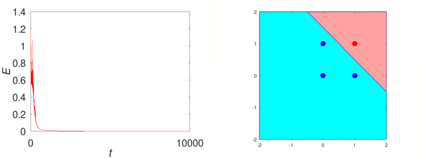
y^n	0	0	0	1
$\hat{p}^n$	0.0002	0.0043	0.0038	0.9927
$\hat{y}^n$	0	0	0	1

$t_{\mathrm{MAX}} = 2000,\ \eta = 0.1$
$E = 0.0684,\ acc = 1$



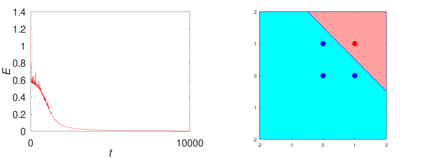
y^n	0	0	0	1
$\hat{p}^n$	0.0080	0.0772	0.0635	0.8873
$\hat{y}^n$	0	0	0	1

$t_{\mathrm{MAX}} = 10000,\ \eta = 0.5$
$E = 0.0005,\ acc = 1$



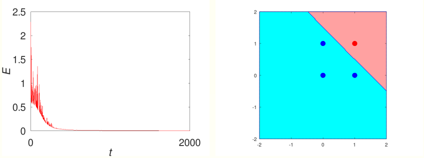
y^n	0	0	0	1
$\hat{p}^n$	0.0000	0.0006	0.0005	0.9990
$\hat{y}^n$	0	0	0	1

$t_{\mathrm{MAX}} = 10000,\ \eta = 0.1$
$E = 0.0041,\ acc = 1$



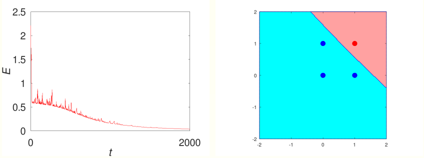
y^n	0	0	0	1
$\hat{p}^n$	0.0002	0.0046	0.0038	0.9922
$\hat{y}^n$	0	0	0	1

$t_{\text{MAX}} = 2000,\ \eta = 0.5$
$E = 0.0027,\ acc = 1$



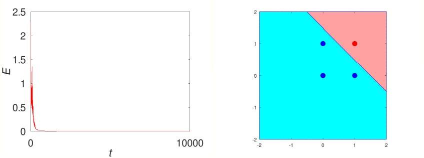
y^n	0	0	0	1
$\hat{p}^n$	0.0000	0.0024	0.0023	0.9940
$\hat{y}^n$	0	0	0	1

$t_{\text{MAX}} = 2000,\ \eta = 0.1$
$E = 0.0388,\ acc = 1$



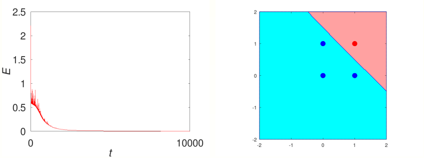
y^n	0	0	0	1
$\hat{p}^n$	0.0003	0.0320	0.0281	0.9105
$\hat{y}^n$	0	0	0	1

$t_{\text{MAX}} = 10000,\ \eta = 0.5$
$E = 0.0004,\ acc = 1$



y^n	0	0	0	1
$\hat{p}^n$	0.0000	0.0004	0.0003	0.9993
$\hat{y}^n$	0	0	0	1

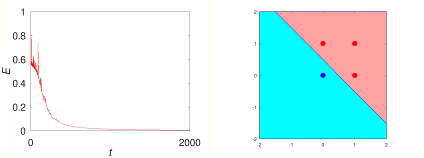
$t_{\text{MAX}} = 10000,\ \eta = 0.1$
$E = 0.0027,\ acc = 1$



y^n	0	0	0	1
$\hat{p}^n$	0.0000	0.0028	0.0026	0.9945
$\hat{y}^n$	0	0	0	1

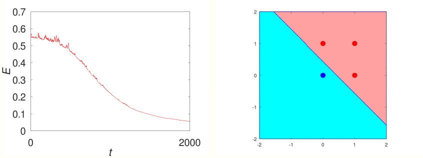
ex2 (OR) | RNA1_1–MGE | $J = 2$

$t_{\mathrm{MAX}} = 2000,\ \eta = 0.5$
$E = 0.0058,\ acc = 1$



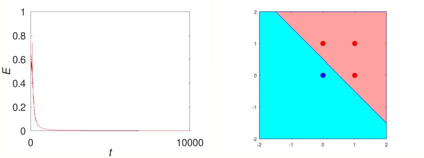
y^n	0	1	1	1
$\hat{p}^n$	0.0142	0.9963	0.9961	0.9986
$\hat{y}^n$	0	1	1	1

$t_{\mathrm{MAX}} = 2000,\ \eta = 0.1$
$E = 0.0562,\ acc = 1$



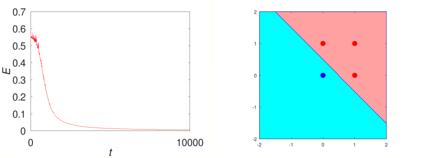
y^n	0	1	1	1
$\hat{p}^n$	0.1219	0.9614	0.9613	0.9843
$\hat{y}^n$	0	1	1	1

$t_{\mathrm{MAX}} = 10000,\ \eta = 0.5$
$E = 0.0010,\ acc = 1$



y^n	0	1	1	1
$\hat{p}^n$	0.0024	0.9994	0.9994	0.9998
$\hat{y}^n$	0	1	1	1

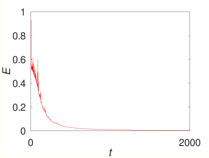
$t_{\mathrm{MAX}} = 10000,\ \eta = 0.1$
$E = 0.0061,\ acc = 1$

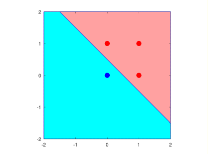


y^n	0	1	1	1
$\hat{p}^n$	0.0147	0.9960	0.9961	0.9981
$\hat{y}^n$	0	1	1	1

ex2 (OR) | RNA1_1–MGE | $J = 3$

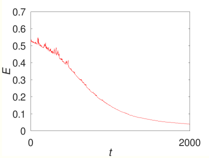
$t_{\mathrm{MAX}} = 2000,\ \eta = 0.5$
$E = 0.0041,\ acc = 1$

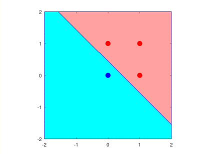




y^n	0	1	1	1
$\hat{p}^n$	0.0104	0.9973	0.9972	0.9995
$\hat{y}^n$	0	1	1	1

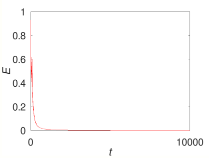
$t_{\mathrm{MAX}} = 2000,\ \eta = 0.1$
$E = 0.0399,\ acc = 1$

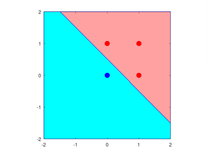




y^n	0	1	1	1
$\hat{p}^n$	0.0919	0.9724	0.9708	0.9944
$\hat{y}^n$	0	1	1	1

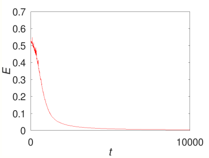
$t_{\mathrm{MAX}} = 10000,\ \eta = 0.5$
$E = 0.0007,\ acc = 1$

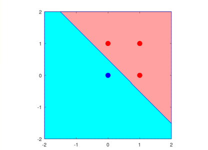




y^n	0	1	1	1
$\hat{p}^n$	0.0017	0.9996	0.9996	0.9999
$\hat{y}^n$	0	1	1	1

$t_{\mathrm{MAX}} = 10000,\ \eta = 0.1$
$E = 0.0043,\ acc = 1$

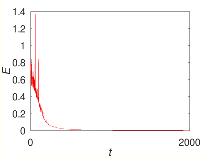


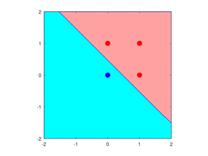


y^n	0	1	1	1
$\hat{p}^n$	0.0108	0.9972	0.9971	0.9995
$\hat{y}^n$	0	1	1	1

ex2 (OR) | RNA1_1–MGE | $J = 10$

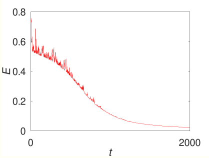
$t_{\mathrm{MAX}} = 2000,\ \eta = 0.5$
$E = 0.0019,\ acc = 1$

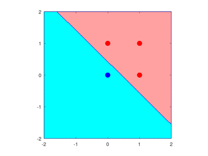




y^n	0	1	1	1
$\hat{p}^n$	0.0046	0.9985	0.9985	1.0000
$\hat{y}^n$	0	1	1	1

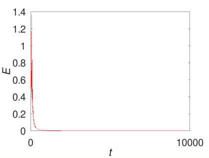
$t_{\mathrm{MAX}} = 2000,\ \eta = 0.1$
$E = 0.0231,\ acc = 1$

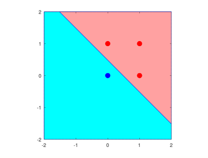




y^n	0	1	1	1
$\hat{p}^n$	0.0549	0.9827	0.9824	0.9994
$\hat{y}^n$	0	1	1	1

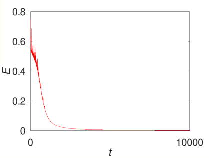
$t_{\mathrm{MAX}} = 10000,\ \eta = 0.5$
$E = 0.0003,\ acc = 1$

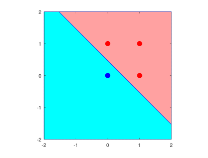




y^n	0	1	1	1
$\hat{p}^n$	0.0007	0.9998	0.9998	1.0000
$\hat{y}^n$	0	1	1	1

$t_{\mathrm{MAX}} = 10000,\ \eta = 0.1$
$E = 0.0020,\ acc = 1$

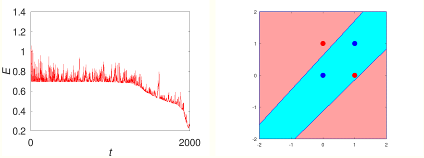




y^n	0	1	1	1
$\hat{p}^n$	0.0046	0.9983	0.9983	1.0000
$\hat{y}^n$	0	1	1	1

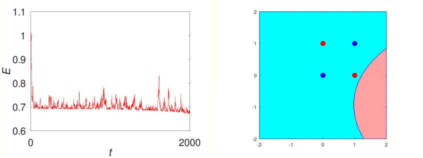
ex3 (XOR) | RNA1_1–MGE | $J = 2$

$t_{\mathrm{MAX}} = 2000,\ \eta = 0.5$
$E = 0.2786,\ acc = 0.75$



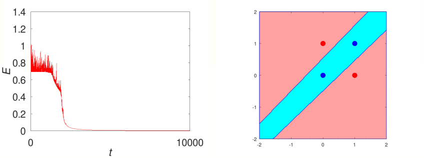
y^n	0	1	1	0
$\hat{p}^n$	0.0803	0.9065	0.4248	0.0733
$\hat{y}^n$	0	1	0	0

$t_{\mathrm{MAX}} = 2000,\ \eta = 0.1$
$E = 0.6857,\ acc = 0.50$



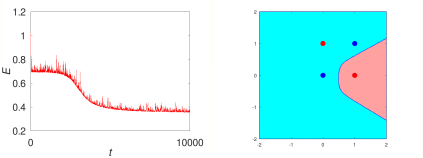
y^n	0	1	1	0
$\hat{p}^n$	0.3848	0.3836	0.4807	0.4322
$\hat{y}^n$	0	0	0	0

$t_{\mathrm{MAX}} = 10000,\ \eta = 0.5$
$E = 0.0023,\ acc = 1$



y^n	0	1	1	0
$\hat{p}^n$	0.0023	0.9979	0.9970	0.0020
$\hat{y}^n$	0	1	1	0

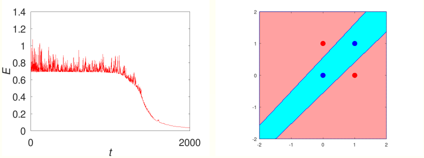
$t_{\mathrm{MAX}} = 10000,\ \eta = 0.1$
$E = 0.3601,\ acc = 0.75$



y^n	0	1	1	0
$\hat{p}^n$	0.0096	0.4339	0.9878	0.4421
$\hat{y}^n$	0	0	1	0

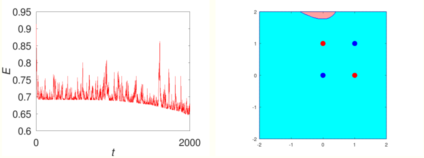
ex3 (XOR) | RNA1_1–MGE | $J = 3$

$t_{\text{MAX}} = 2000,\ \eta = 0.5$
$E = 0.0365,\ acc = 1$



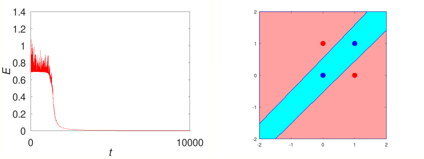
y^n	0	1	1	0
$\hat{p}^n$	0.0412	0.9625	0.9682	0.0327
$\hat{y}^n$	0	1	1	0

$t_{\text{MAX}} = 2000,\ \eta = 0.1$
$E = 0.6584,\ acc = 0.50$



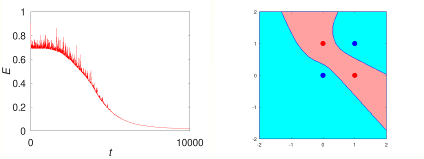
y^n	0	1	1	0
$\hat{p}^n$	0.3146	0.4608	0.4385	0.4814
$\hat{y}^n$	0	0	0	0

$t_{\text{MAX}} = 10000,\ \eta = 0.5$
$E = 0.0022,\ acc = 1$



y^n	0	1	1	0
$\hat{p}^n$	0.0030	0.9980	0.9981	0.0019
$\hat{y}^n$	0	1	1	0

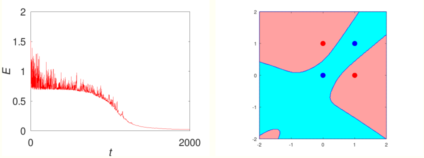
$t_{\text{MAX}} = 10000,\ \eta = 0.1$
$E = 0.0175,\ acc = 1$



y^n	0	1	1	0
$\hat{p}^n$	0.0113	0.9927	0.9739	0.0247
$\hat{y}^n$	0	1	1	0

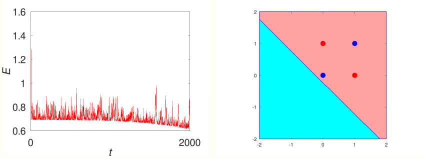
ex3 (XOR) | RNA1_1–MGE | $J = 10$

$t_{\mathrm{MAX}} = 2000,\ \eta = 0.5$
$E = 0.0228,\ acc = 1$



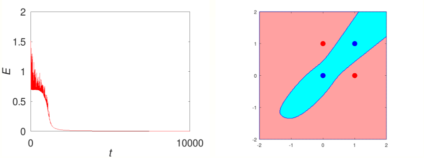
y^n	0	1	1	0
$\hat{p}^n$	0.0103	0.9813	0.9845	0.0454
$\hat{y}^n$	0	1	1	0

$t_{\mathrm{MAX}} = 2000,\ \eta = 0.1$
$E = 0.7587,\ acc = 0.50$



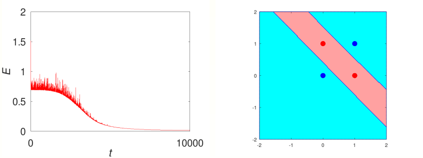
y^n	0	1	1	0
$\hat{p}^n$	0.5859	0.7934	0.7921	0.8153
$\hat{y}^n$	1	1	1	1

$t_{\mathrm{MAX}} = 10000,\ \eta = 0.5$
$E = 0.0020,\ acc = 1$



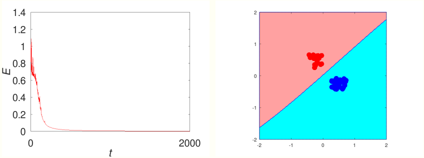
y^n	0	1	1	0
$\hat{p}^n$	0.0006	0.9982	0.9982	0.0039
$\hat{y}^n$	0	1	1	0

$t_{\mathrm{MAX}} = 10000,\ \eta = 0.1$
$E = 0.0150,\ acc = 1$

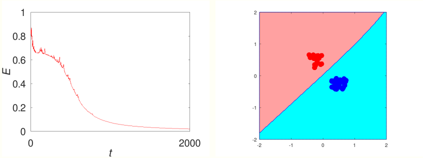


y^n	0	1	1	0
$\hat{p}^n$	0.0182	0.9853	0.9872	0.0138
$\hat{y}^n$	0	1	1	0

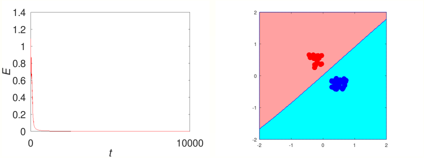
$t_{\text{MAX}} = 2000,\ \eta = 0.5$
$E = 0.0026,\ acc = 1$



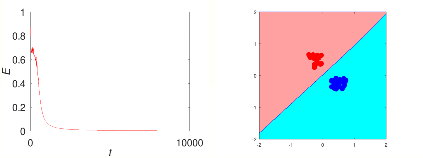
$t_{\text{MAX}} = 2000,\ \eta = 0.1$
$E = 0.0210,\ acc = 1$



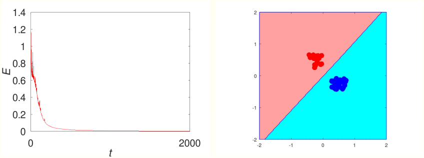
$t_{\text{MAX}} = 10000,\ \eta = 0.5$
$E = 0.0005,\ acc = 1$



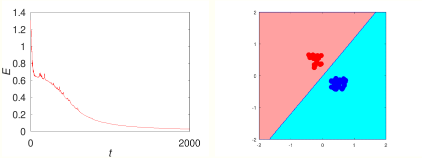
$t_{\text{MAX}} = 10000,\ \eta = 0.1$
$E = 0.0026,\ acc = 1$



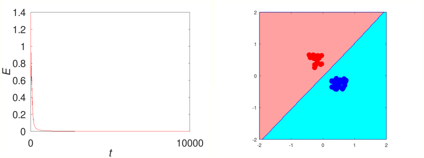
$t_{\mathrm{MAX}} = 2000,\ \eta = 0.5$
$E = 0.0030,\ acc = 1$



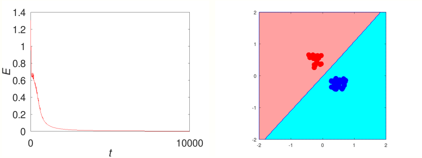
$t_{\mathrm{MAX}} = 2000,\ \eta = 0.1$
$E = 0.0275,\ acc = 1$



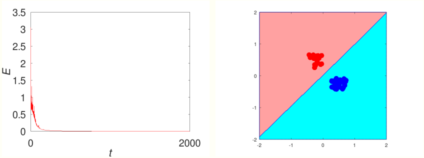
$t_{\mathrm{MAX}} = 10000,\ \eta = 0.5$
$E = 0.0005,\ acc = 1$



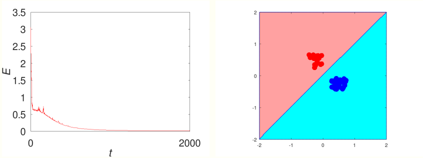
$t_{\mathrm{MAX}} = 10000,\ \eta = 0.1$
$E = 0.0032,\ acc = 1$



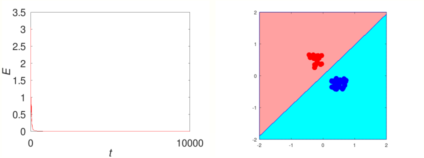
$t_{\mathrm{MAX}} = 2000,\ \eta = 0.5$
$E = 0.0016,\ acc = 1$



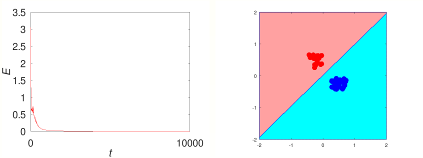
$t_{\mathrm{MAX}} = 2000,\ \eta = 0.1$
$E = 0.0125,\ acc = 1$



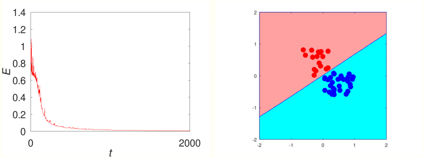
$t_{\mathrm{MAX}} = 10000,\ \eta = 0.5$
$E = 0.0003,\ acc = 1$



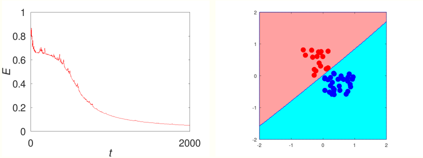
$t_{\mathrm{MAX}} = 10000,\ \eta = 0.1$
$E = 0.0016,\ acc = 1$



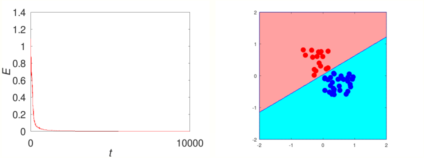
$t_{\mathrm{MAX}} = 2000,\ \eta = 0.5$
$E = 0.0072,\ acc = 1$



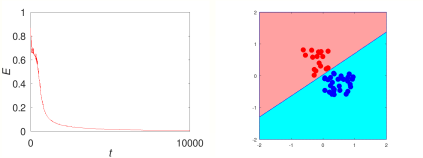
$t_{\mathrm{MAX}} = 2000,\ \eta = 0.1$
$E = 0.0515,\ acc = 1$



$t_{\mathrm{MAX}} = 10000,\ \eta = 0.5$
$E = 0.0010,\ acc = 1$

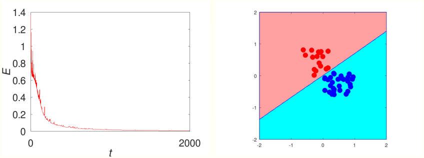


$t_{\mathrm{MAX}} = 10000,\ \eta = 0.1$
$E = 0.0073,\ acc = 1$

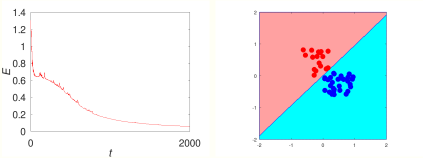


ex5 | RNA1_1–MGE | $J = 3$

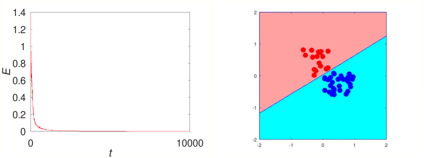
$t_{\mathrm{MAX}} = 2000,\ \eta = 0.5$
$E = 0.0081,\ acc = 1$



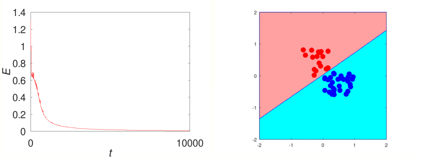
$t_{\mathrm{MAX}} = 2000,\ \eta = 0.1$
$E = 0.0595,\ acc = 1$



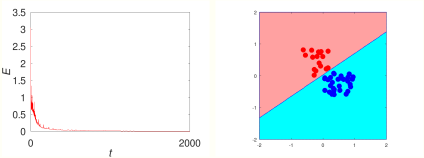
$t_{\mathrm{MAX}} = 10000,\ \eta = 0.5$
$E = 0.0011,\ acc = 1$



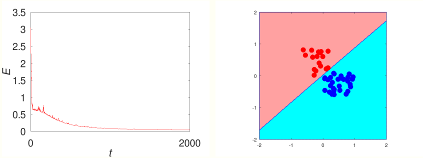
$t_{\mathrm{MAX}} = 10000,\ \eta = 0.1$
$E = 0.0083,\ acc = 1$



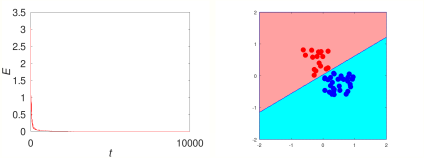
$t_{\text{MAX}} = 2000,\ \eta = 0.5$
$E = 0.0061,\ acc = 1$



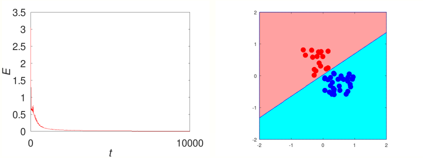
$t_{\text{MAX}} = 2000,\ \eta = 0.1$
$E = 0.0409,\ acc = 1$



$t_{\text{MAX}} = 10000,\ \eta = 0.5$
$E = 0.0009,\ acc = 1$

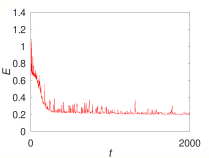


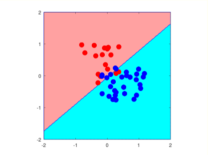
$t_{\text{MAX}} = 10000,\ \eta = 0.1$
$E = 0.0068,\ acc = 1$



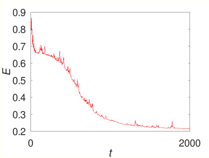
ex6 | RNA1_1–MGE | $J = 2$

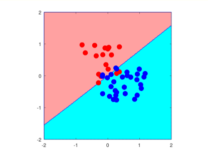
$t_{\mathrm{MAX}} = 2000,\ \eta = 0.5$
$E = 0.2195,\ acc = 0.83$



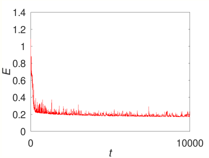


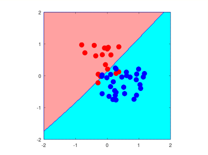
$t_{\mathrm{MAX}} = 2000,\ \eta = 0.1$
$E = 0.2202,\ acc = 0.88$



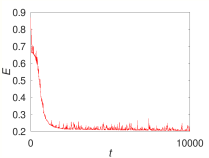


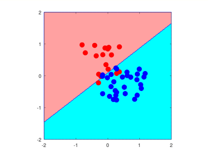
$t_{\mathrm{MAX}} = 10000,\ \eta = 0.5$
$E = 0.1697,\ acc = 0.90$





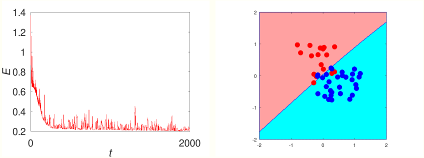
$t_{\mathrm{MAX}} = 10000,\ \eta = 0.1$
$E = 0.2035,\ acc = 0.90$



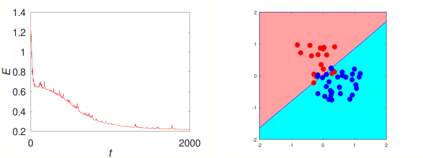


ex6 | RNA1_1–MGE | $J = 3$

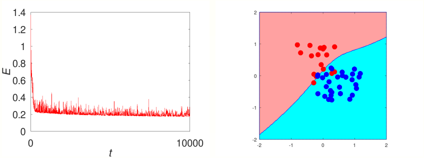
$t_{\mathrm{MAX}} = 2000,\ \eta = 0.5$
$E = 0.2056,\ acc = 0.92$



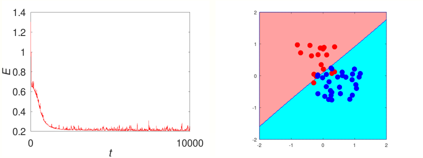
$t_{\mathrm{MAX}} = 2000,\ \eta = 0.1$
$E = 0.2199,\ acc = 0.90$



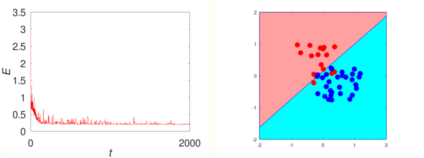
$t_{\mathrm{MAX}} = 10000,\ \eta = 0.5$
$E = 0.1747,\ acc = 0.90$



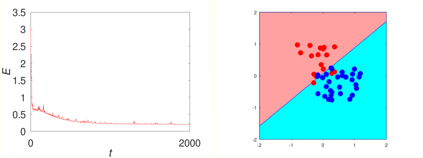
$t_{\mathrm{MAX}} = 10000,\ \eta = 0.1$
$E = 0.2078,\ acc = 0.90$



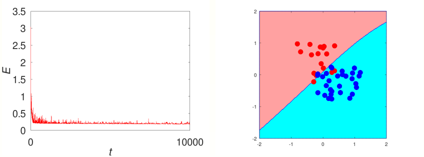
$t_{\mathrm{MAX}} = 2000, \ \eta = 0.5$
$E = 0.2081, \ acc = 0.92$



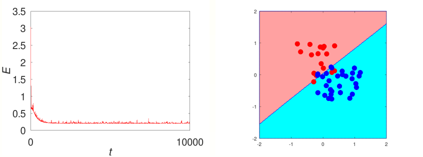
$t_{\mathrm{MAX}} = 2000, \ \eta = 0.1$
$E = 0.2056, \ acc = 0.90$



$t_{\mathrm{MAX}} = 10000, \ \eta = 0.5$
$E = 0.1835, \ acc = 0.90$

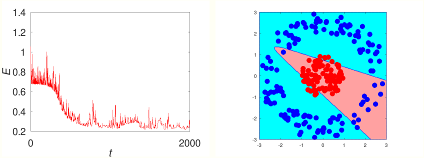


$t_{\mathrm{MAX}} = 10000, \ \eta = 0.1$
$E = 0.2034, \ acc = 0.88$

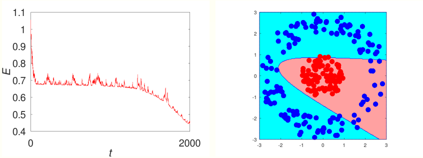


ex8 | RNA1_1–MGE | $J = 2$

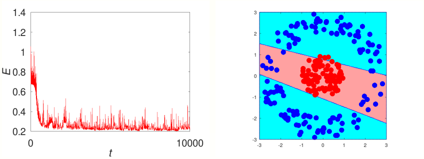
$t_{\text{MAX}} = 2000,\ \eta = 0.5$
$E = 0.3021,\ acc = 0.88$



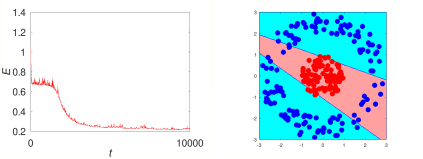
$t_{\text{MAX}} = 2000,\ \eta = 0.1$
$E = 0.4432,\ acc = 0.93$



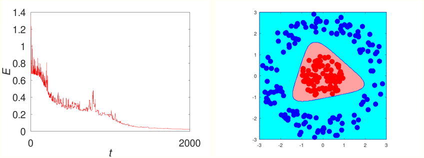
$t_{\text{MAX}} = 10000,\ \eta = 0.5$
$E = 0.2321,\ acc = 0.92$



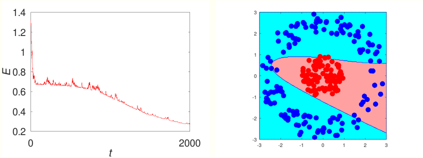
$t_{\text{MAX}} = 10000,\ \eta = 0.1$
$E = 0.2233,\ acc = 0.91$



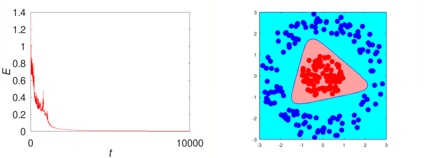
$t_{\mathrm{MAX}} = 2000,\ \eta = 0.5$
$E = 0.0252,\ acc = 1$



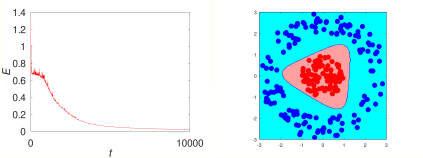
$t_{\mathrm{MAX}} = 2000,\ \eta = 0.1$
$E = 0.2717,\ acc = 0.94$



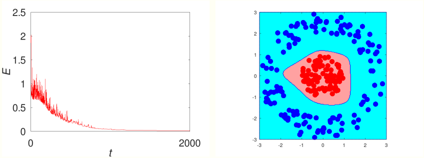
$t_{\mathrm{MAX}} = 10000,\ \eta = 0.5$
$E = 0.0034,\ acc = 1$



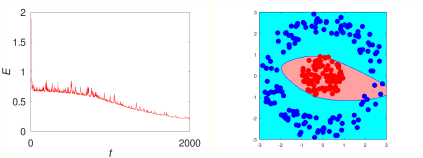
$t_{\mathrm{MAX}} = 10000,\ \eta = 0.1$
$E = 0.0218,\ acc = 1$



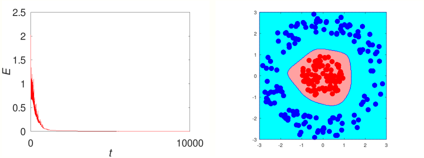
$t_{\text{MAX}} = 2000,\ \eta = 0.5$
$E = 0.0132,\ acc = 1$



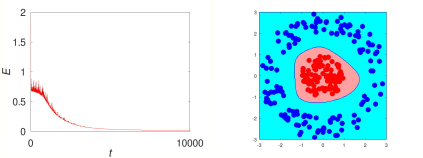
$t_{\text{MAX}} = 2000,\ \eta = 0.1$
$E = 0.2054,\ acc = 0.98$



$t_{\text{MAX}} = 10000,\ \eta = 0.5$
$E = 0.0017,\ acc = 1$



$t_{\text{MAX}} = 10000,\ \eta = 0.1$
$E = 0.0133,\ acc = 1$



Exercícios

Exercício 1. Sejam:

$$f(z) = \sigma(z),$$
$$g(z) = \sigma(z),$$
$$E_n(\hat{p}^n; y^n) = -y^n \ln(\hat{p}^n) - (1 - y^n) \ln(1 - \hat{p}^n).$$

(a) Seja $j \in \{0, \ldots, J\}$. Mostre que:

$$\frac{\partial E_n}{\partial \tilde{v}_j}(\hat{p}^n; y^n) = (\hat{p}^n - y^n)\, \tilde{h}_j.$$

(b) Sejam $i \in \{0, \ldots, J\}$ e $j \in \{1, \ldots, J\}$. Mostre que:

$$\frac{\partial E_n}{\partial \tilde{u}_{ij}}(\hat{p}^n; y^n) = (\hat{p}^n - y^n)\, v_j h_j (1 - h_j) \tilde{x}_i.$$

Exercícios

Exercício 2. Considere a base de dados binária $D = (x^n, y^n)_{n=1}^6$ $(I = 2)$ com

$$\begin{array}{ll} x^1 = (-1, 1)^\top & y^1 = 0, \\ x^2 = (-1, -1)^\top & y^2 = 1, \\ x^3 = (0, 0)^\top & y^3 = 1, \\ x^4 = (1, 1)^\top & y^4 = 1, \\ x^5 = (-1, 0)^\top & y^5 = 0, \\ x^6 = (1, -1)^\top & y^6 = 0. \end{array}$$

Aplique o Algoritmo RNA1_1–MGE e $J = 2$, $t_{\max} = 3$, $\eta = 0.1$ e n dado pela sequência 3, 2, 5 à base de dados D, indicando a *accuracy* que se obteve. Inicialize as matrizes pesos com matrizes nulas.

Exercício 3. Implemente o algoritmo RNA1_1–MGE.