Bienvenue!

ÉCOLE D'HIVER FRANCOPHONE EN APPRENTISSAGE PROFOND

5 - 9 mars 2018

IVADO
HEC Montréal
Polytechnique Montréal
Université de Montréal

MILA

Apprentissage automatique

Alain Tapp

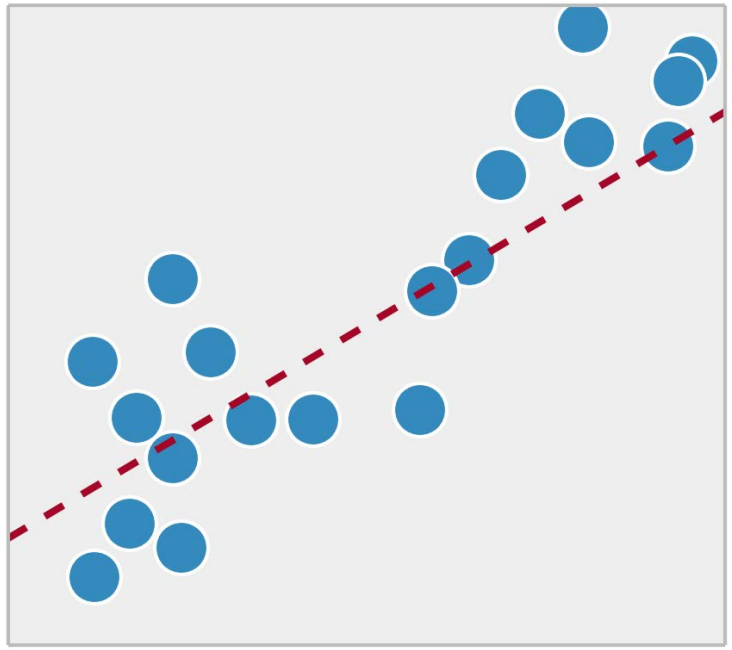
Intelligence artificielle

- Classification
- Régression
- Regroupement *(Clustering)*
- Modèles génératifs
- Apprentissage par renforcement

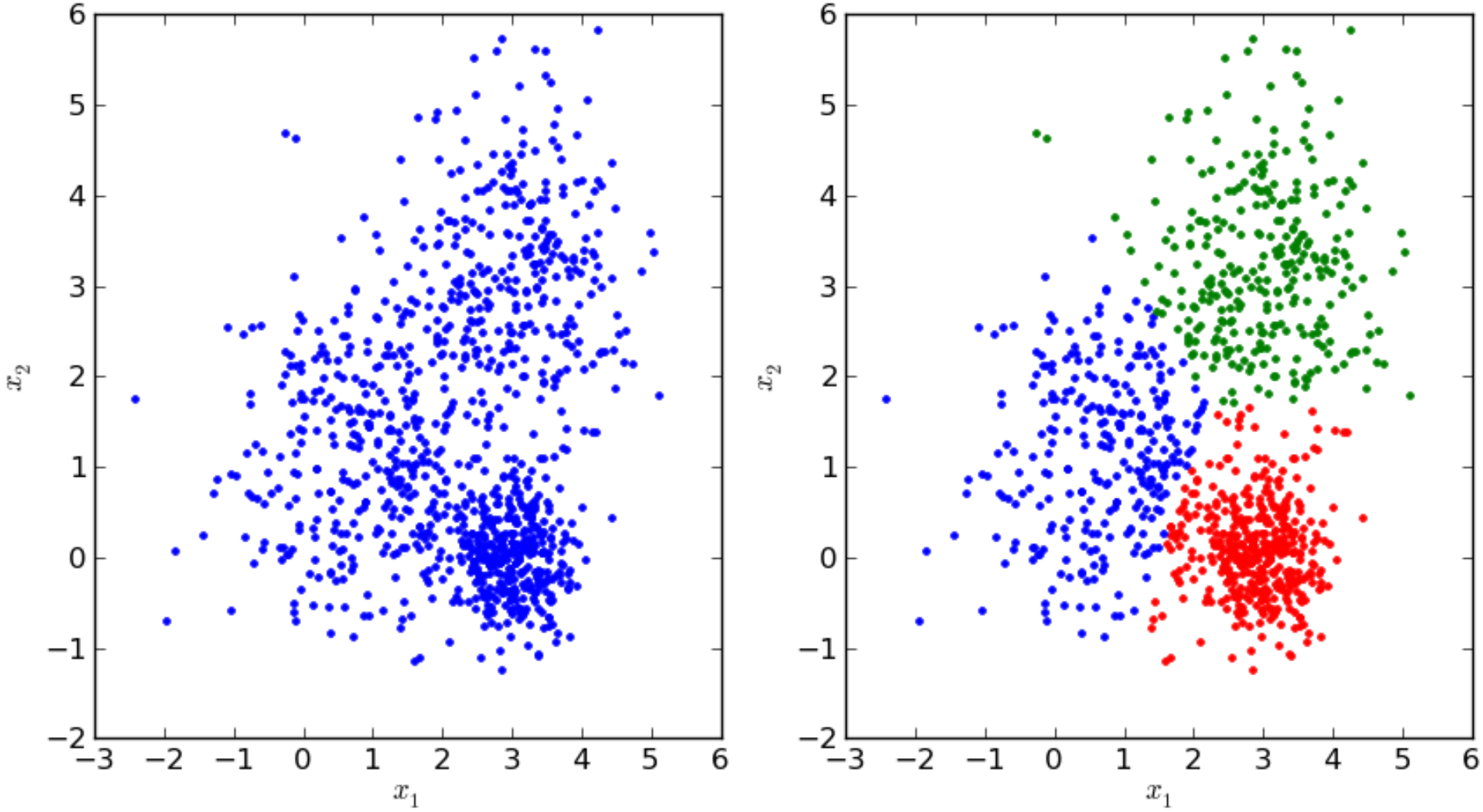
Classification

VS

Régression

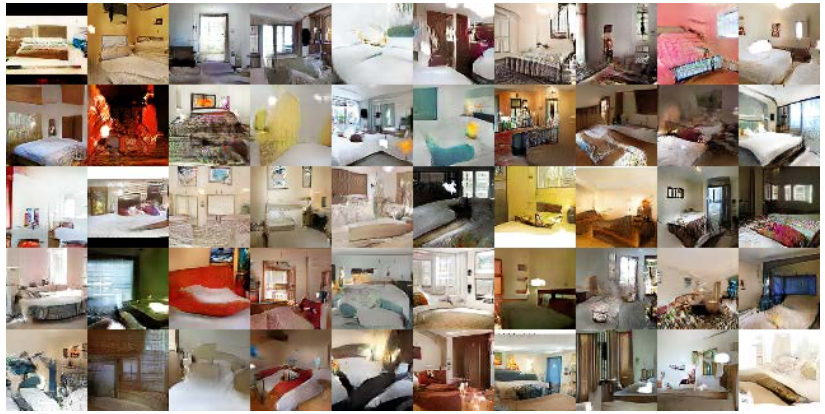


Regroupement (*Clustering*)



Modèles génératifs

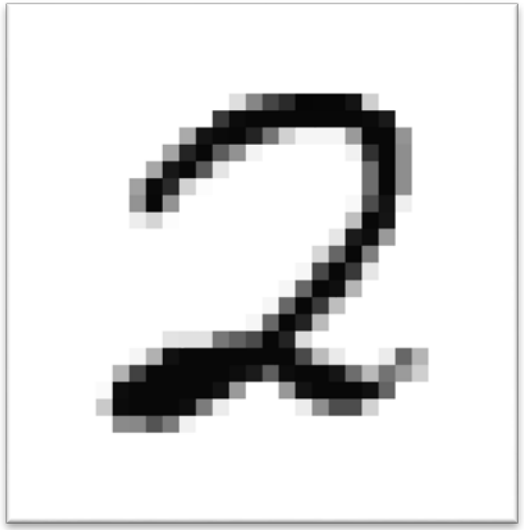
Modèles génératifs



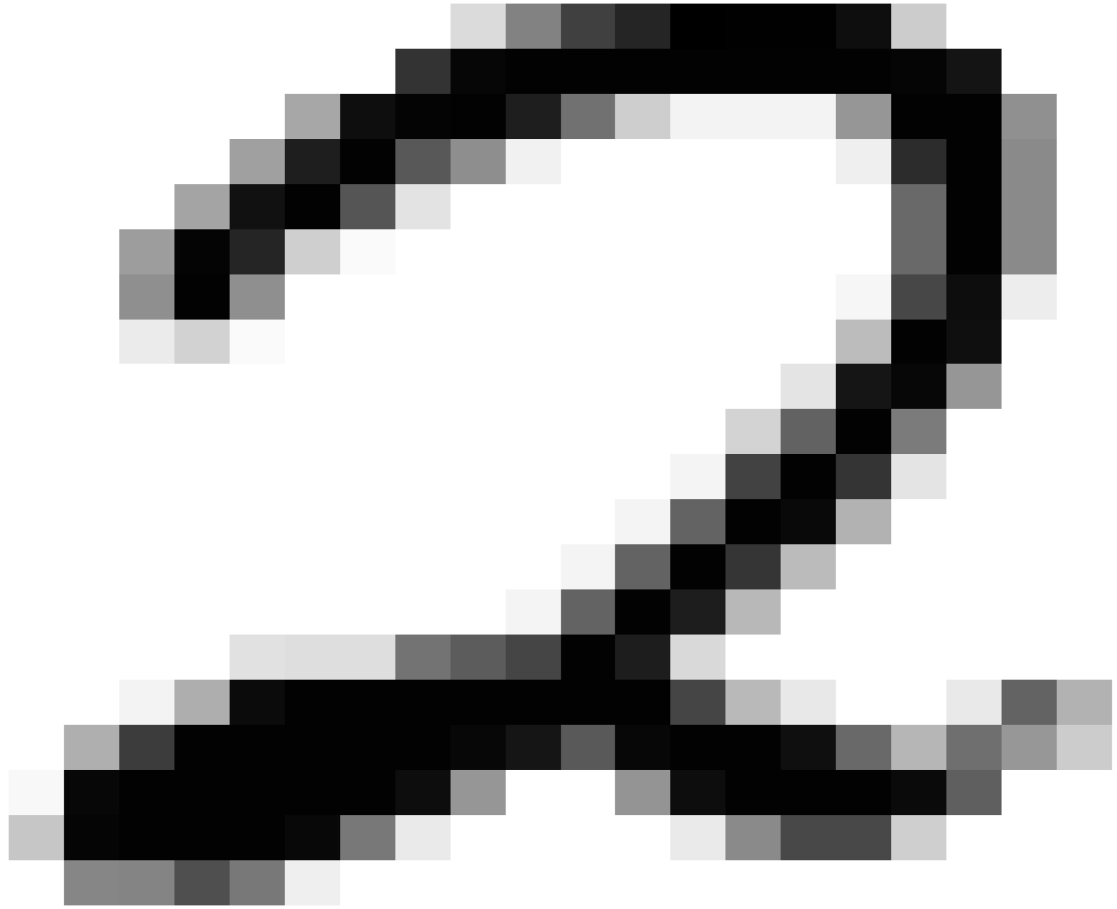
Modèles génératifs

Renforcement

Classification

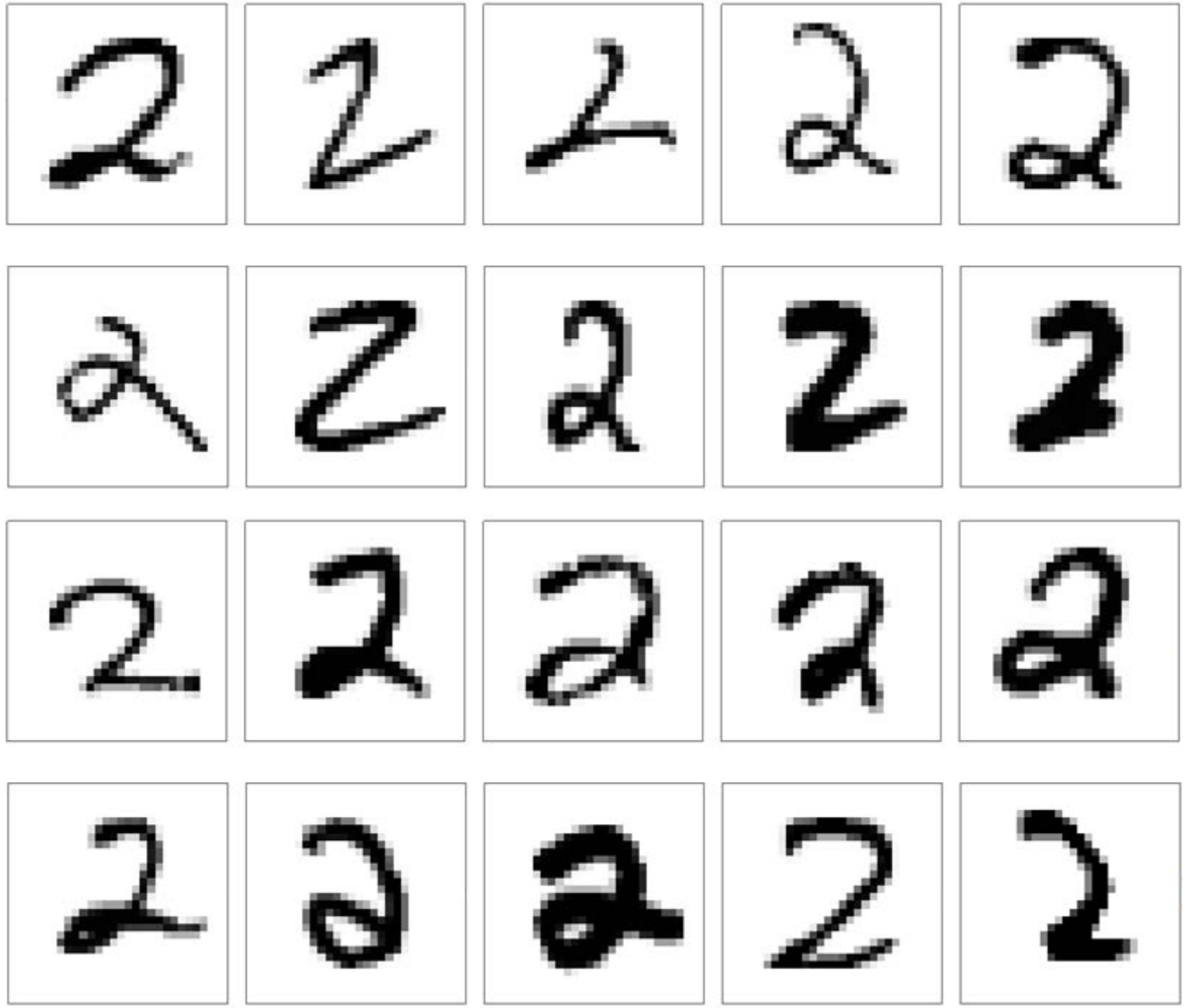


= ?

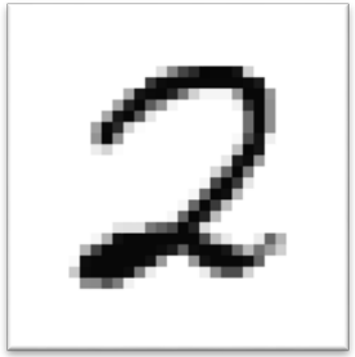


$$
\begin{aligned}
A = (&0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, \\
&0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, \\
&0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, \\
&0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, \\
&0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 36, 125, 191, 218, 255, 254, 254, 241, 51, 0, 0, 0, 0, 0, 0, 0, \\
&0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 204, 249, 253, 253, 253, 253, 253, 253, 253, 250, 235, 0, 0, 0, 0, 0, 0, \\
&0, 0, 0, 0, 0, 0, 0, 0, 0, 88, 241, 251, 253, 225, 142, 49, 12, 12, 12, 105, 253, 253, 111, 0, 0, 0, 0, 0, \\
&0, 0, 0, 0, 0, 0, 0, 0, 95, 225, 253, 167, 113, 14, 0, 0, 0, 0, 0, 16, 211, 253, 117, 0, 0, 0, 0, 0, \\
&0, 0, 0, 0, 0, 0, 0, 91, 238, 253, 170, 28, 0, 0, 0, 0, 0, 0, 0, 0, 150, 253, 117, 0, 0, 0, 0, 0, \\
&0, 0, 0, 0, 0, 0, 98, 251, 218, 48, 5, 0, 0, 0, 0, 0, 0, 0, 0, 0, 150, 253, 117, 0, 0, 0, 0, 0, \\
&0, 0, 0, 0, 0, 0, 112, 253, 112, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 9, 184, 242, 18, 0, 0, 0, 0, 0, \\
&0, 0, 0, 0, 0, 0, 20, 45, 5, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 67, 253, 240, 0, 0, 0, 0, 0, 0, \\
&0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 27, 234, 248, 105, 0, 0, 0, 0, 0, 0, \\
&0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 44, 157, 253, 132, 0, 0, 0, 0, 0, 0, 0, \\
&0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 11, 189, 253, 203, 27, 0, 0, 0, 0, 0, 0, 0, \\
&0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 11, 156, 253, 246, 77, 0, 0, 0, 0, 0, 0, 0, 0, \\
&0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 11, 156, 253, 202, 68, 0, 0, 0, 0, 0, 0, 0, 0, 0, \\
&0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 11, 156, 253, 226, 71, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, \\
&0, 0, 0, 0, 0, 0, 0, 0, 30, 33, 33, 140, 163, 186, 253, 226, 38, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, \\
&0, 0, 0, 0, 0, 0, 12, 81, 244, 253, 253, 253, 253, 253, 253, 253, 186, 70, 23, 0, 0, 22, 156, 77, 0, 0, 0, 0, \\
&0, 0, 0, 0, 0, 80, 195, 253, 253, 253, 253, 253, 248, 234, 166, 248, 253, 253, 240, 150, 73, 144, 104, 51, 0, 0, 0, 0, \\
&0, 0, 0, 0, 7, 248, 253, 253, 253, 253, 253, 242, 105, 0, 0, 107, 242, 253, 253, 253, 245, 160, 0, 0, 0, 0, 0, 0, \\
&0, 0, 0, 0, 57, 250, 253, 253, 253, 247, 135, 21, 0, 0, 0, 0, 21, 117, 183, 183, 48, 0, 0, 0, 0, 0, 0, 0, \\
&0, 0, 0, 0, 0, 121, 123, 176, 135, 16, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, \\
&0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, \\
&0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, \\
&0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, \\
&0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)
\end{aligned}
$$

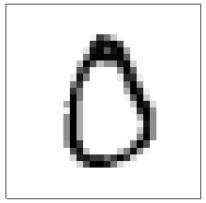
$$A = (a_1, a_2, a_3, a_4, a_5, \cdots, a_{780}, a_{781}, a_{782}, a_{783}, a_{784})$$



60000 exemples d'apprentissage

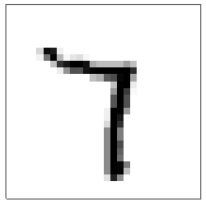


= ?



=0

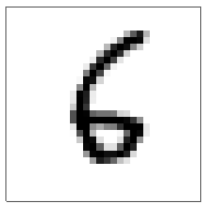
=2



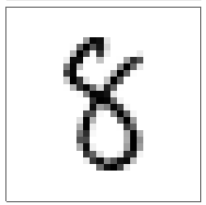
=7

=9

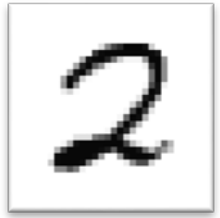
=2

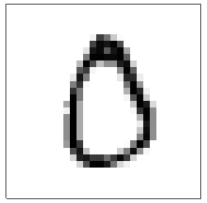


=6



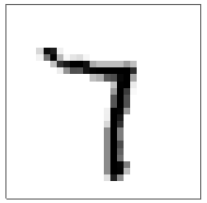
=8





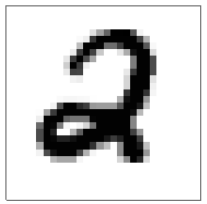
=0

=2

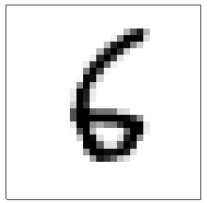


=7

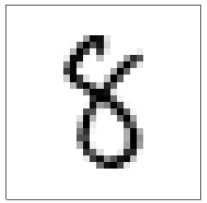
=9



=2



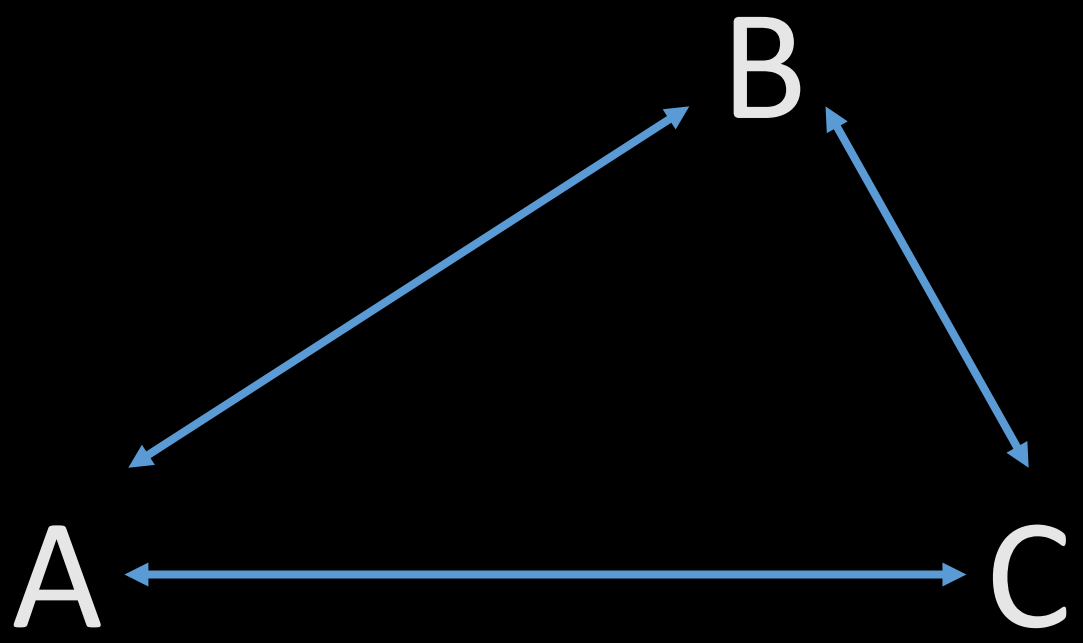
=6



=8

Distance

Positive: $D(A,B) \geq 0$
Symétrie: $D(A,B) = D(B,A)$
Séparation: $D(A,B) = 0 \rightarrow A = B$
Inégalité triangulaire: $D(A,C) \leq D(A,B) + D(B,C)$



Distance Euclidienne

$D(A, B)$

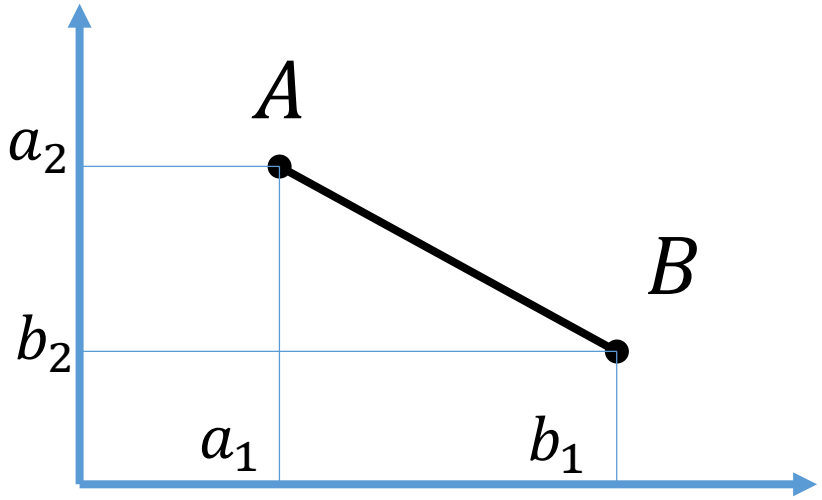
Distance Euclidienne

$$D(A,B) = b_1 - a_1$$

Distance Euclidienne

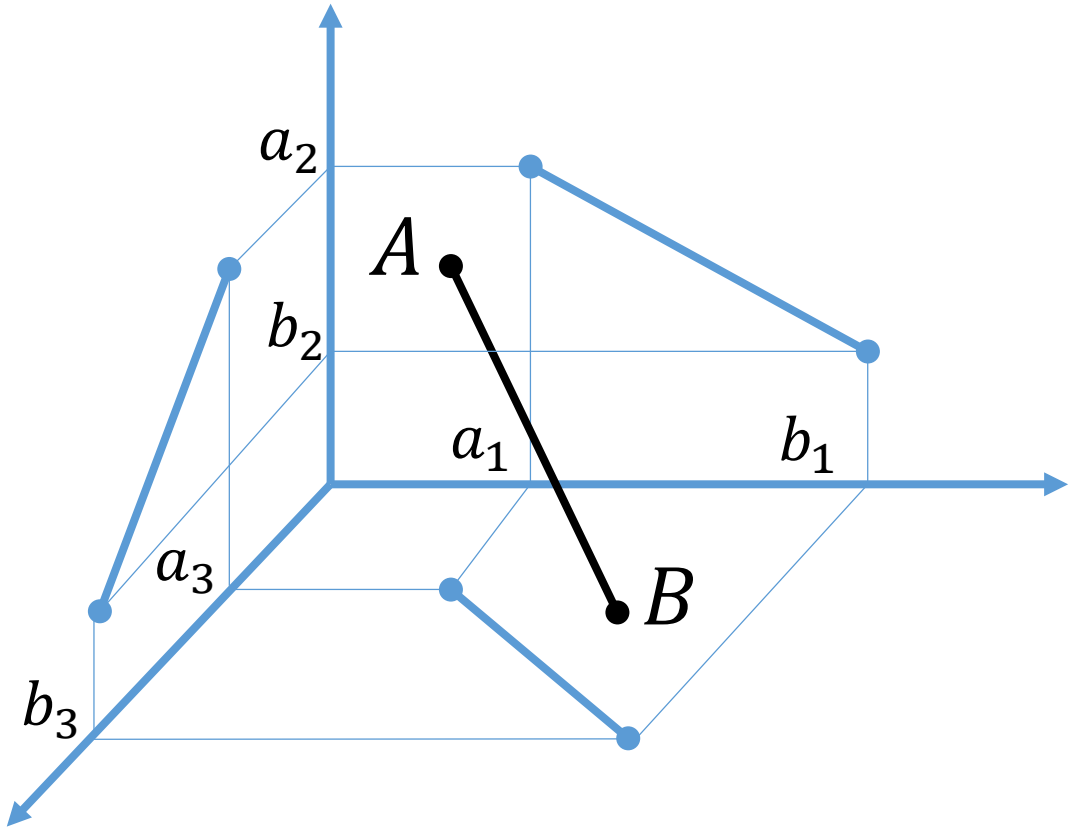
$$D(A,B) = \sqrt{(b_1 - a_1)^2}$$

Distance Euclidienne



$$D(A,B) = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2}$$

Distance Euclidienne



$$D(A,B) = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2 + (b_3 - a_3)^2}$$

$$A = (a_1, a_2, a_3) \qquad B = (b_1, b_2, b_3)$$

$$D(A,B) = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2 + (b_3 - a_3)^2}$$

$$A = (a_1, a_2, a_3, a_4) \qquad B = (b_1, b_2, b_3, b_4)$$

$$D(A,B) = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2 + (b_3 - a_3)^2 + (b_4 - a_4)^2}$$

$$A = (a_1, a_2, a_3, \cdots, a_{784}) \quad B = (b_1, b_2, b_3, \cdots, b_{784})$$

$$D(A,B) = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2 + \cdots + (b_{784} - a_{784})^2}$$

A	B	$D(A,B)$
		**2688.85**
		**2862.24**
		**2700.58**
		**2951.47**
		**2171.14**
		**2672.08**
		**2633.81**

Plus proches voisins

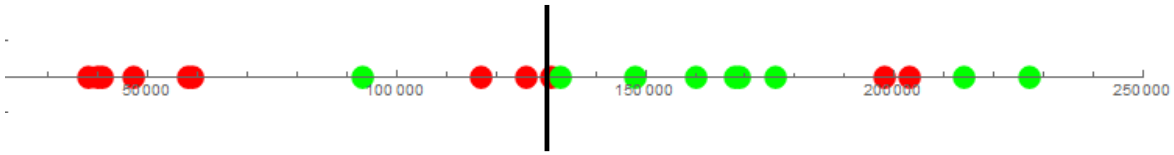
60000 exemples de référence

$$\boldsymbol{E = 3.8\%}$$

Age	Salaire	Scolarité	Diligent
46	169000	18	OUI
68	168000	19	OUI
61	160000	14	OUI
53	117000	16	NON
31	198000	12	NON
30	126000	13	NON
43	131000	13	NON
48	227000	21	OUI
71	148000	18	OUI
60	47000	19	NON
56	214000	18	OUI
57	133000	16	OUI
41	203000	9	NON
68	176000	17	OUI
71	38000	9	NON
54	58000	15	NON
49	40000	11	NON
75	93000	18	OUI
59	59000	10	NON
57	41000	13	NON

Modéliser et prédire : Diligence

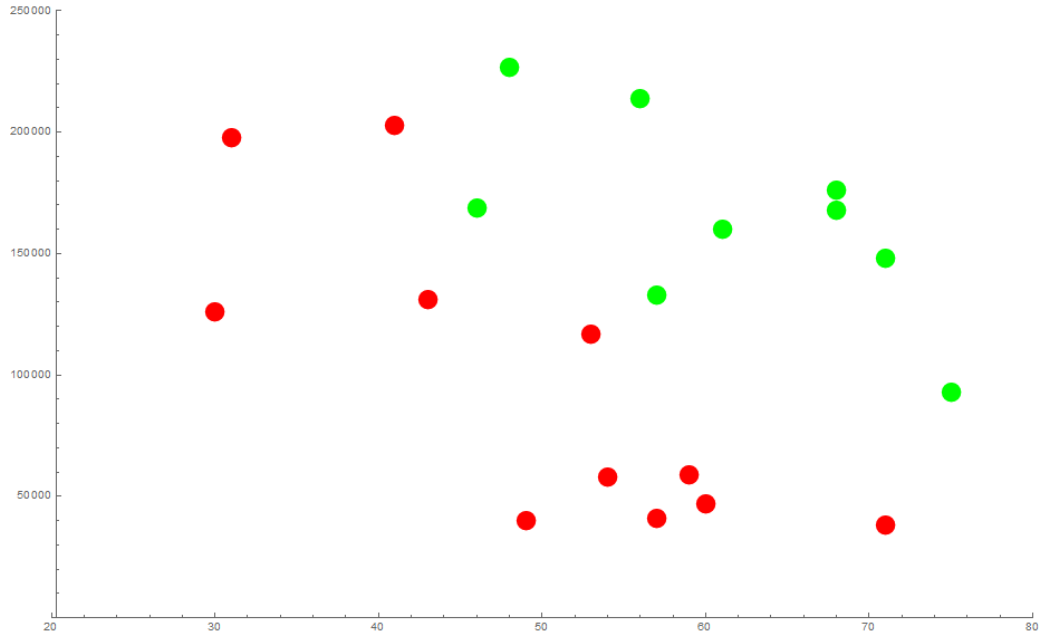
Age	Salaire	Scolarité	Diligent
46	169000	18	OUI
68	168000	19	OUI
61	160000	14	OUI
53	117000	16	NON
31	198000	12	NON
30	126000	13	NON
43	131000	13	NON
48	227000	21	OUI
71	148000	18	OUI
60	47000	19	NON
56	214000	18	OUI
57	133000	16	OUI
41	203000	9	NON
68	176000	17	OUI
71	38000	9	NON
54	58000	15	NON
49	40000	11	NON
75	93000	18	OUI
59	59000	10	NON
57	41000	13	NON



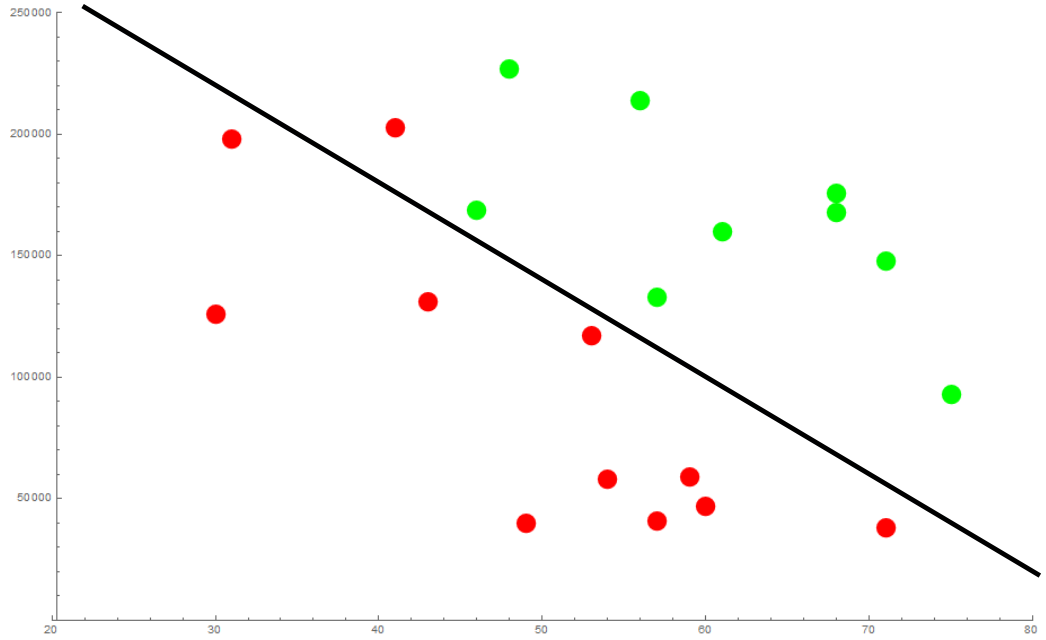
Salaire > 132000

$$\boldsymbol{E = 15\%}$$

Age	Salaire	Scolarité	Diligent
46	169000	18	OUI
68	168000	19	OUI
61	160000	14	OUI
53	117000	16	NON
31	198000	12	NON
30	126000	13	NON
43	131000	13	NON
48	227000	21	OUI
71	148000	18	OUI
60	47000	19	NON
56	214000	18	OUI
57	133000	16	OUI
41	203000	9	NON
68	176000	17	OUI
71	38000	9	NON
54	58000	15	NON
49	40000	11	NON
75	93000	18	OUI
59	59000	10	NON
57	41000	13	NON



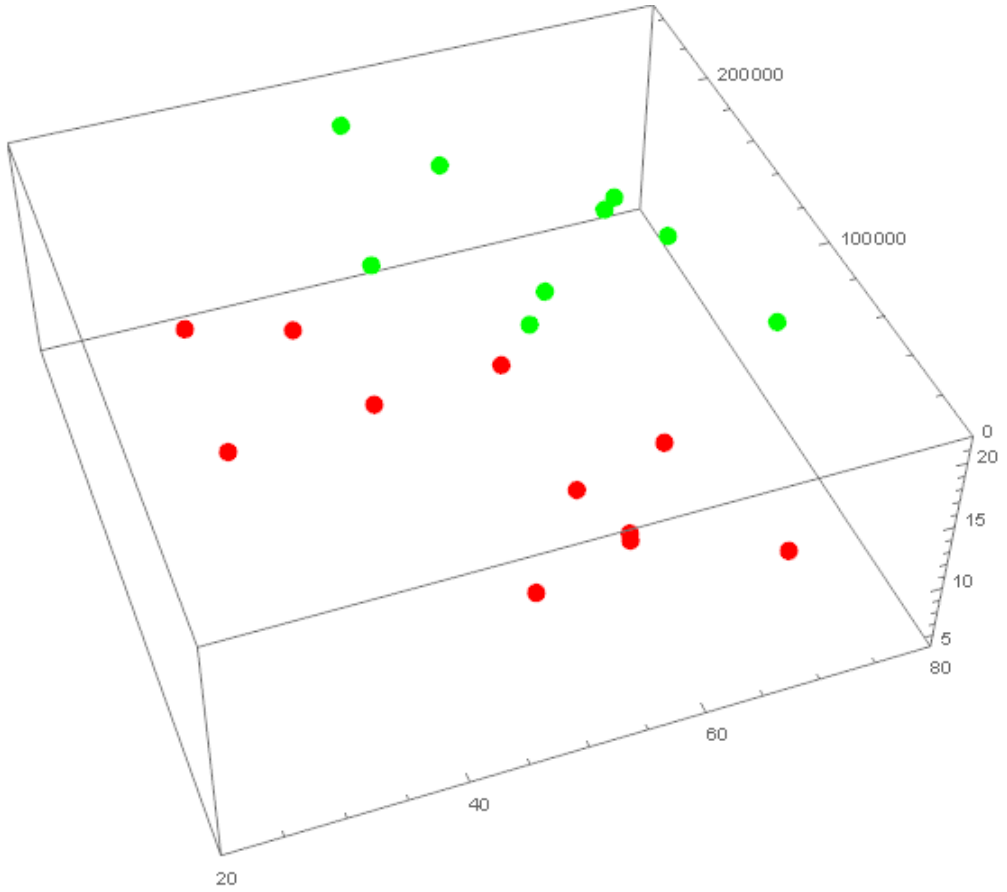
Age	Salaire	Scolarité	Diligent
46	169000	18	OUI
68	168000	19	OUI
61	160000	14	OUI
53	117000	16	NON
31	198000	12	NON
30	126000	13	NON
43	131000	13	NON
48	227000	21	OUI
71	148000	18	OUI
60	47000	19	NON
56	214000	18	OUI
57	133000	16	OUI
41	203000	9	NON
68	176000	17	OUI
71	38000	9	NON
54	58000	15	NON
49	40000	11	NON
75	93000	18	OUI
59	59000	10	NON
57	41000	13	NON



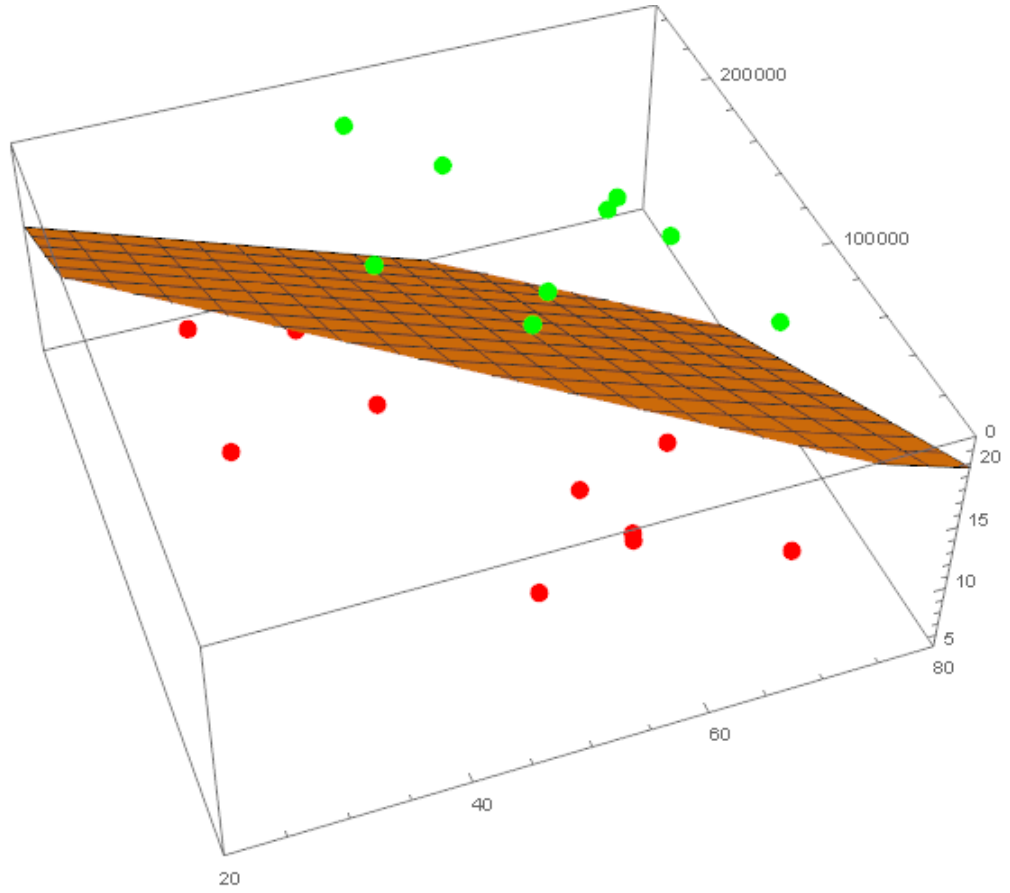
$$4000\,\mathrm{Age} + \mathrm{Salaire} > 340000$$

$$\boldsymbol{E = 5\%}$$

Age	Salaire	Scolarité	Diligent
46	169000	18	OUI
68	168000	19	OUI
61	160000	14	OUI
53	117000	16	NON
31	198000	12	NON
30	126000	13	NON
43	131000	13	NON
48	227000	21	OUI
71	148000	18	OUI
60	47000	19	NON
56	214000	18	OUI
57	133000	16	OUI
41	203000	9	NON
68	176000	17	OUI
71	38000	9	NON
54	58000	15	NON
49	40000	11	NON
75	93000	18	OUI
59	59000	10	NON
57	41000	13	NON



Age	Salaire	Scolarité	Diligent
46	169000	18	OUI
68	168000	19	OUI
61	160000	14	OUI
53	117000	16	NON
31	198000	12	NON
30	126000	13	NON
43	131000	13	NON
48	227000	21	OUI
71	148000	18	OUI
60	47000	19	NON
56	214000	18	OUI
57	133000	16	OUI
41	203000	9	NON
68	176000	17	OUI
71	38000	9	NON
54	58000	15	NON
49	40000	11	NON
75	93000	18	OUI
59	59000	10	NON
57	41000	13	NON



$$11900\ \text{Scolarité} + 3250\ \text{Age} + \text{Salaire} > 499000$$

$$\boldsymbol{E = 0\%}$$

$$\text{Salaire} > 132000$$
$$4000\ \text{Age} + \text{Salaire} > 340000$$
$$11900\ \text{Scolarité} + 3250\ \text{Age} + \text{Salaire} > 499000$$

$$a_1 > 132000$$
$$4000a_2 + a_1 > 340000$$
$$11900a_3 + 3250a_2 + a_1 > 499000$$

$$1789a_4 + 1190a_3 + 4360a_2 + a_1 > 273456$$
$$2719a_5 + 5799a_4 + 9018a_3 + 3477a_2 + a_1 > 345678$$

$$m_5a_5 + m_4a_4 + m_3a_3 + m_2a_2 + m_1a_1 > \text{s}$$

$$\langle M, A \rangle > \text{s}$$

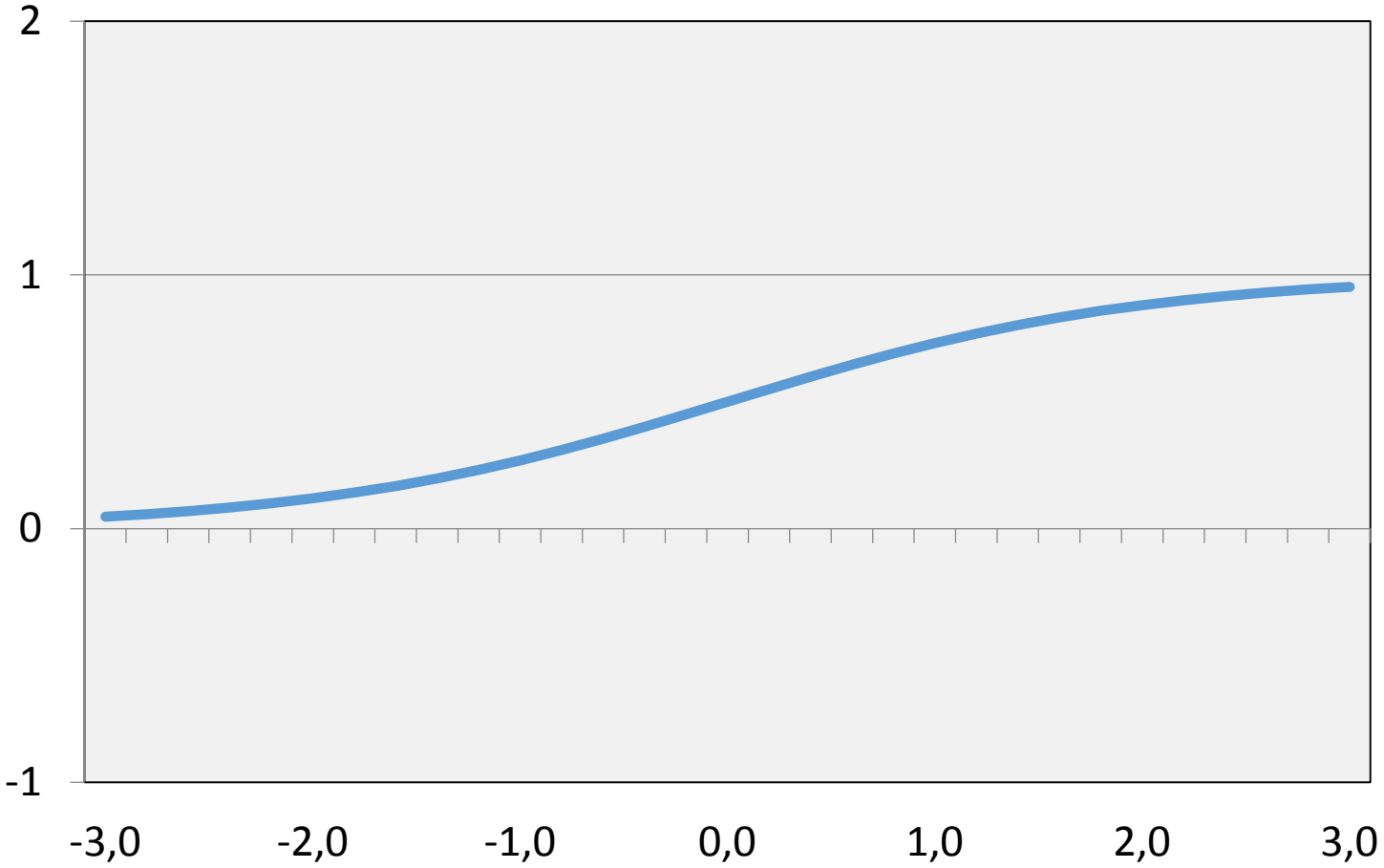
Hyperplan

$$\boldsymbol{E = 7.4\%}$$

La Sigmoïde

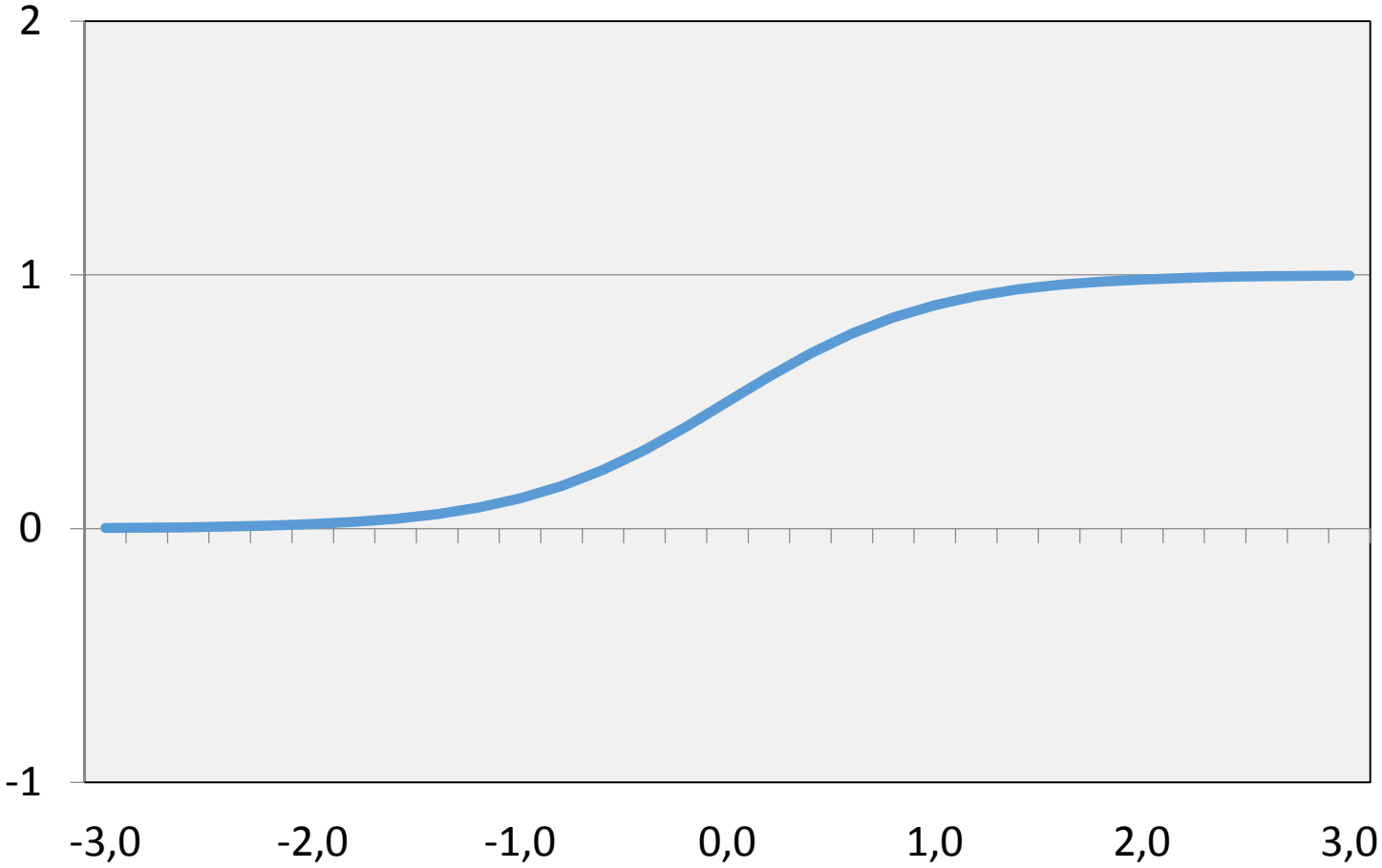
$$\varphi(x) = \frac{1}{1 + e^{-x}}$$

Sigmoïde



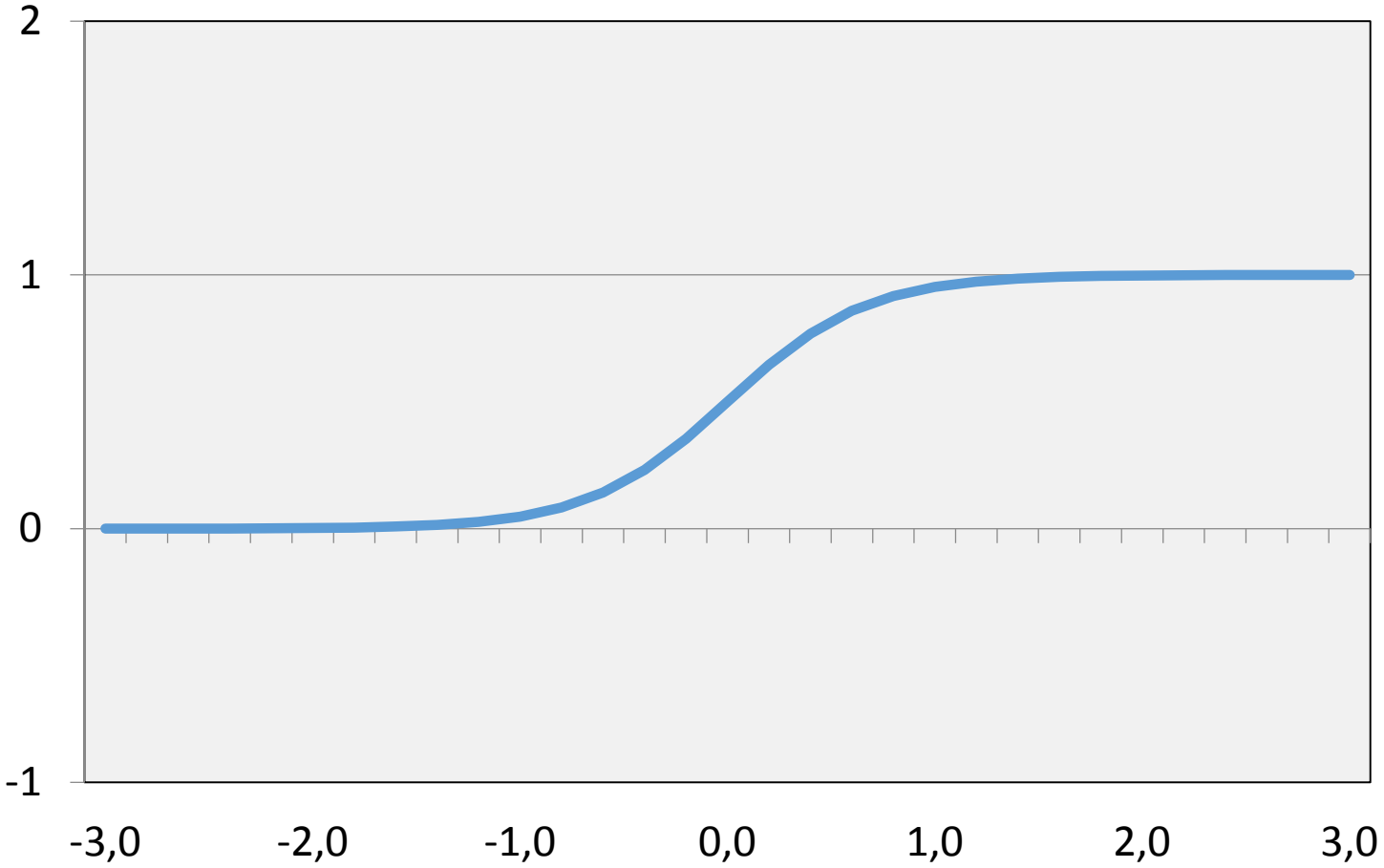
$$\varphi(x) = \frac{1}{1 + e^{-1x}}$$

Sigmoïde



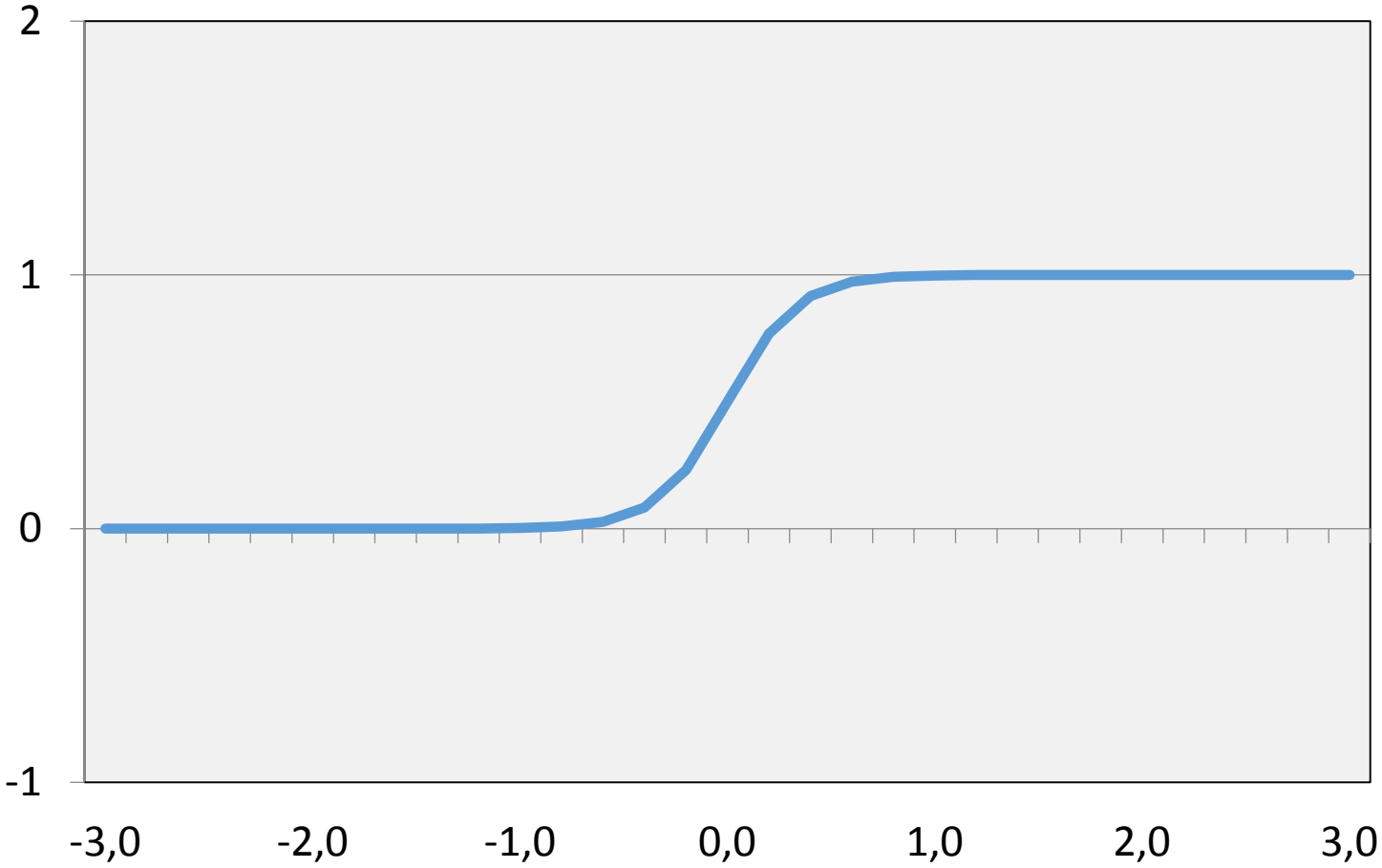
$$\varphi(x) = \frac{1}{1 + e^{-2x}}$$

Sigmoïde



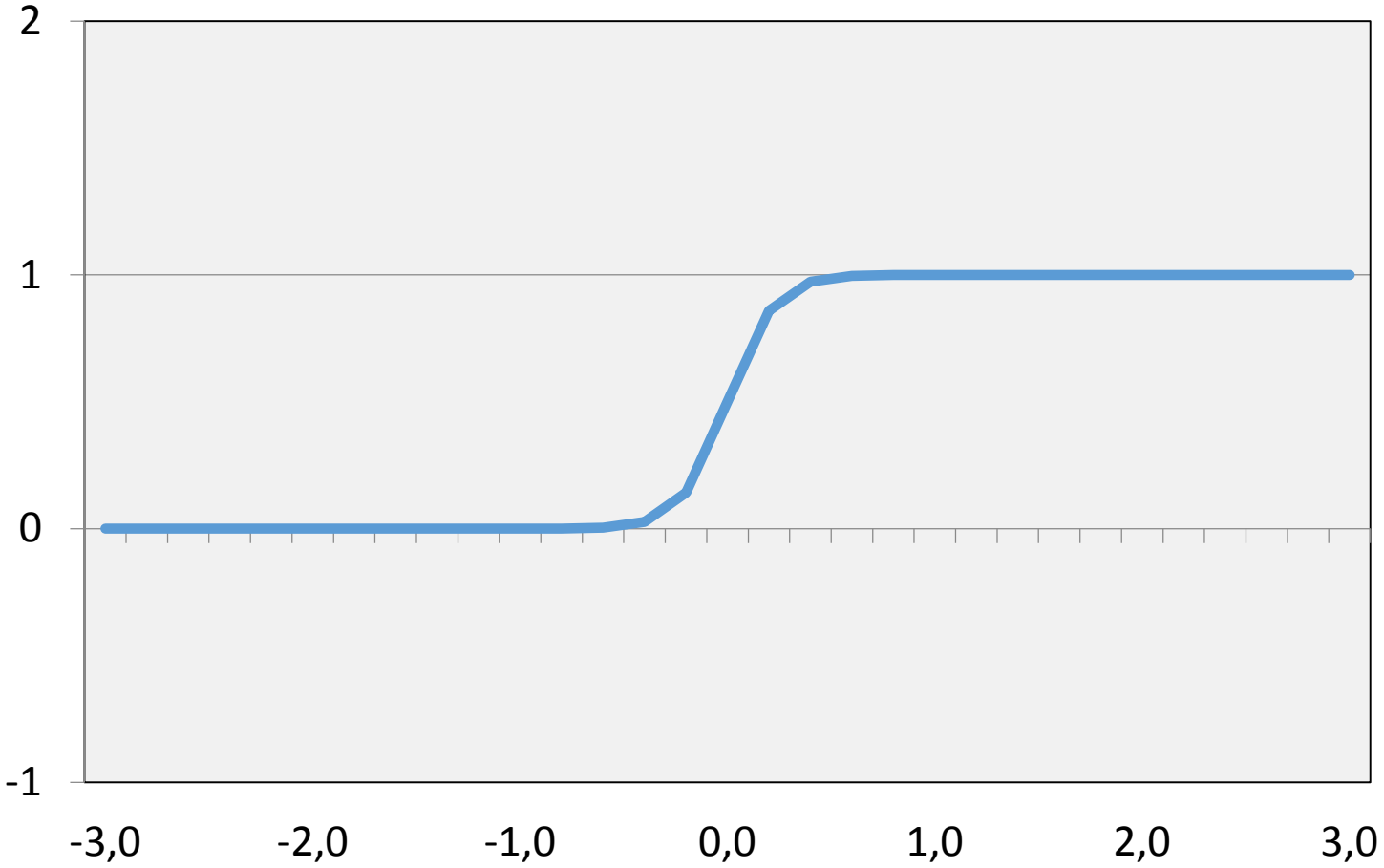
$$\varphi(x) = \frac{1}{1 + e^{-3x}}$$

Sigmoïde



$$\varphi(x) = \frac{1}{1 + e^{-6x}}$$

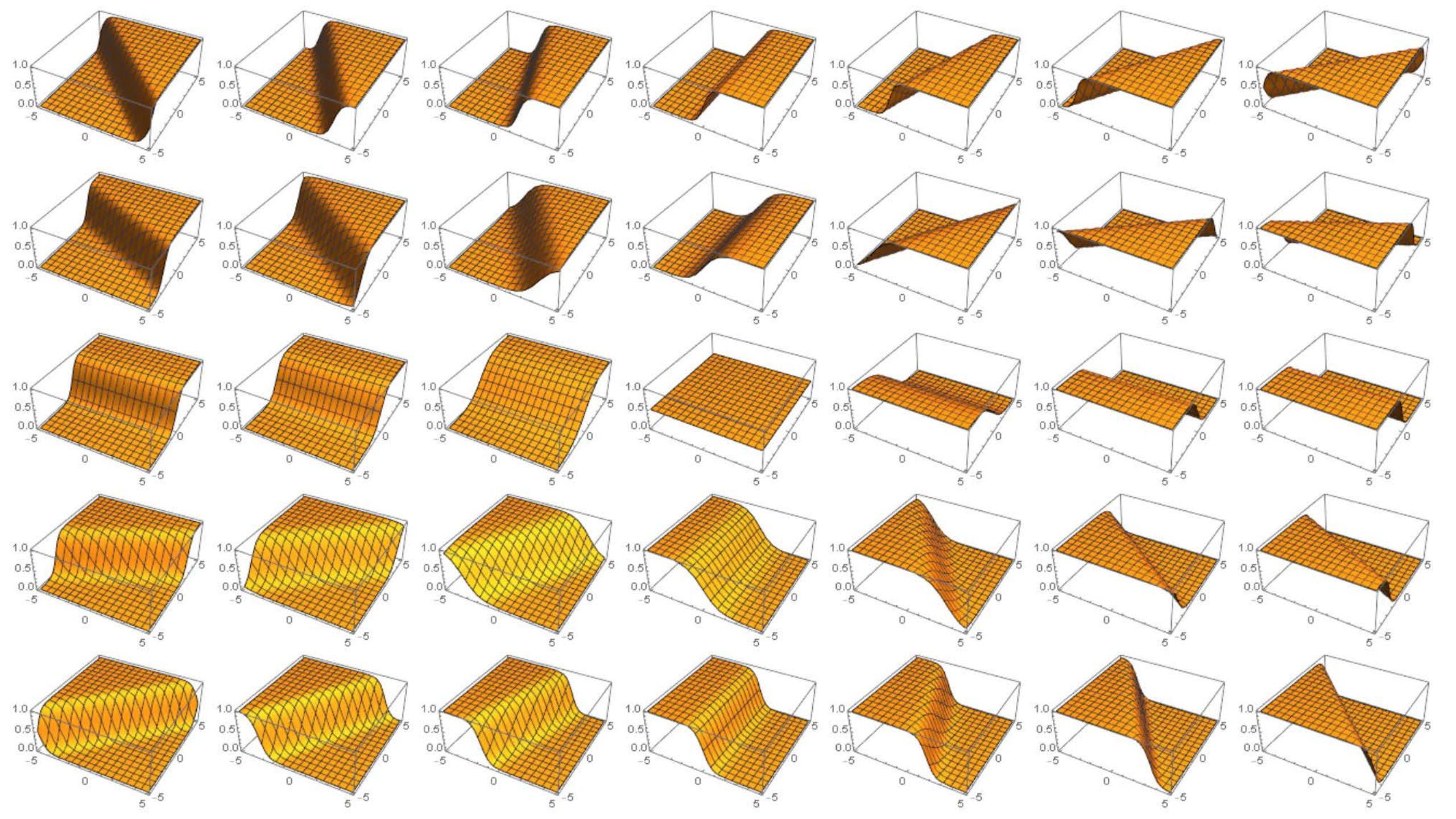
Sigmoïde



$$\varphi(x) = \frac{1}{1 + e^{-9x}}$$

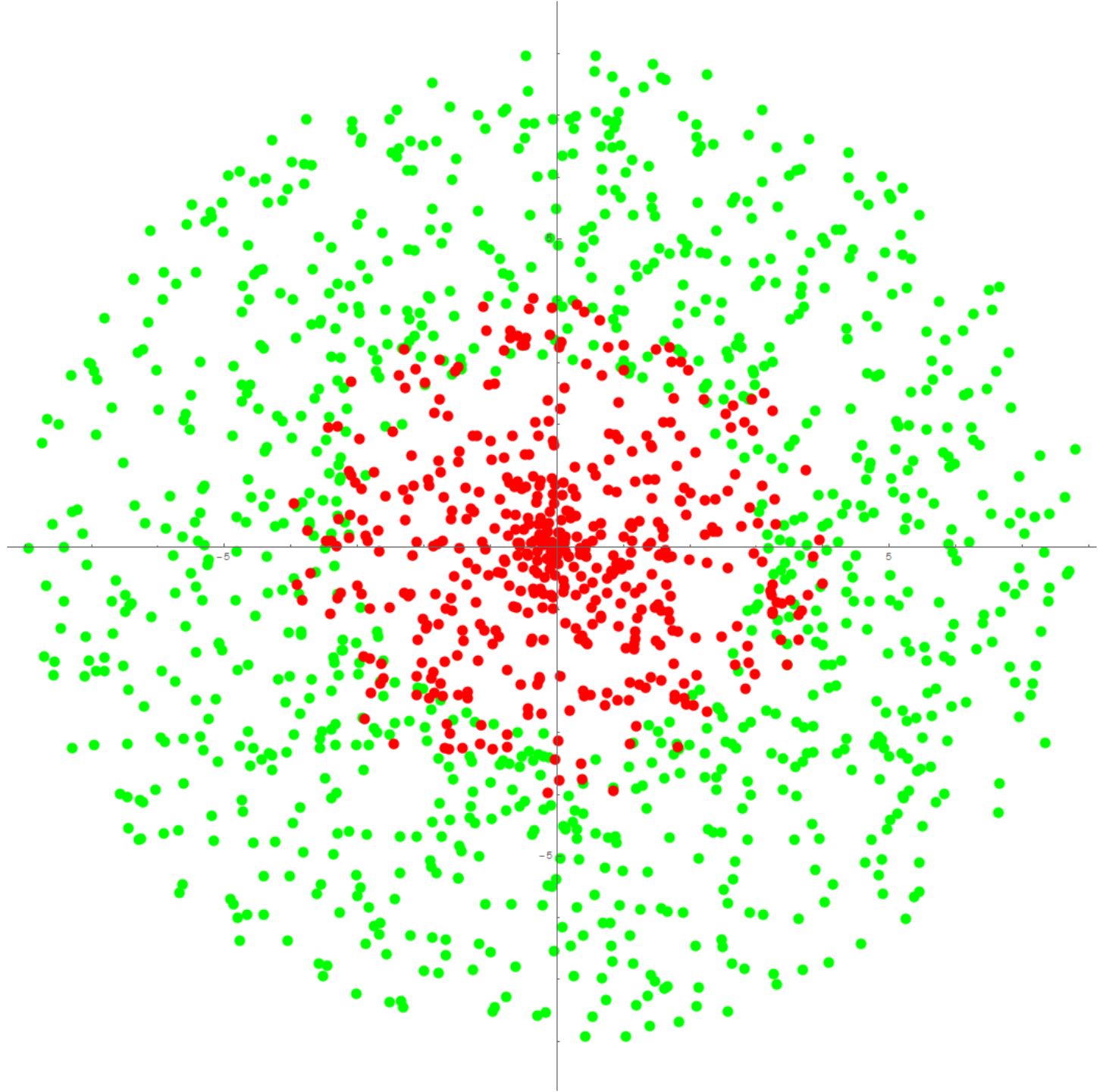
Classification Logistique

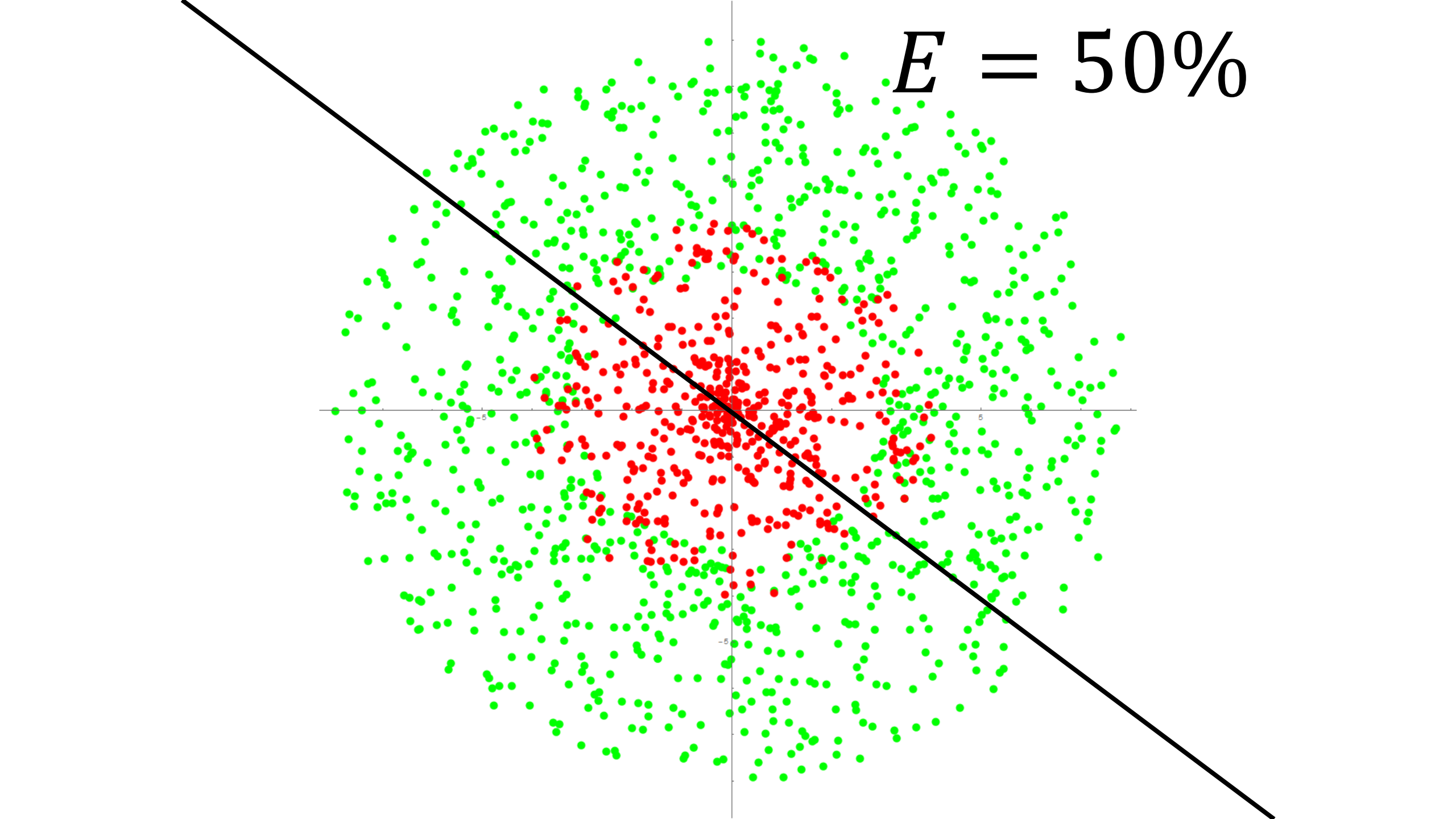
$$P_\theta(x) = \frac{1}{1 + e^{-\langle \theta, x \rangle}}$$

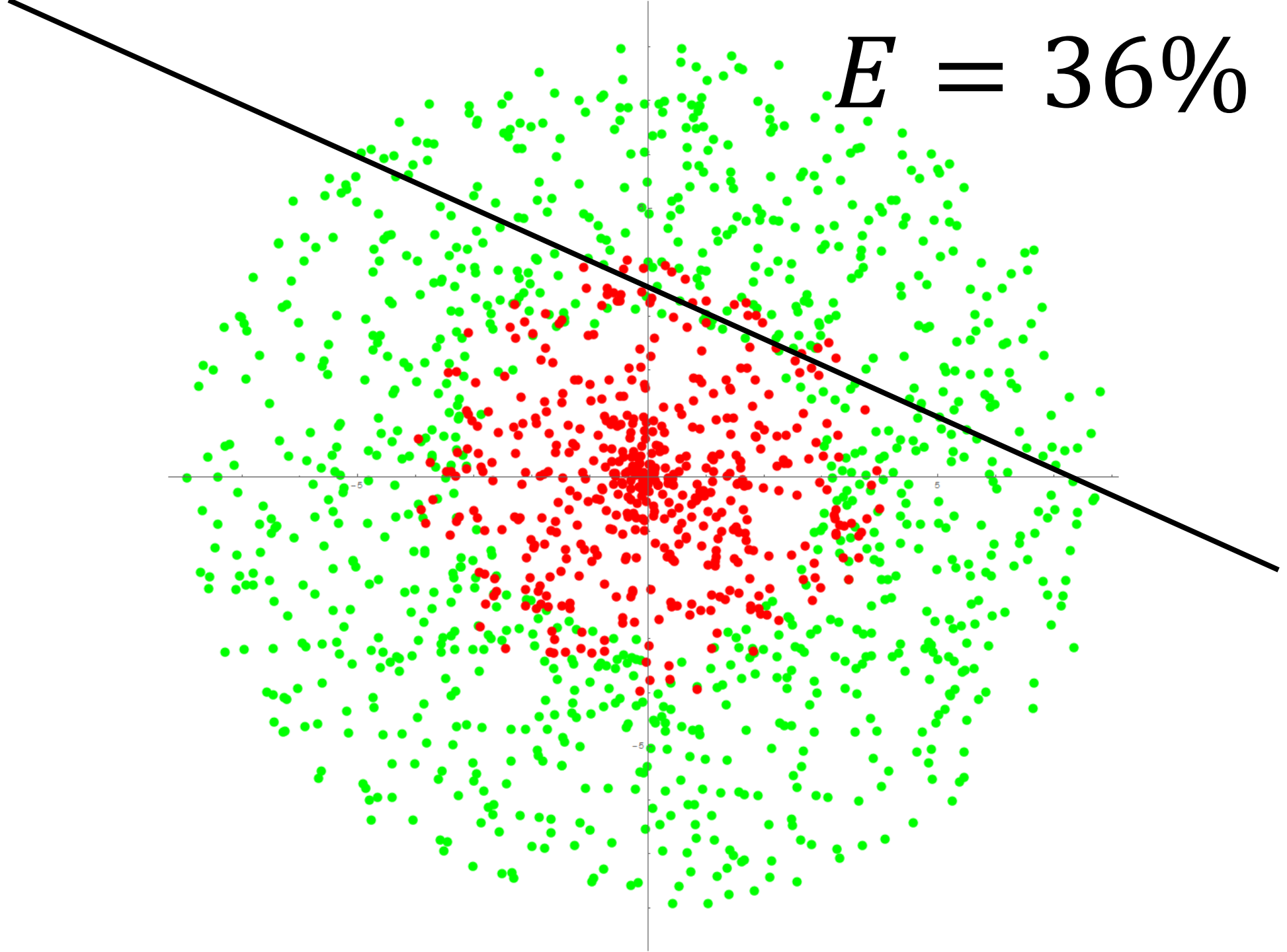

1.0
0.5
0.0
-5
0
5
-5
0
5

Δx	Δy	OK
-3.85	4.81	OUI
-1.53	-7.57	OUI
-0.25	-0.920	NON
-1.04	-0.890	NON
0.14	-0.033	NON
0.80	7.88	OUI
2.51	0.395	NON
6.89	-0.21	OUI
2.94	4.35	OUI
-1.56	-0.847	NON
-1.01	0.057	NON
2.34	-5.52	OUI
-1.45	-5.23	OUI
-4.68	0.50	OUI
4.27	-3.28	OUI
1.24	1.97	NON
-0.29	0.85	NON
1.35	1.86	NON
-1.57	0.05	NON
5.42	-2.24	OUI
…	…	…

Δx	Δy	OK
-3.85	4.81	OUI
-1.53	-7.57	OUI
-0.25	-0.920	NON
-1.04	-0.890	NON
0.14	-0.033	NON
0.80	7.88	OUI
2.51	0.395	NON
6.89	-0.21	OUI
2.94	4.35	OUI
-1.56	-0.847	NON
-1.01	0.057	NON
2.34	-5.52	OUI
-1.45	-5.23	OUI
-4.68	0.50	OUI
4.27	-3.28	OUI
1.24	1.97	NON
-0.29	0.85	NON
1.35	1.86	NON
-1.57	0.05	NON
5.42	-2.24	OUI
...	...	...

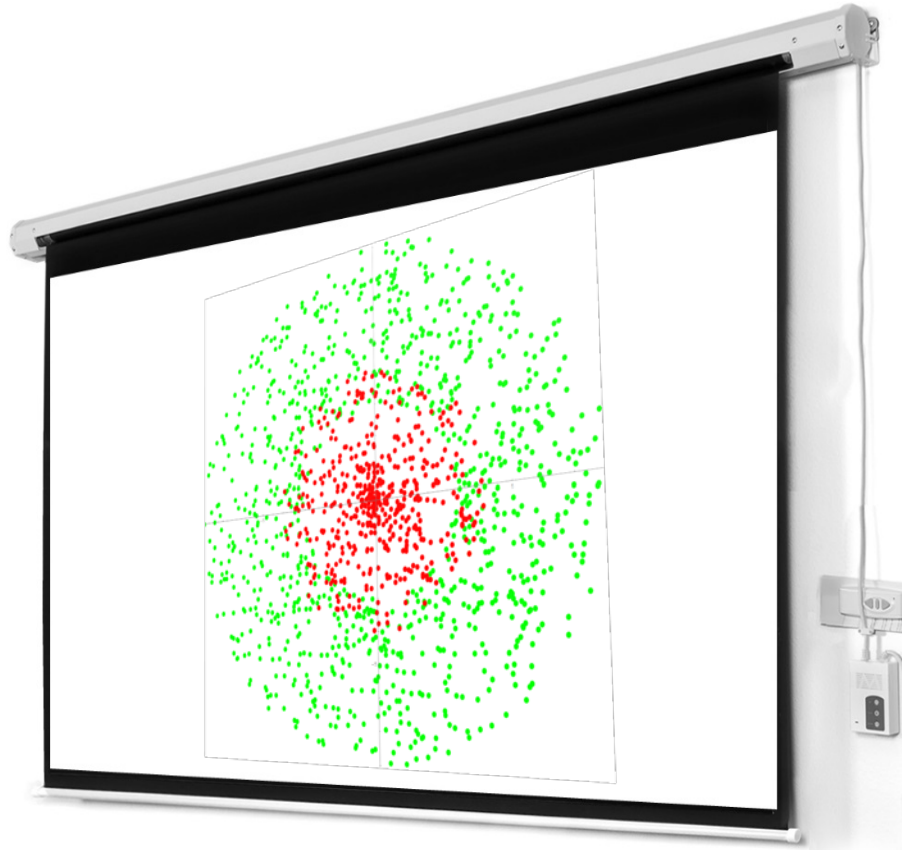

-5
5
-5

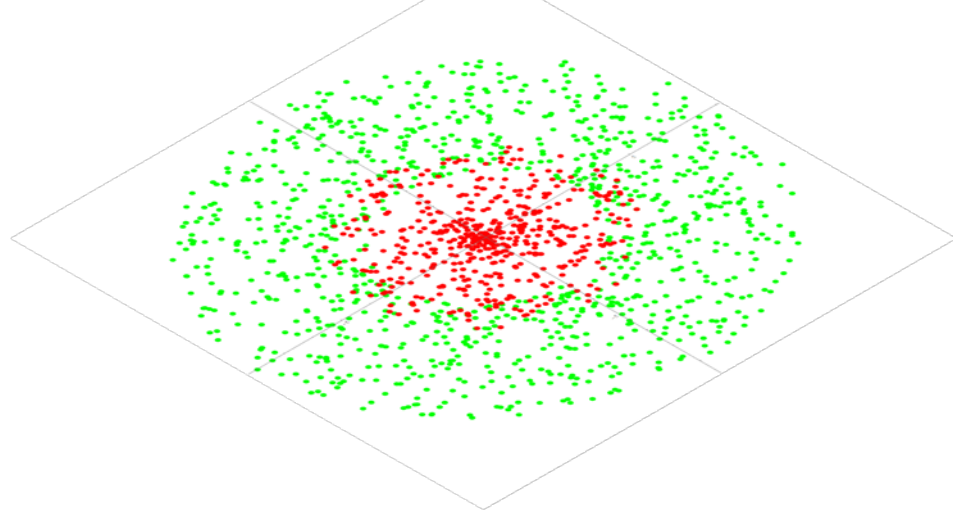

$E = 50\%$

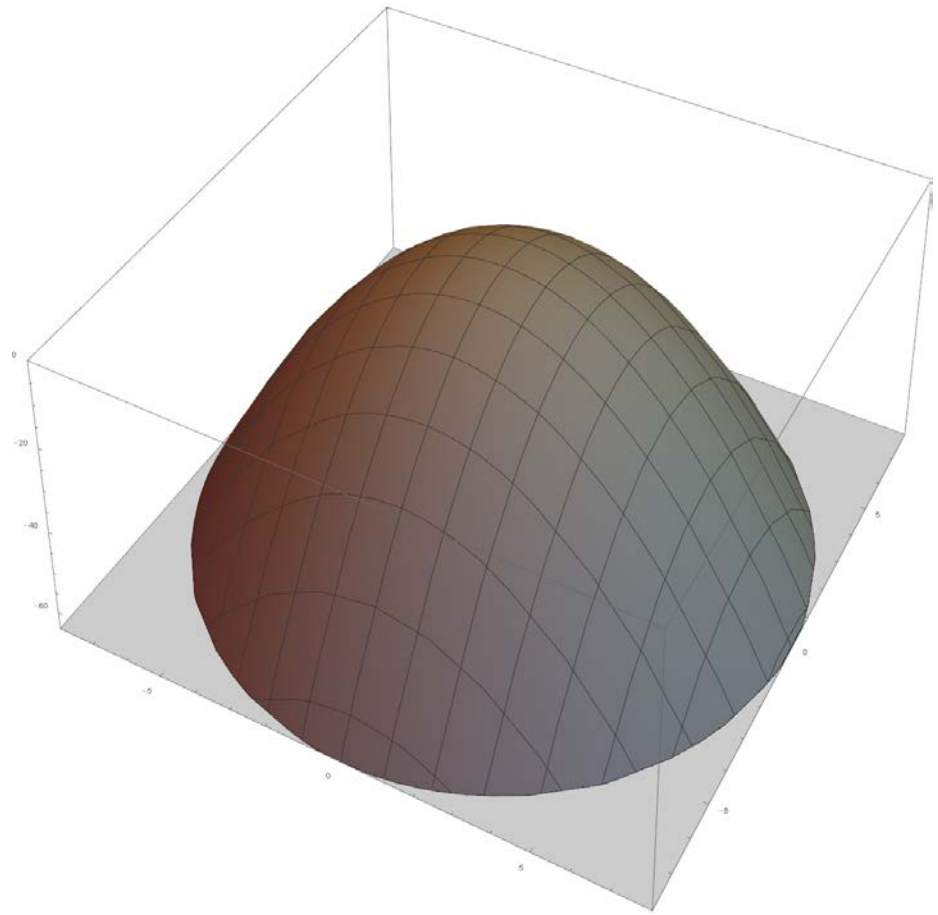

$E = 36\%$

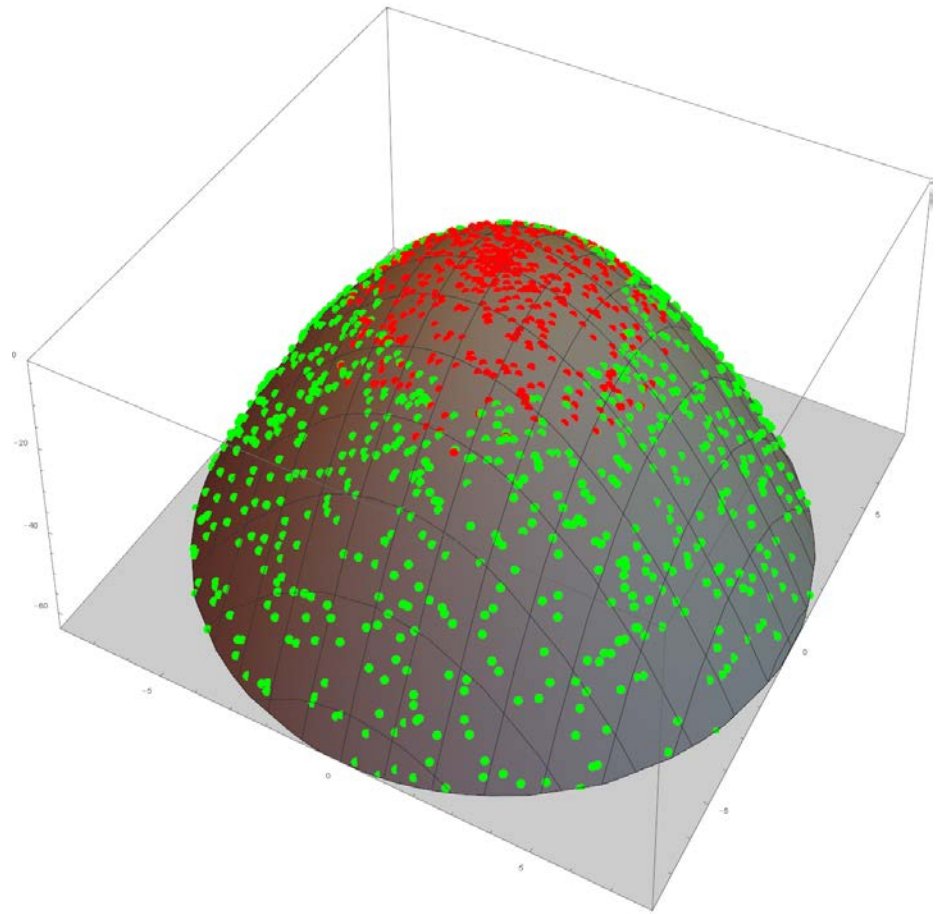
Projection

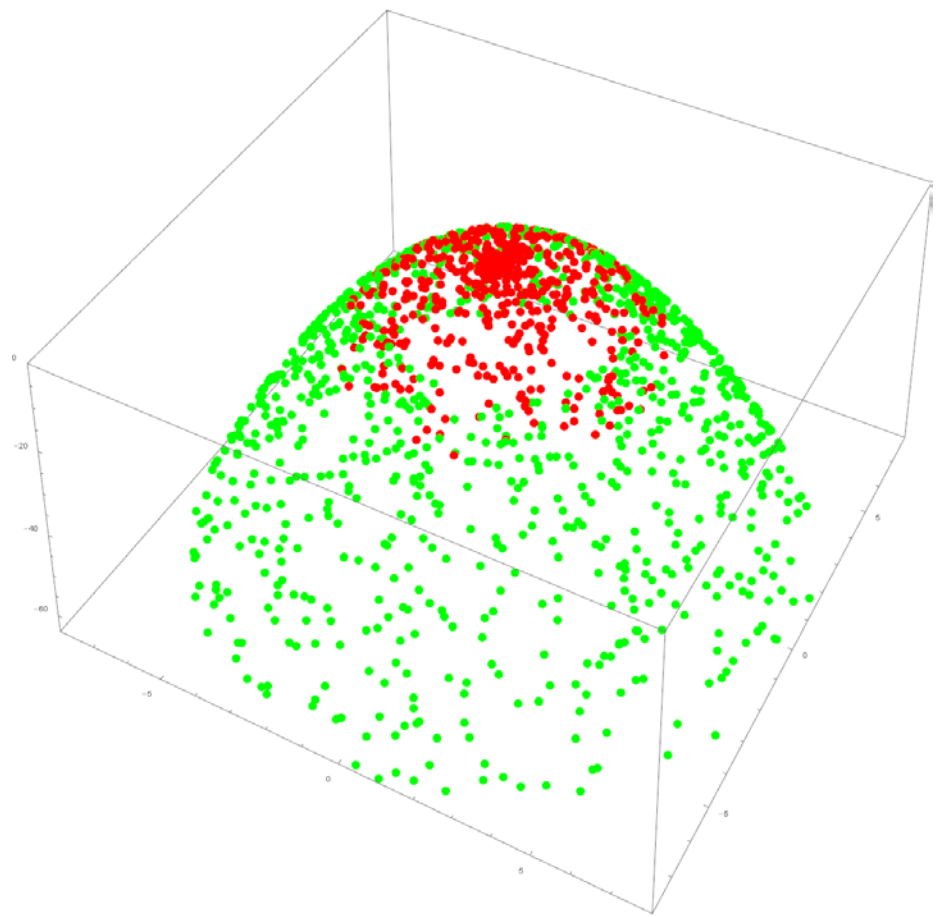
$$(a_1, a_2) \rightarrow (a_1, a_2, a_1^2 + a_2^2)$$

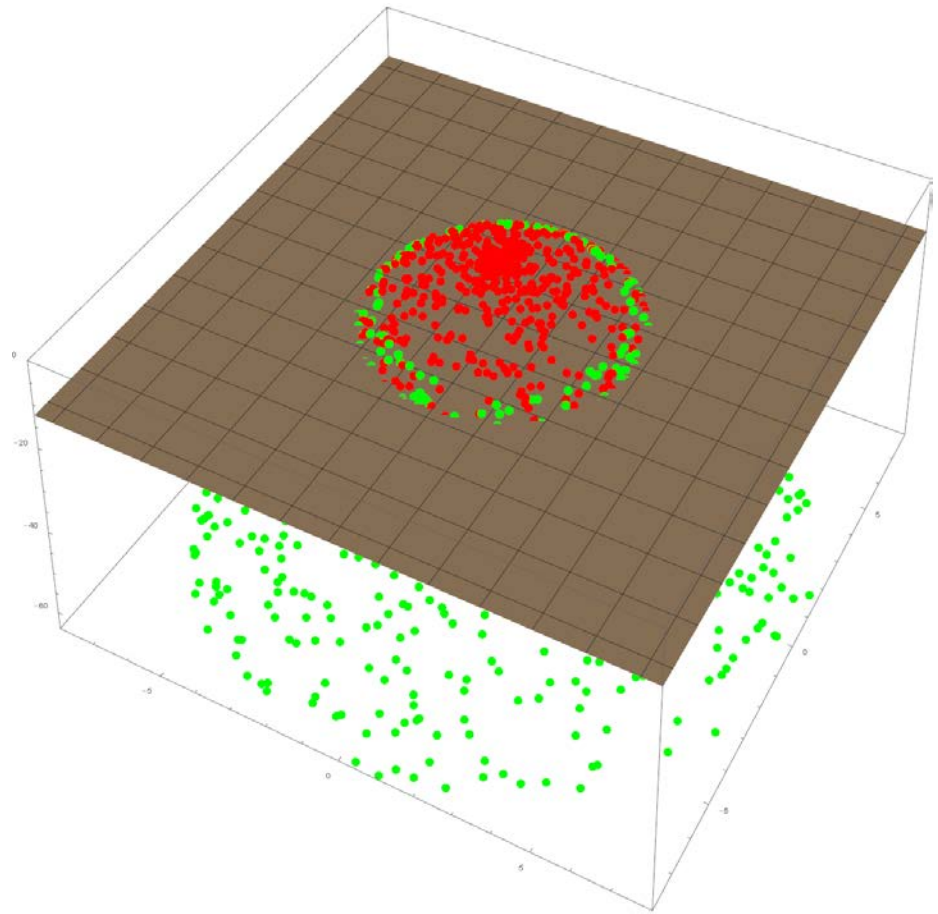


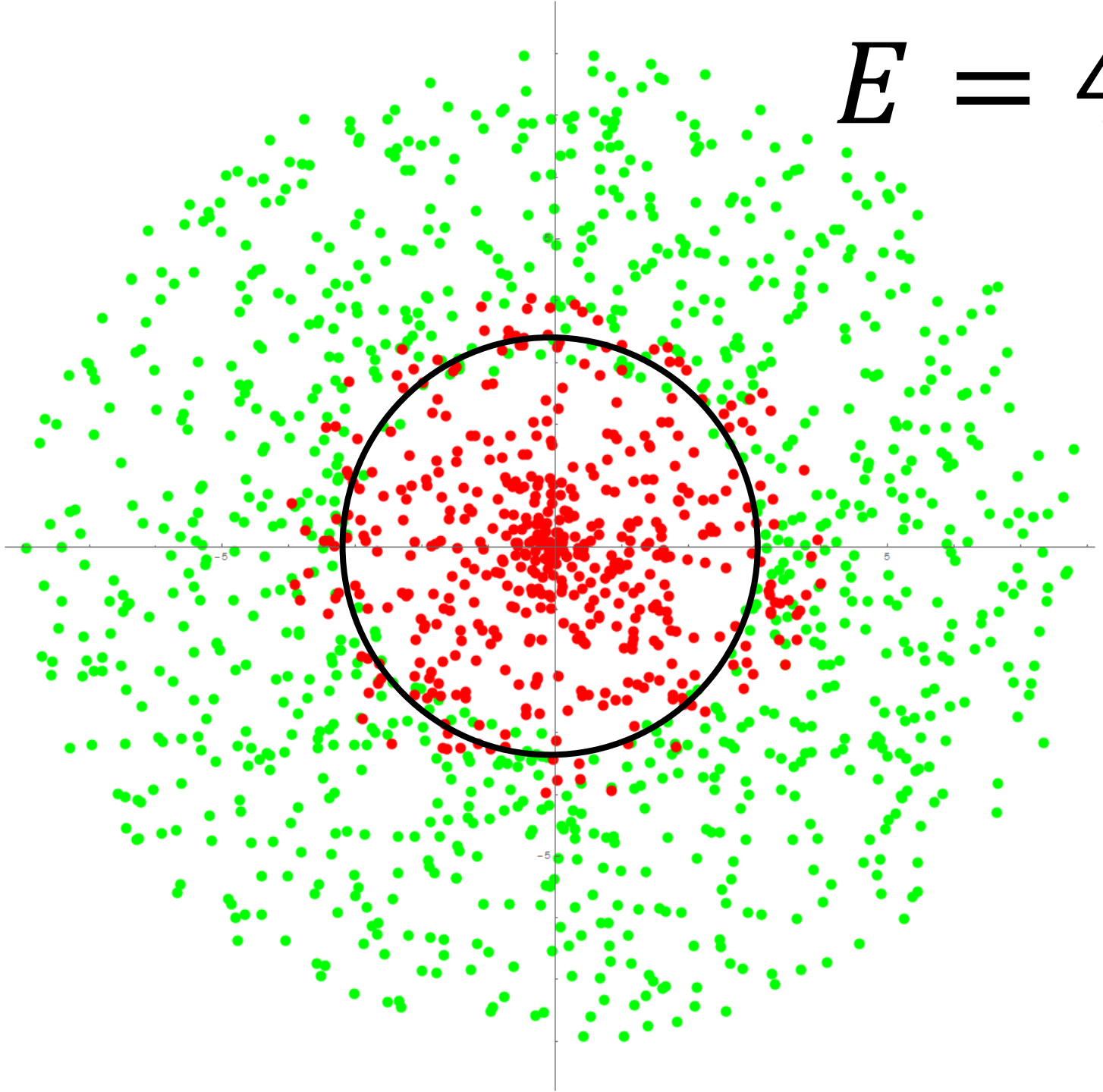



0
-20
-40
-60
-5
0
5
5
0
-5


0
-20
-40
-60
-5
0
5
5
0
-5


0
-20
-40
-60
-5
0
5
5
0
-5


0
-20
-40
-60
-5
0
5
5
0
-5


$E = 4\%$

Noyau quadratique

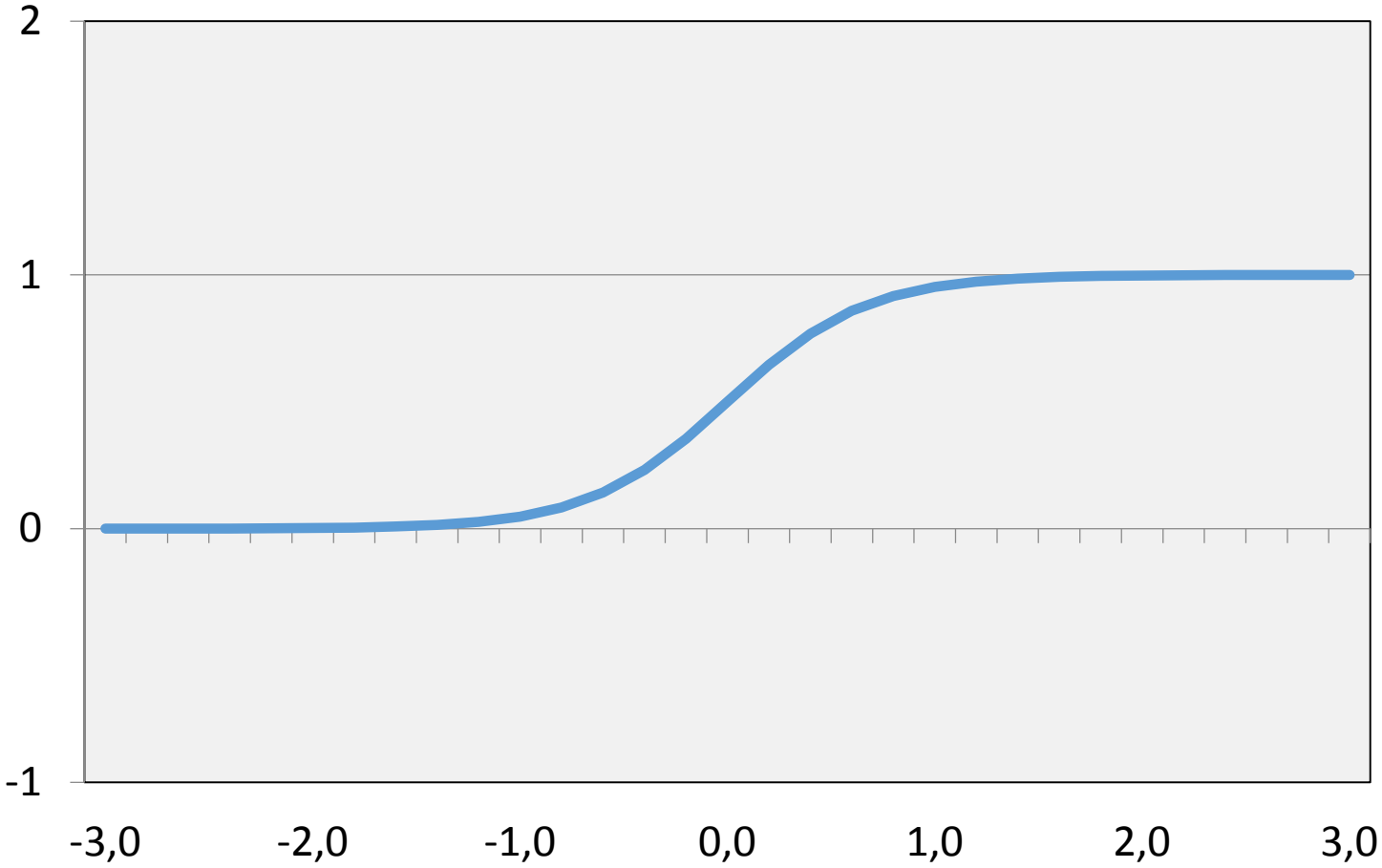
$$(a_1, a_2) \rightarrow (a_1, a_2, a_1 a_2, a_1^2, a_2^2)$$

Hyperplan avec noyau quadratique

$$E = 1.9\%$$

Frank Rosenblatt 1958
Perceptron

Sigmoïde

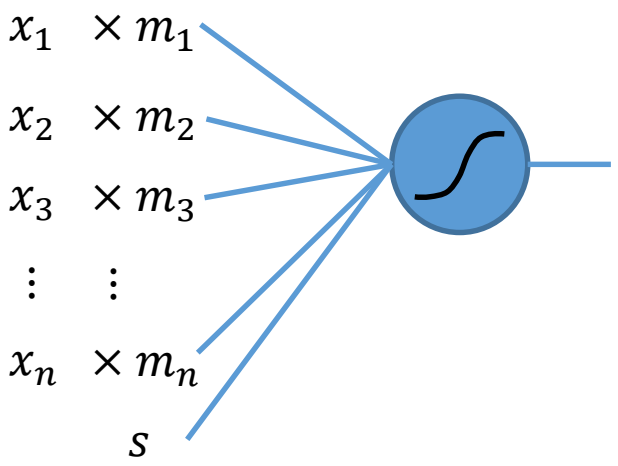


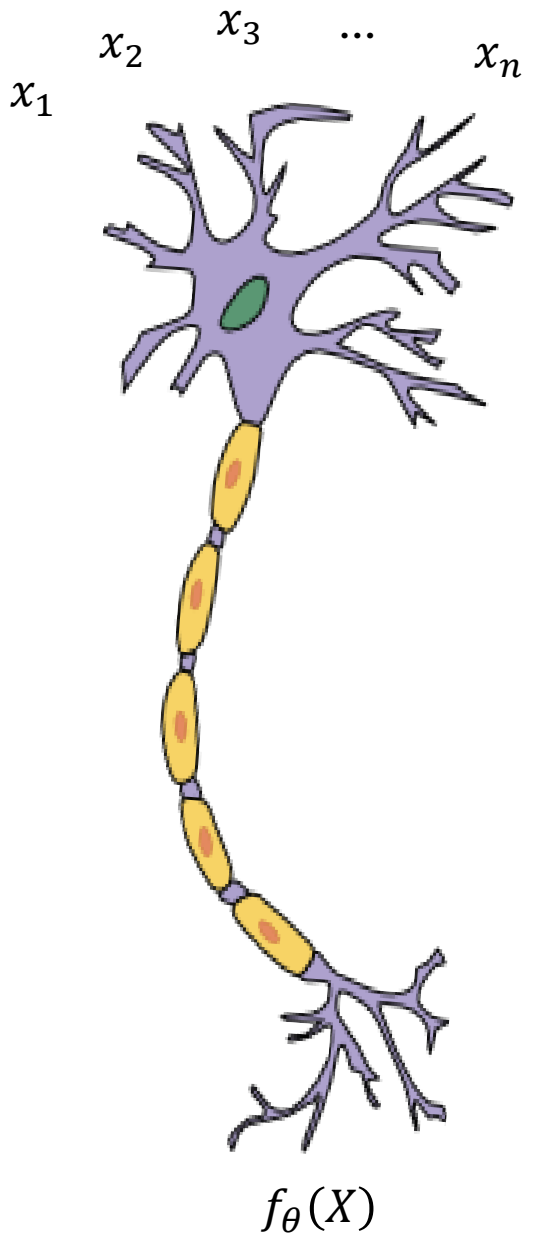
$$\varphi(x) = \frac{1}{1 + e^{-3x}}$$

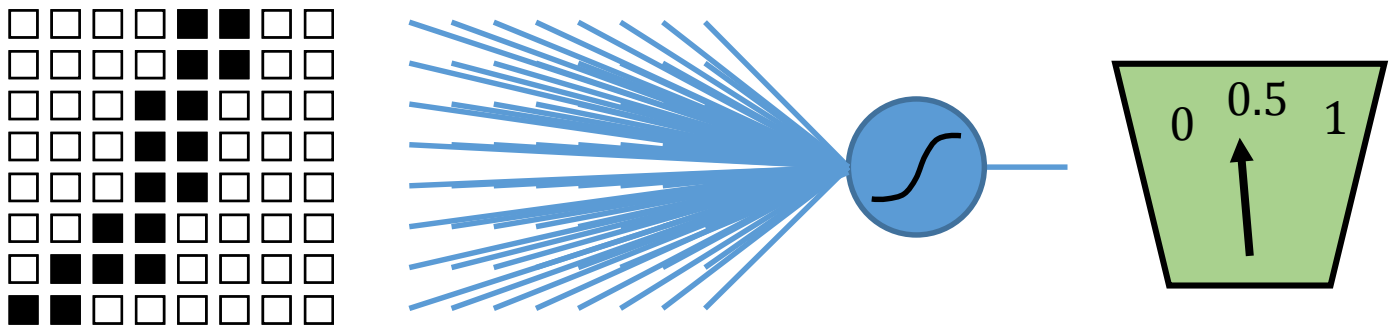
Neurones artificiels

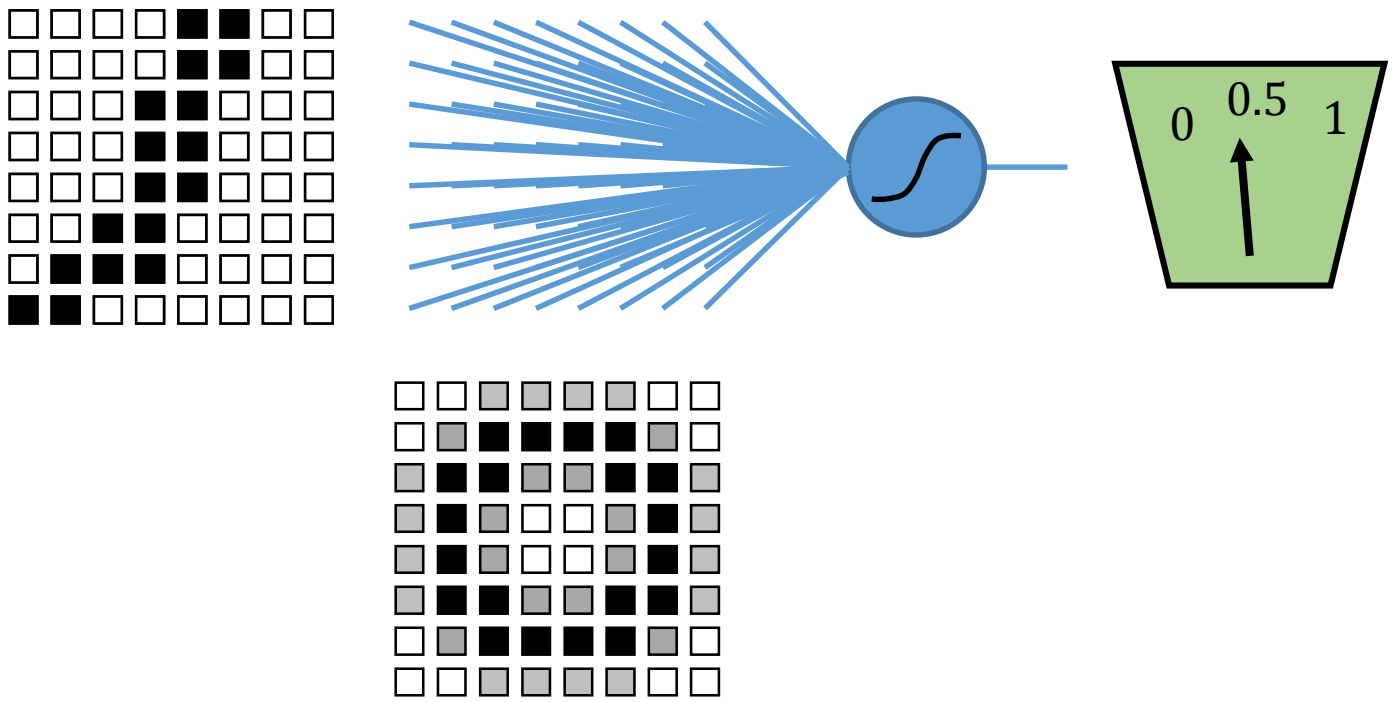
$$\theta = (M, s)$$

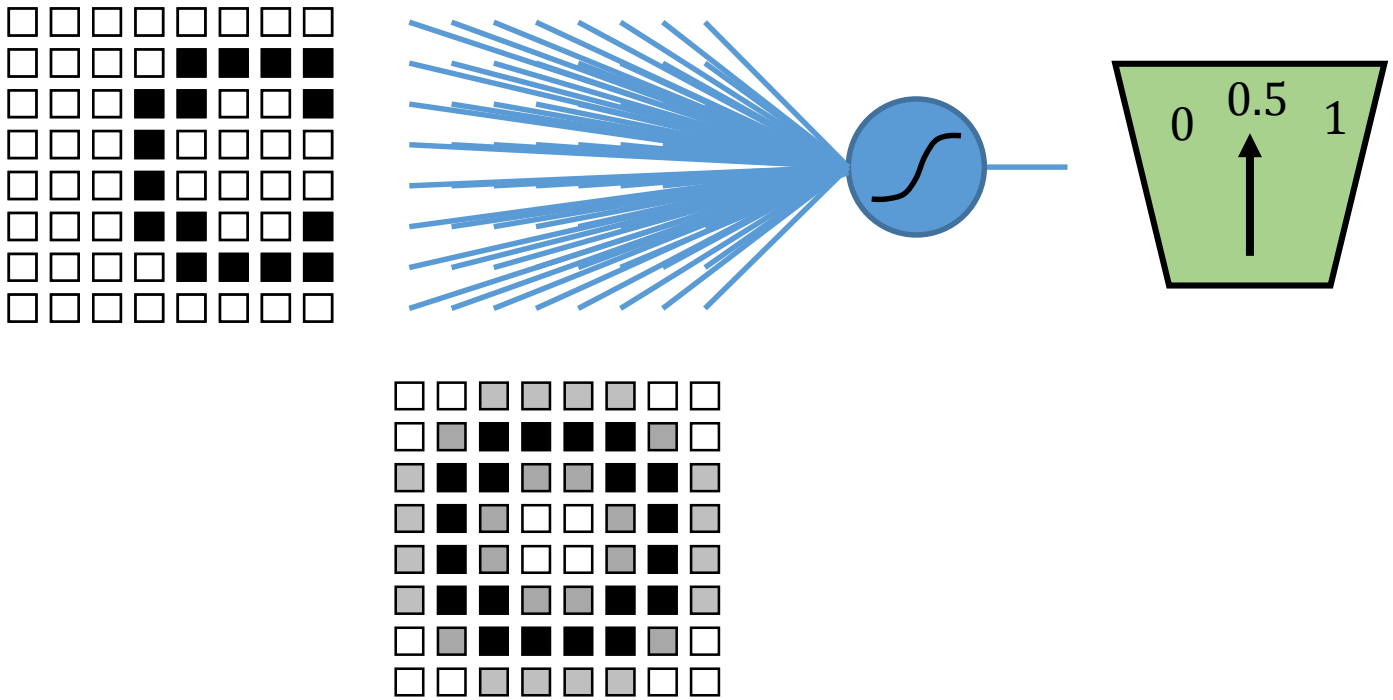
$$f_\theta(X) = \frac{1}{1 + e^{\langle M, X \rangle + s}}$$

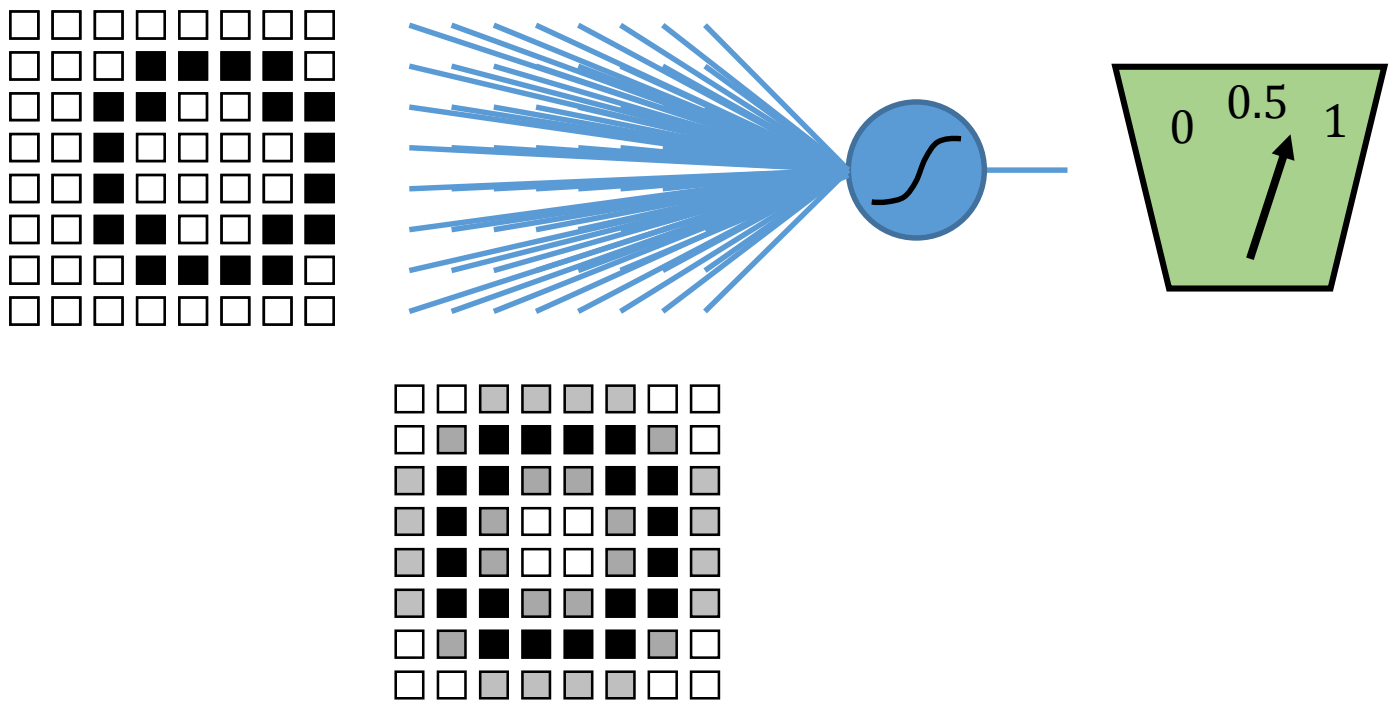


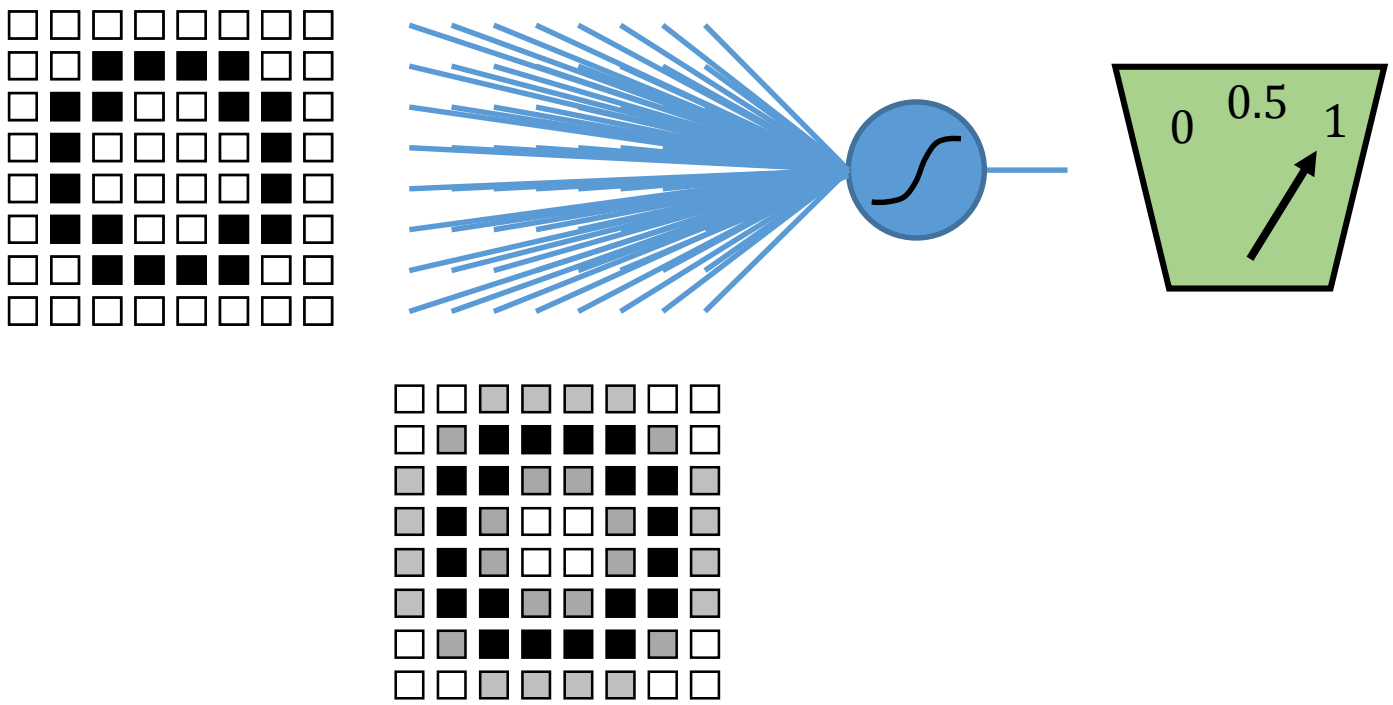



0
0.5
1

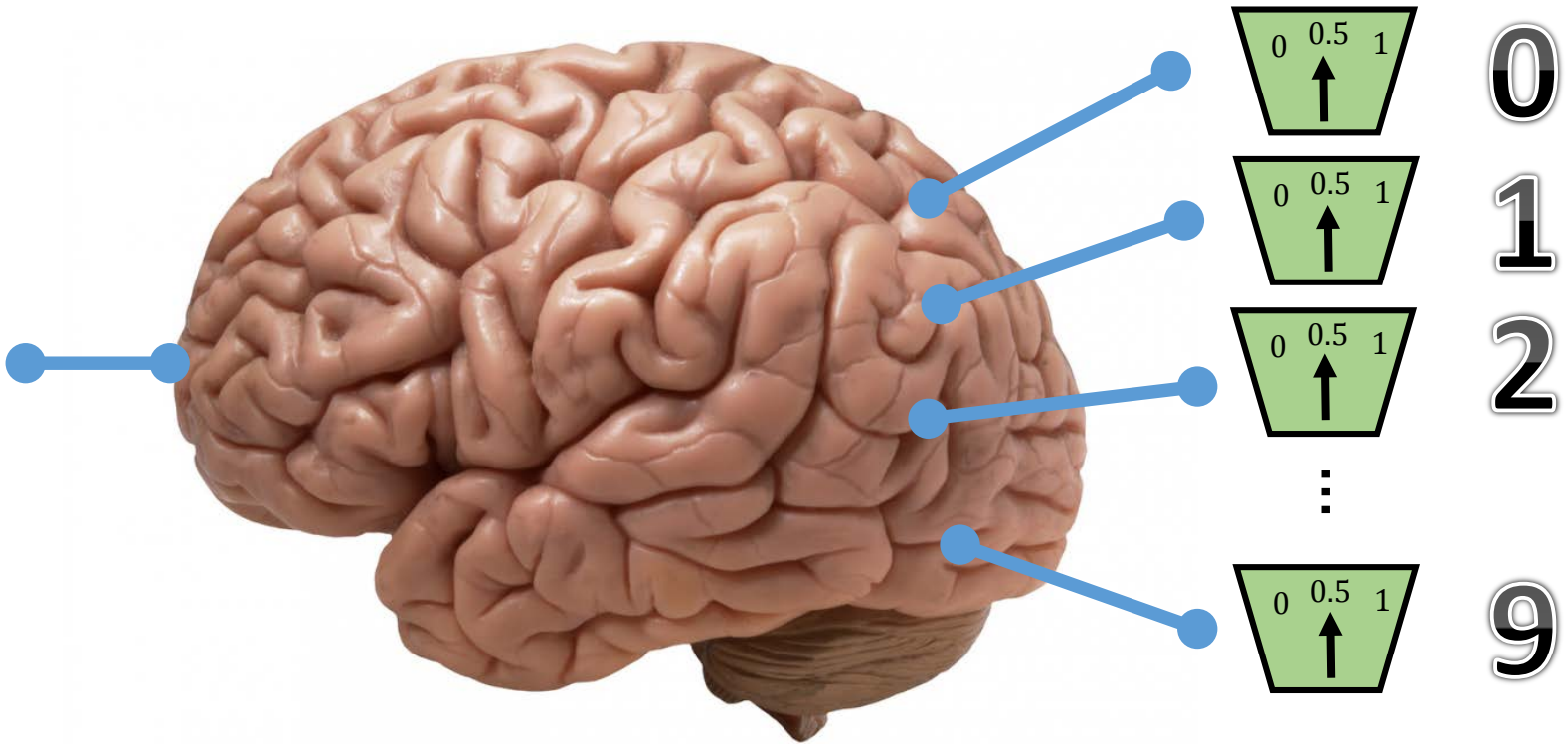

0
0.5
1


0
0.5
1

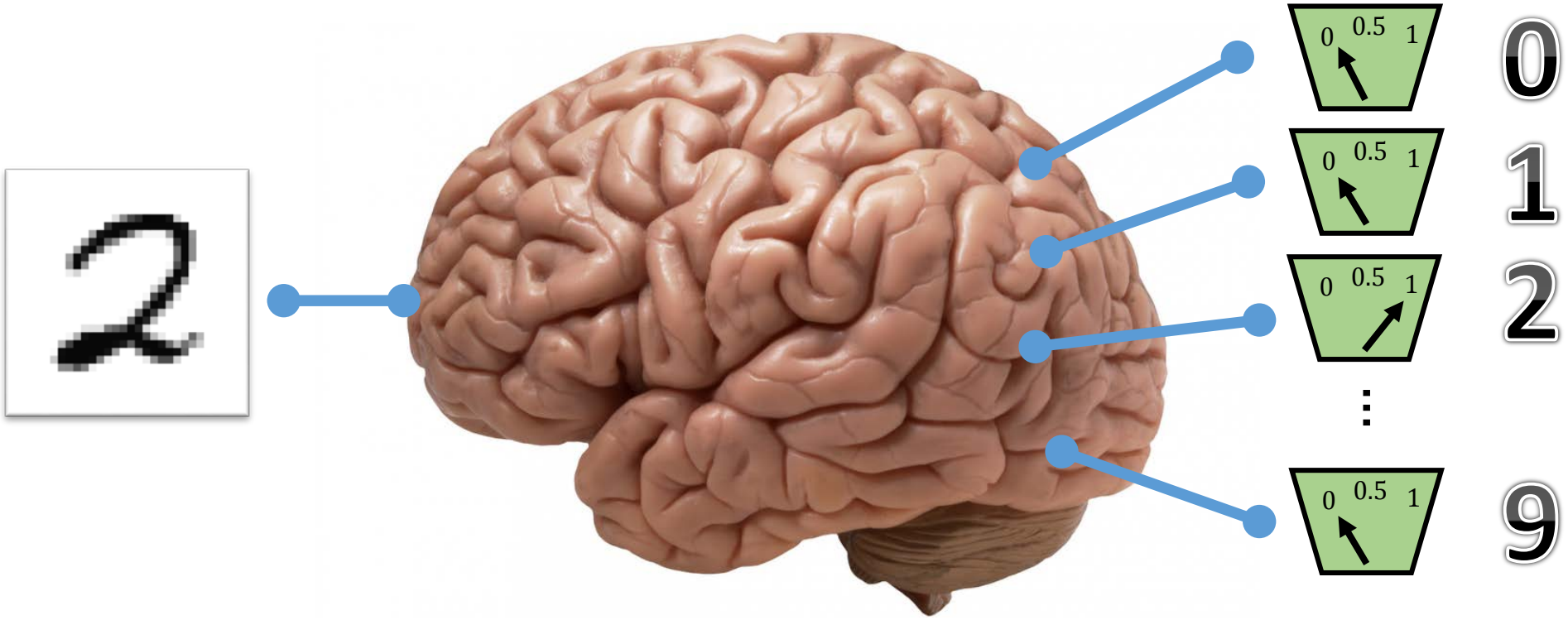

0
0.5
1


0
0.5
1

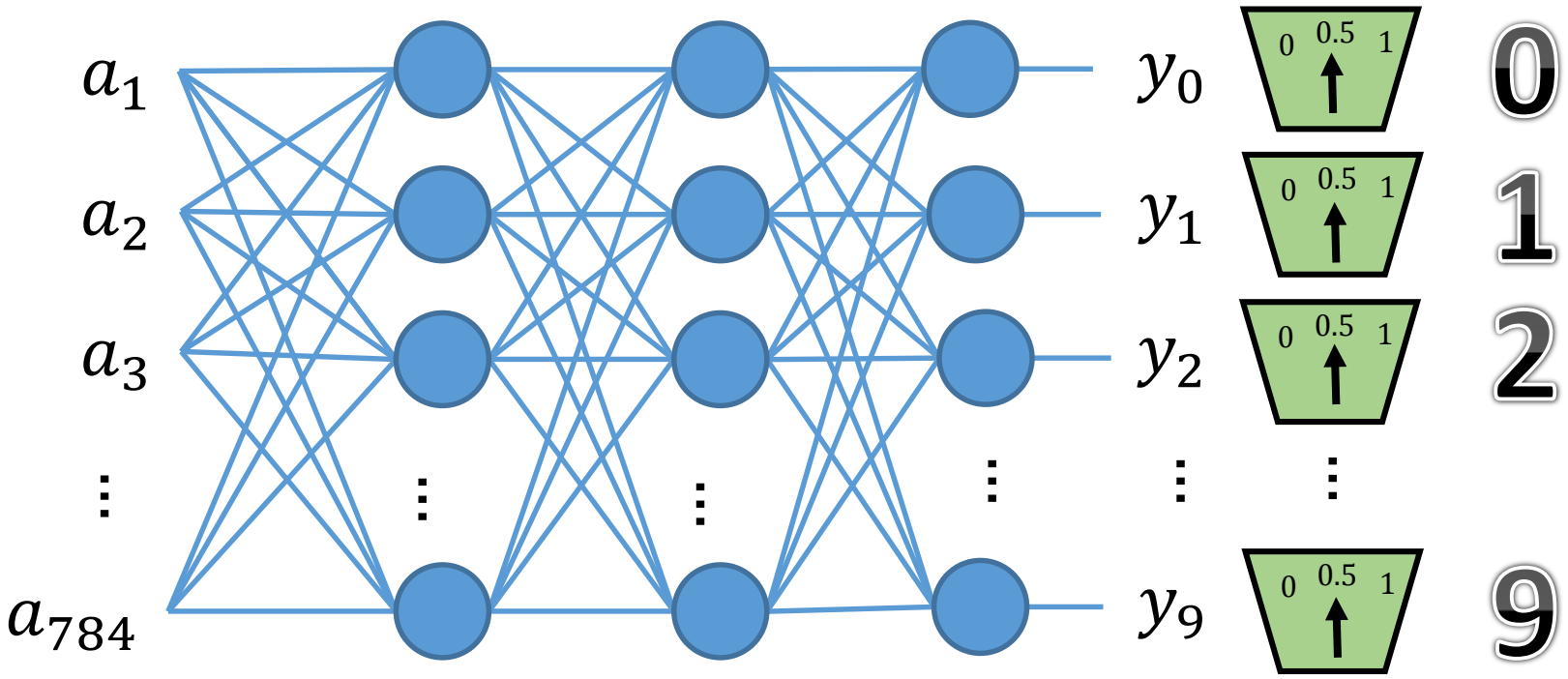
Réseau neuronal artificiel



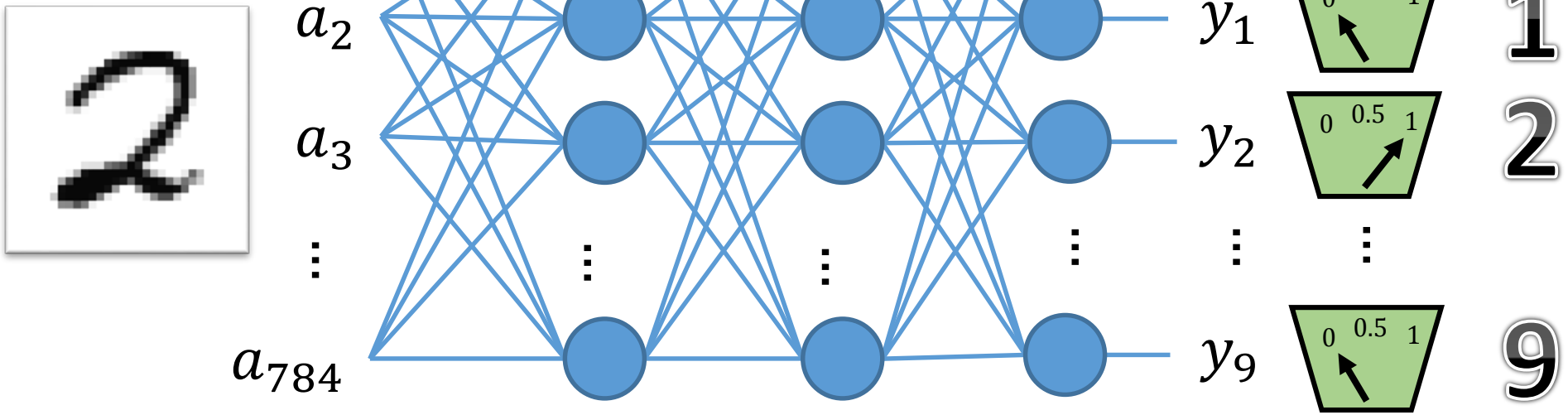
Réseau neuronal artificiel



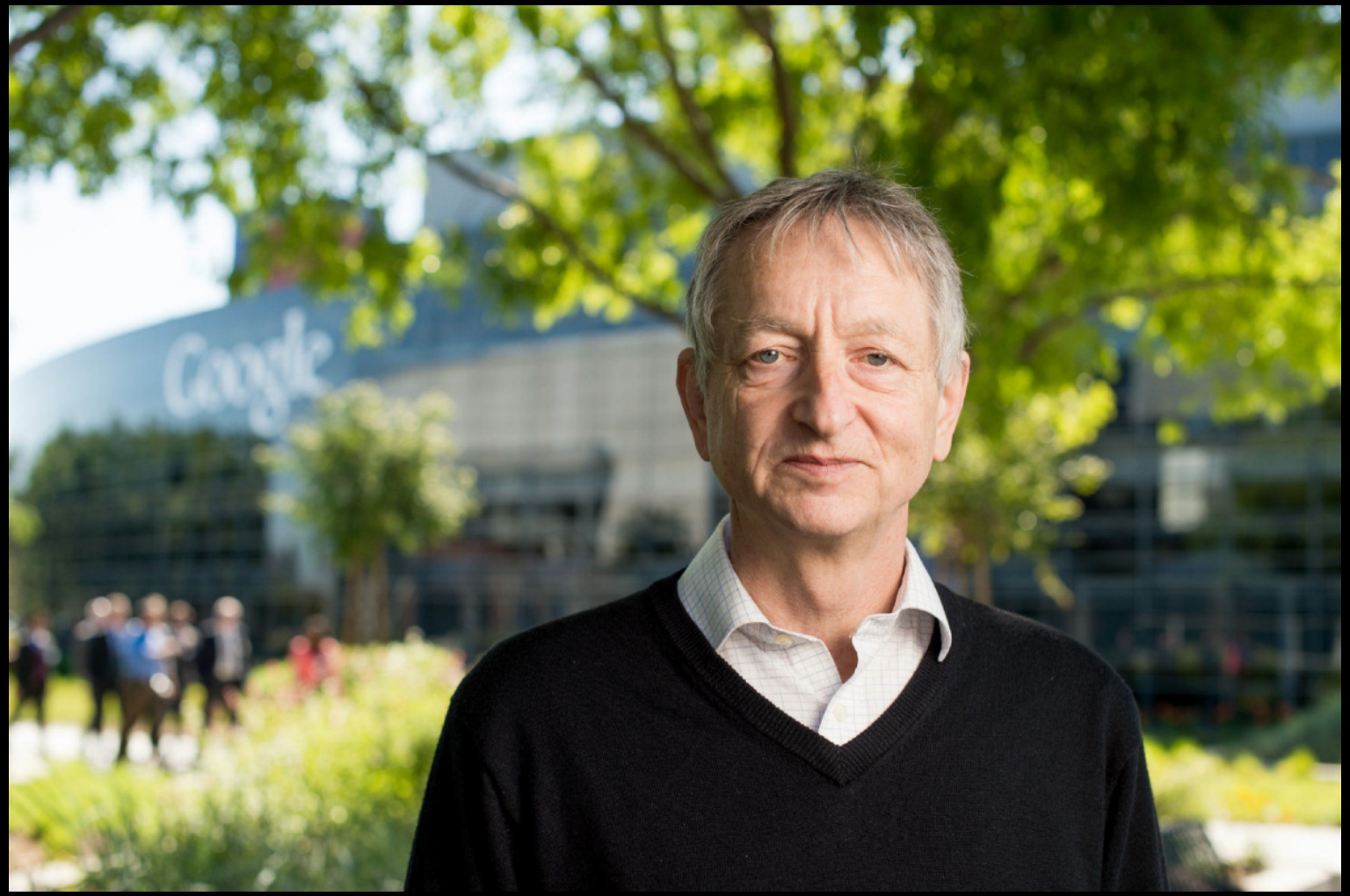
Réseau neuronal artificiel



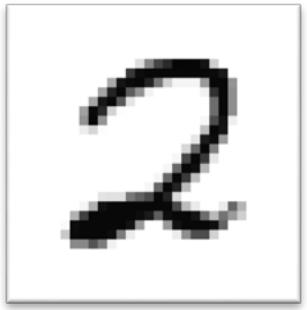
Réseau neuronal artificiel

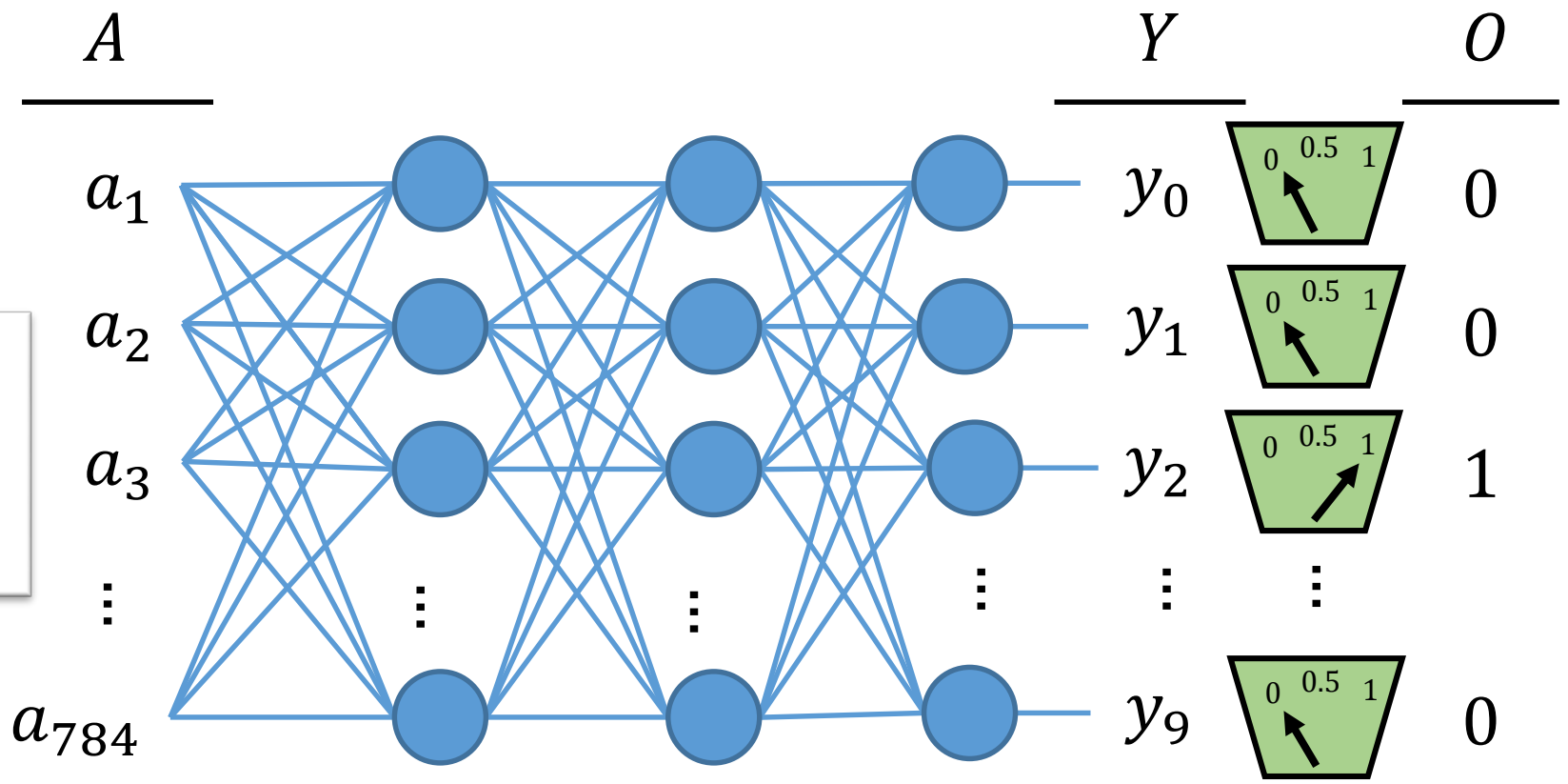


Geoffrey Hinton
Rétropropagation du gradient



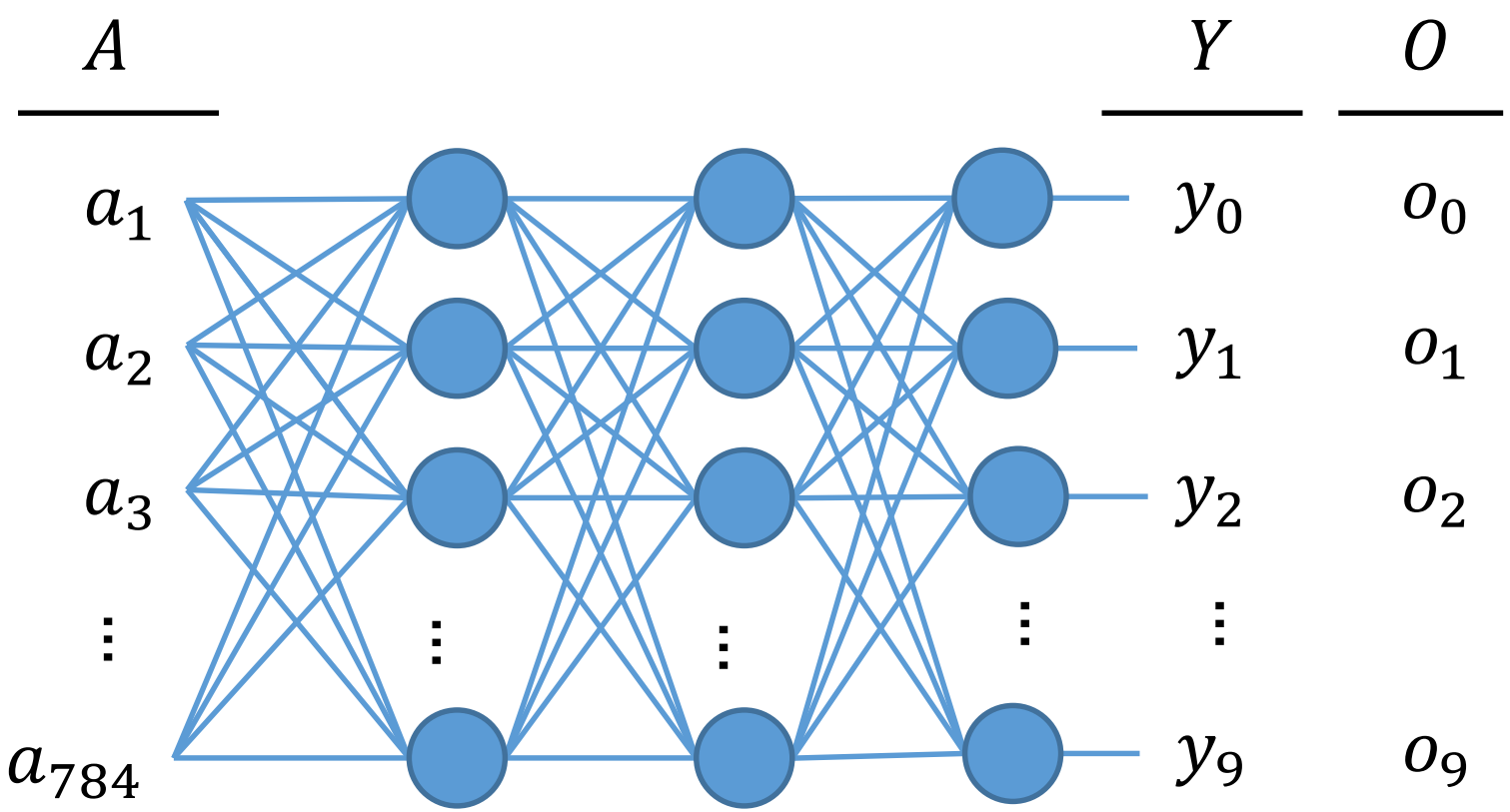
Réseau neuronal artificiel



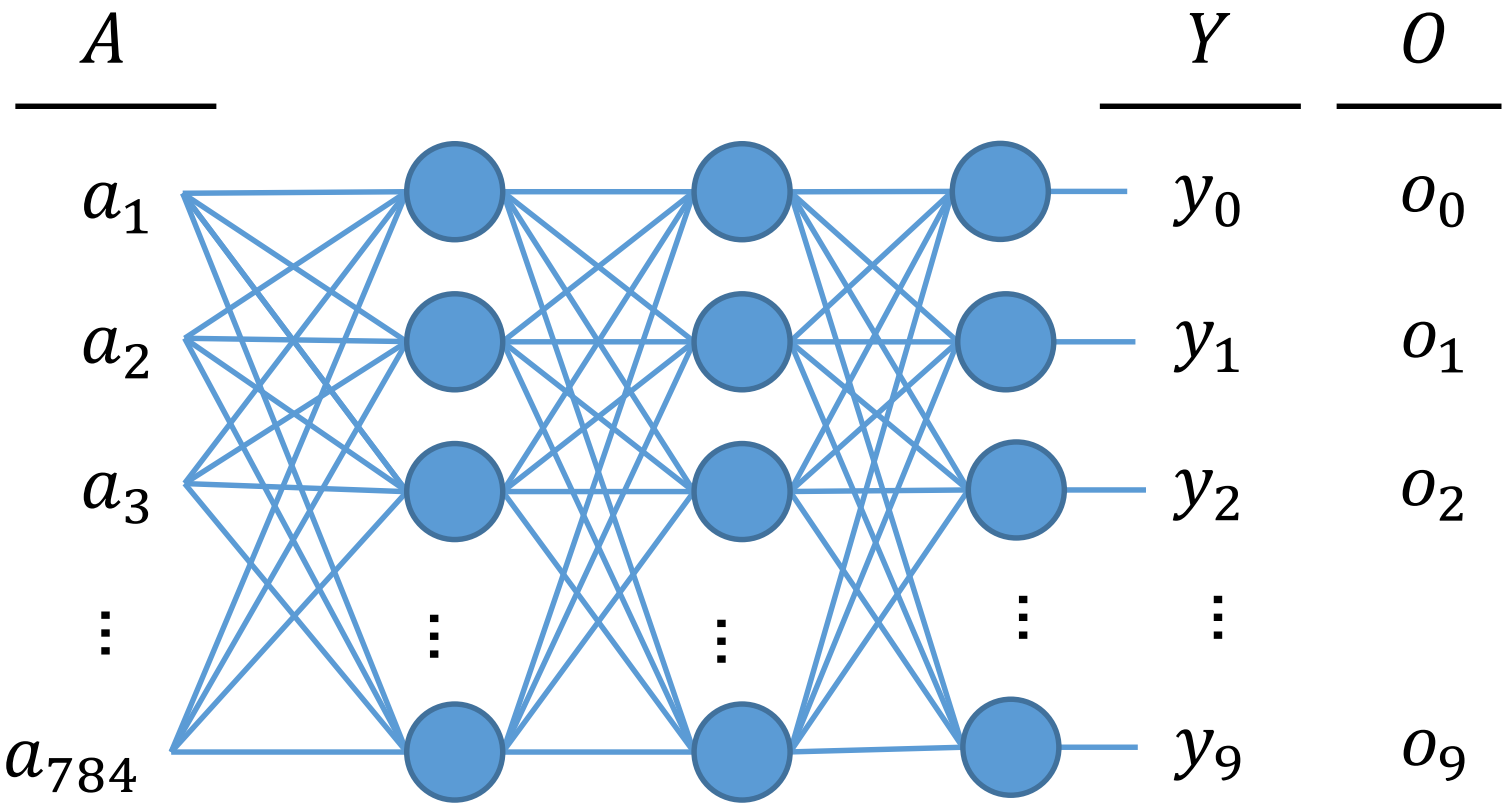


$$\boldsymbol{E}_A = D(Y, O)$$

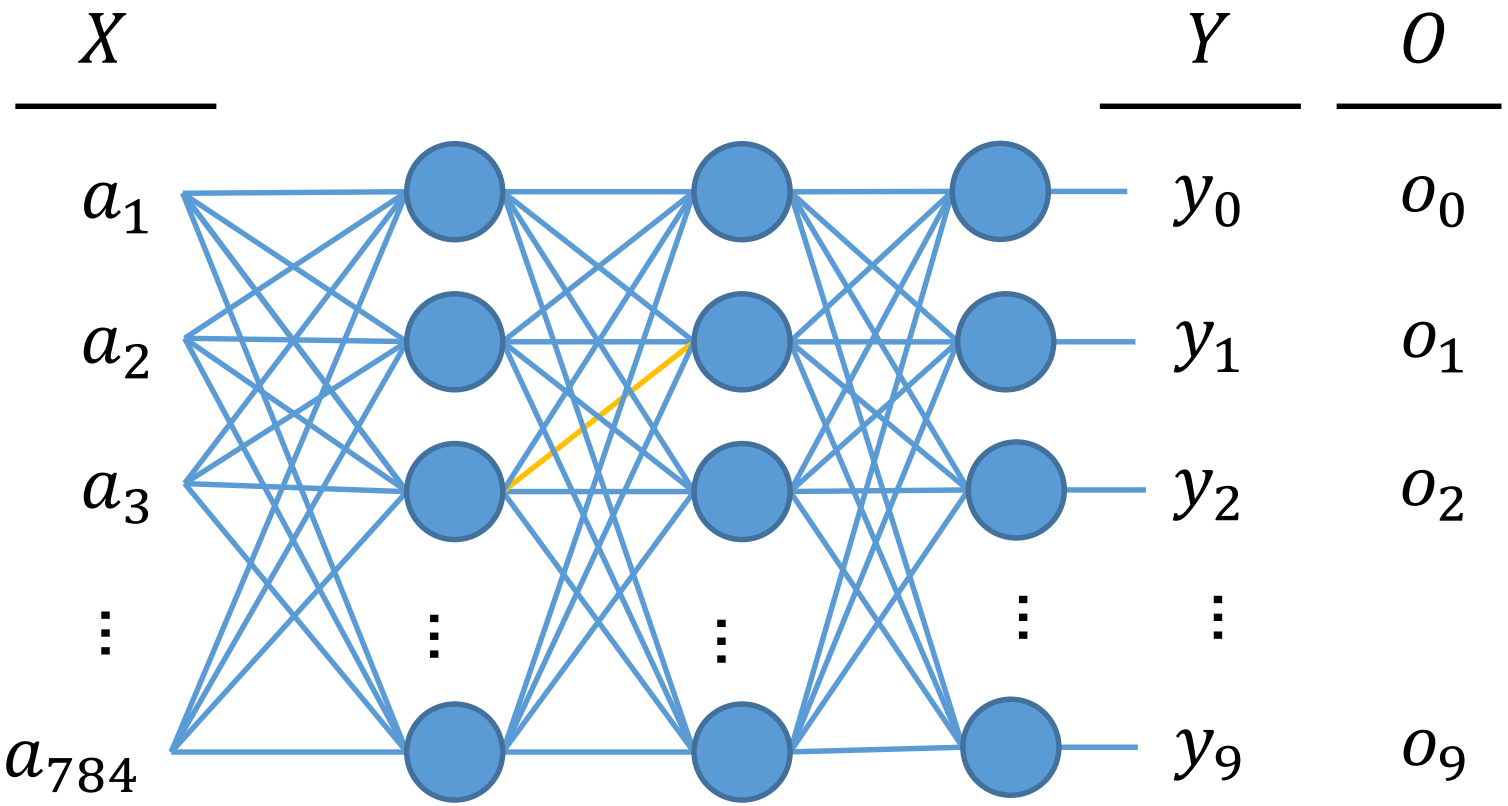
Réseau neuronal artificiel

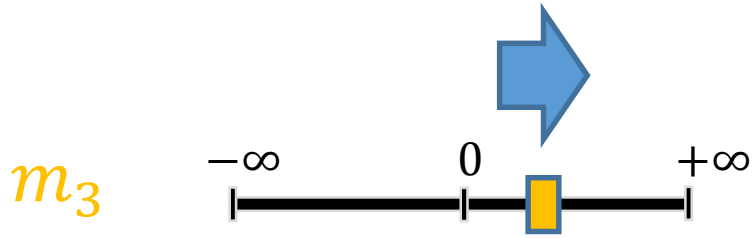


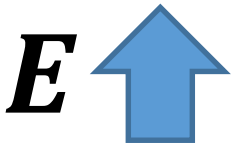
$$\boldsymbol{E} = \boldsymbol{E}_{A_1} + \boldsymbol{E}_{A_2} + \boldsymbol{E}_{A_3} + \cdots + \boldsymbol{E}_{A_{60000}}$$

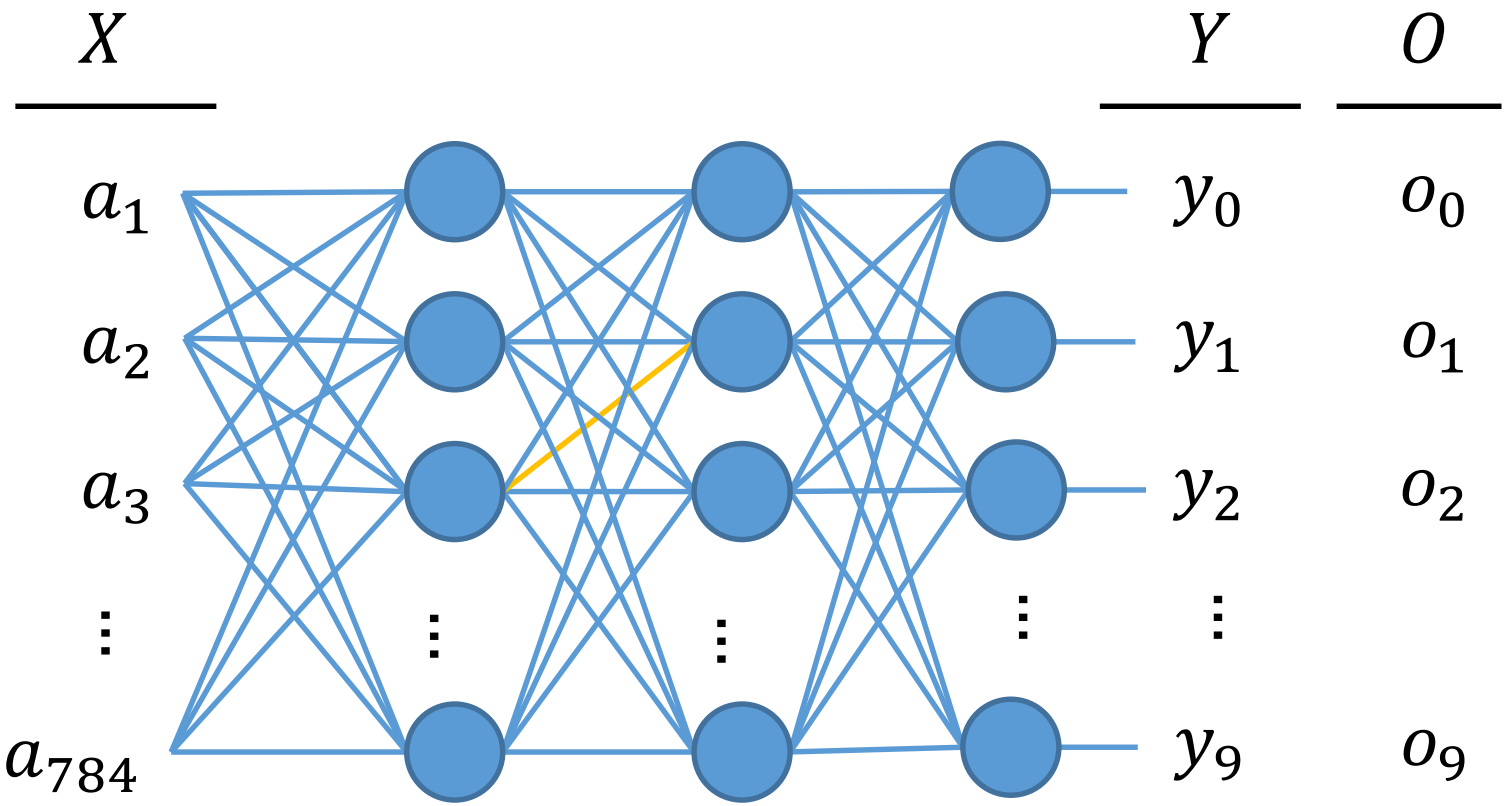


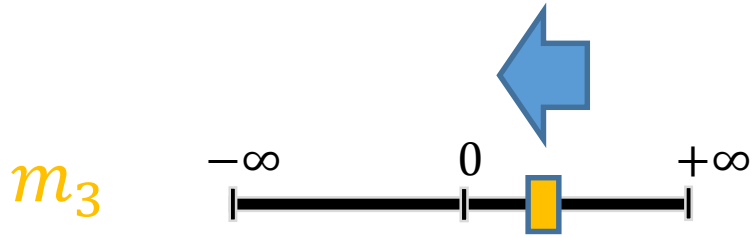
$\boldsymbol{E}$

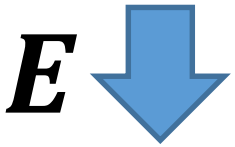

X
Y
O
a_1
a_2
a_3
a_{784}
y_0
y_1
y_2
y_9
o_0
o_1
o_2
o_9


m_3
$-\infty$
0
$+\infty$


$\boldsymbol{E}$

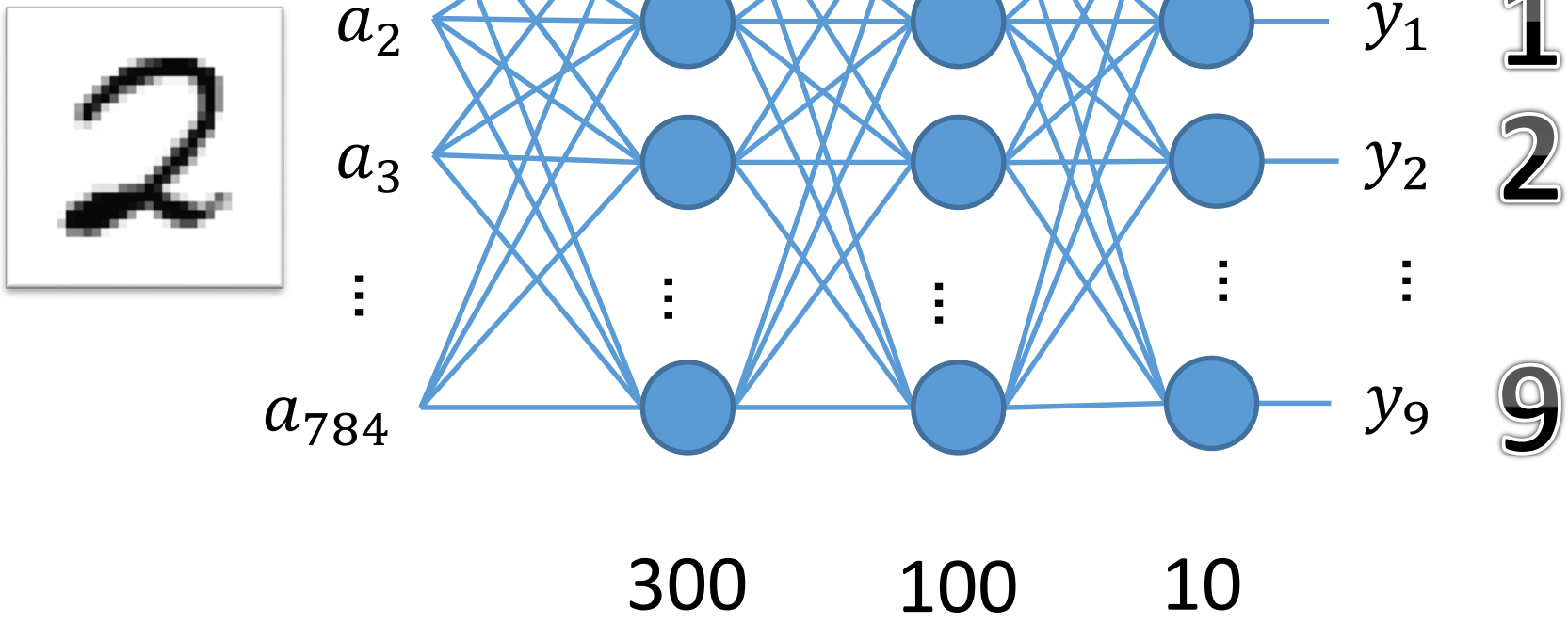

X
Y
O
a_1
a_2
a_3
a_{784}
y_0
y_1
y_2
y_9
o_0
o_1
o_2
o_9


m_3
$-\infty$
0
$+\infty$


$\boldsymbol{E}$

410 neurones

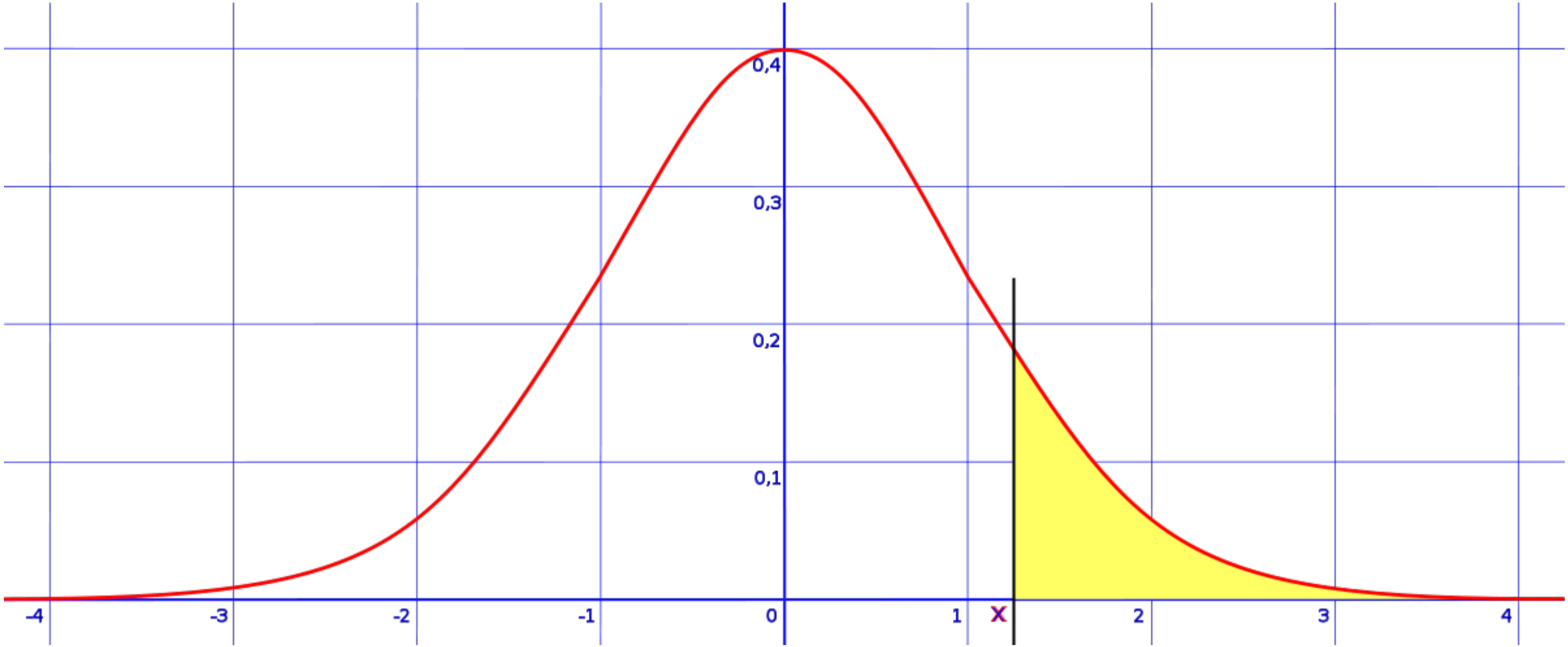
266 200 connexions

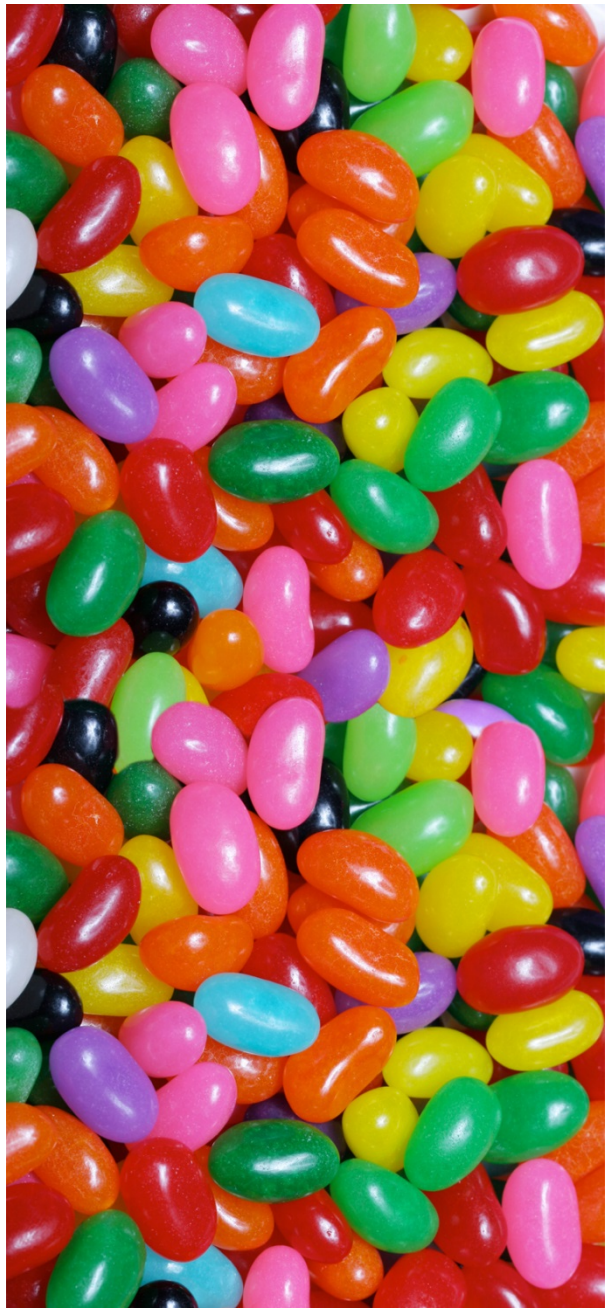


RN à 3 niveaux (300,100)

$$\boldsymbol{E = 3.1\%}$$

Valeur-p





Les bonbons donnent des boutons?

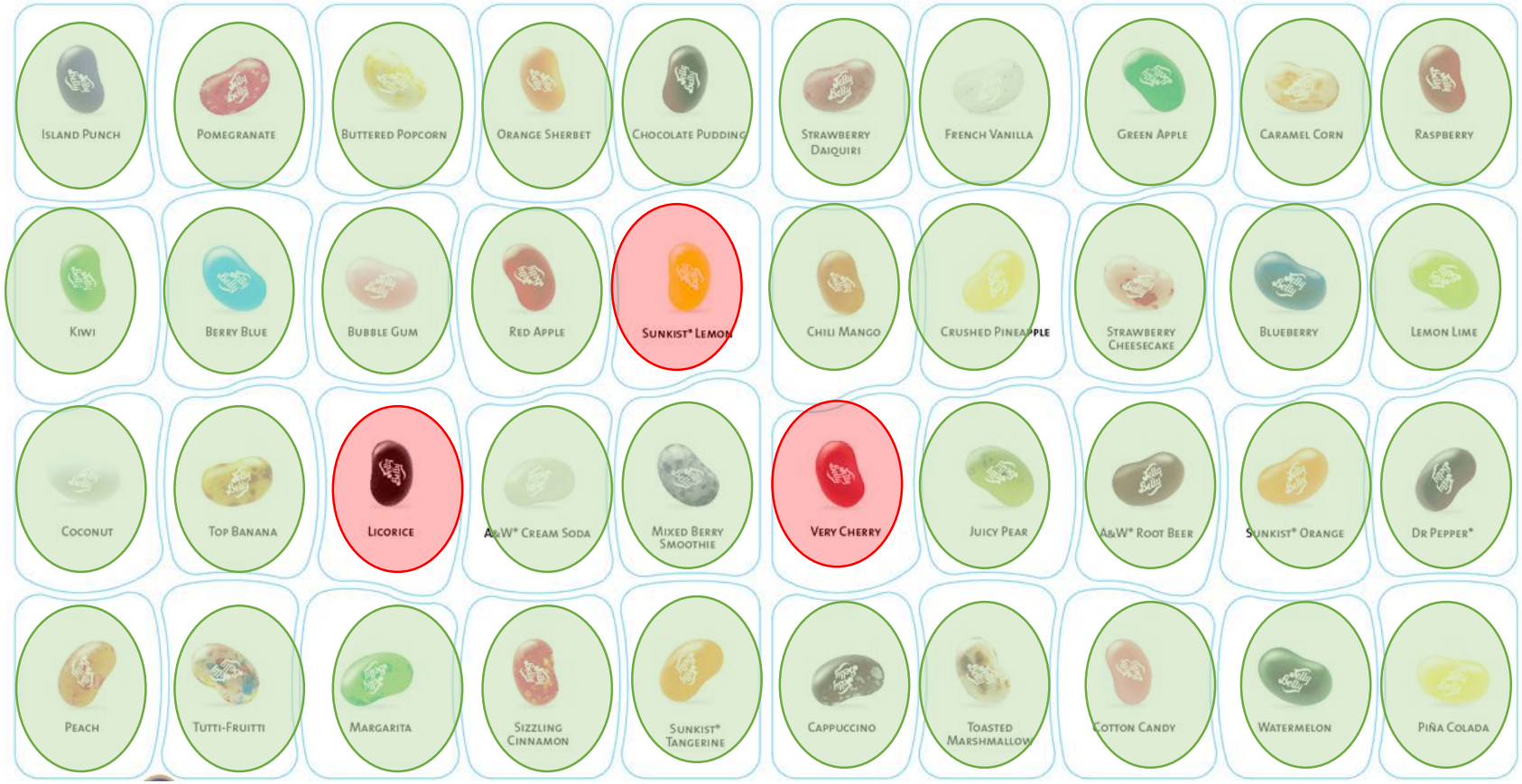
?

Non!

$p > 0.05$

Peut-être une couleur en particulier?

$p > 0.05$ $p < 0.05$



Election 2016

TIME

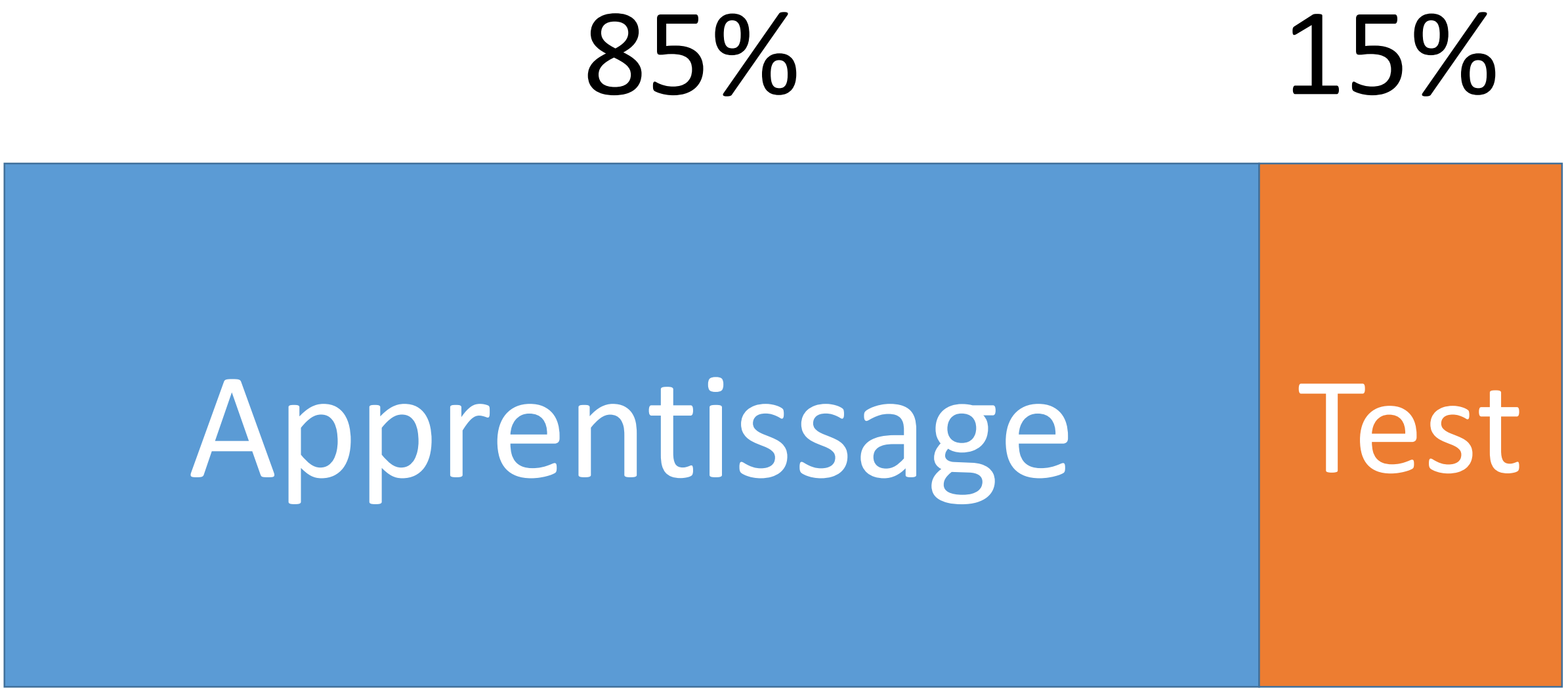
Donald Trump

Will make america great again?

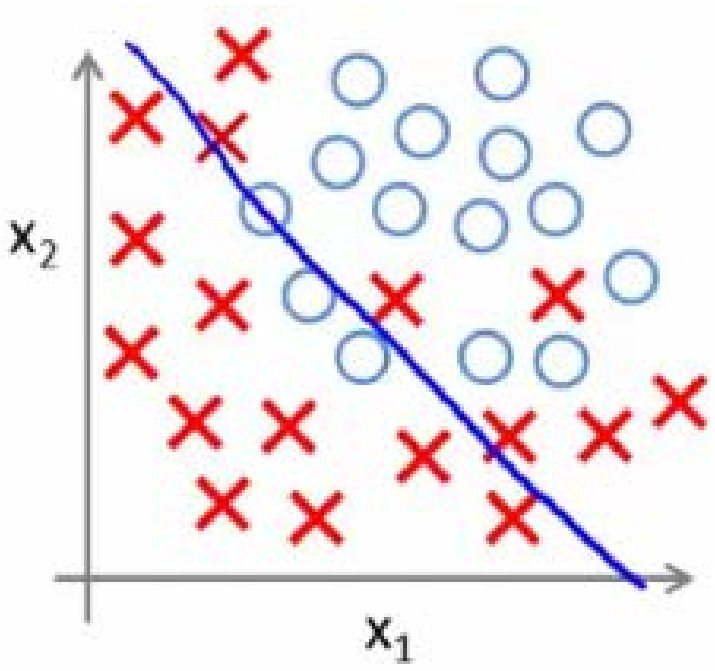
New study shows Jelly Beans gives pimples!

100%

Exemples

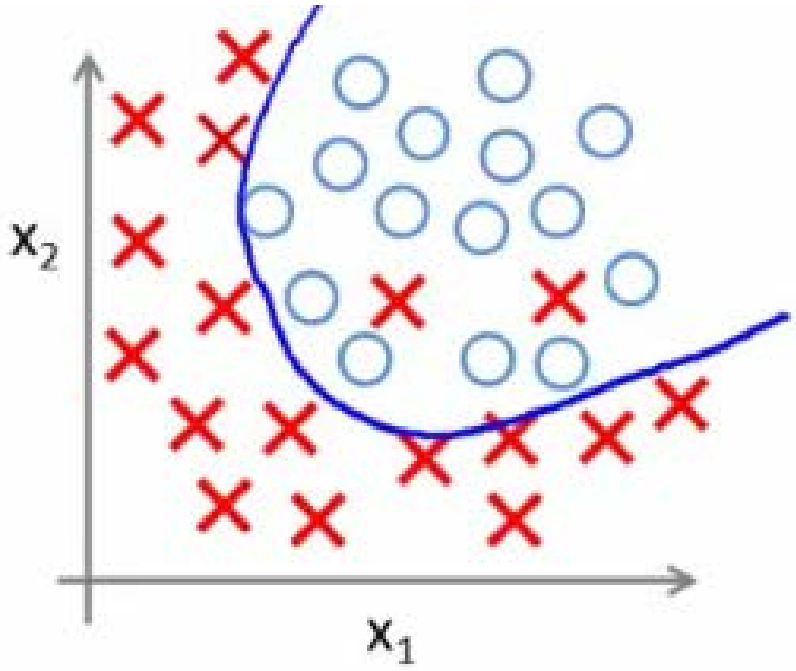

85%
15%
Apprentissage
Test

Sur-apprentisage?

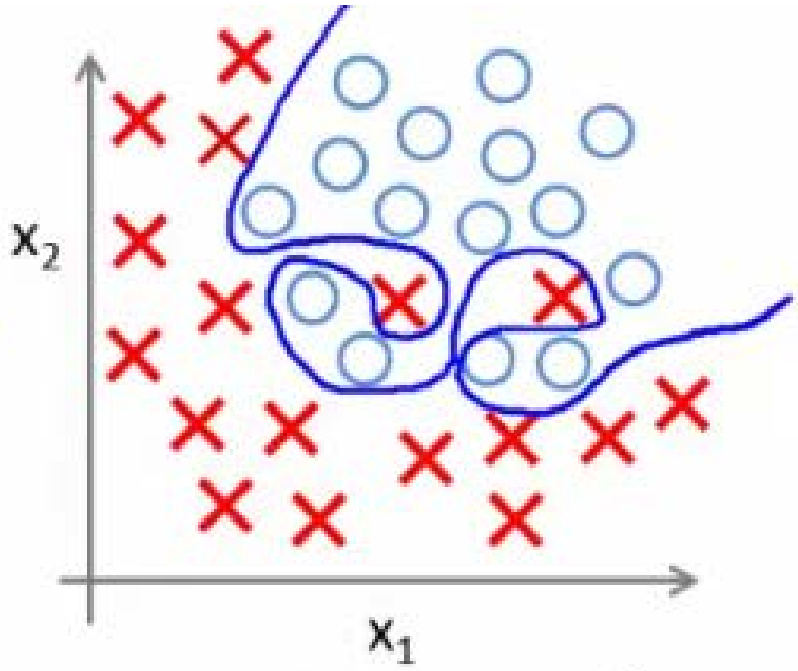


$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$

(g = sigmoid function)



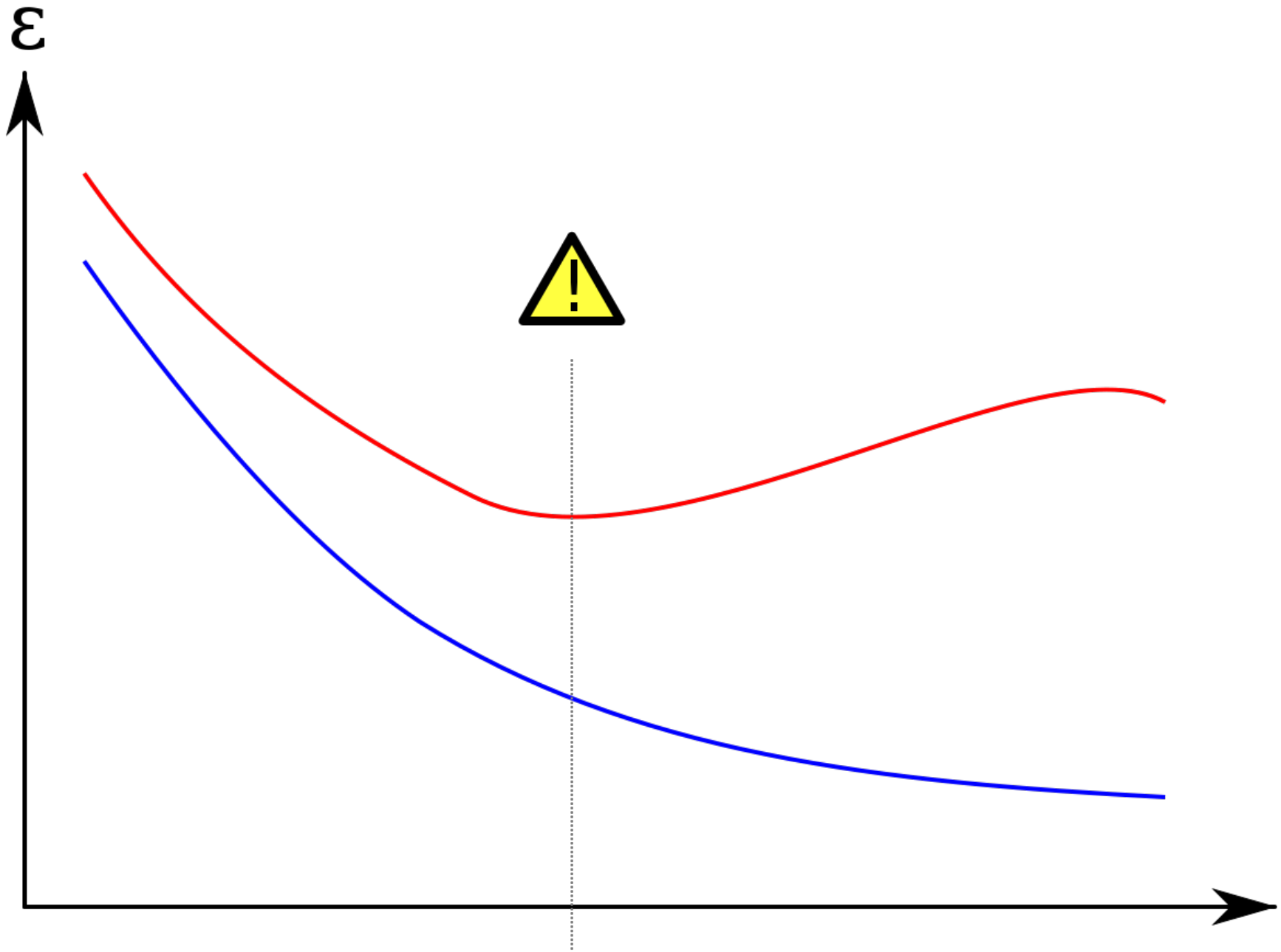
$g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2 + \theta_5 x_1 x_2)$

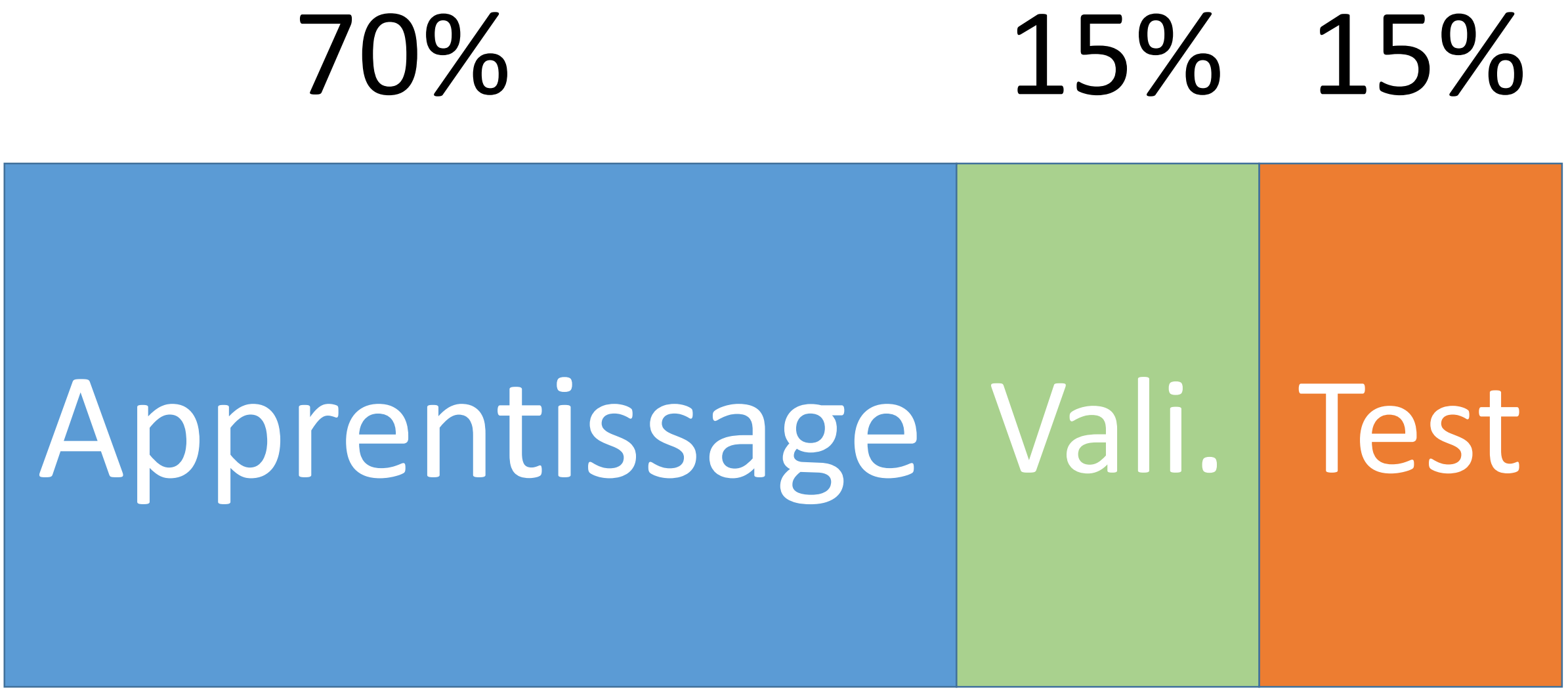


$g(\theta_0 + \theta_1 x_1 + \theta_2 x_1^2 + \theta_3 x_1^2 x_2 + \theta_4 x_1^2 x_2^2 + \theta_5 x_1^2 x_2^3 + \theta_6 x_1^3 x_2 + \ldots$

Paramètres et hyper paramètres

- Capacité
  - Nombre de paramètres
  - Degré
  - Complexité
- Temps d’entrainement
- Tolérance aux erreurs
- Régularisation


ε

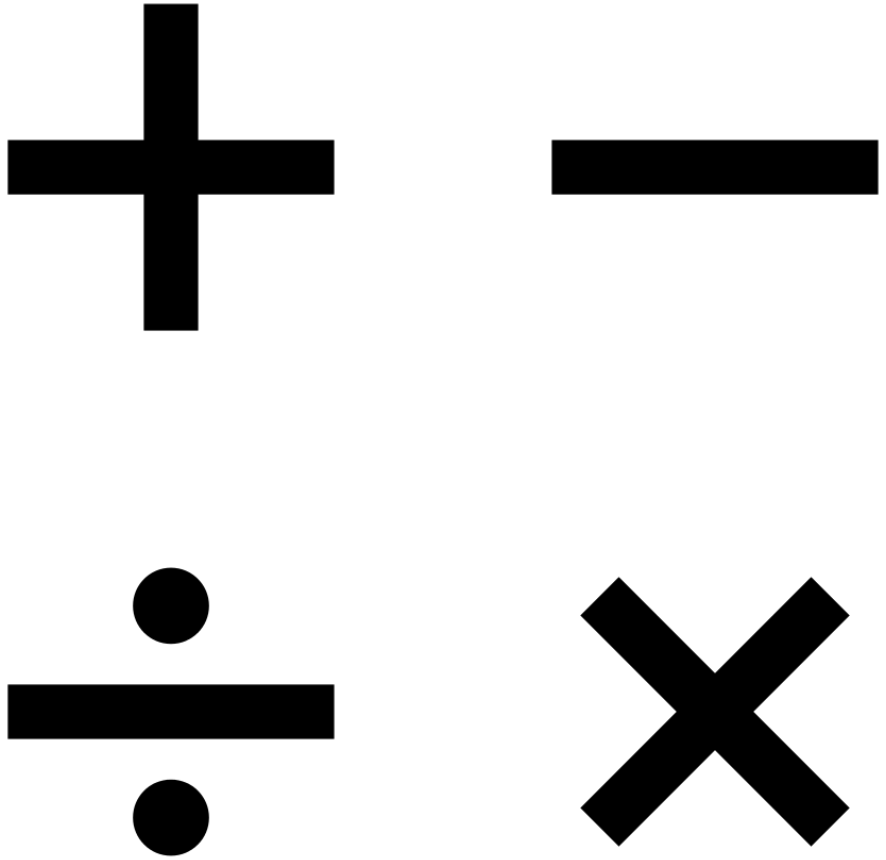

70%
15%
15%
Apprentissage
Vali.
Test

Puissance, Capacité, Force brute…

FLOPS

FLoating-point Operations Per Second

Opérations en virgule flottante par seconde

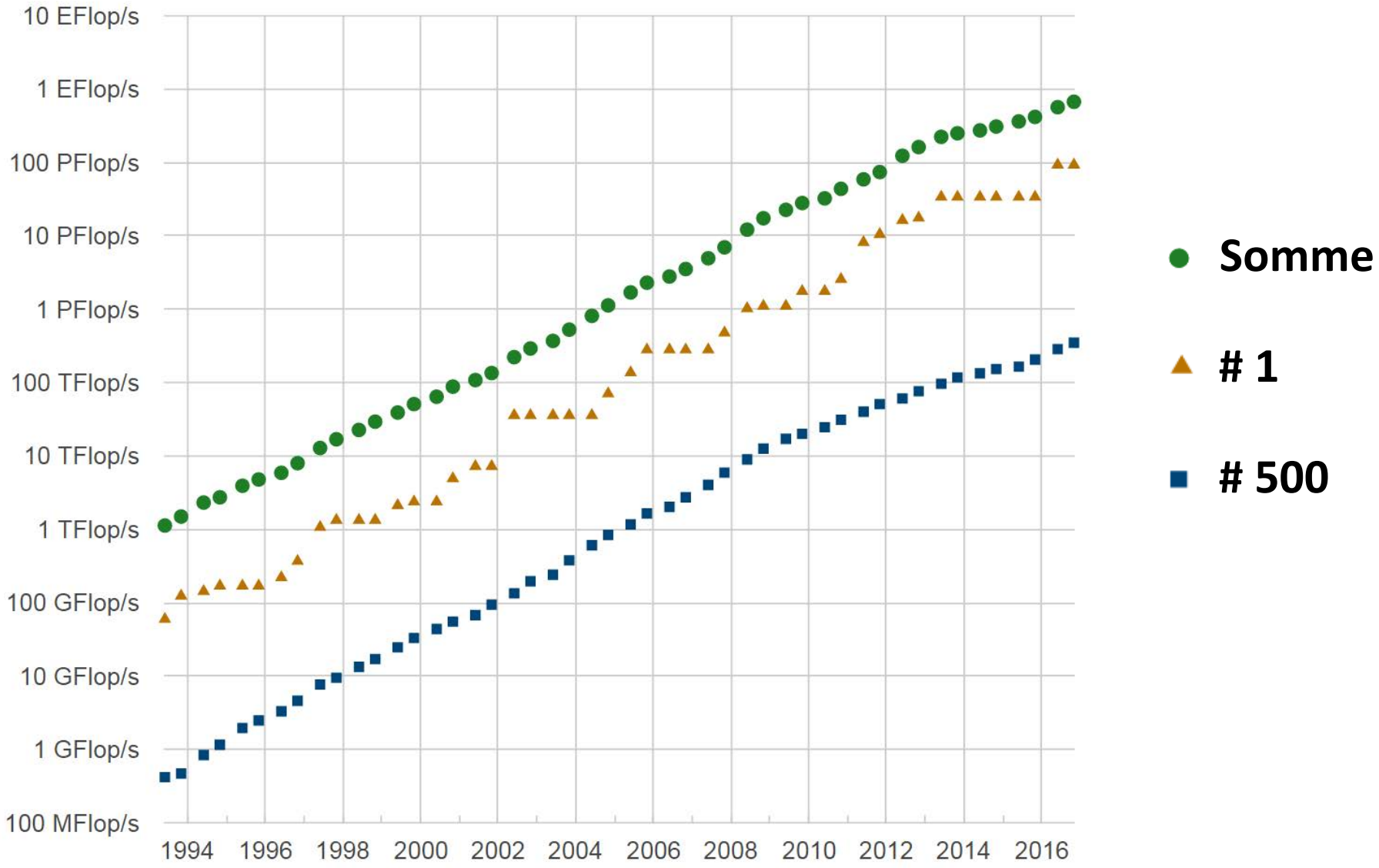


Intel Core i7
90 *giga*-FLOPS
$400

Nvidia GTX 1080 Ti
10 *tera*-FLOPS
$900

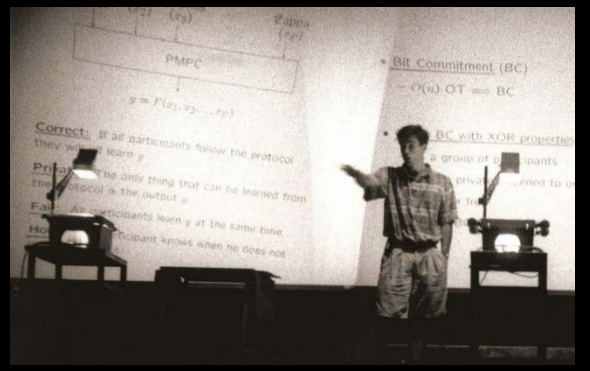
Tianhe-2
90 *peta*-FLOPS
$520 000 000

Top 500!


10 EFlop/s
1 EFlop/s
100 PFlop/s
10 PFlop/s
1 PFlop/s
100 TFlop/s
10 TFlop/s
1 TFlop/s
100 GFlop/s
10 GFlop/s
1 GFlop/s
100 MFlop/s
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
Somme
1
500

2018 ⟷ 1995

⟷



1 seconde ⟷ 4 mois

1 nuit ⟷ 10 millénaires

Le cerveau

1.4 kg
90G neurones
jusqu’à 10 000 connexions
Jusqu’à 100 Hz

86G neurones x 10K connexions x 100 Hz

=

86 *peta*-Connexions/*s*

=

86 PC/*s* x 12 FLOP = 1 *exa*-FLOPS

Cerveau = 1 *exa*-FLOPS