



Bienvenue!

**ÉCOLE D'ÉTÉ FRANCOPHONE
EN APPRENTISSAGE PROFOND**

21-25 août 2017



IVADO

HEC Montréal
Polytechnique Montréal
Université de Montréal



MILA

A futuristic robot with a yellow hard hat and a tablet, symbolizing automation in training. The robot is shown from the chest up, wearing a yellow hard hat with a headlamp. Its body is a mix of yellow and silver, with visible mechanical joints and components. It is holding a large, rectangular tablet or screen in front of it with both hands. The background is a blurred industrial or factory setting with various structures and lights.

Apprentissage automatique

Alain Tapp

Intelligence artificielle

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

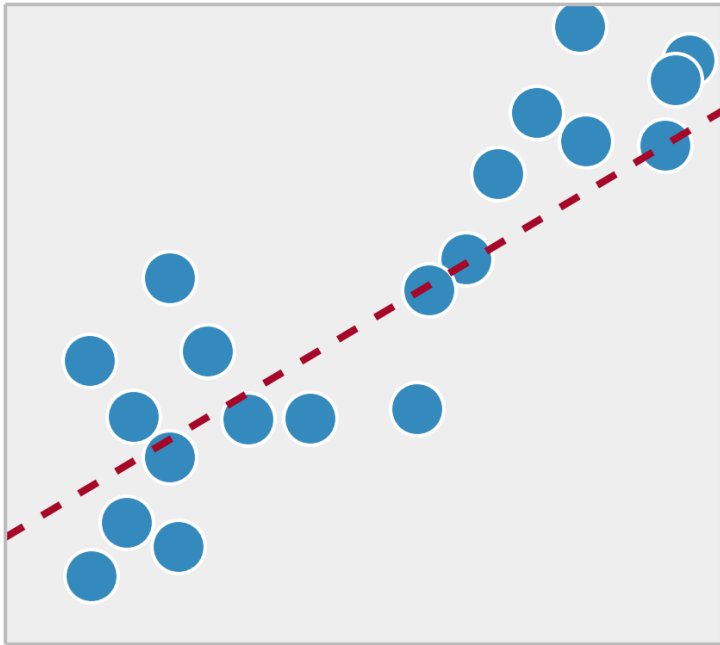
Classification



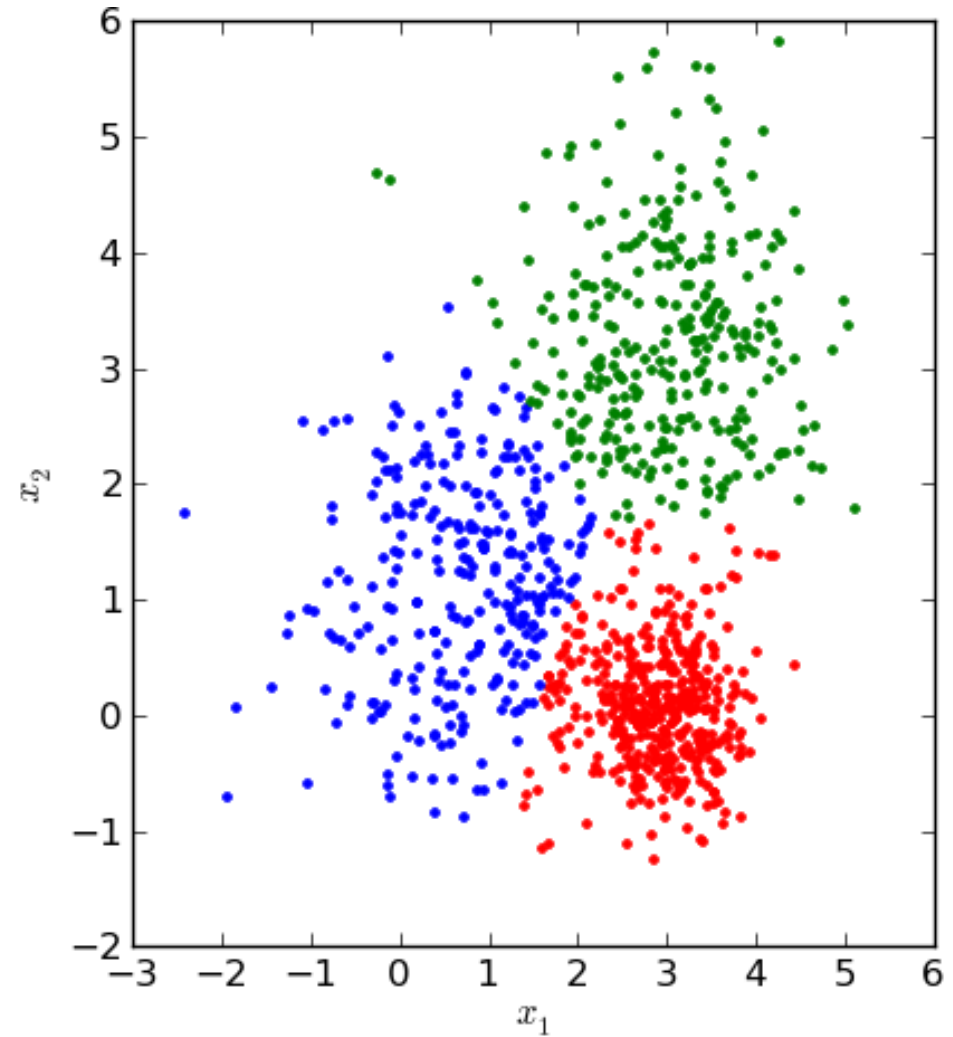
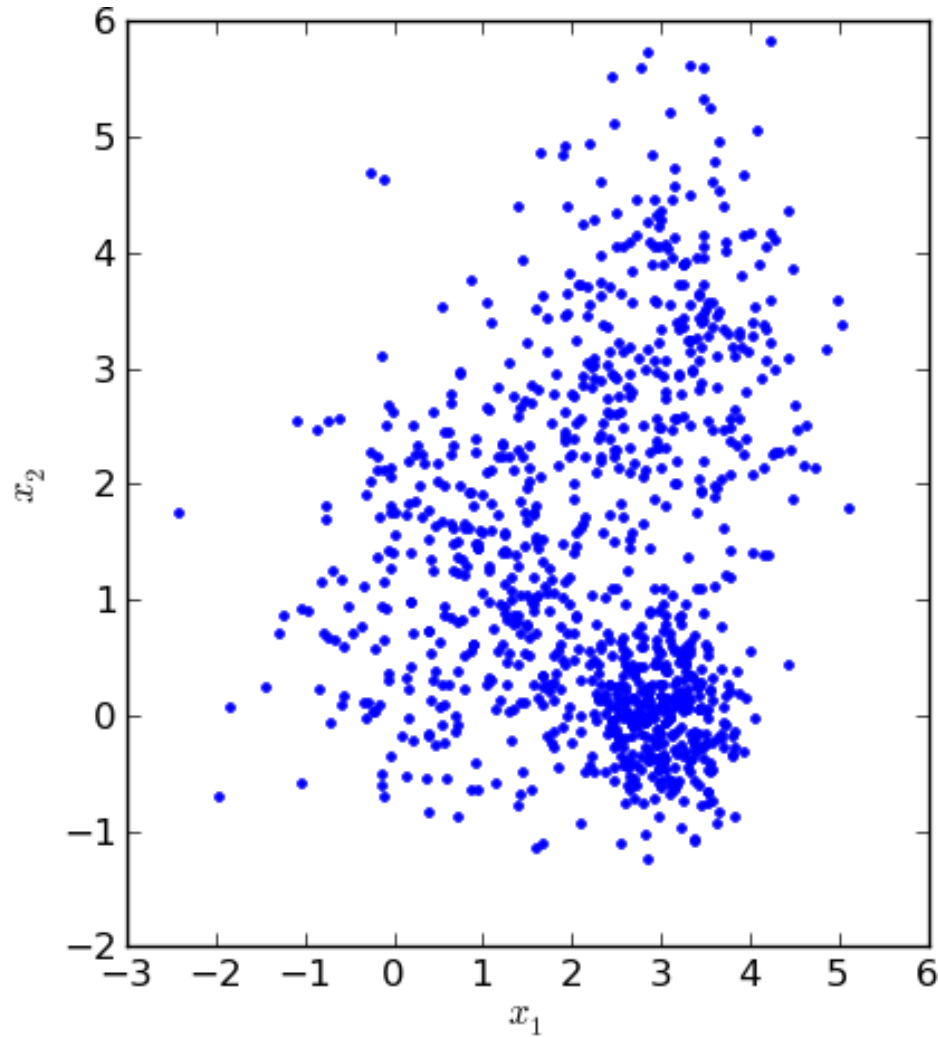
VS



Régression



Regroupement (*Clustering*)



Modèles génératifs



Modèles génératifs



Modèles génératifs









Renforcement



Google DeepMind

Challenge Match

8 - 15 March 2017



ALPHAGO
00:05:30

LEE SEDOL
00:28:28

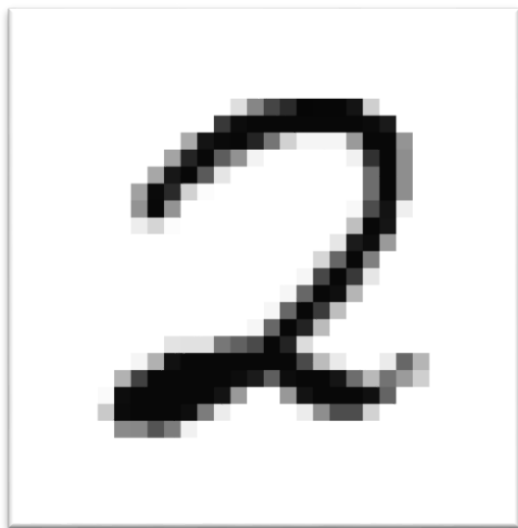


AlphaGo Lee Sedol

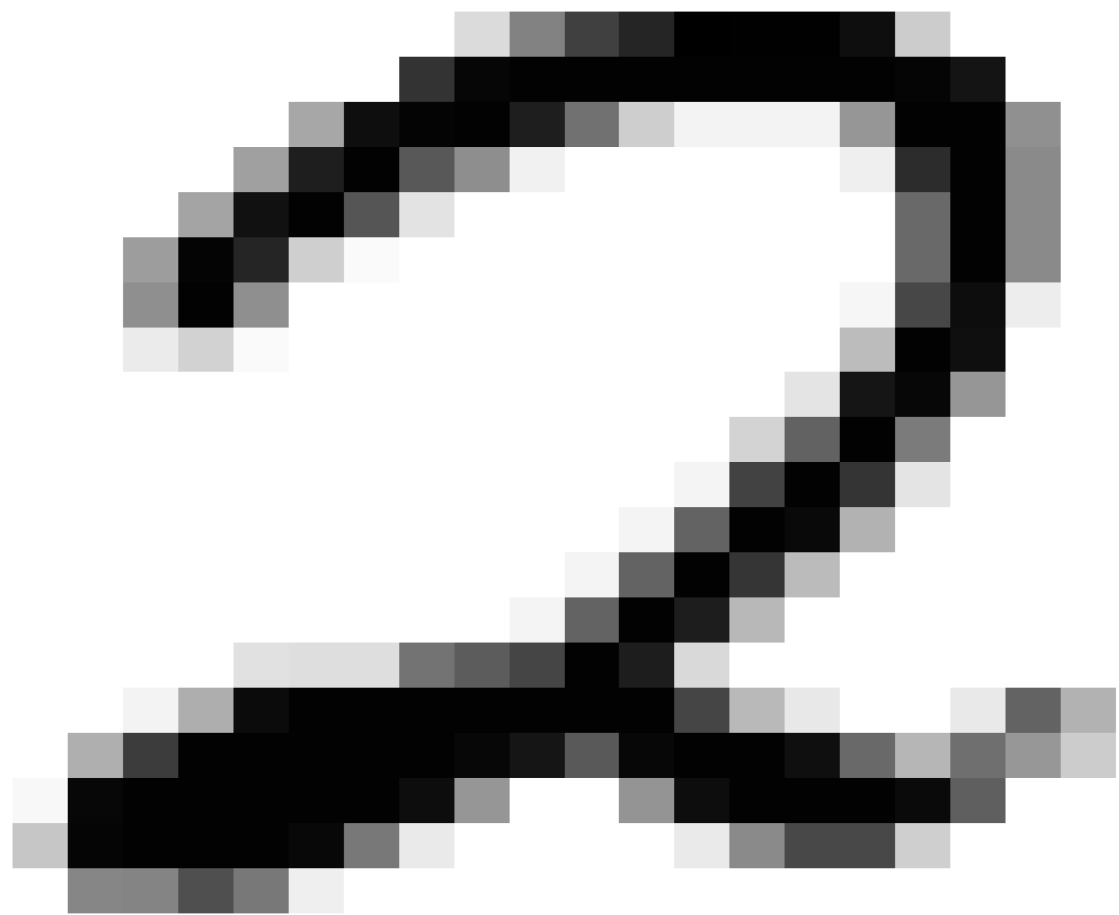


3	4	2	1	9	5	6	2	1	8
8	9	1	2	5	0	0	6	6	4
6	7	0	1	6	3	6	3	7	0
3	7	7	9	4	6	6	1	8	2
2	9	3	4	3	9	8	7	2	5
1	5	9	8	3	6	5	7	2	3
9	3	1	9	1	5	8	0	8	4
5	6	2	6	8	5	8	8	9	9
3	7	7	0	9	4	8	5	4	3
7	9	6	4	7	0	6	9	2	3

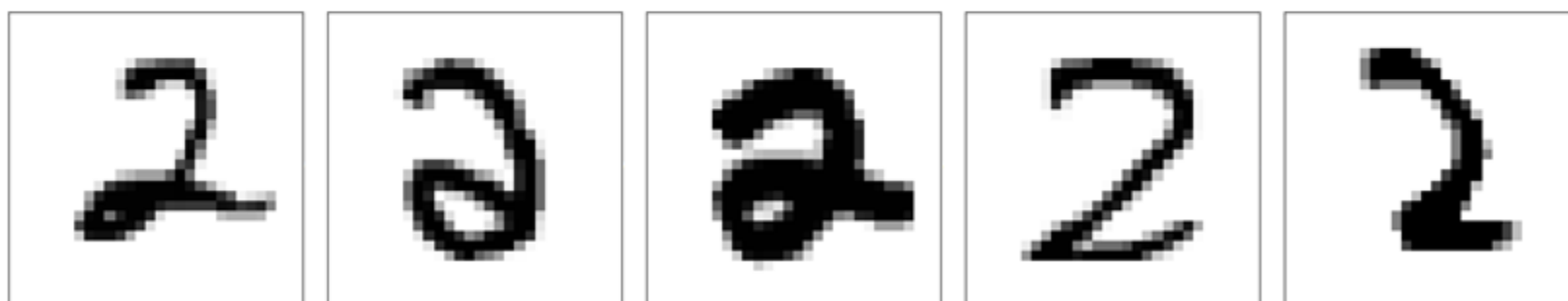
Classification



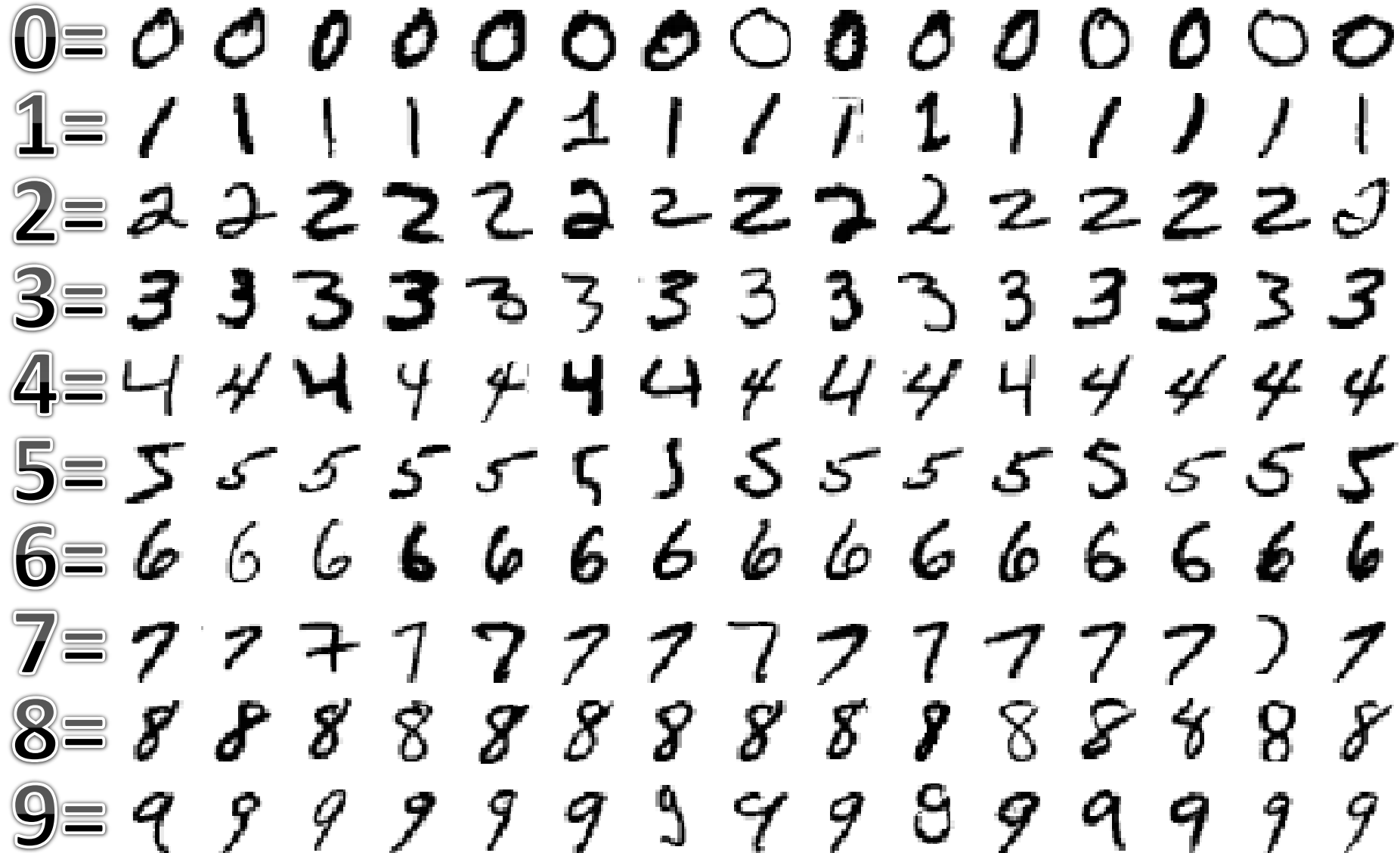
= ?

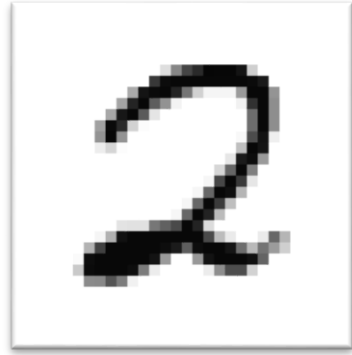


$$A = (a_1, a_2, a_3, a_4, a_5, \dots, a_{780}, a_{781}, a_{782}, a_{783}, a_{784})$$

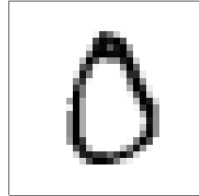


60000 exemples d'apprentissage

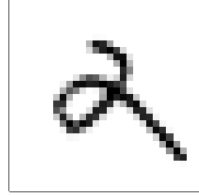




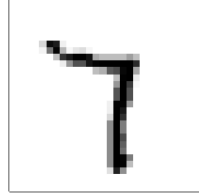
= ?



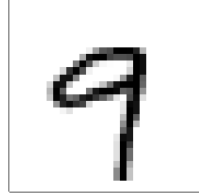
=0



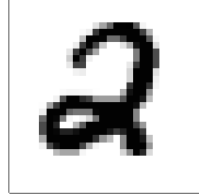
=2



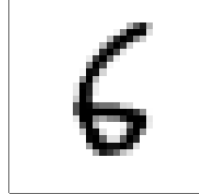
=7



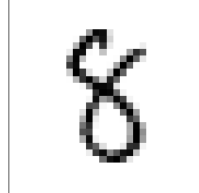
=9



=2



=6



=8

2

0

=0

2

=2

7

=7

9

=9

2

=2

6

=6

8

=8

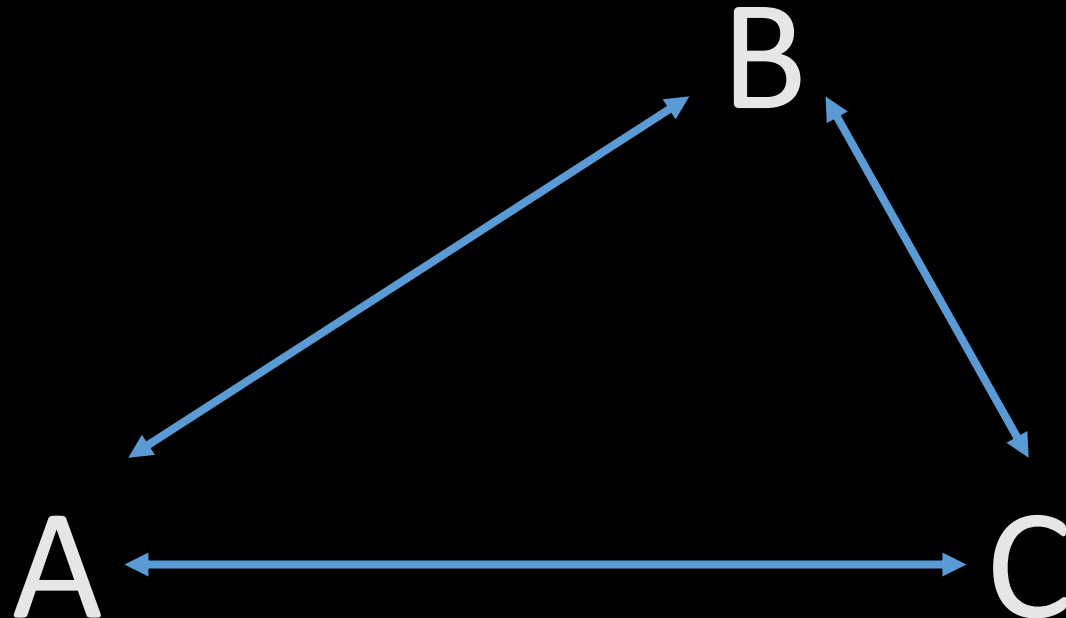
Distance

Positive: $D(A, B) \geq 0$

Symétrie: $D(A, B) = D(B, A)$

Séparation: $D(A, B) = 0 \rightarrow A = B$

Inégalité triangulaire: $D(A, C) \leq D(A, B) + D(B, C)$

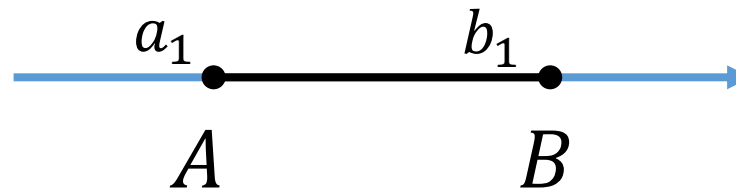


Distance Euclidienne



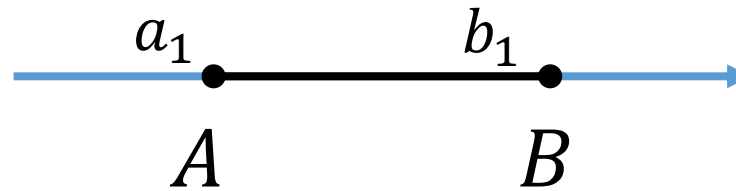
$$D(A, B)$$

Distance Euclidienne



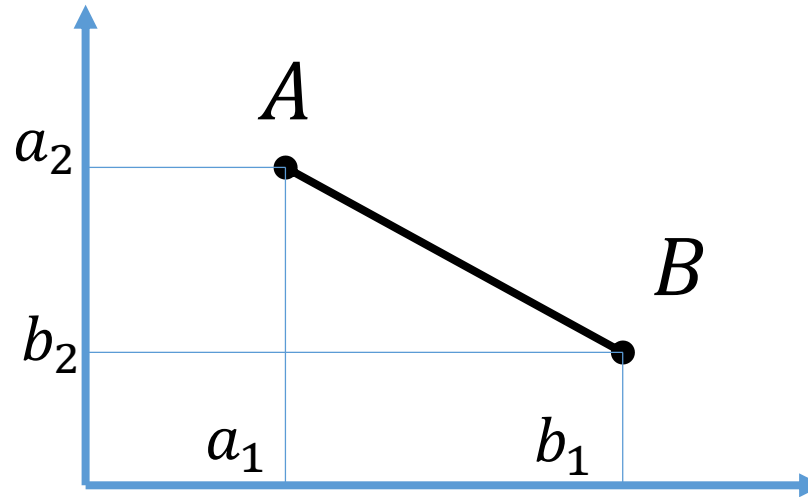
$$D(A, B) = b_1 - a_1$$

Distance Euclidienne



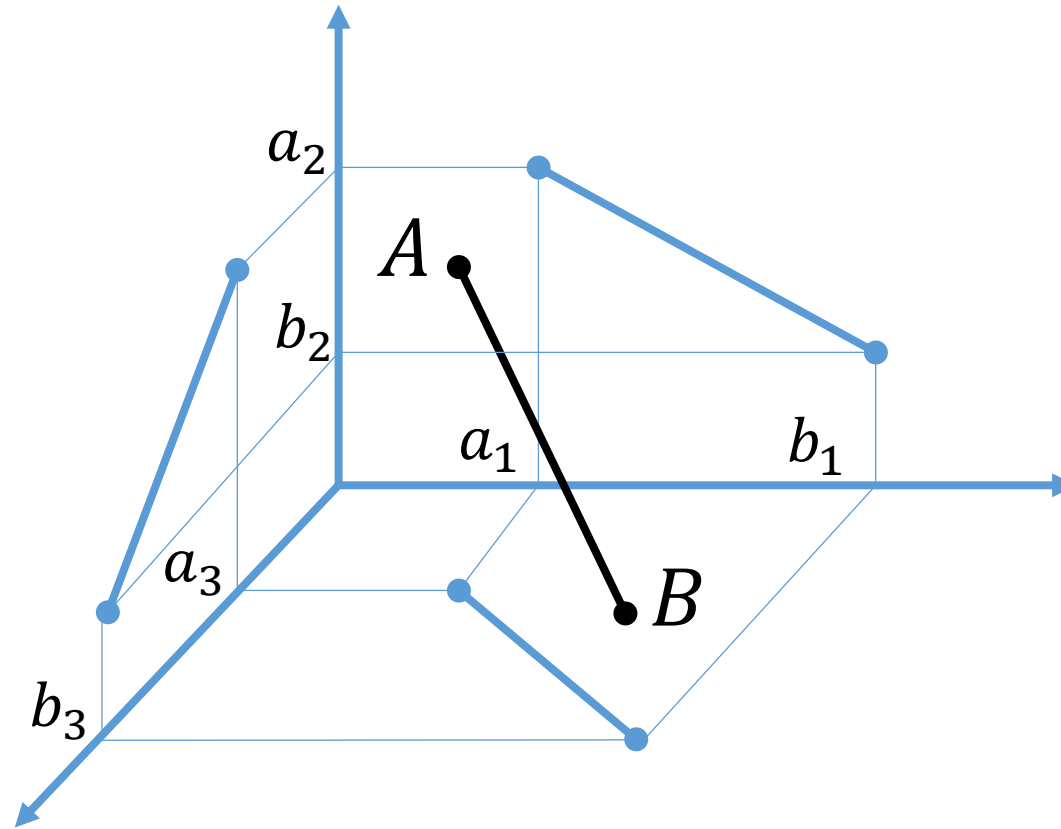
$$D(A, B) = \sqrt{(b_1 - a_1)^2}$$

Distance Euclidienne



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

Distance Euclidienne



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

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

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

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

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

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

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

A B $D(A, B)$

2	0	2688.85
2	2	2862.24
2	7	2700.58
2	9	2951.47
2	2	2171.14
2	6	2672.08
2	8	2633.81

Plus proches voisins



60000 exemples de référence

$$E = 3.8\%$$

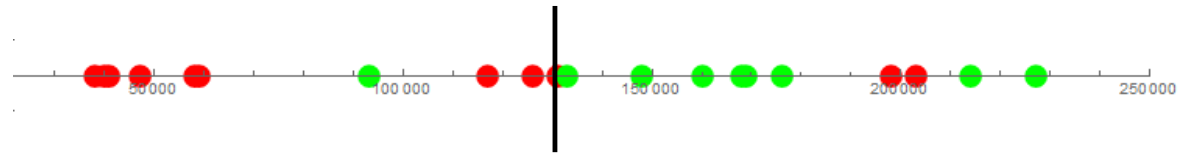
Nos données

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

Nos données

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

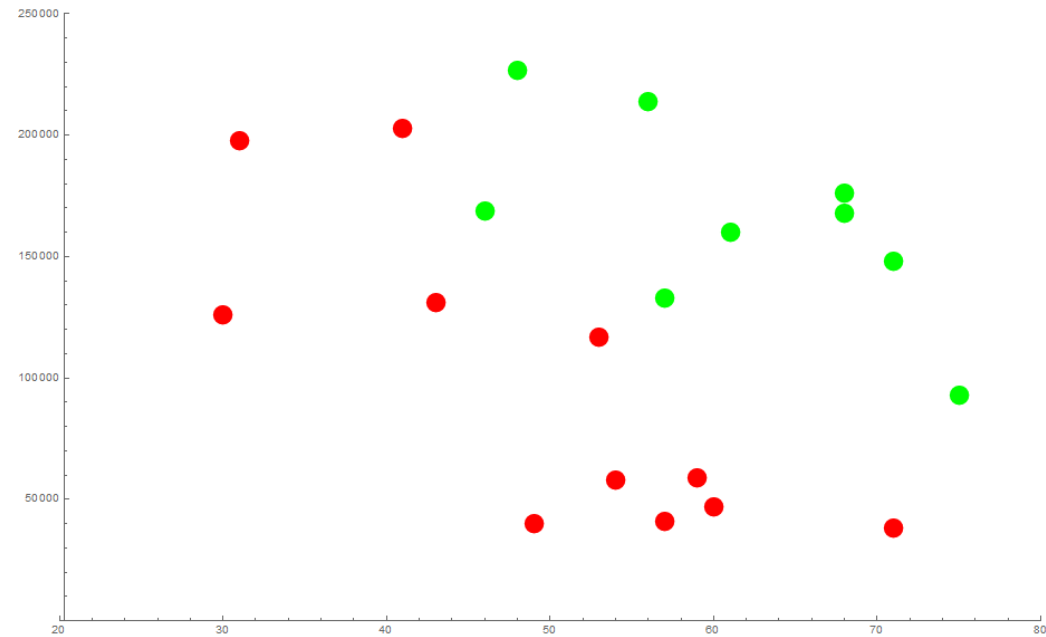


Salaire $>$ 132000

$$E = 15\%$$

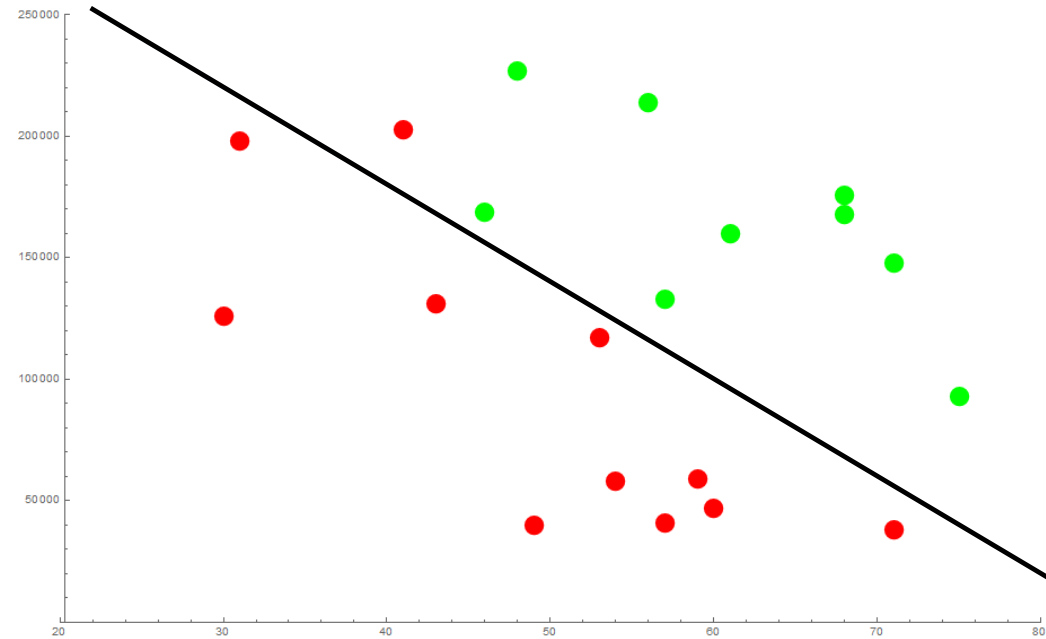
Nos données

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



Nos données

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

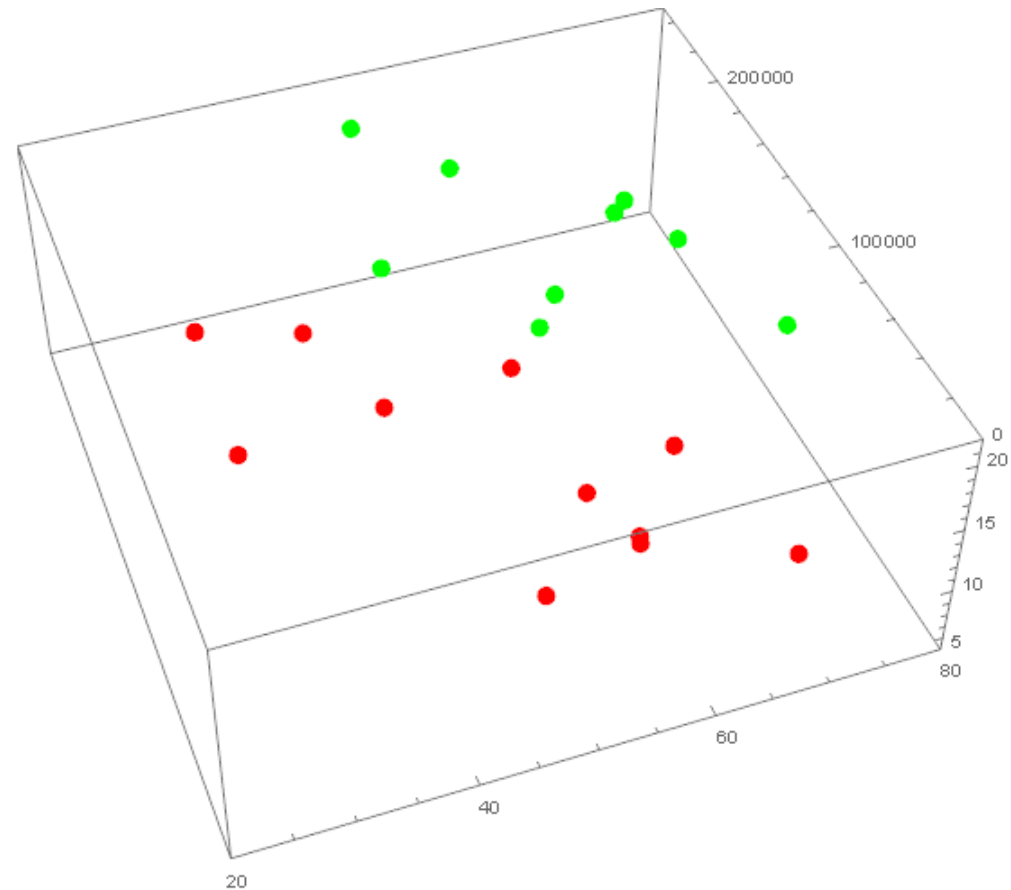


$$4000 \text{ Age} + \text{Salaire} > 340000$$

$$E = 5\%$$

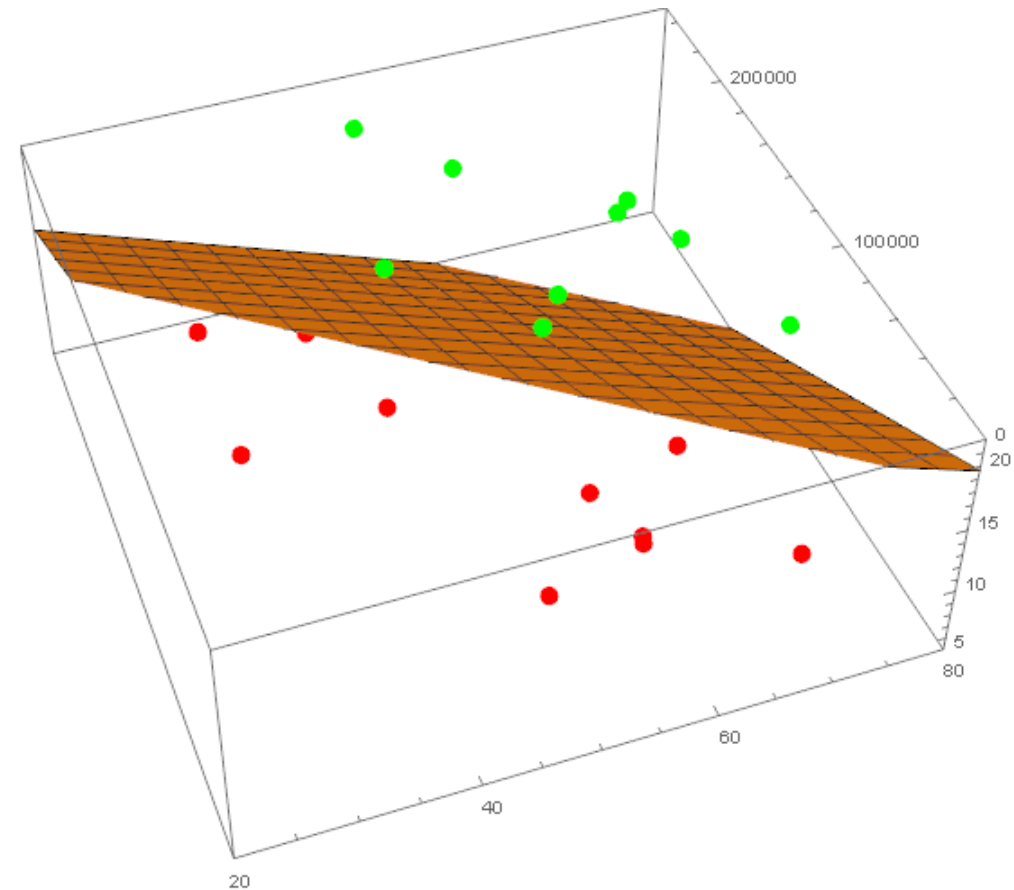
Nos données

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



Nos données

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



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

$$E = 0\%$$

$$\text{Salaire} > 132000$$

$$4000 \text{ Age} + \text{Salaire} > 340000$$

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

$$a_1 > 132000$$

$$4000a_2 + a_1 > 340000$$

$$11900a_3 + 3250a_2 + a_1 > 499000$$

$$1789a_4 + 1190a_3 + 4360a_2 + a_1 > 273456$$

$$2719a_5 + 5799a_4 + 9018a_3 + 3477a_2 + a_1 > 345678$$

$$m_5a_5 + m_4a_4 + m_3a_3 + m_2a_2 + m_1a_1 > s$$

$$\langle M, A \rangle > s$$

Hyperplan

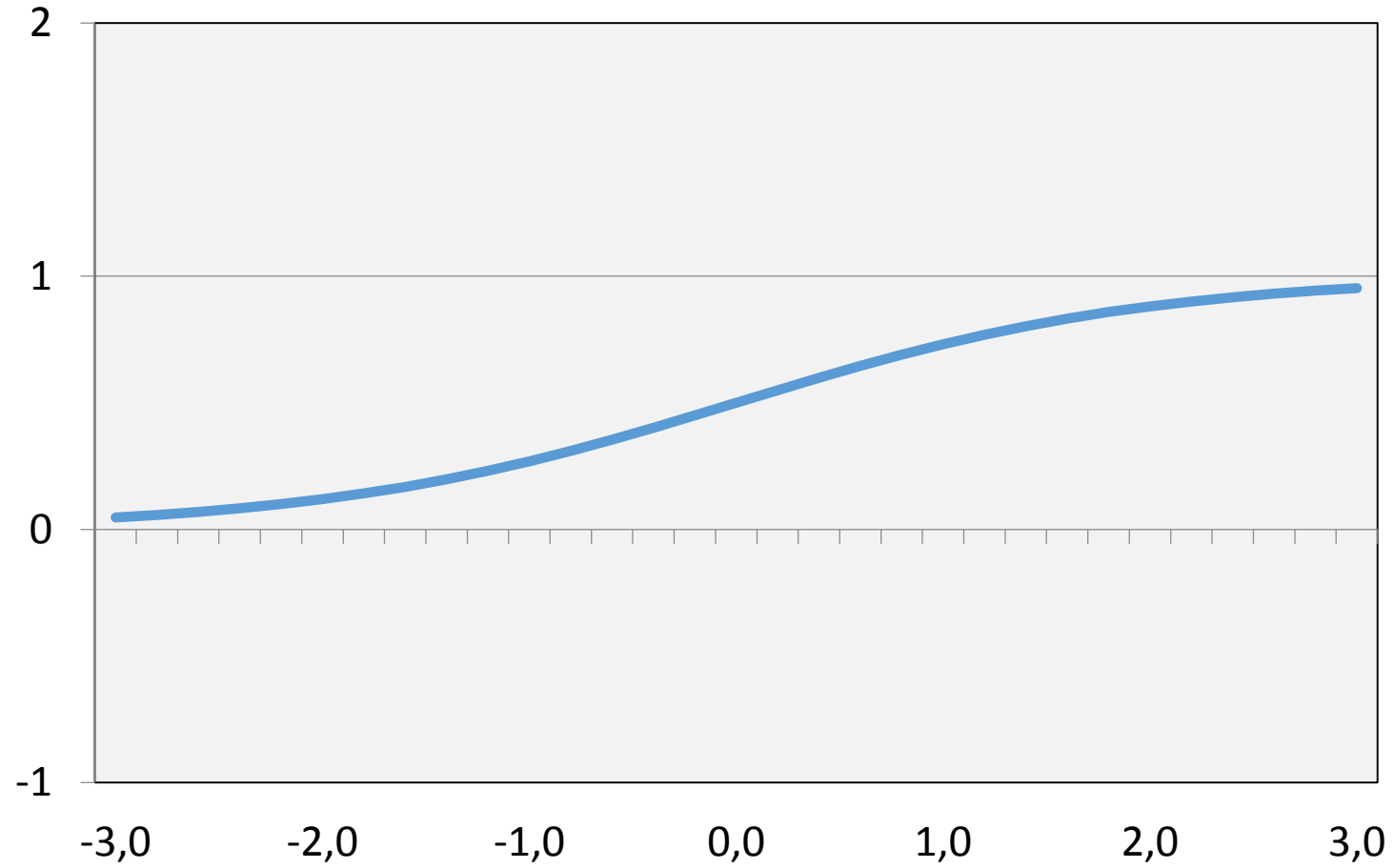
0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 1 1 1 1 1 1 1 1 1 1 1 1 1
2 2 2 2 2 2 2 2 2 2 2 2 2 2
3 3 3 3 3 3 3 3 3 3 3 3 3 3
4 4 4 4 4 4 4 4 4 4 4 4 4 4
5 5 5 5 5 5 5 5 5 5 5 5 5 5
6 6 6 6 6 6 6 6 6 6 6 6 6 6
7 7 7 7 7 7 7 7 7 7 7 7 7 7
8 8 8 8 8 8 8 8 8 8 8 8 8 8
9 9 9 9 9 9 9 9 9 9 9 9 9 9

$$E = 7.4\%$$

La Sigmoid

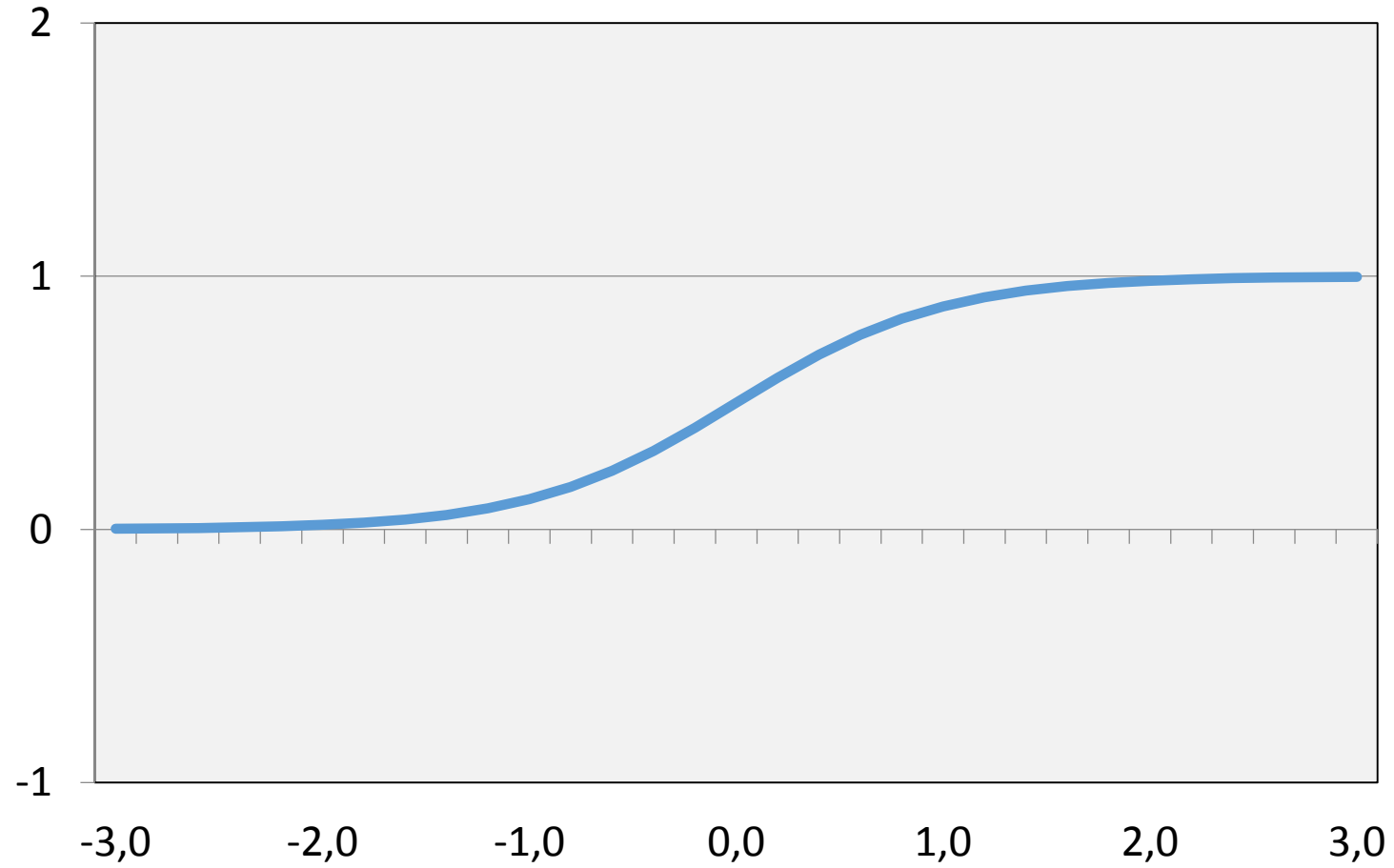
$$\varphi(x) = \frac{1}{1 + e^{-x}}$$

Sigmoïde



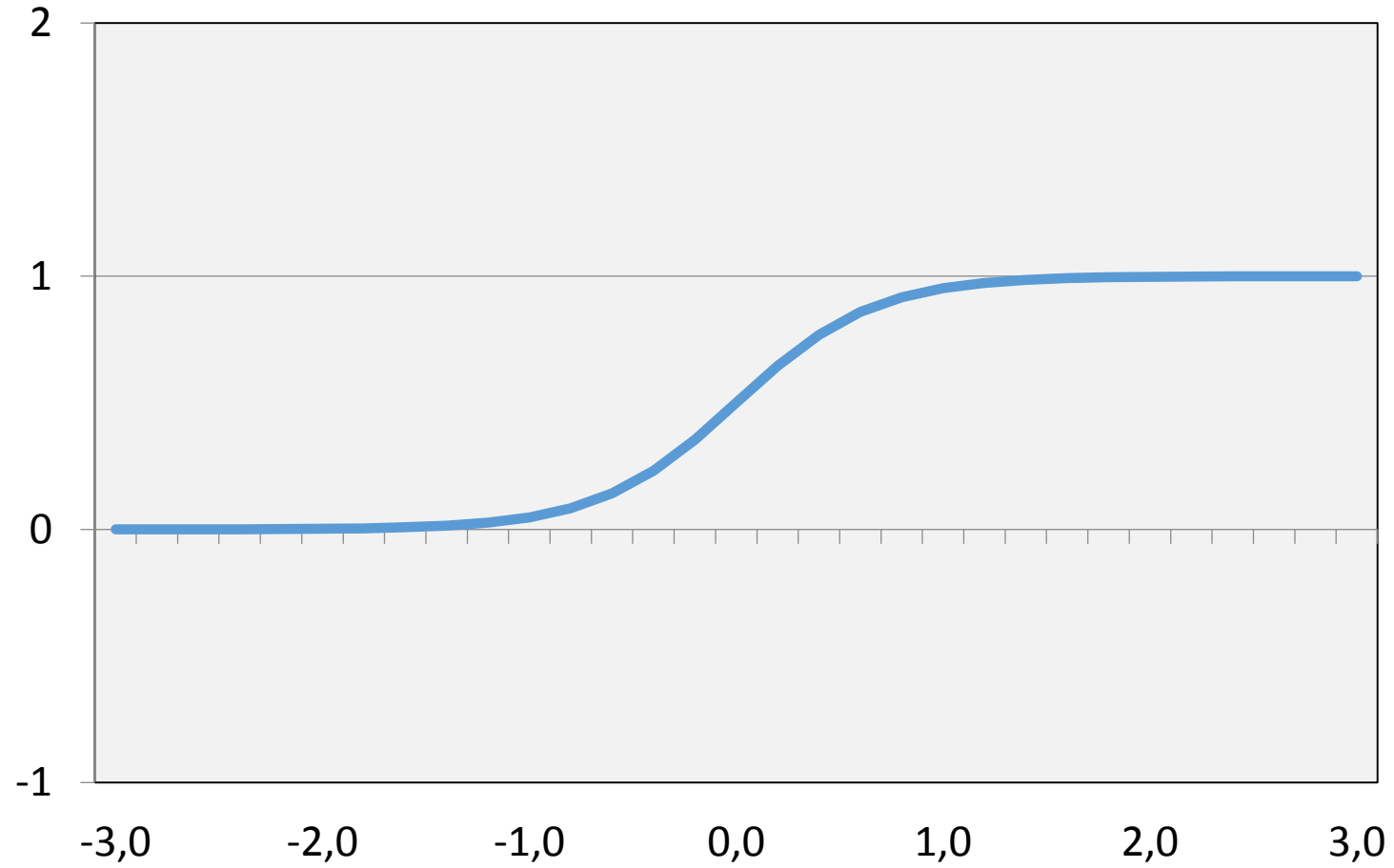
$$\varphi(x) = \frac{1}{1 + e^{-1x}}$$

Sigmoïde



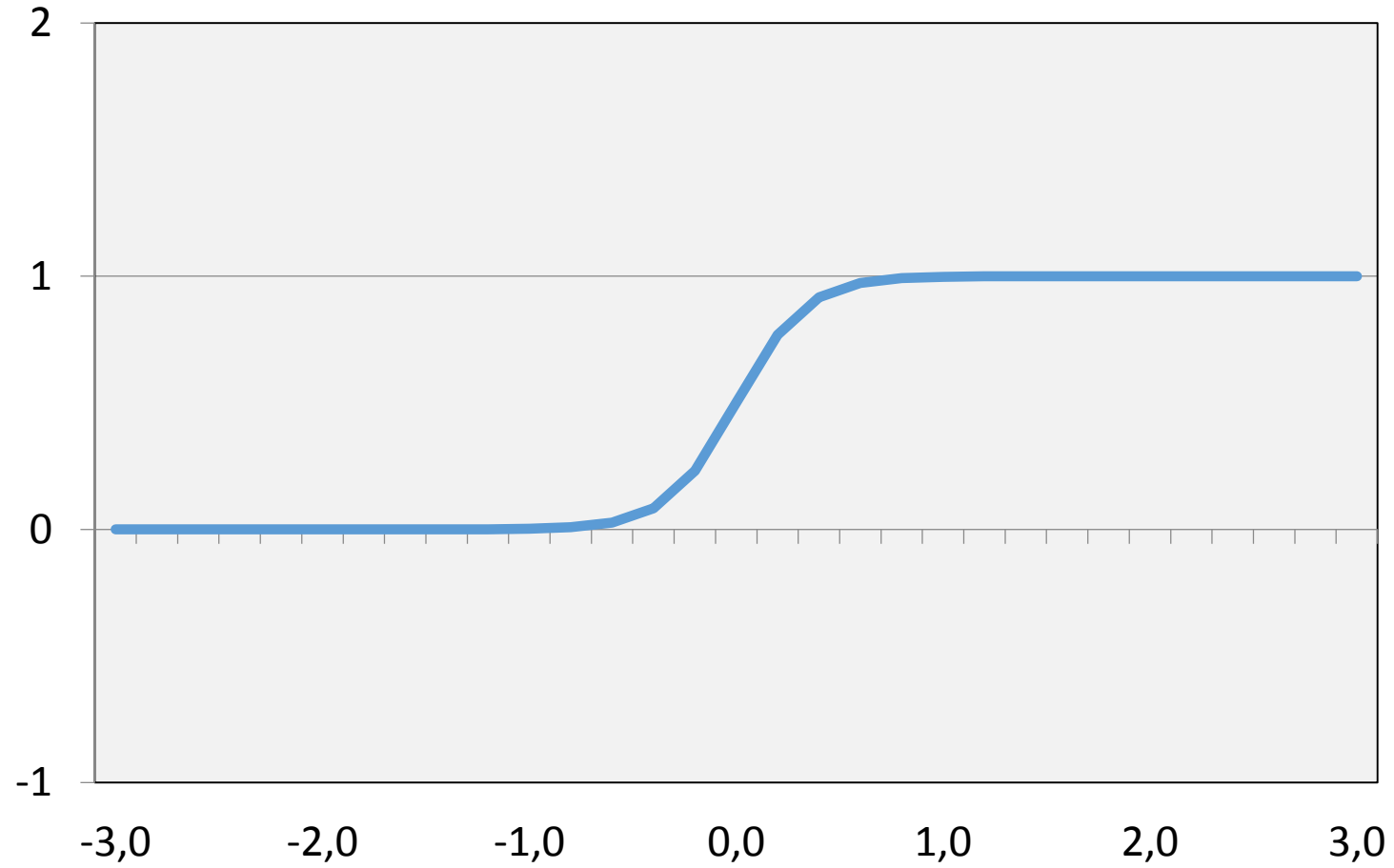
$$\varphi(x) = \frac{1}{1 + e^{-2x}}$$

Sigmoïde



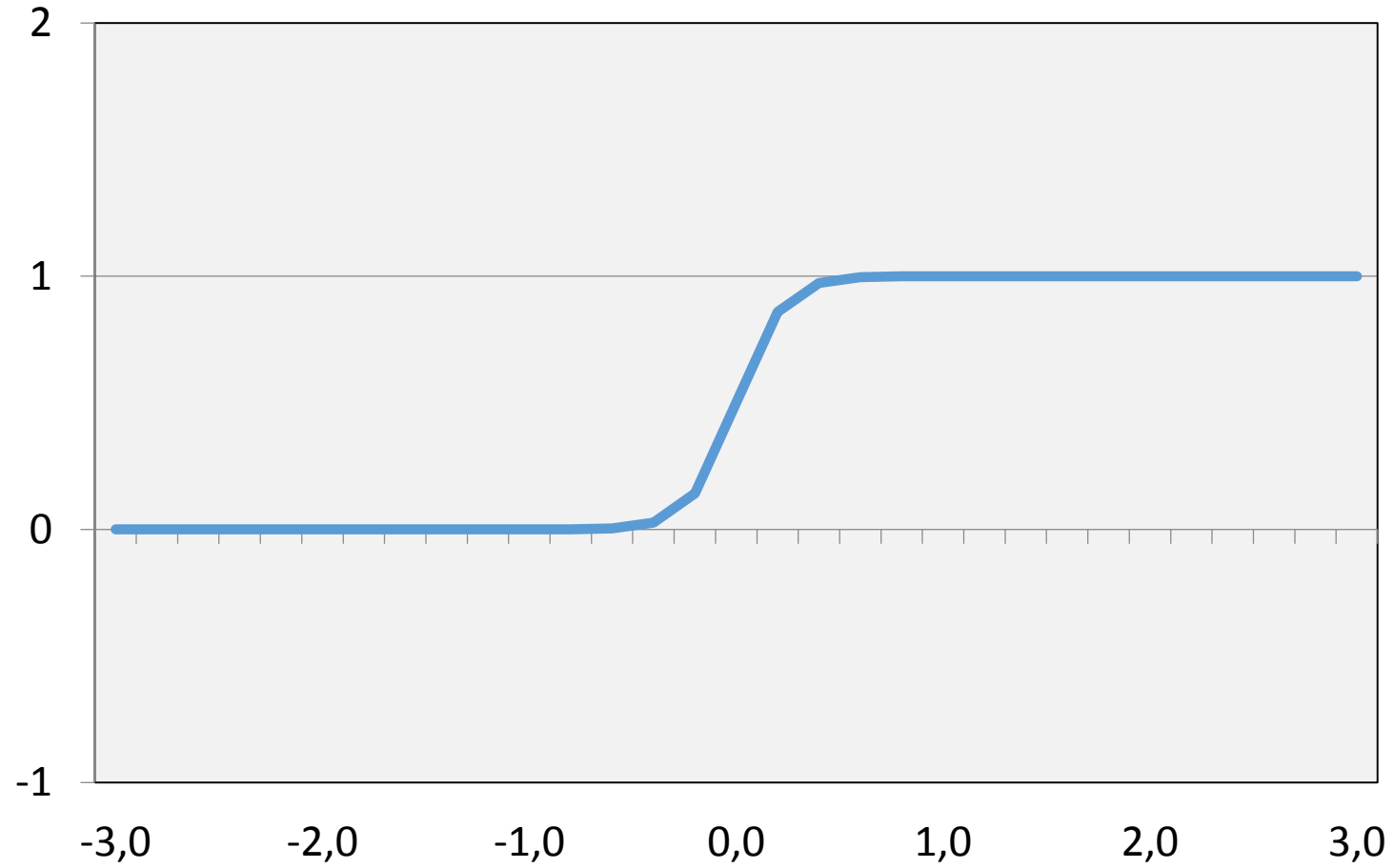
$$\varphi(x) = \frac{1}{1 + e^{-3x}}$$

Sigmoïde



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

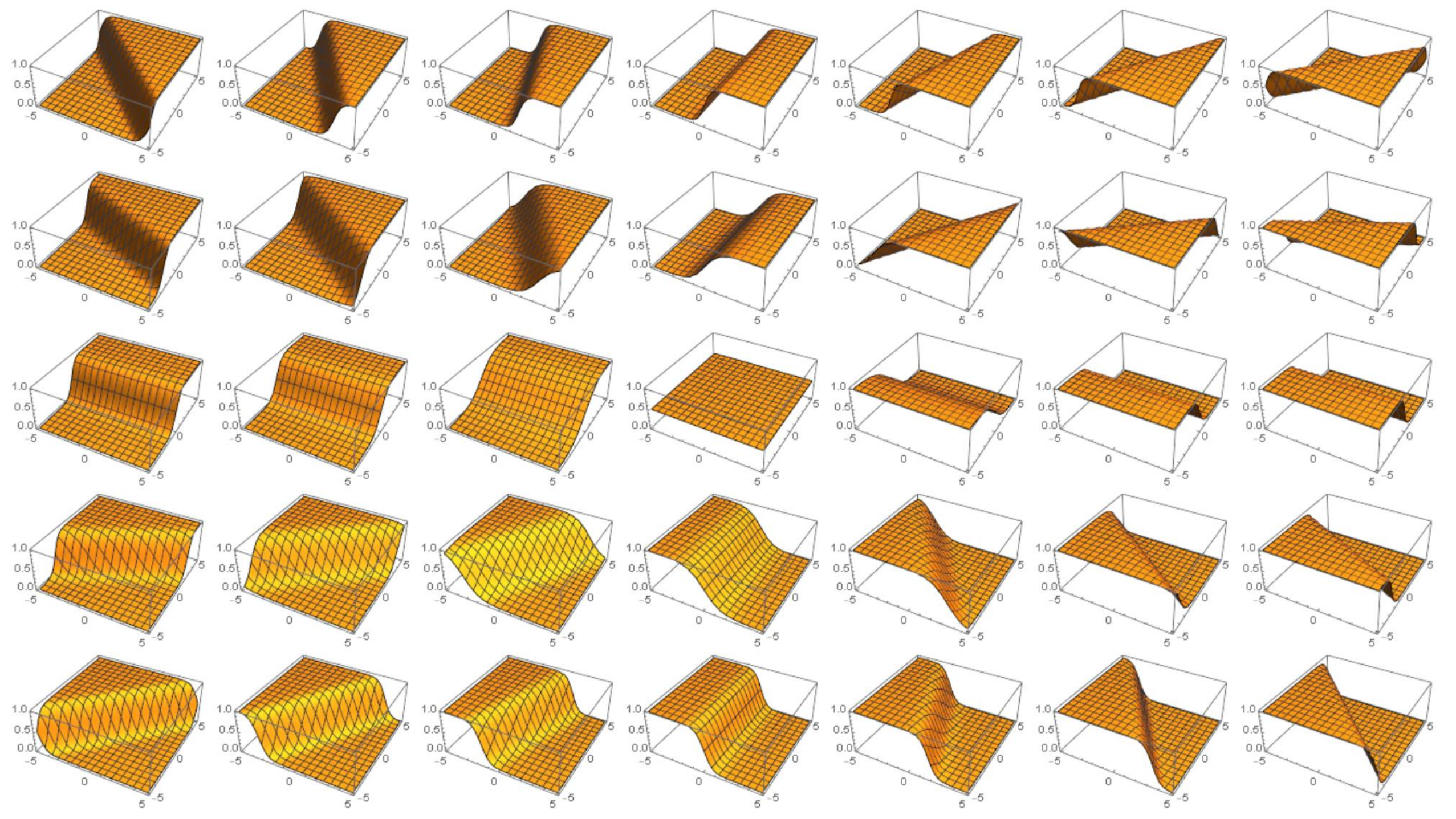
Sigmoïde



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

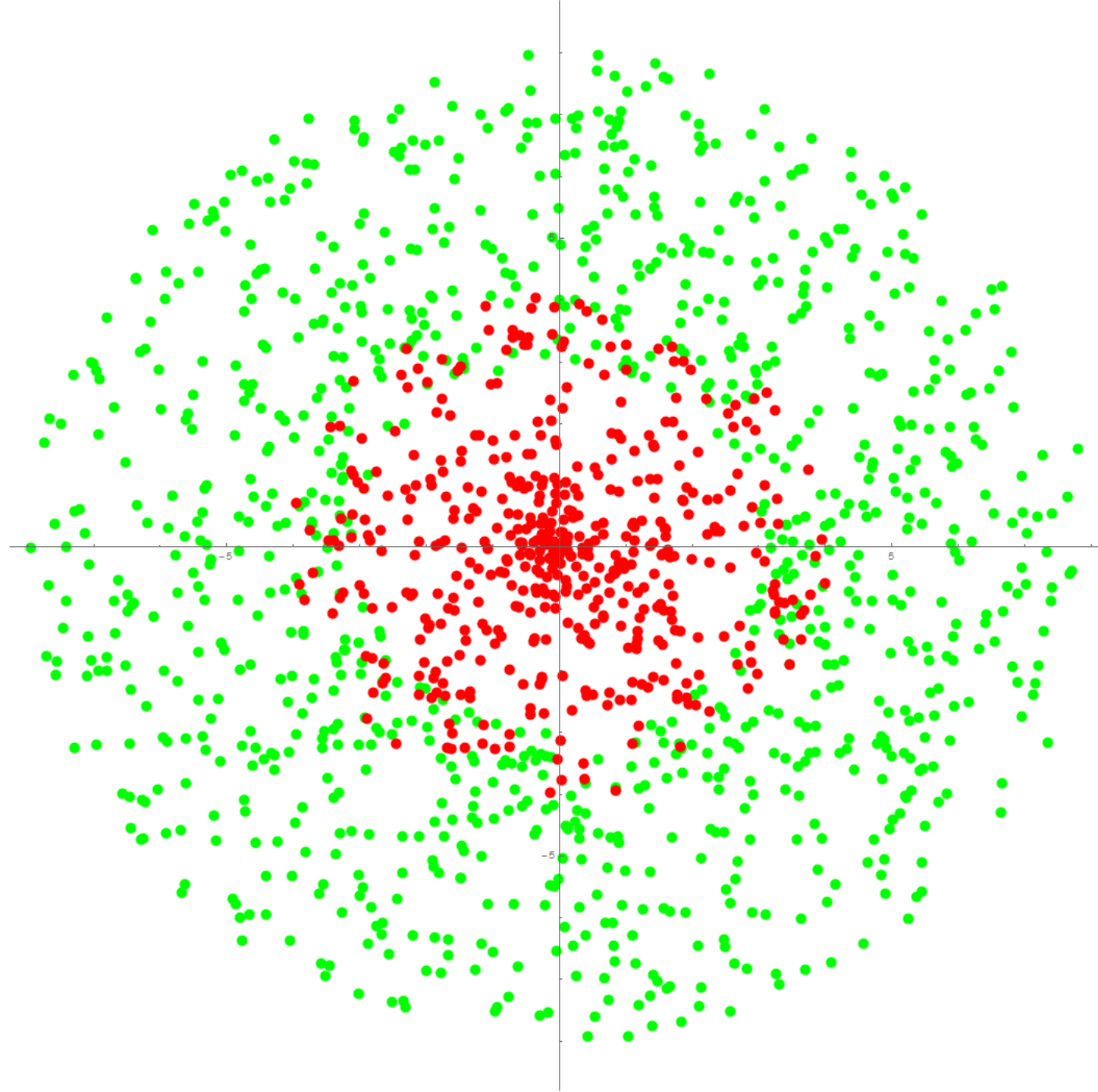
Classification Logistique

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

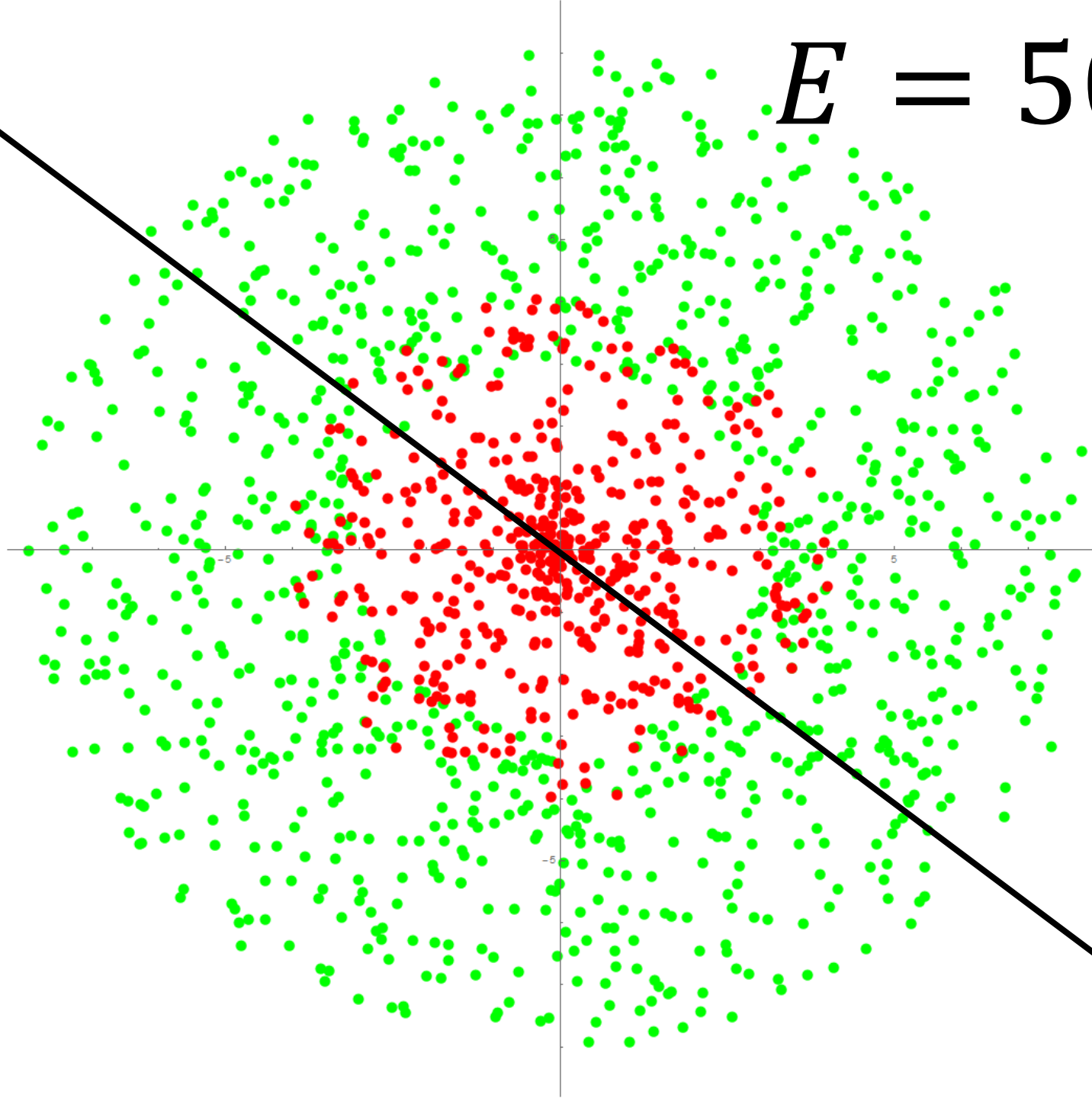


Δx	Δy	OK
-3.85	4.81	OUI
-1.53	-7.57	OUI
-0.25	-0.920	NON
-1.04	-0.890	NON
0.14	-0.033	NON
0.80	7.88	OUI
2.51	0.395	NON
6.89	-0.21	OUI
2.94	4.35	OUI
-1.56	-0.847	NON
-1.01	0.057	NON
2.34	-5.52	OUI
-1.45	-5.23	OUI
-4.68	0.50	OUI
4.27	-3.28	OUI
1.24	1.97	NON
-0.29	0.85	NON
1.35	1.86	NON
-1.57	0.05	NON
5.42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

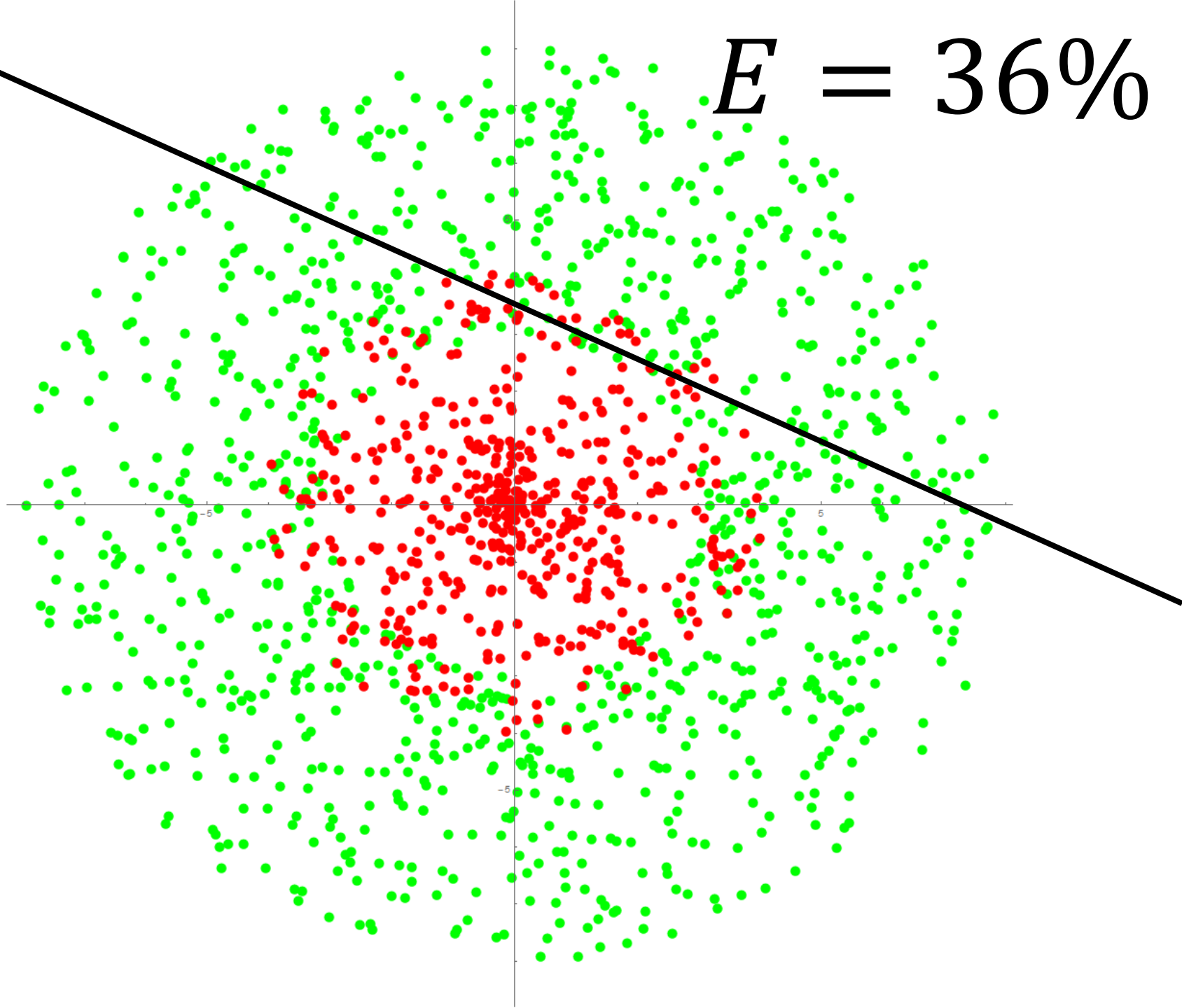
Δx	Δy	OK
-3.85	4.81	OUI
-1.53	-7.57	OUI
-0.25	-0.920	NON
-1.04	-0.890	NON
0.14	-0.033	NON
0.80	7.88	OUI
2.51	0.395	NON
6.89	-0.21	OUI
2.94	4.35	OUI
-1.56	-0.847	NON
-1.01	0.057	NON
2.34	-5.52	OUI
-1.45	-5.23	OUI
-4.68	0.50	OUI
4.27	-3.28	OUI
1.24	1.97	NON
-0.29	0.85	NON
1.35	1.86	NON
-1.57	0.05	NON
5.42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



$E = 50\%$

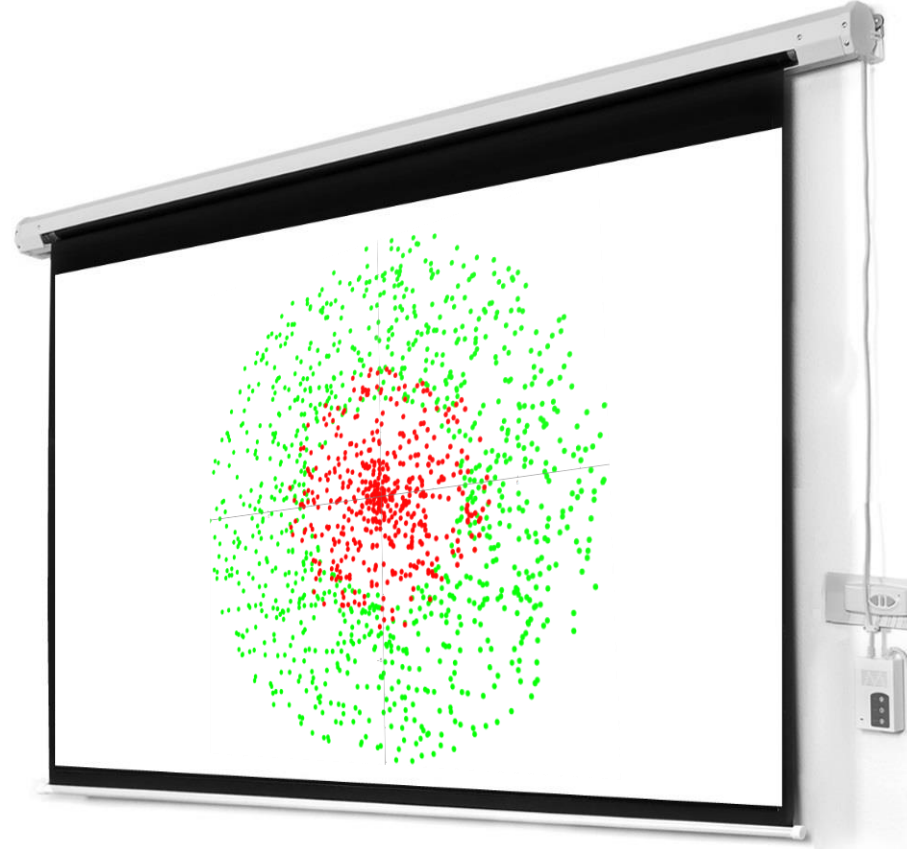


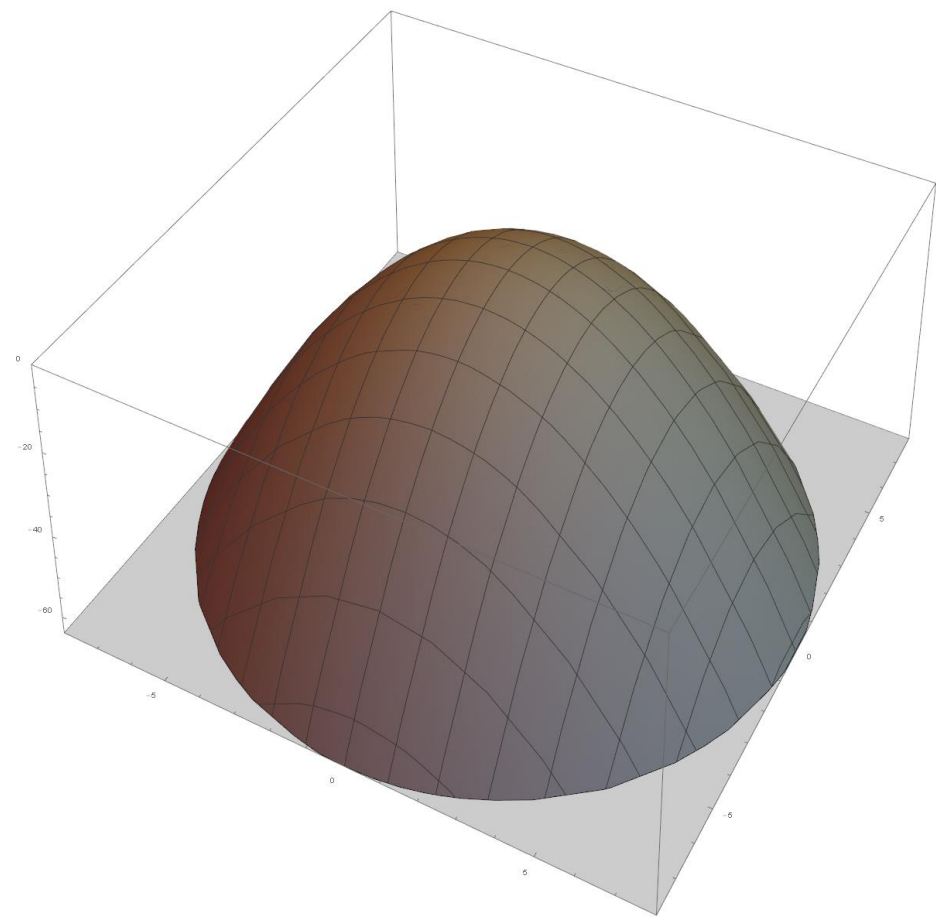
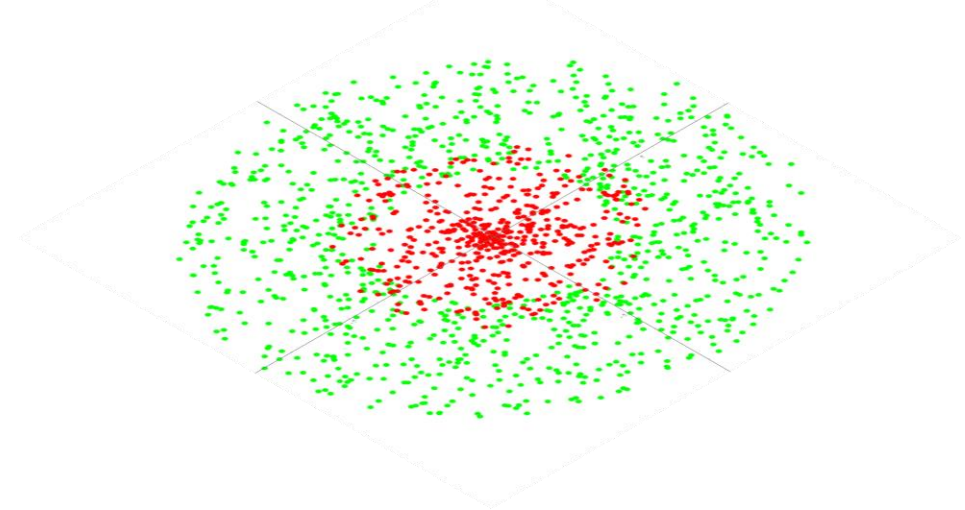
$E = 36\%$

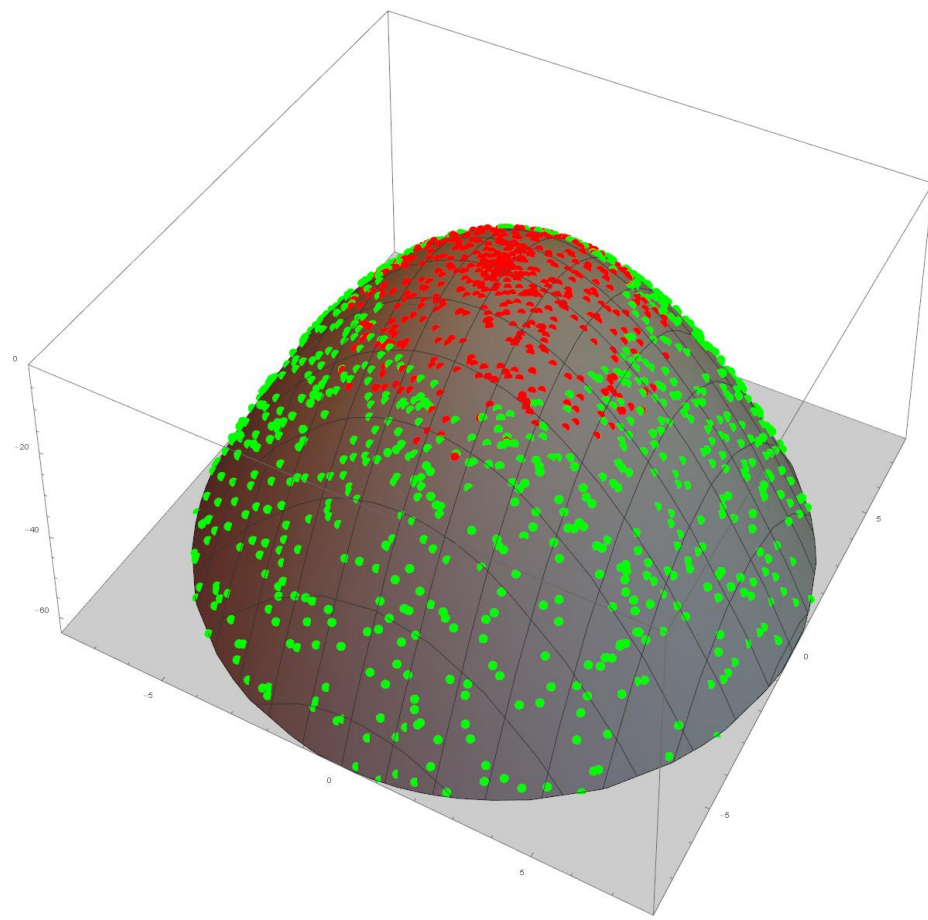


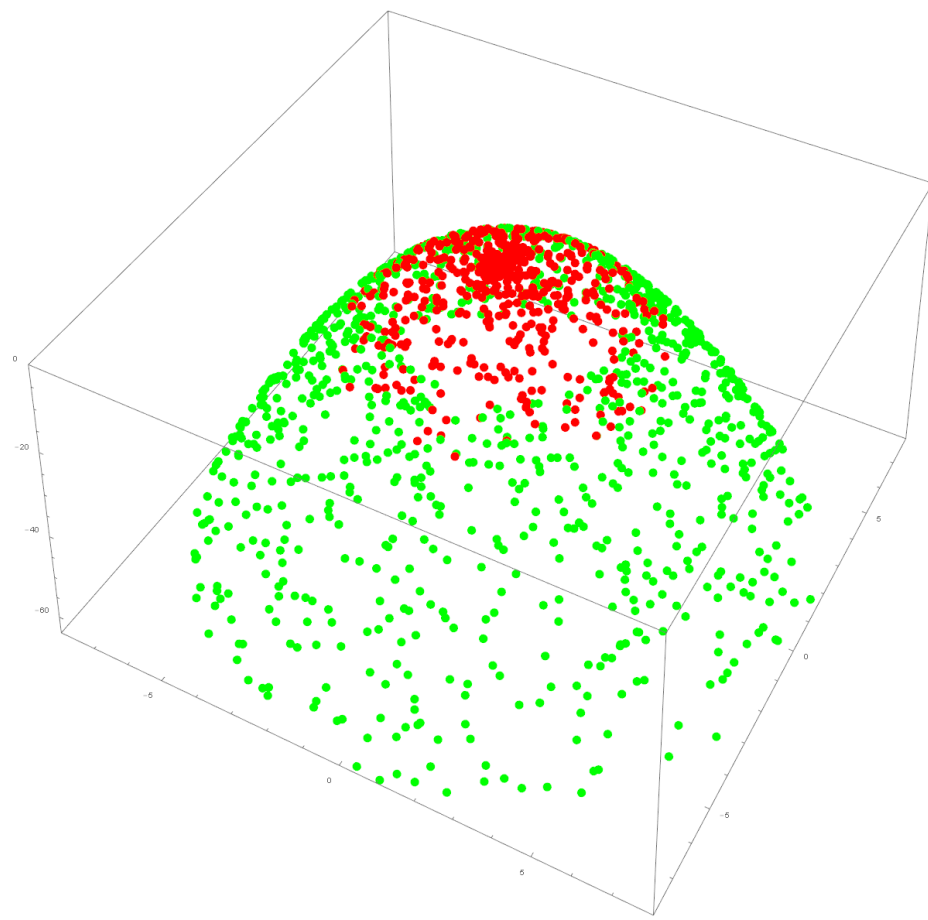
Projection

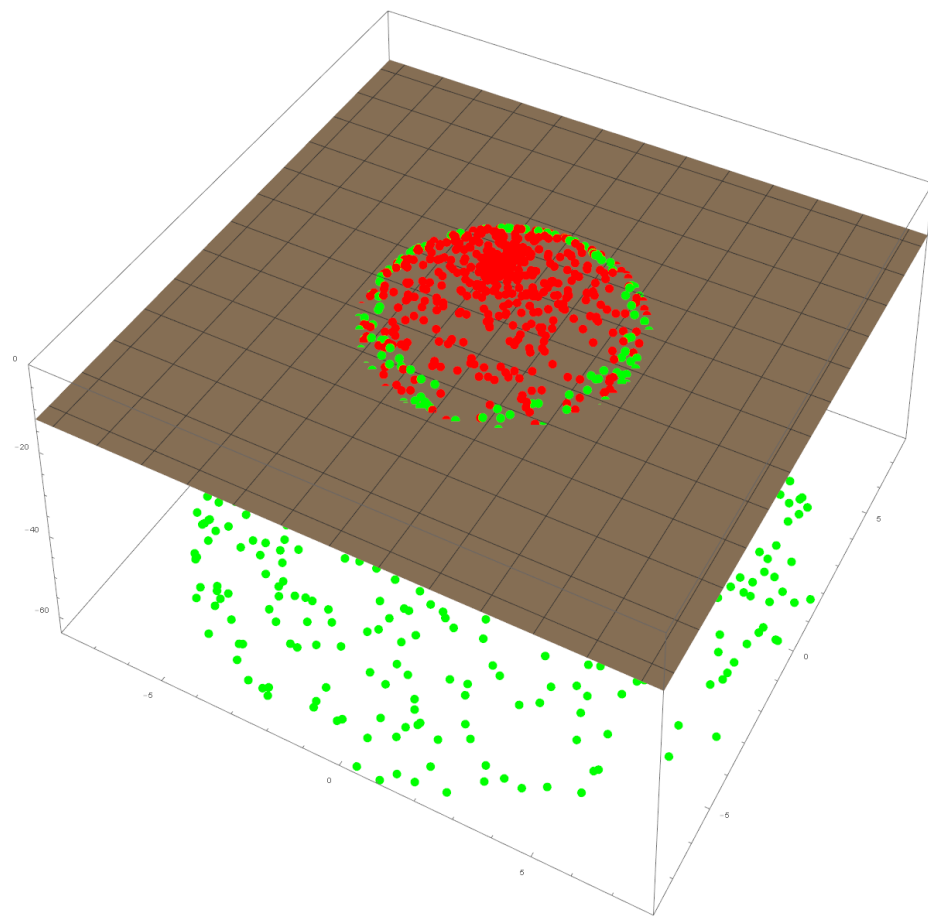
$$(a_1, a_2) \rightarrow (a_1, a_2, a_1^2 + a_2^2)$$



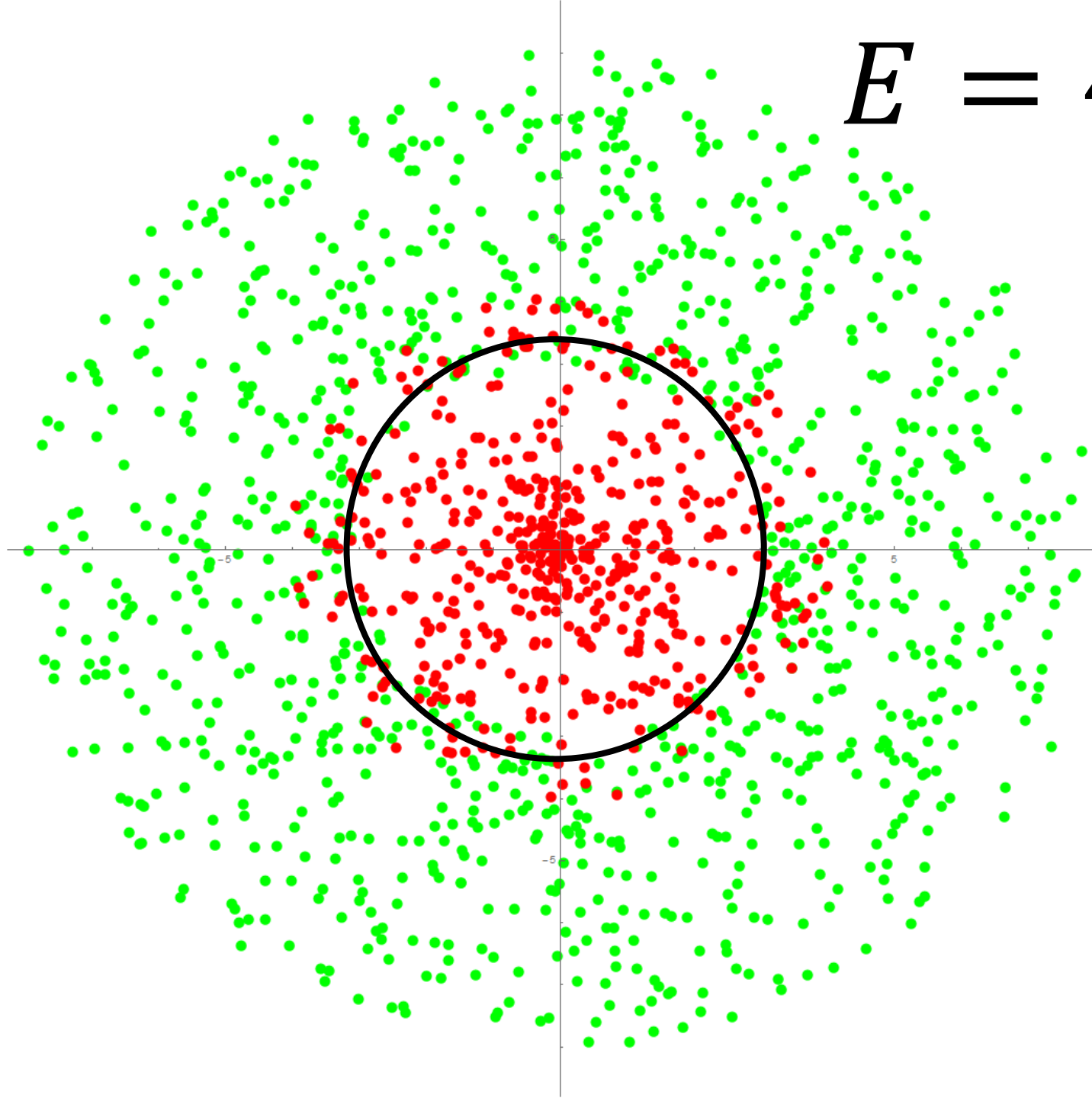








$$E = 4\%$$



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

0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 1 1 1 1 1 1 1 1 1 1 1 1 1
2 2 2 2 2 2 2 2 2 2 2 2 2 2
3 3 3 3 3 3 3 3 3 3 3 3 3 3
4 4 4 4 4 4 4 4 4 4 4 4 4 4
5 5 5 5 5 5 5 5 5 5 5 5 5 5
6 6 6 6 6 6 6 6 6 6 6 6 6 6
7 7 7 7 7 7 7 7 7 7 7 7 7 7
8 8 8 8 8 8 8 8 8 8 8 8 8 8
9 9 9 9 9 9 9 9 9 9 9 9 9 9

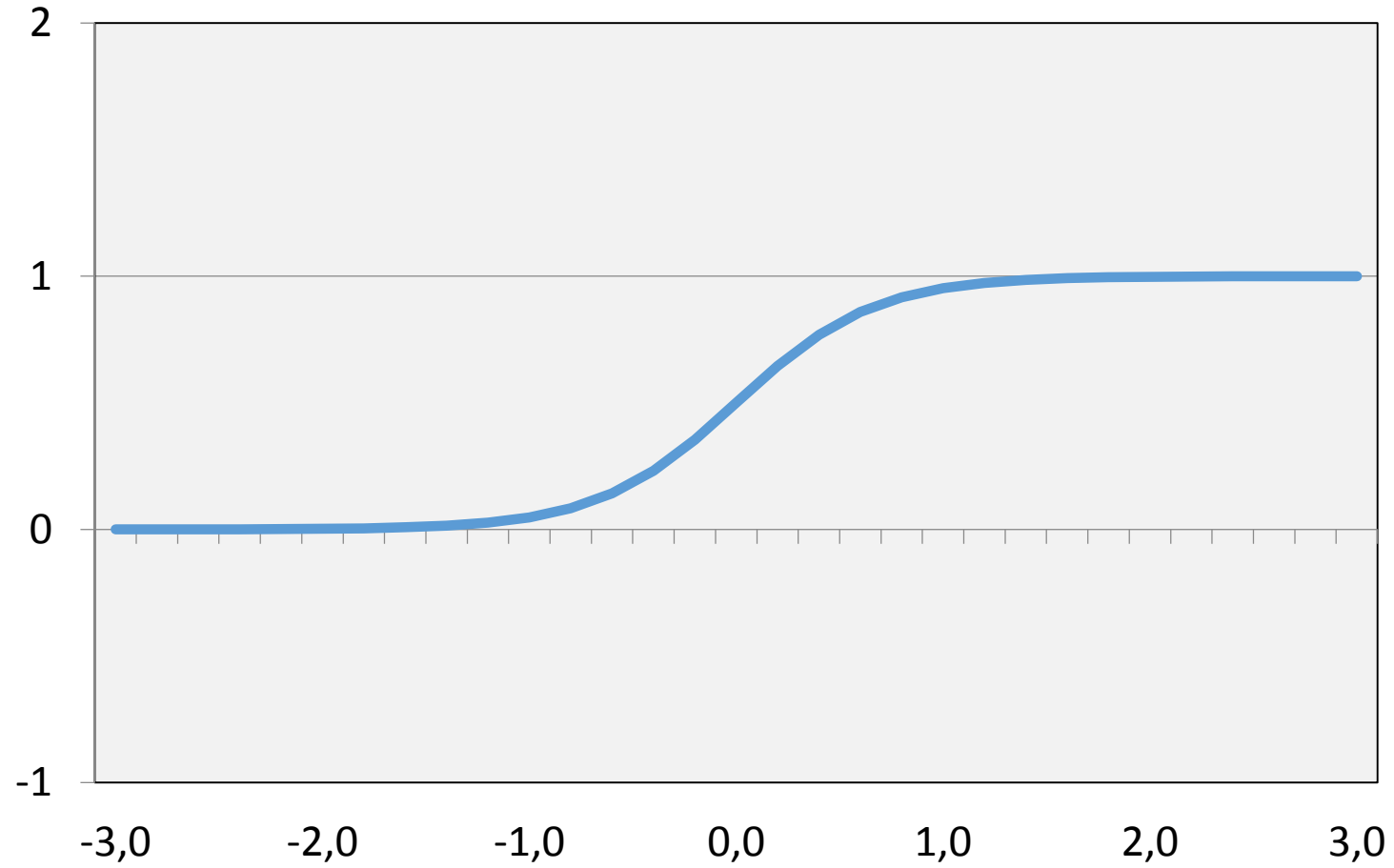
$$E = 1.9\%$$

Frank Rosenblatt 1958

Perceptron



Sigmoïde

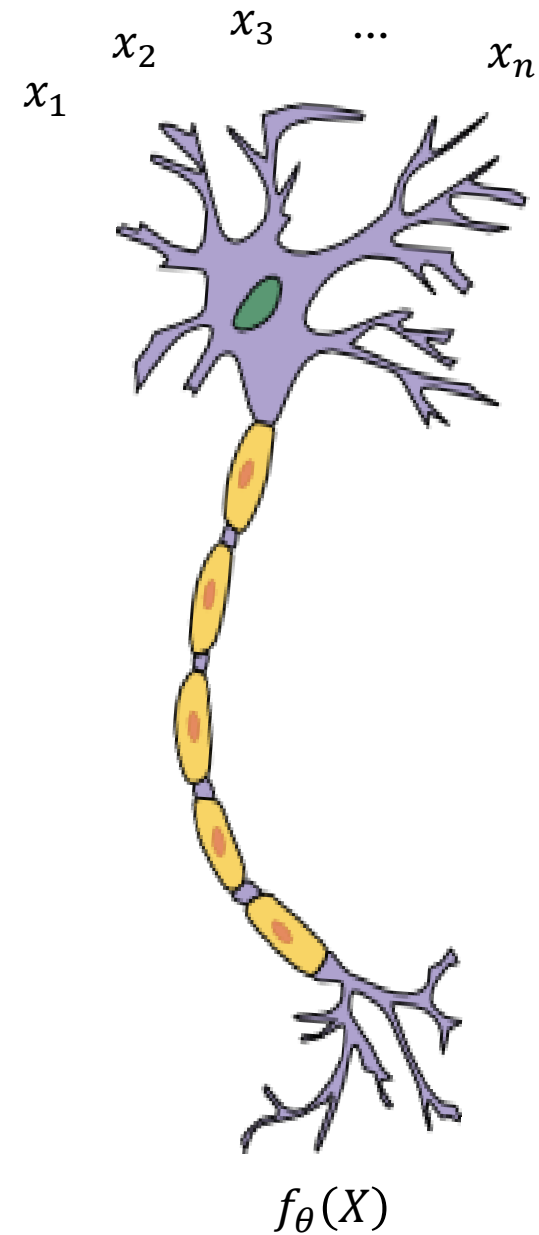
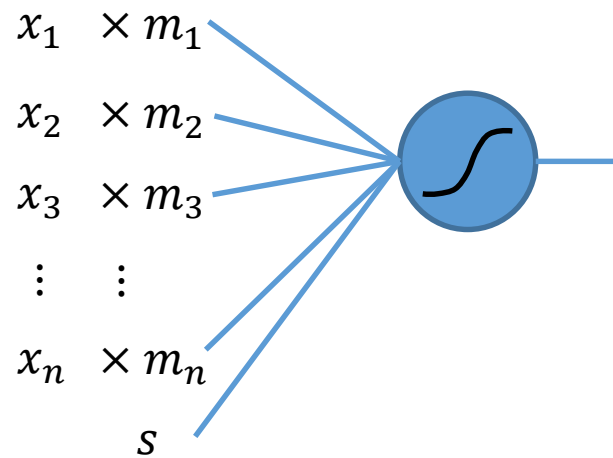


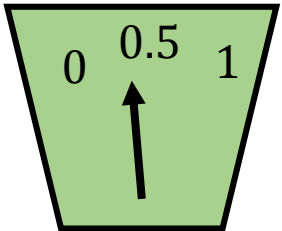
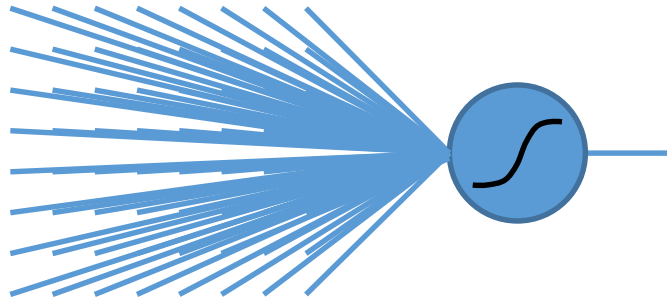
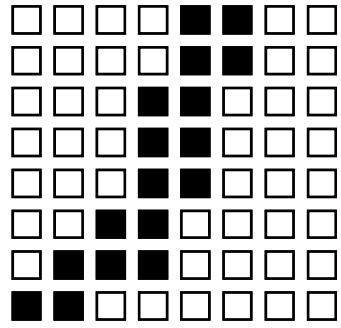
$$\varphi(x) = \frac{1}{1 + e^{-3x}}$$

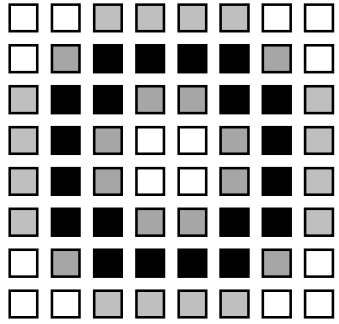
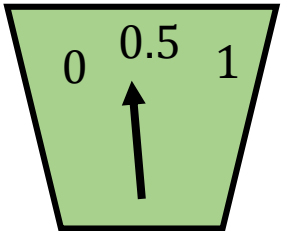
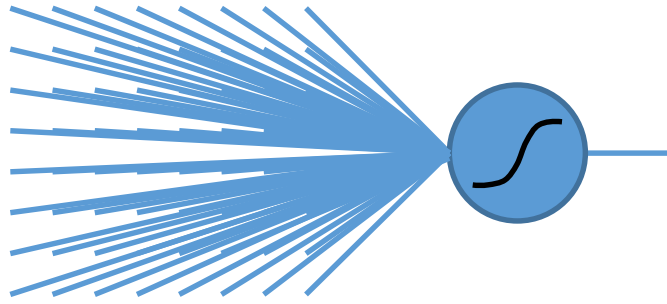
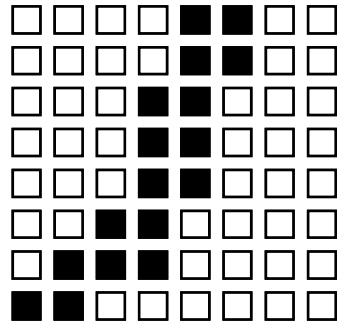
Neurones artificiels

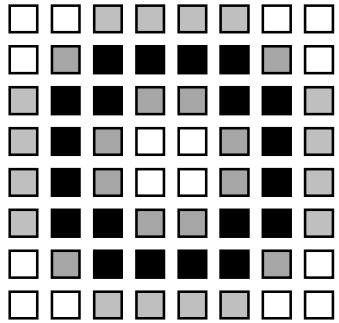
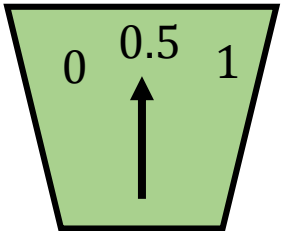
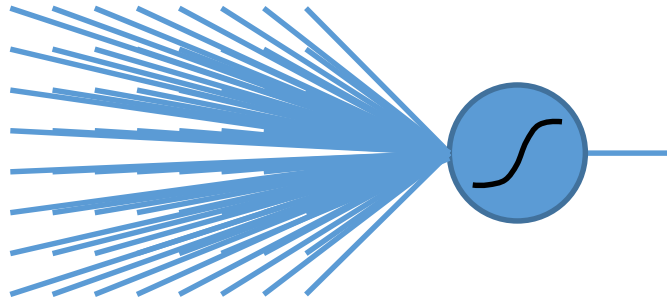
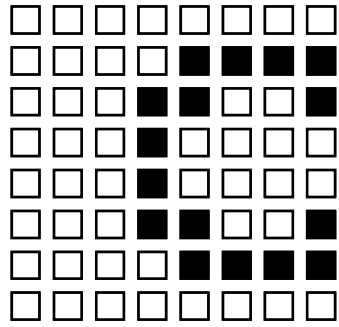
$$\theta = (M, s)$$

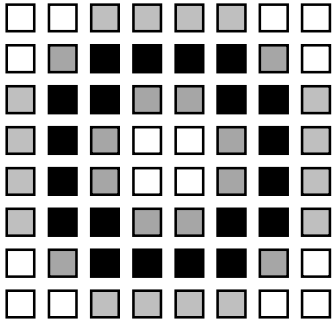
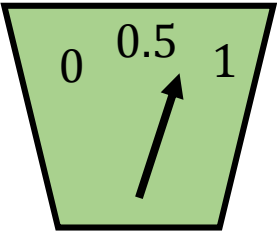
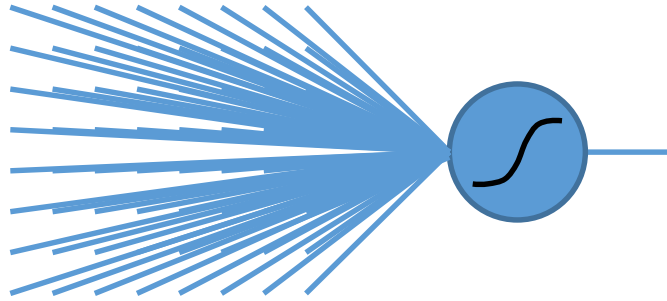
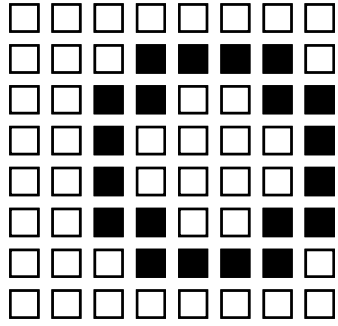
$$f_{\theta}(X) = \frac{1}{1 + e^{\langle M, X \rangle + s}}$$

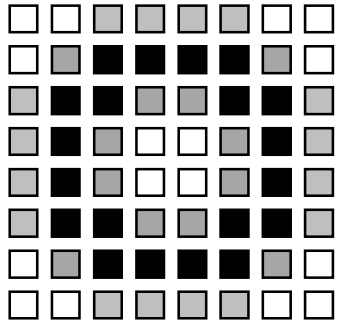
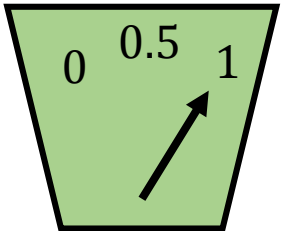
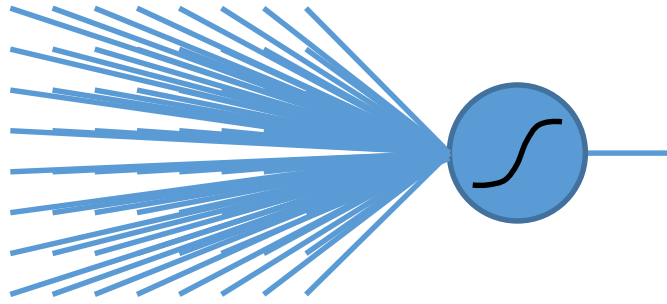
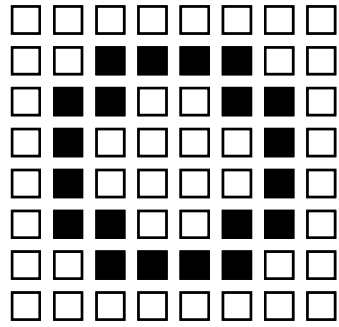




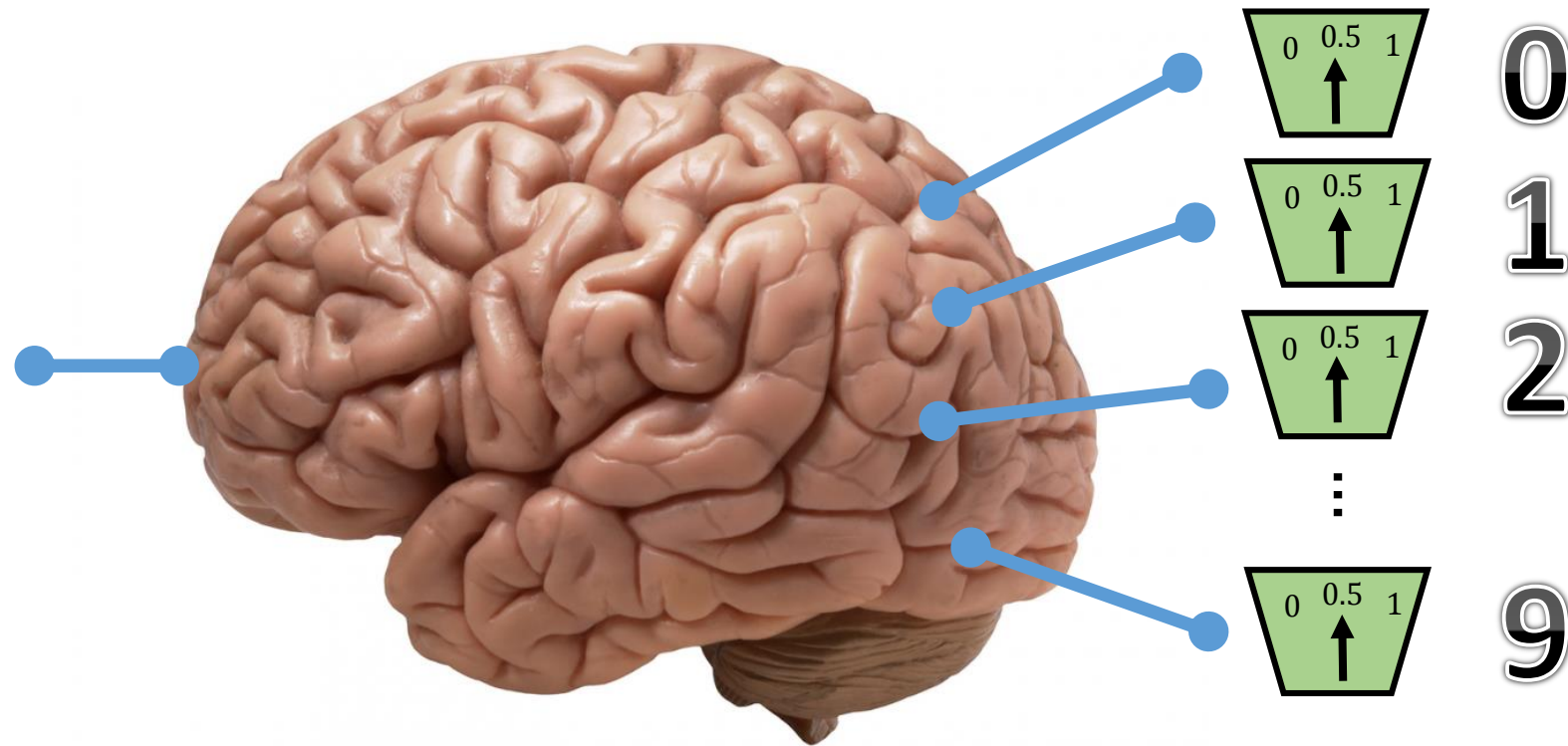




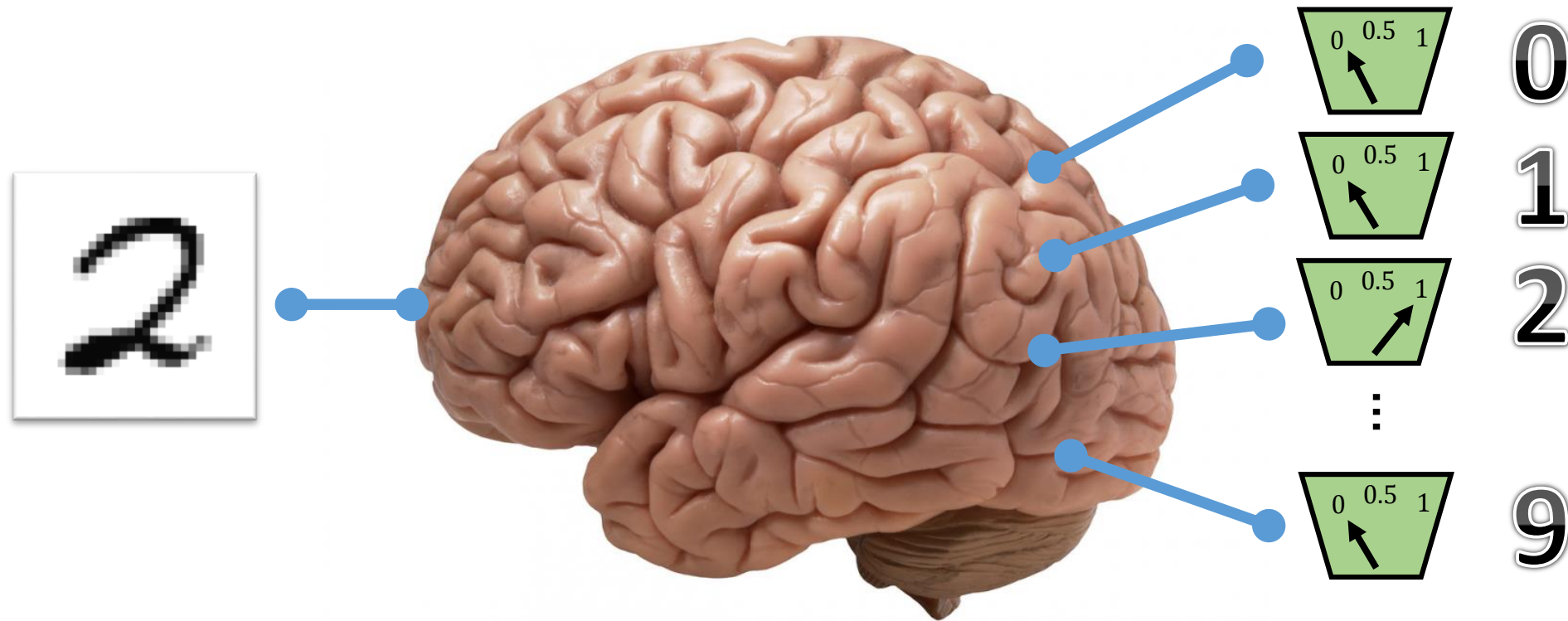




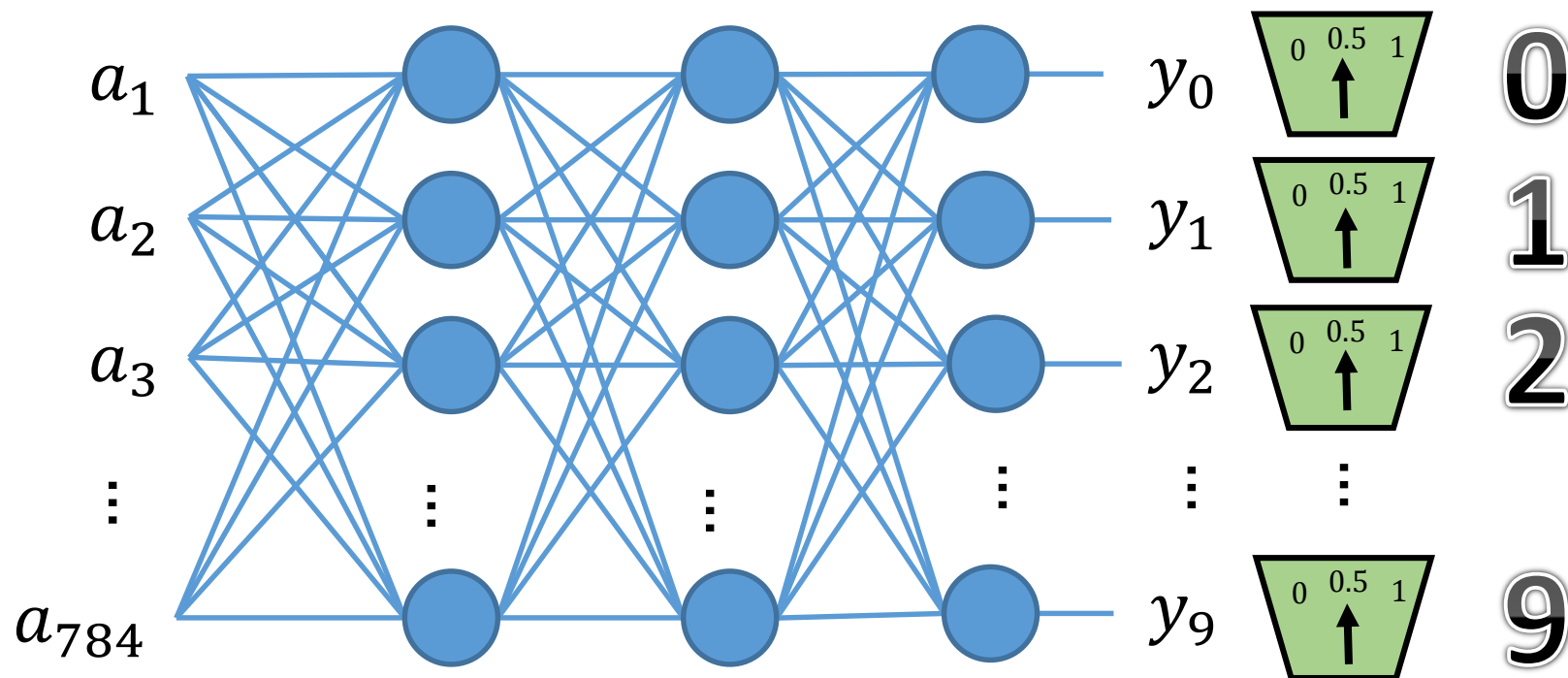
Réseau neuronal artificiel



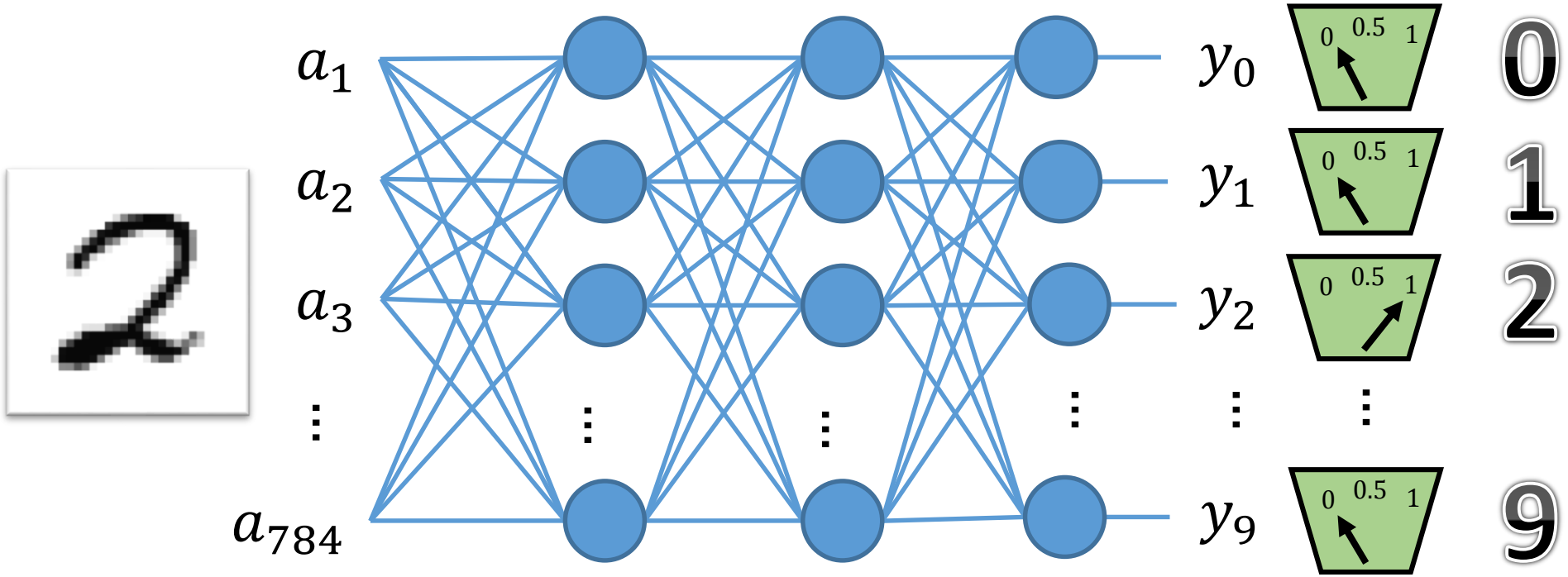
Réseau neuronal artificiel



Réseau neuronal artificiel



Réseau neuronal artificiel

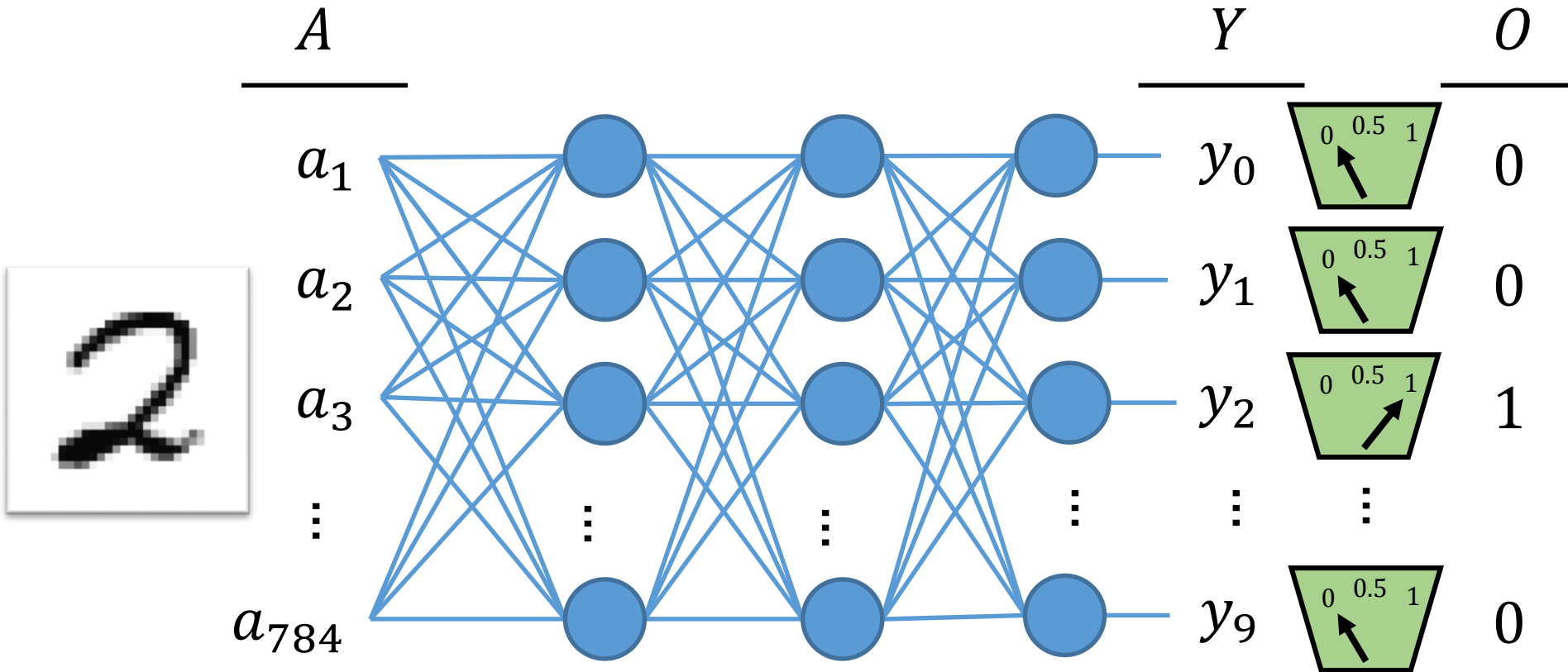


Geoffrey Hinton

Rétropropagation du gradient

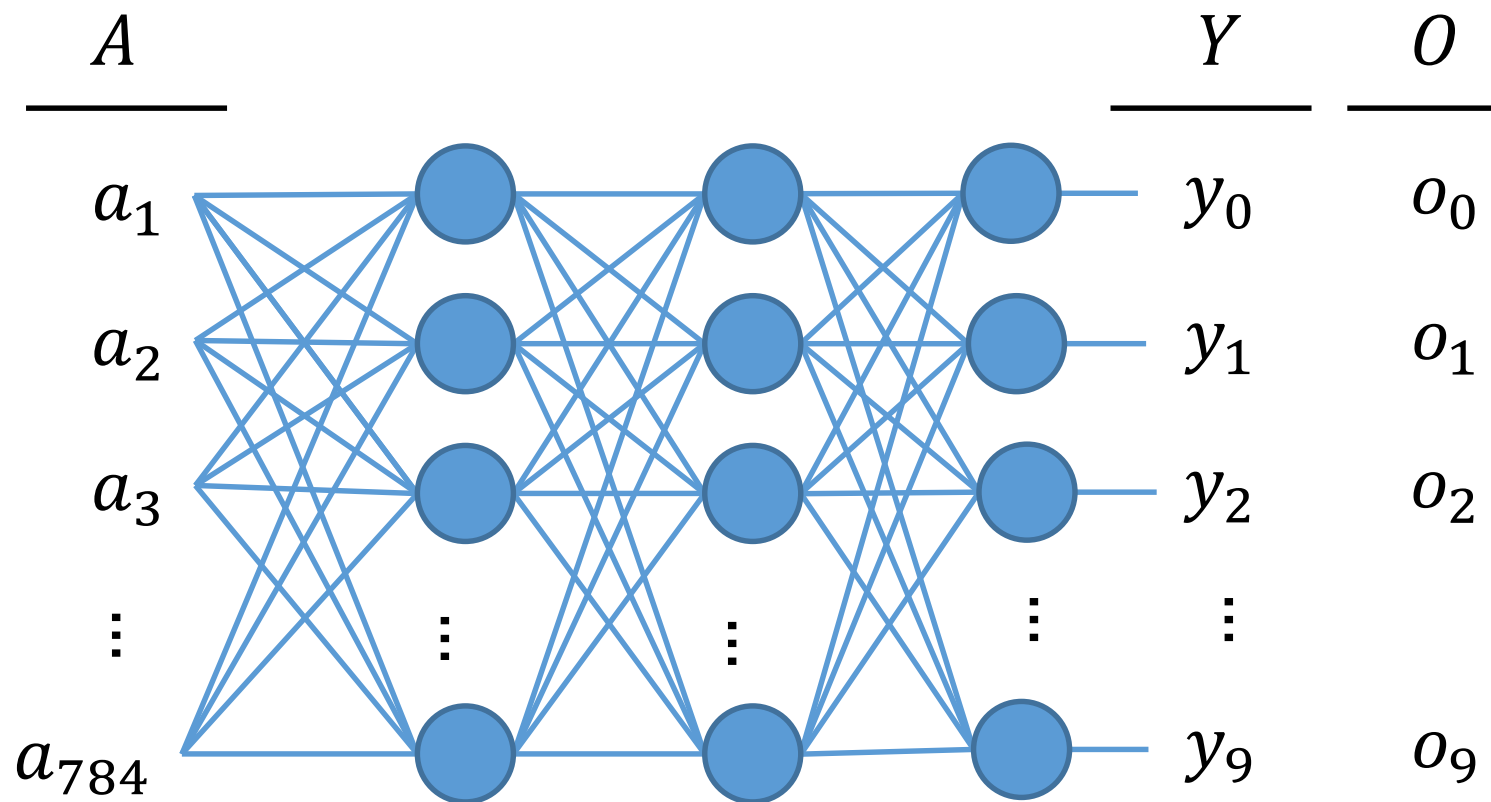


Réseau neuronal artificiel

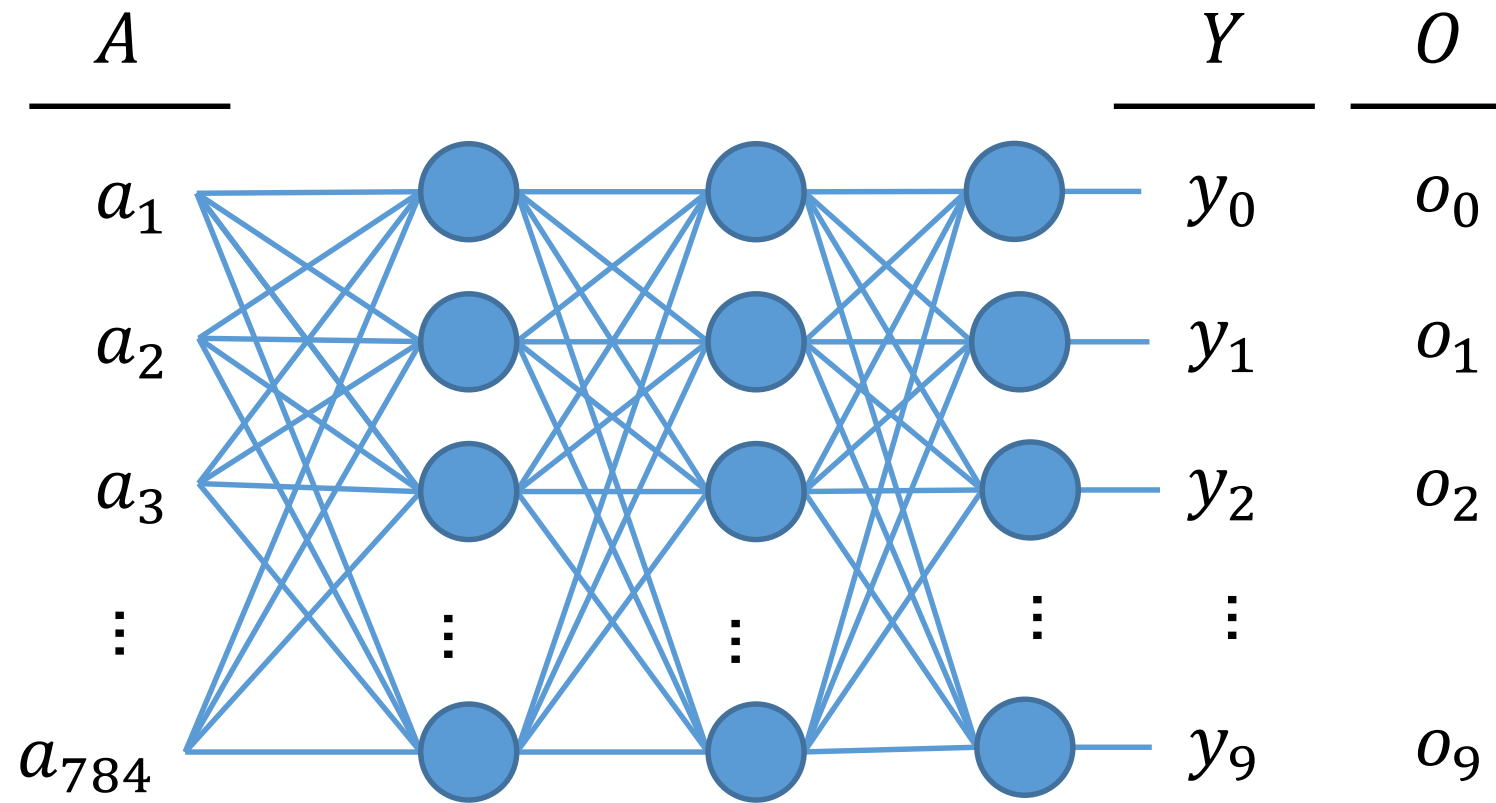


$$E_A = D(Y, O)$$

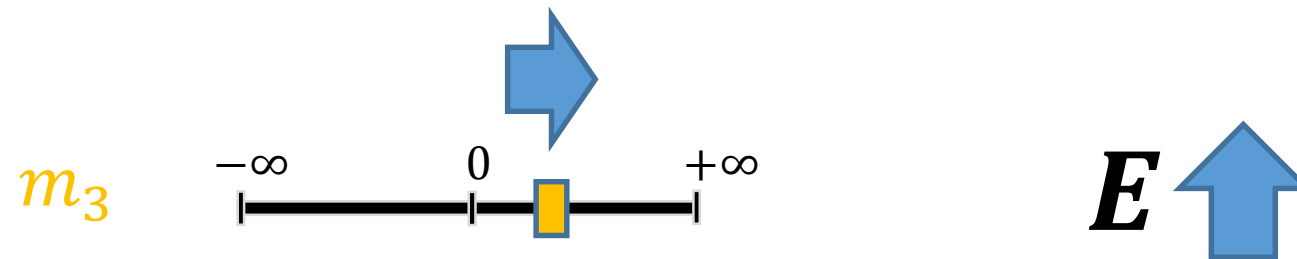
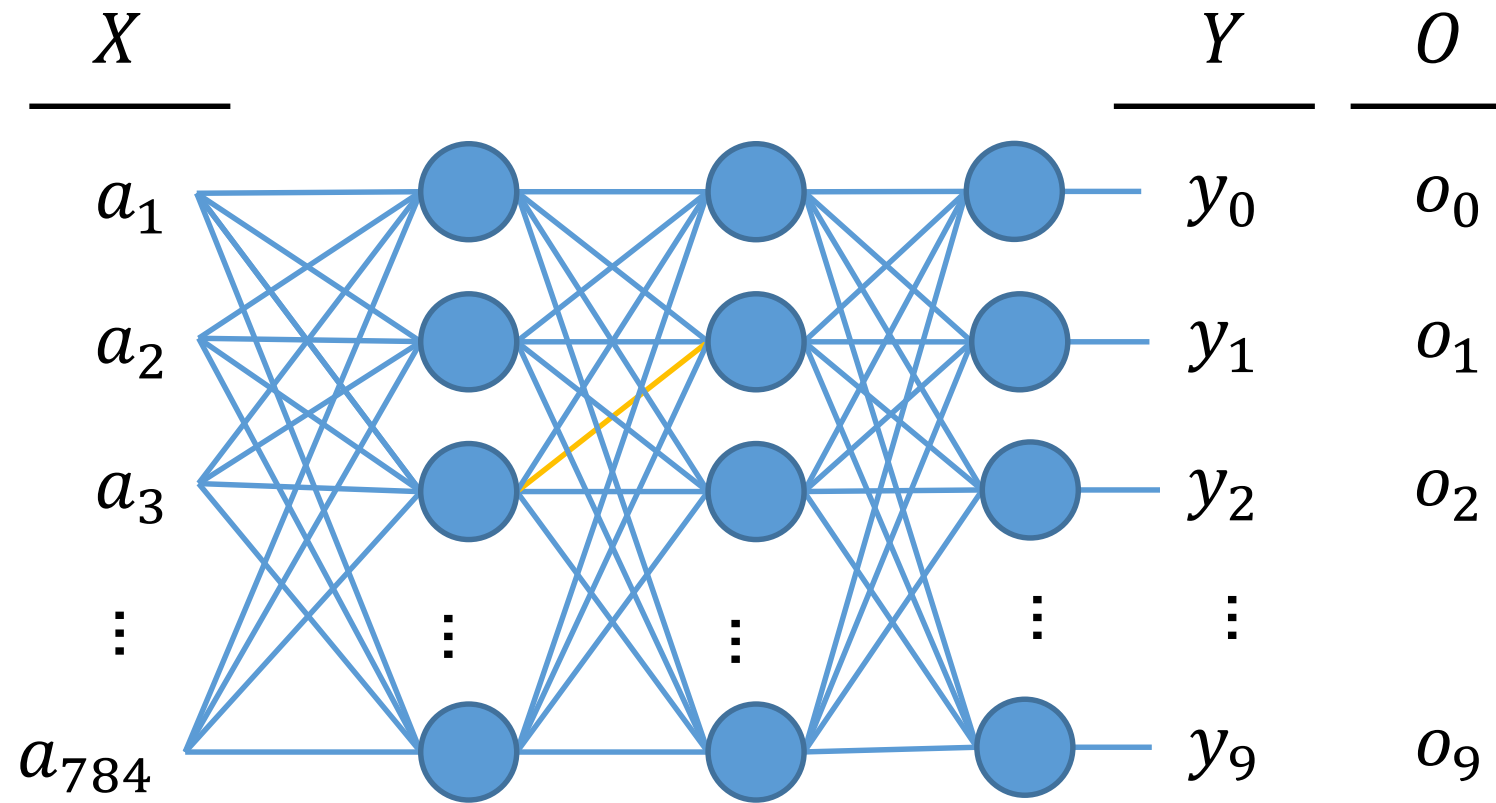
Réseau neuronal artificiel

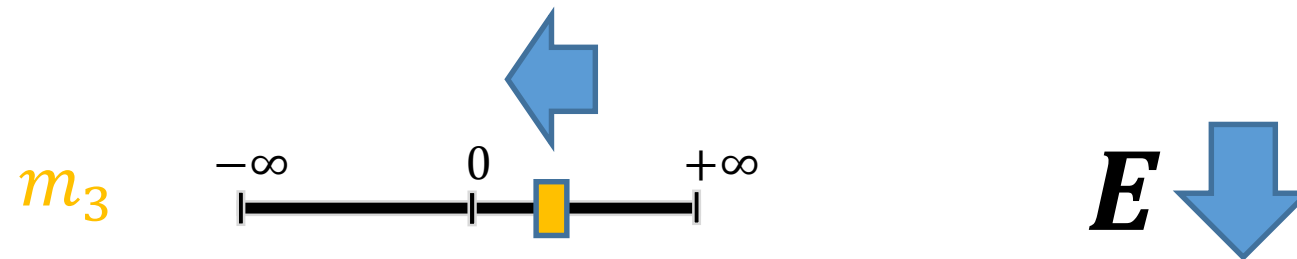
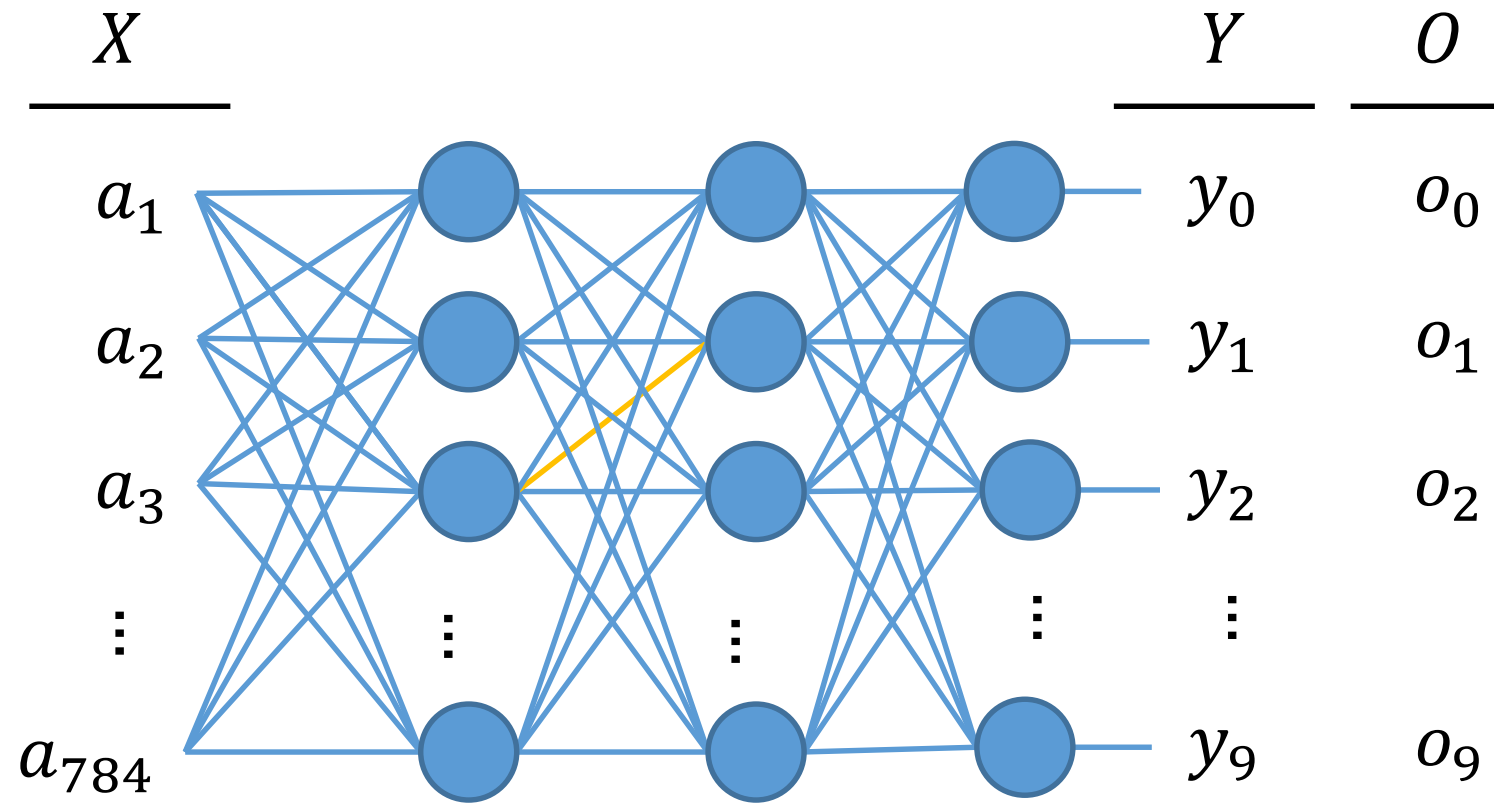


$$\mathbf{E} = \mathbf{E}_{A_1} + \mathbf{E}_{A_2} + \mathbf{E}_{A_3} + \dots + \mathbf{E}_{A_{60000}}$$



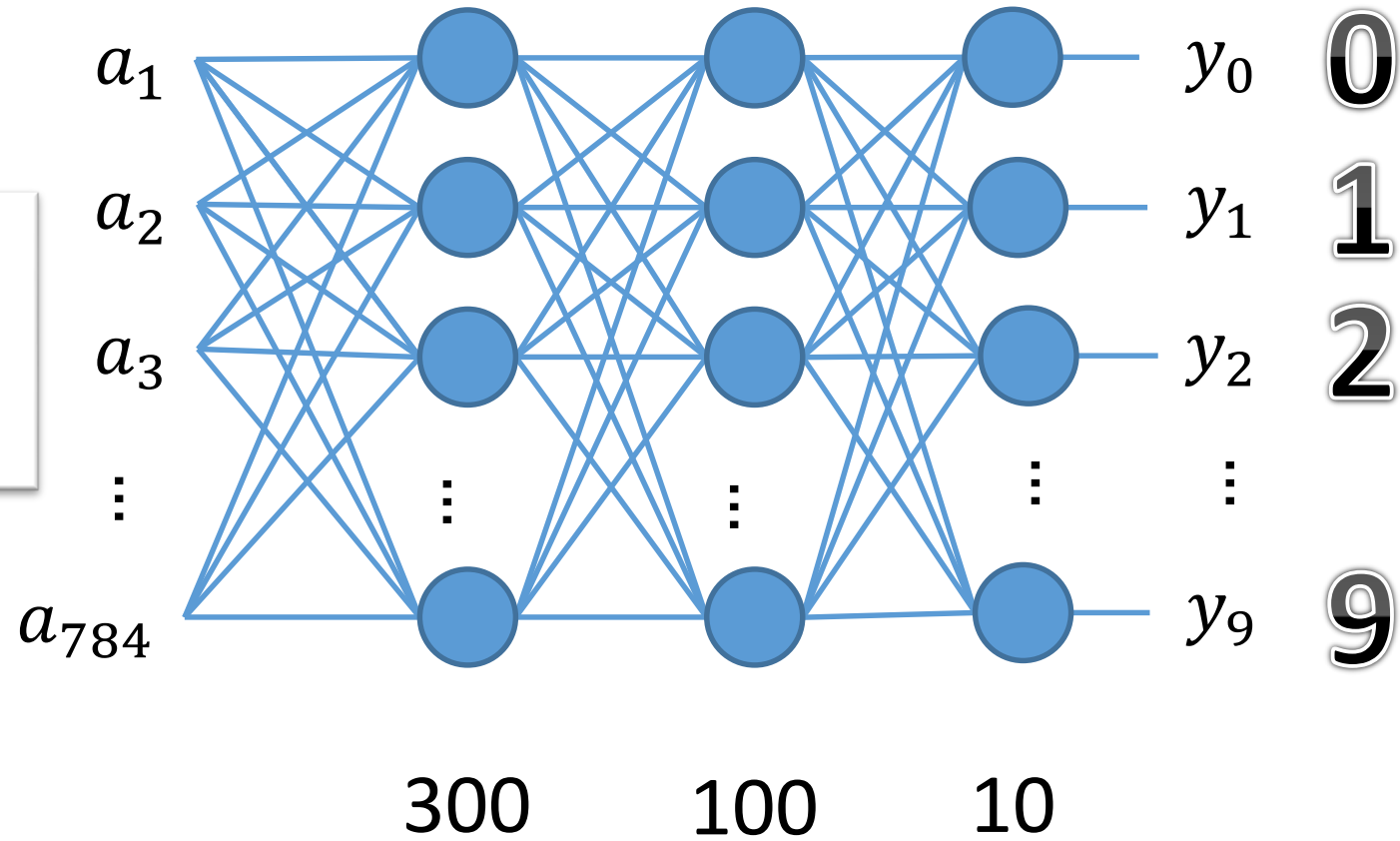
E





410 neurones

266 200 connexions

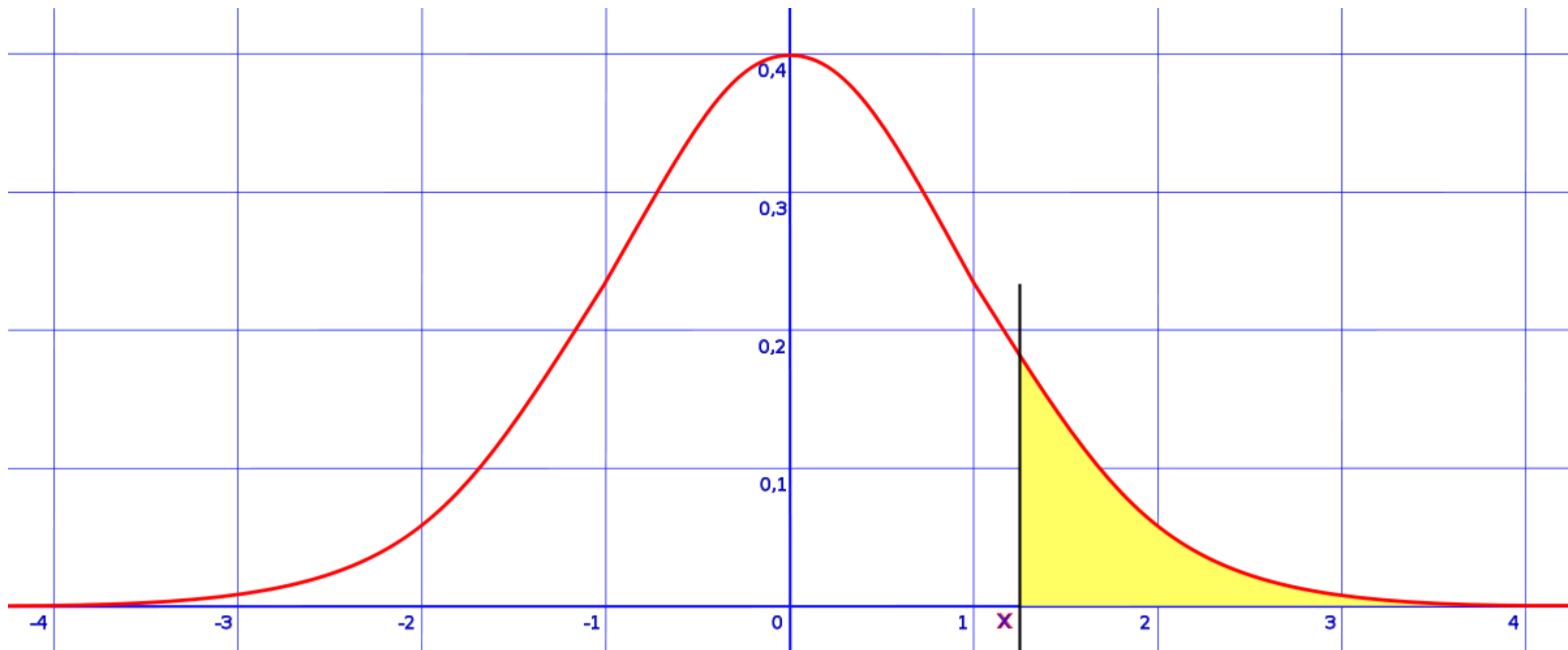


RN à 3 niveaux (300,100)

0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
4 4 4 4 4 4 4 4 4 4 4 4 4 4 4
5 5 5 5 5 5 5 5 5 5 5 5 5 5 5
6 6 6 6 6 6 6 6 6 6 6 6 6 6 6
7 7 7 7 7 7 7 7 7 7 7 7 7 7 7
8 8 8 8 8 8 8 8 8 8 8 8 8 8 8
9 9 9 9 9 9 9 9 9 9 9 9 9 9 9

$$E = 3.1\%$$

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%

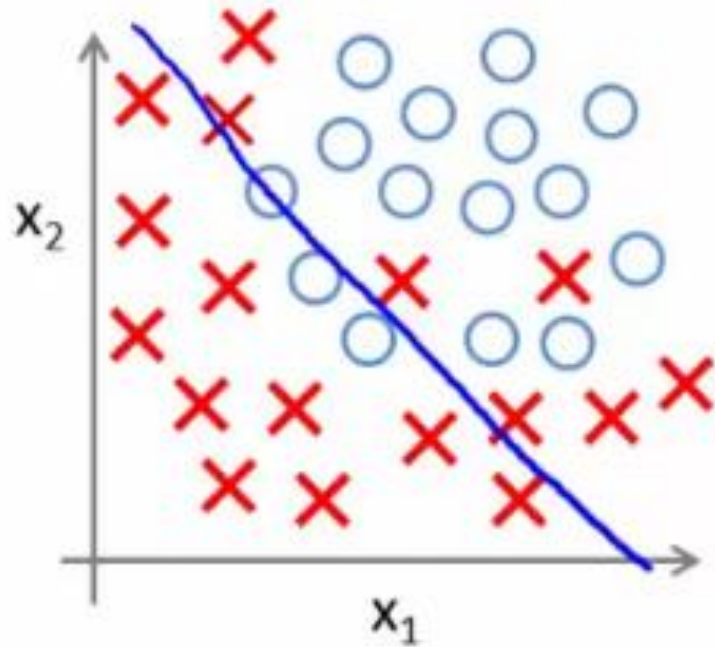
15%

Apprentissage

Test

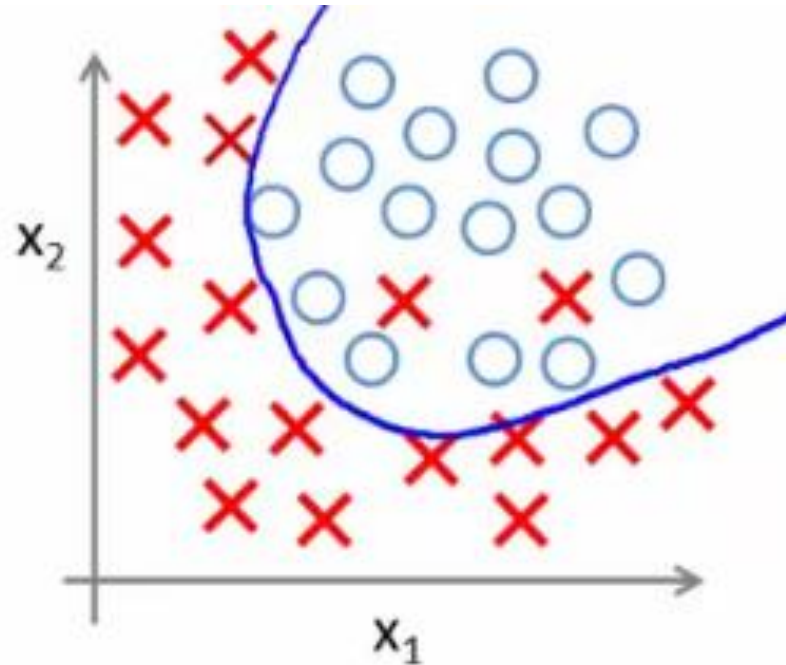


Sur-apprentissage?

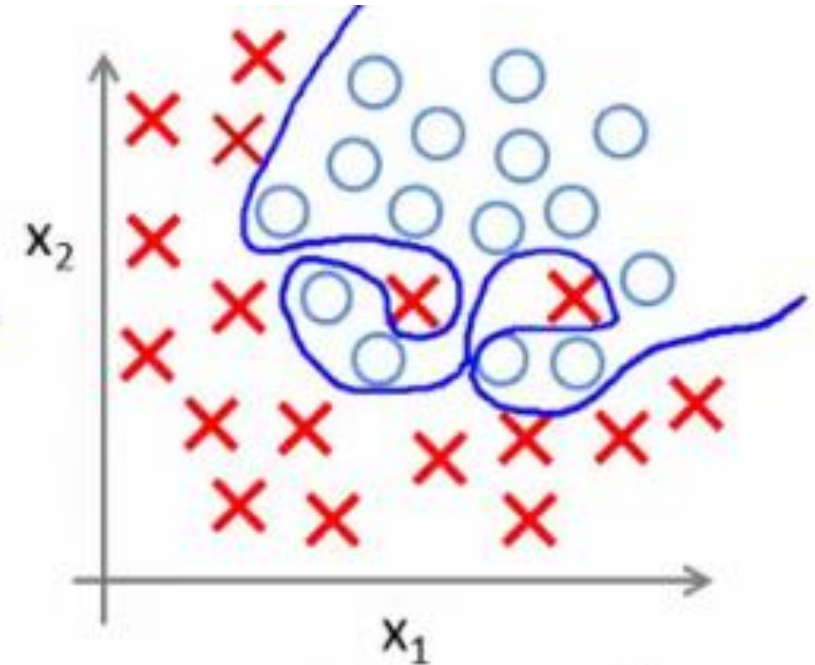


$$h_{\theta}(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$$

(g = sigmoid function)



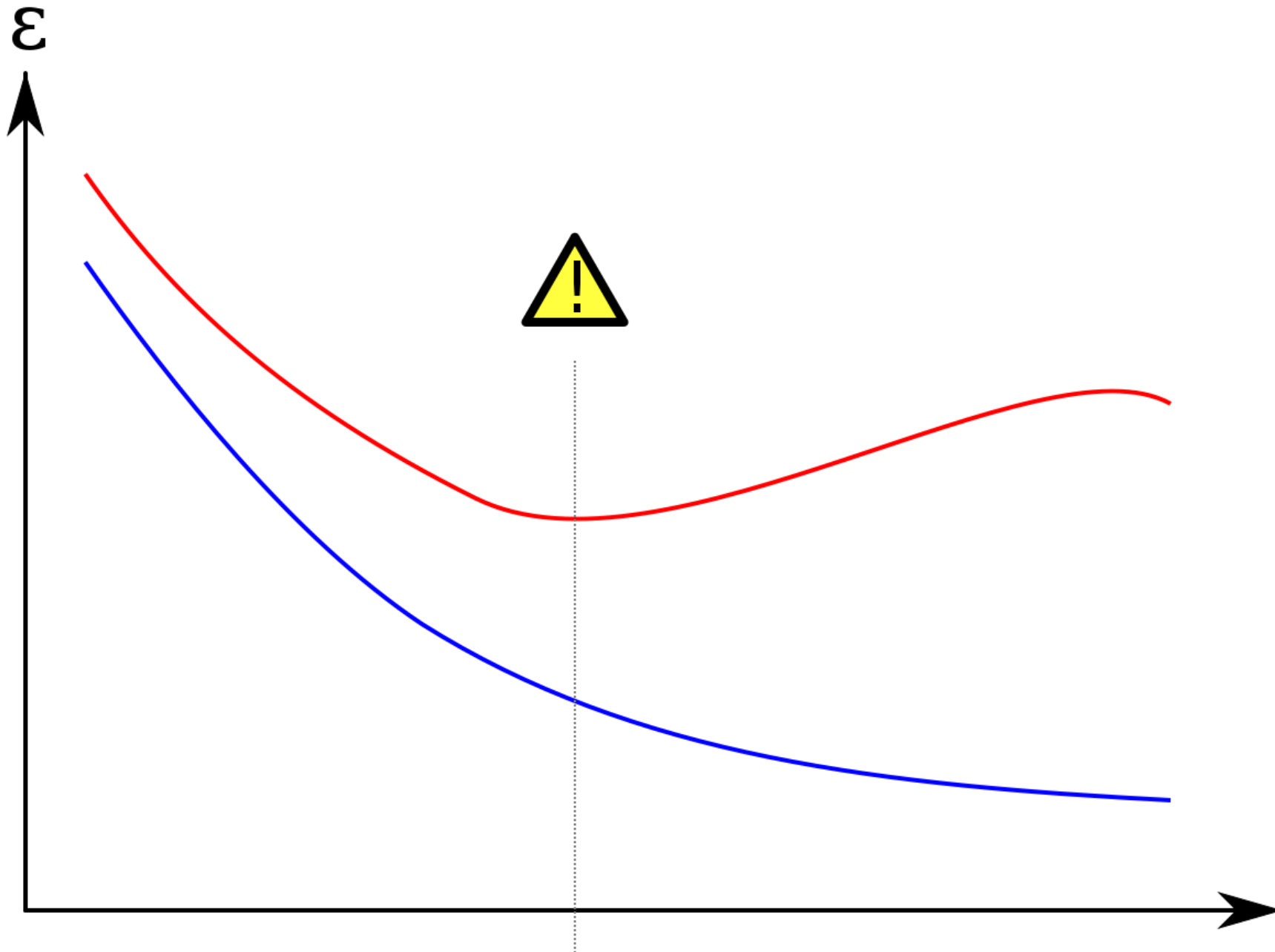
$$g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2 + \theta_5 x_1 x_2)$$



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

Paramètres et hyper paramètres

- Capacité
 - Nombre de paramètres
 - Degré
 - Complexité
- Temps d'entraînement
- Tolérance aux erreurs
- Régularisation



70%

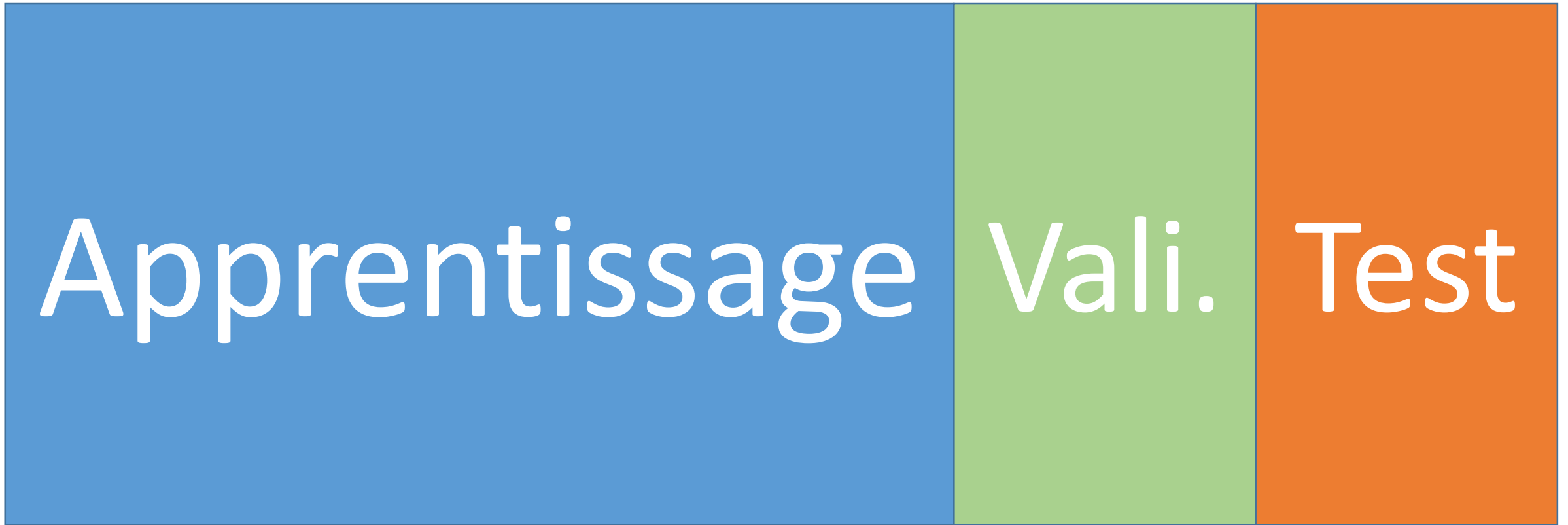
15%

15%

Apprentissage

Vali.

Test

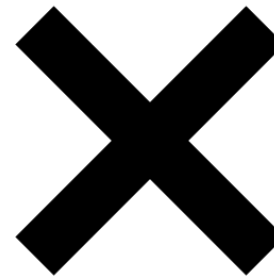
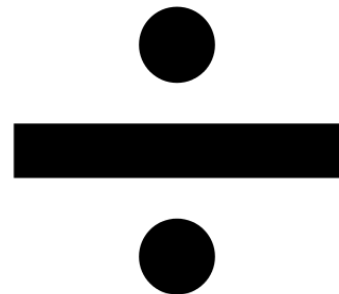
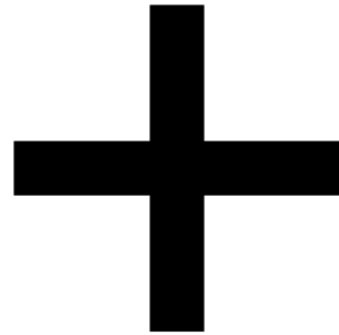


Puissance, Capacité, Force brute...



FLOPS

FLoating-point Operations Per Second
Opérations en virgule flottante par seconde





Intel Core i7
90 *giga*-FLOPS
\$400

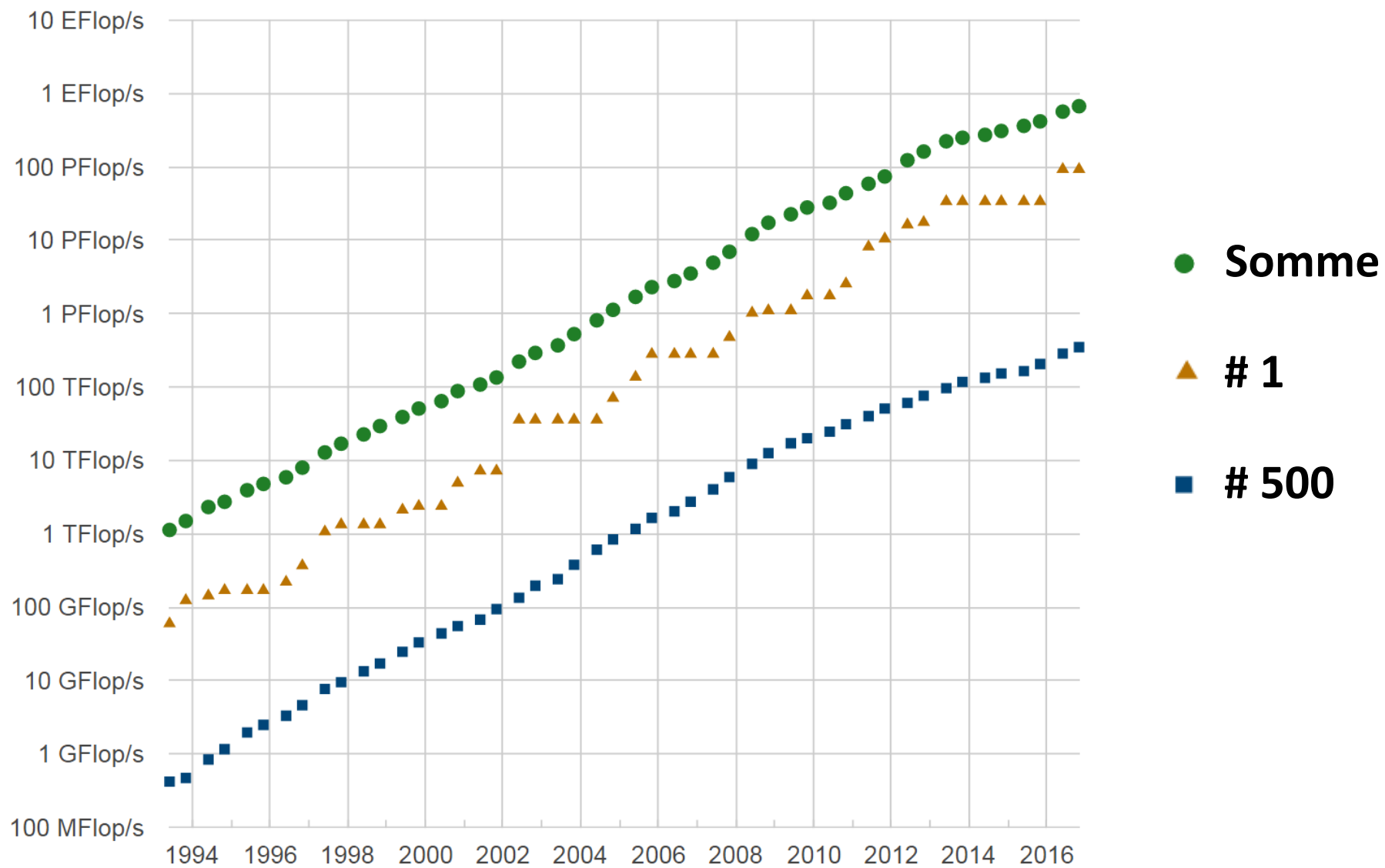


Nvidia GTX 1080 Ti
10 *tera*-FLOPS
\$900



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

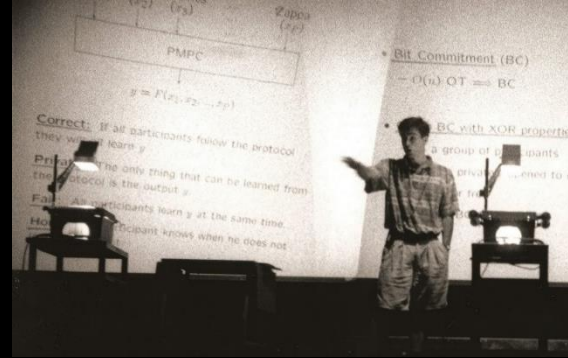
Top 500!



2016



1995



1 seconde



1 mois

1 nuit



2.5 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