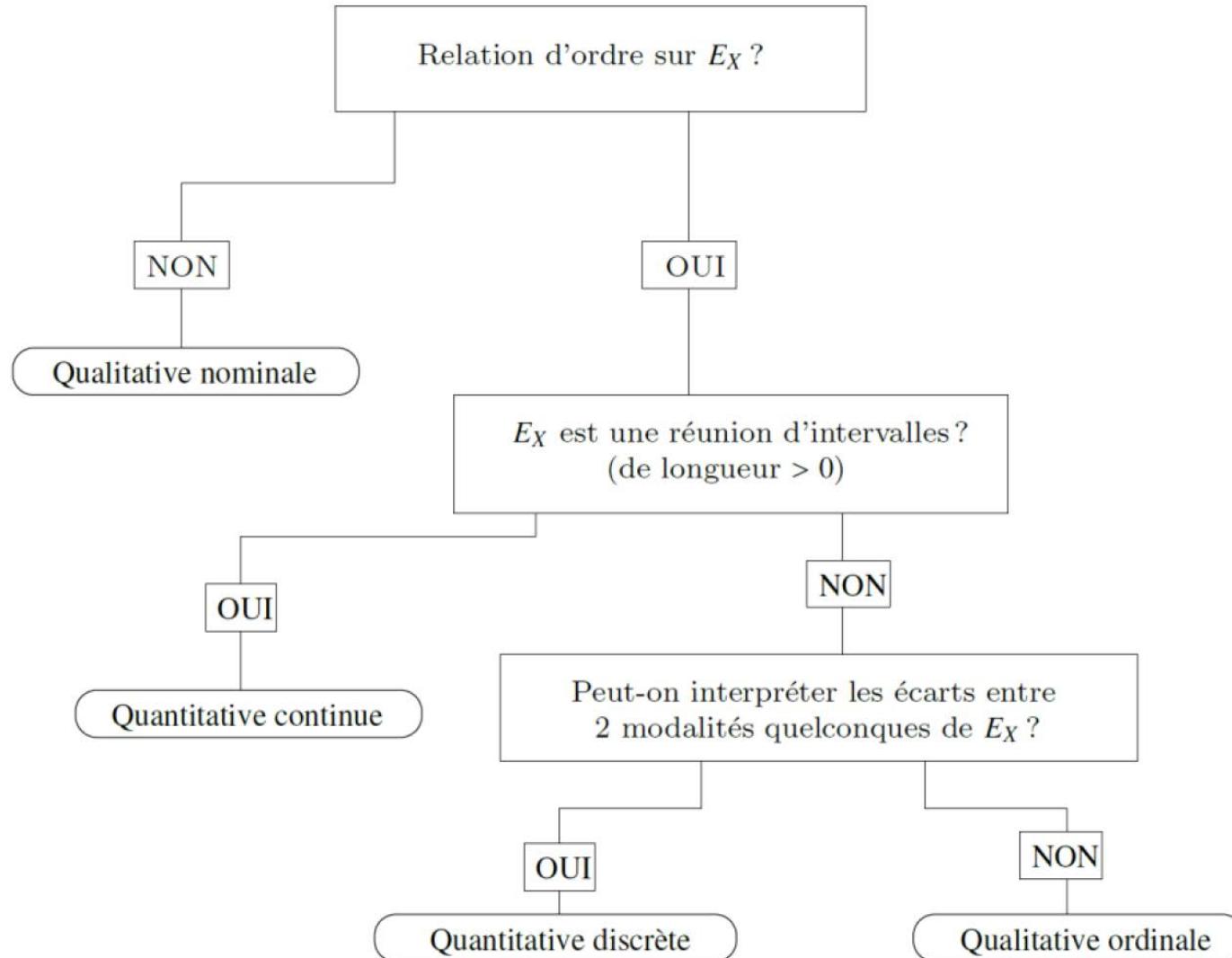
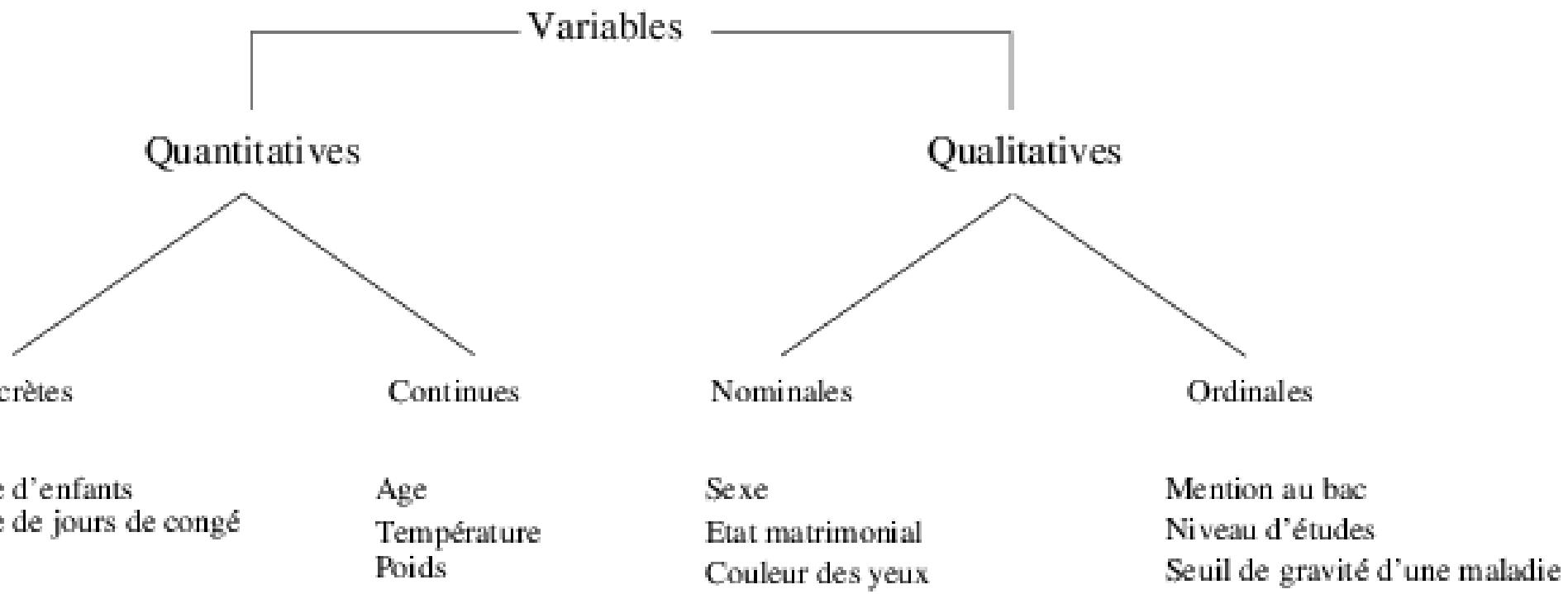


Module 1 : Structuration des variables

- le type d'une variable X se détermine par rapport à l'ensemble E_X des modalités susceptibles d'être observées.





Variables qualitatives	sexe, situation et matgras
Variables ordinaires	viande, poisson, fruit_crus, fruit_legume_cuits et chocolat
Variables quantitatives	discrètes the et cafe
Variables quantitatives	continues taille, poids et age

Variables qualitatives

n°ind.	X	modalité	effectif
1	x_1	1	n_1
:	:	:	:
i	x_i	m	n_m
:	:	:	:
n	x_n	M	n_M

couleur des yeux	bleu	vert	brun	pers	noir
fréquences absolues	66	34	80	15	5
fréquences relatives	33%	17%	40%	7.5%	2.5%

Couleur des yeux de 10 étudiants

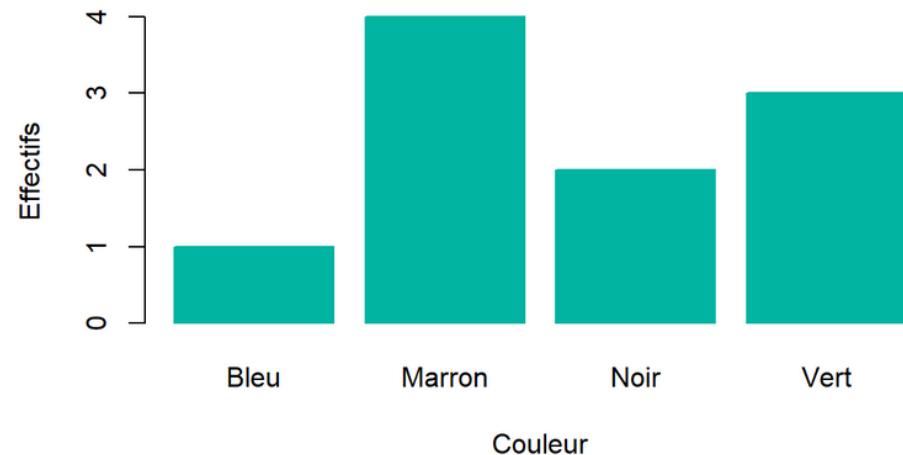


diagramme en barres

Femmes

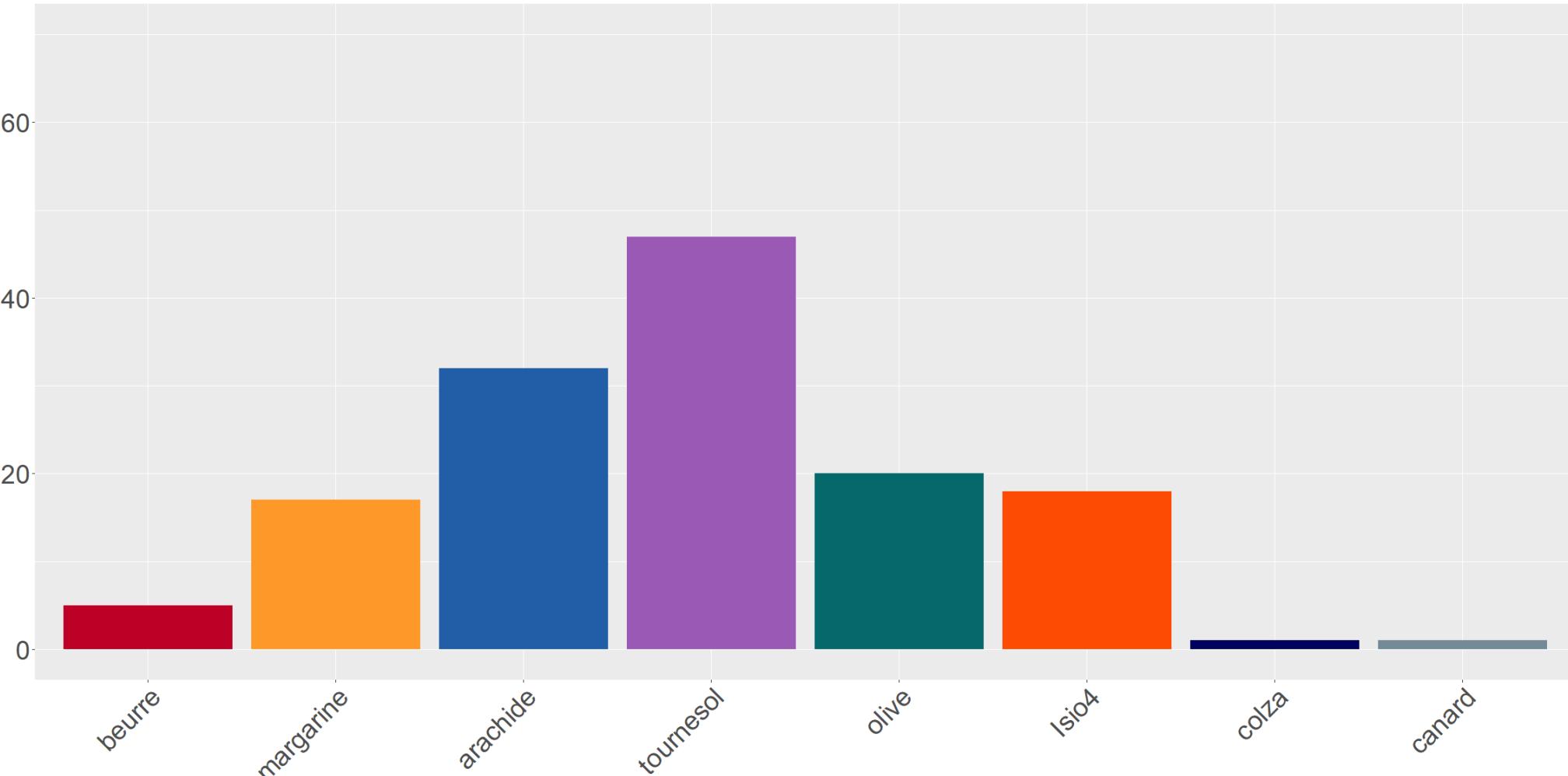
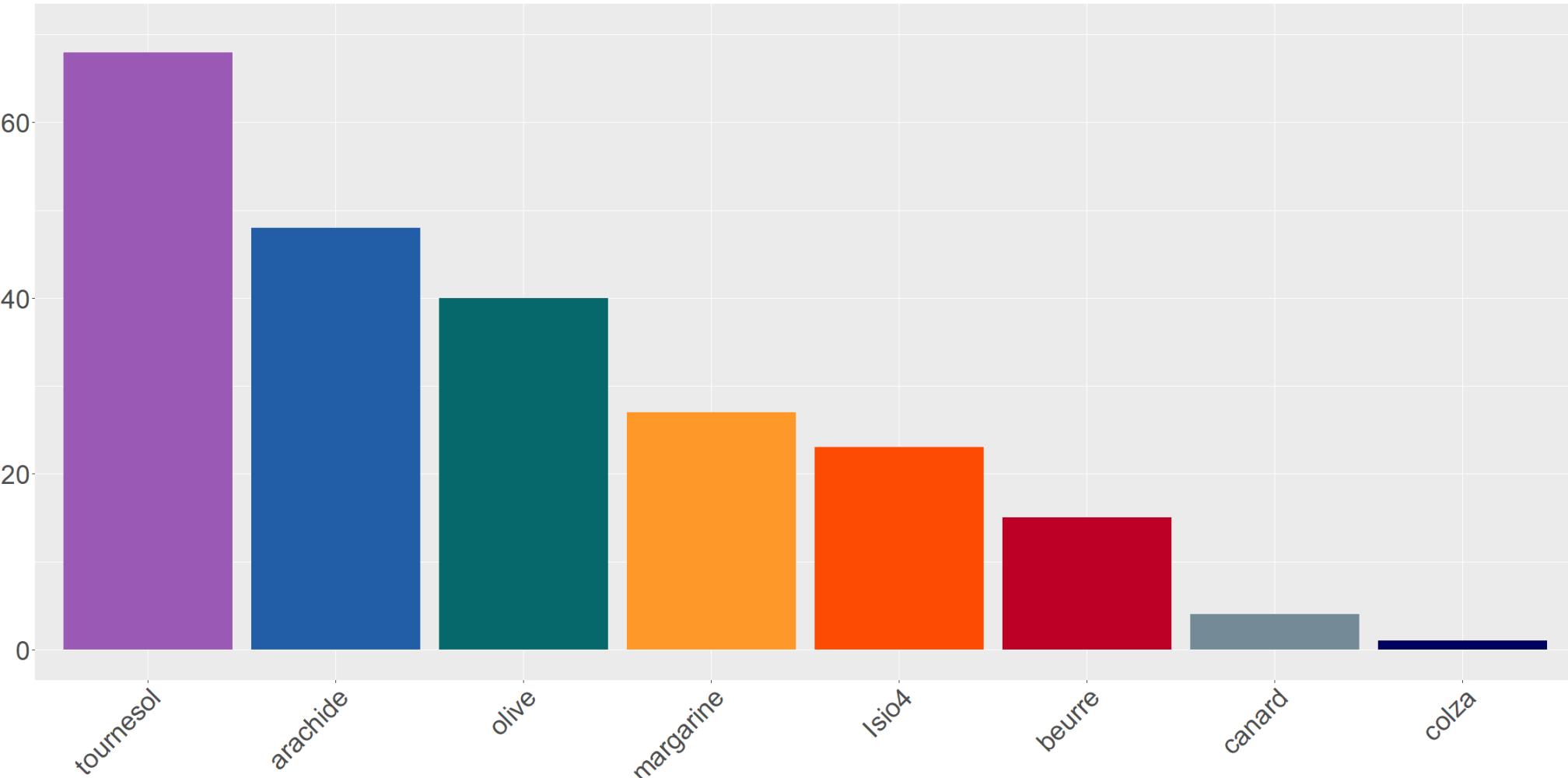


diagramme en barres

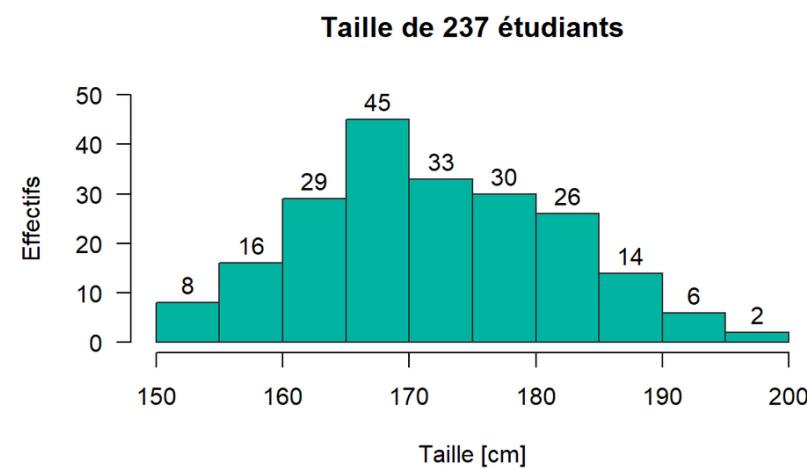
Global sans distinction sexe



Variable quantitative

n°ind.	X	intervalle	effectif
1	x_1	1	n_1
\vdots	\vdots	\vdots	\vdots
i	x_i	$[a, b[$	$n_{[a,b[}$
\vdots	\vdots	\vdots	\vdots
n	x_n	M	n_M

- c_j^- la borne inférieure de la classe j ,
- c_j^+ la borne supérieure de la classe j ,
- $c_j = (c_j^+ + c_j^-)/2$ le centre de la classe j ,
- $a_j = c_j^+ - c_j^-$ l'amplitude de la classe j ,
- n_j l'effectif de la classe j ,
- N_j l'effectif cumulé de la classe j ,
- f_j la fréquence de la classe j ,
- F_j la fréquence cumulée de la classe j .

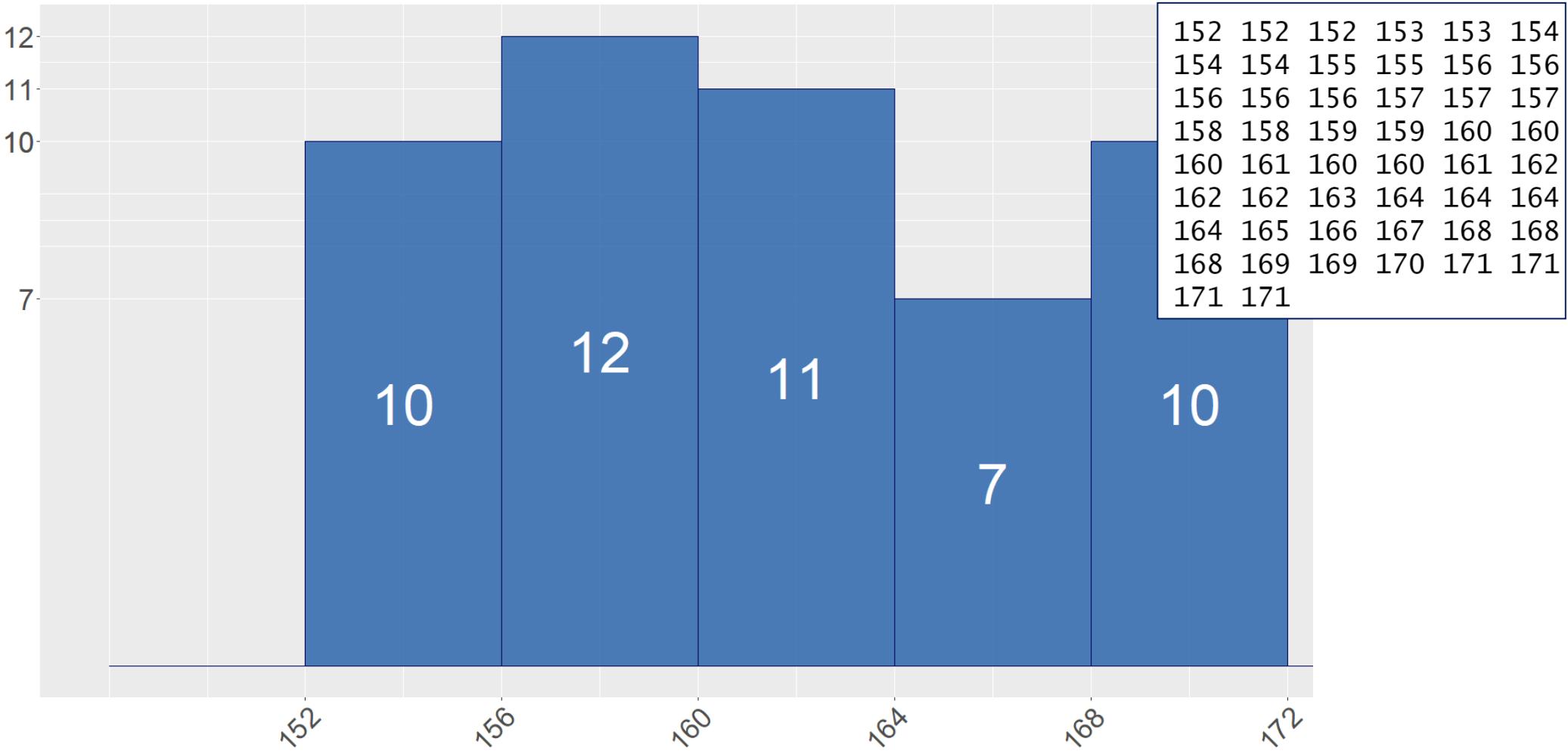


- La règle de Sturge : $J = 1 + (3.3 \log_{10}(n))$.
- La règle de Yule : $J = 2.5 \sqrt[4]{n}$.

histogramme

	densite.effectif	densite.frequence	effectifs.cumule	frequence.cumulee
[152,156)	10	0.20	10	0.20
[156,160)	12	0.24	22	0.44
[160,164)	11	0.22	33	0.66
[164,168)	7	0.14	40	0.80
[168,172]	10	0.20	50	1.00

Histogramme des effectifs

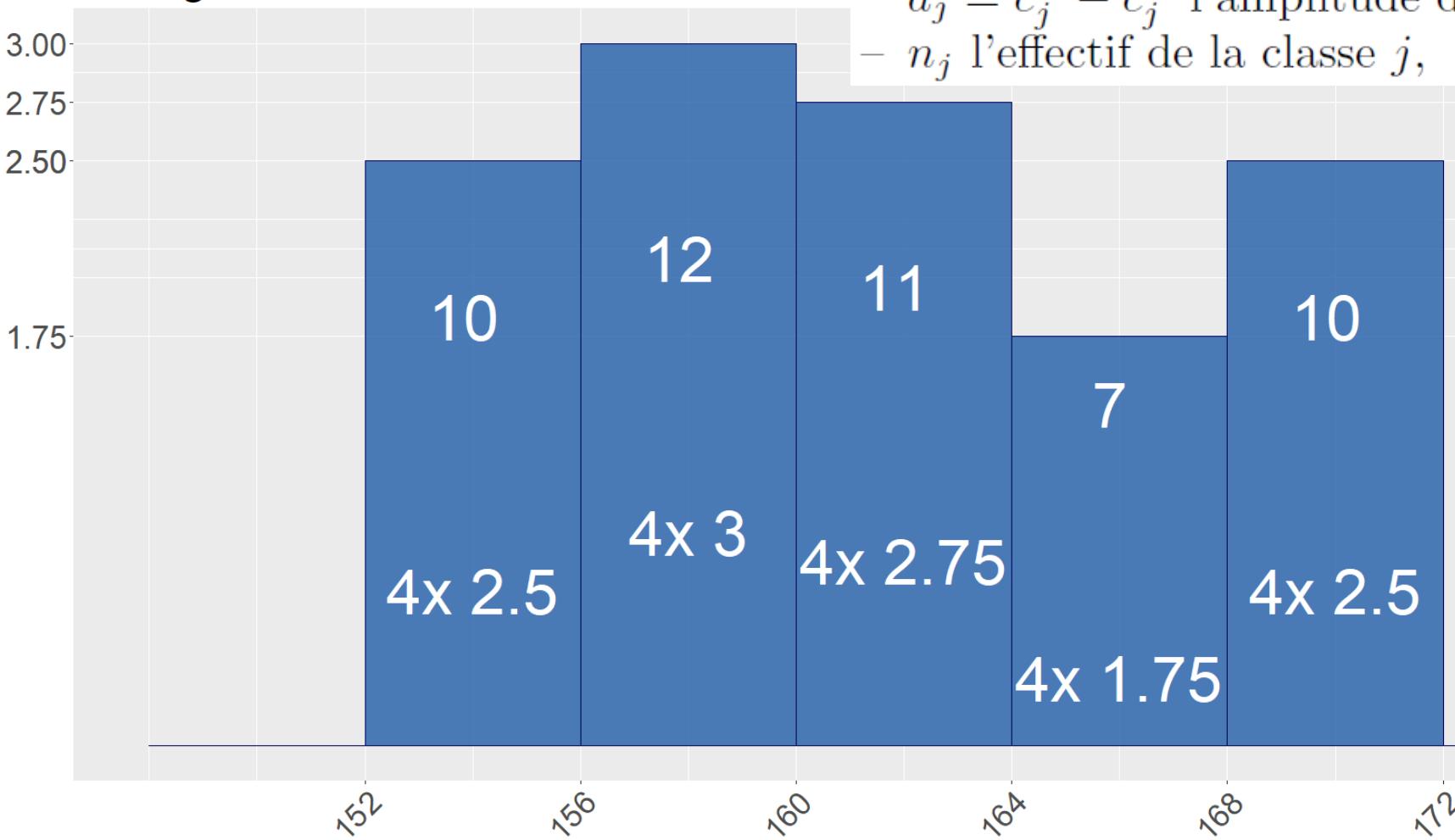


histogramme

$$h_j = \frac{n_j}{a_j}$$

	densite.effectif	densite.frequence	effectifs.cumule	frequence.cumulee
[152,156)	10	0.20	10	0.20
[156,160)	12	0.24	22	0.44
[160,164)	11	0.22	33	0.66
[164,168)	7	0.14	40	0.80
[168,172]	10	0.20	50	1.00

Histogramme de densité d'effectifs



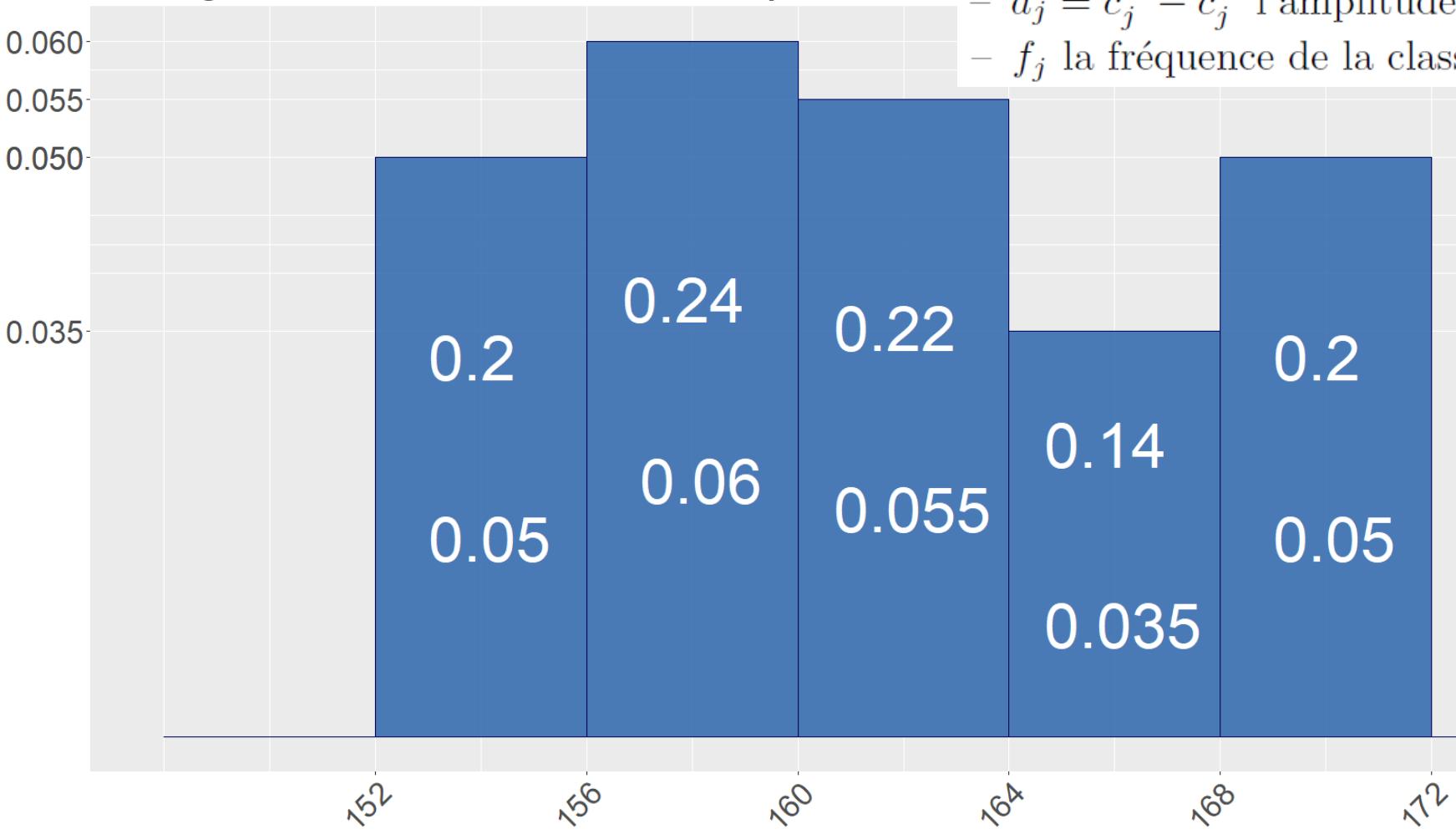
- $a_j = c_j^+ - c_j^-$ l'amplitude de la classe j ,
- n_j l'effectif de la classe j ,

histogramme

$$d_j = \frac{f_j}{a_j}$$

	densite.effectif	densite.frequence	effectifs.cumule	frequence.cumulee
[152,156)	10	0.20	10	0.20
[156,160)	12	0.24	22	0.44
[160,164)	11	0.22	33	0.66
[164,168)	7	0.14	40	0.80
[168,172]	10	0.20	50	1.00

Histogramme de densité des fréquences

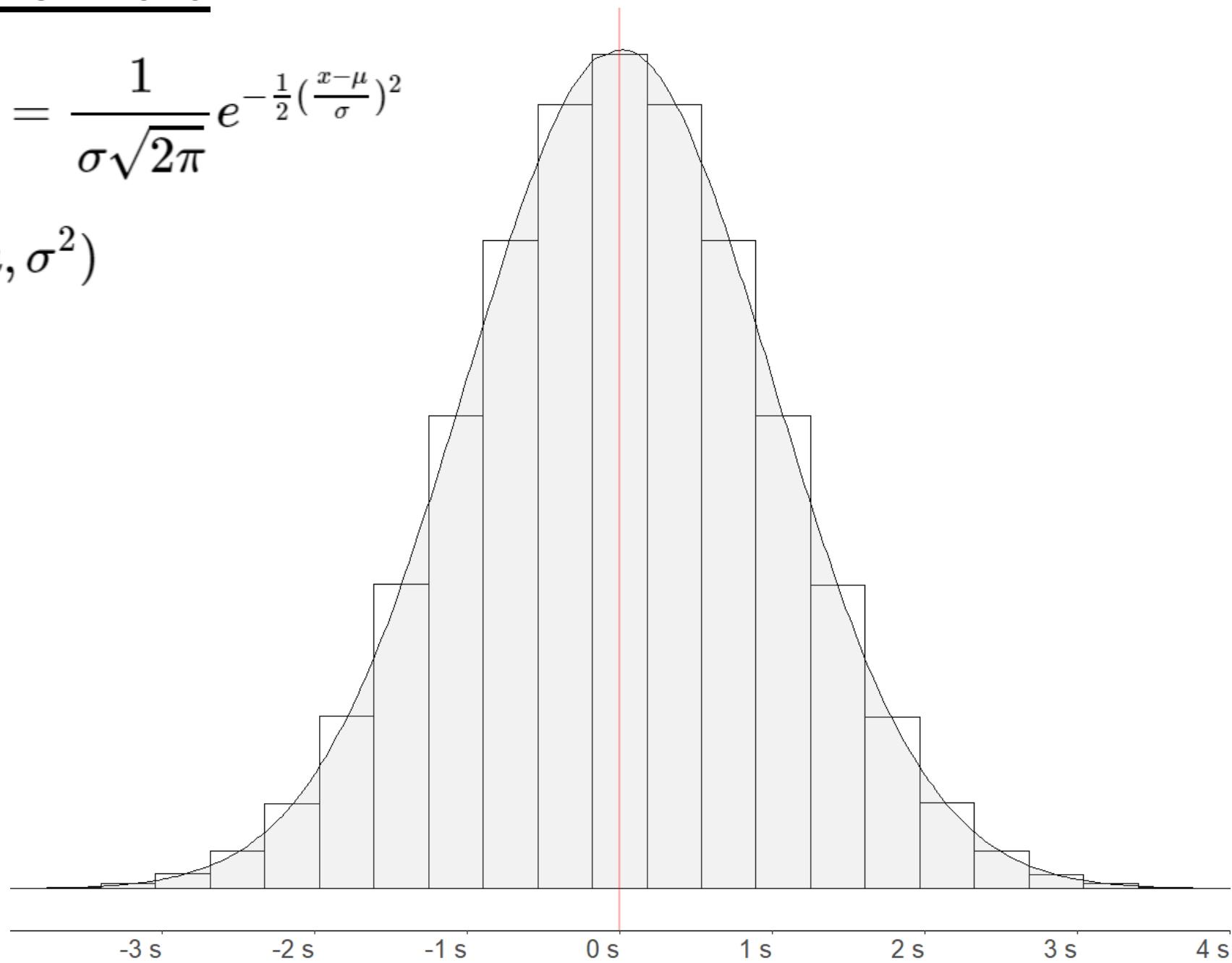


- $a_j = c_j^+ - c_j^-$ l'amplitude de la classe j ,
- f_j la fréquence de la classe j ,

Loi Normale

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

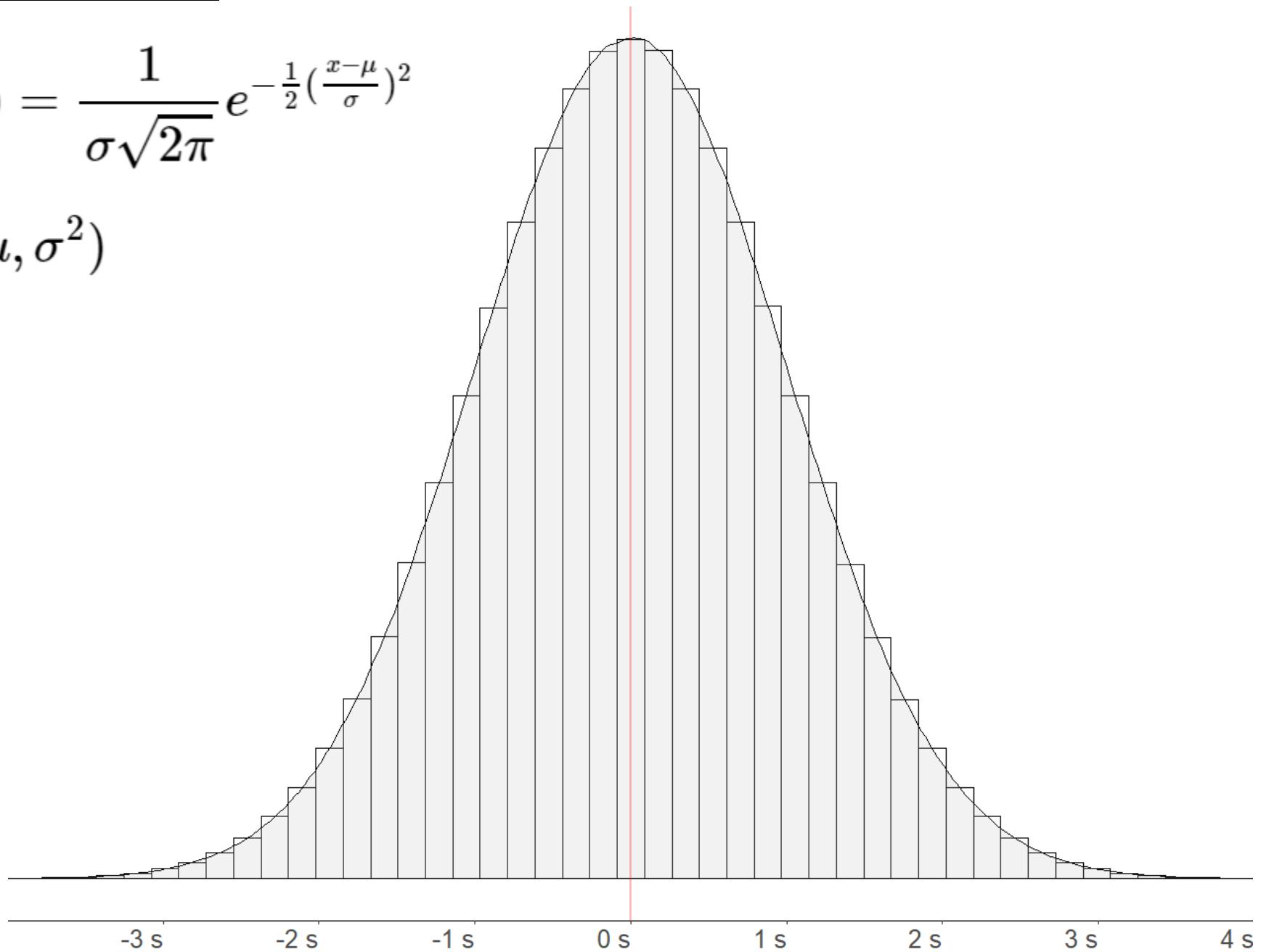
$$\mathcal{N}(\mu, \sigma^2)$$



Loi Normale

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

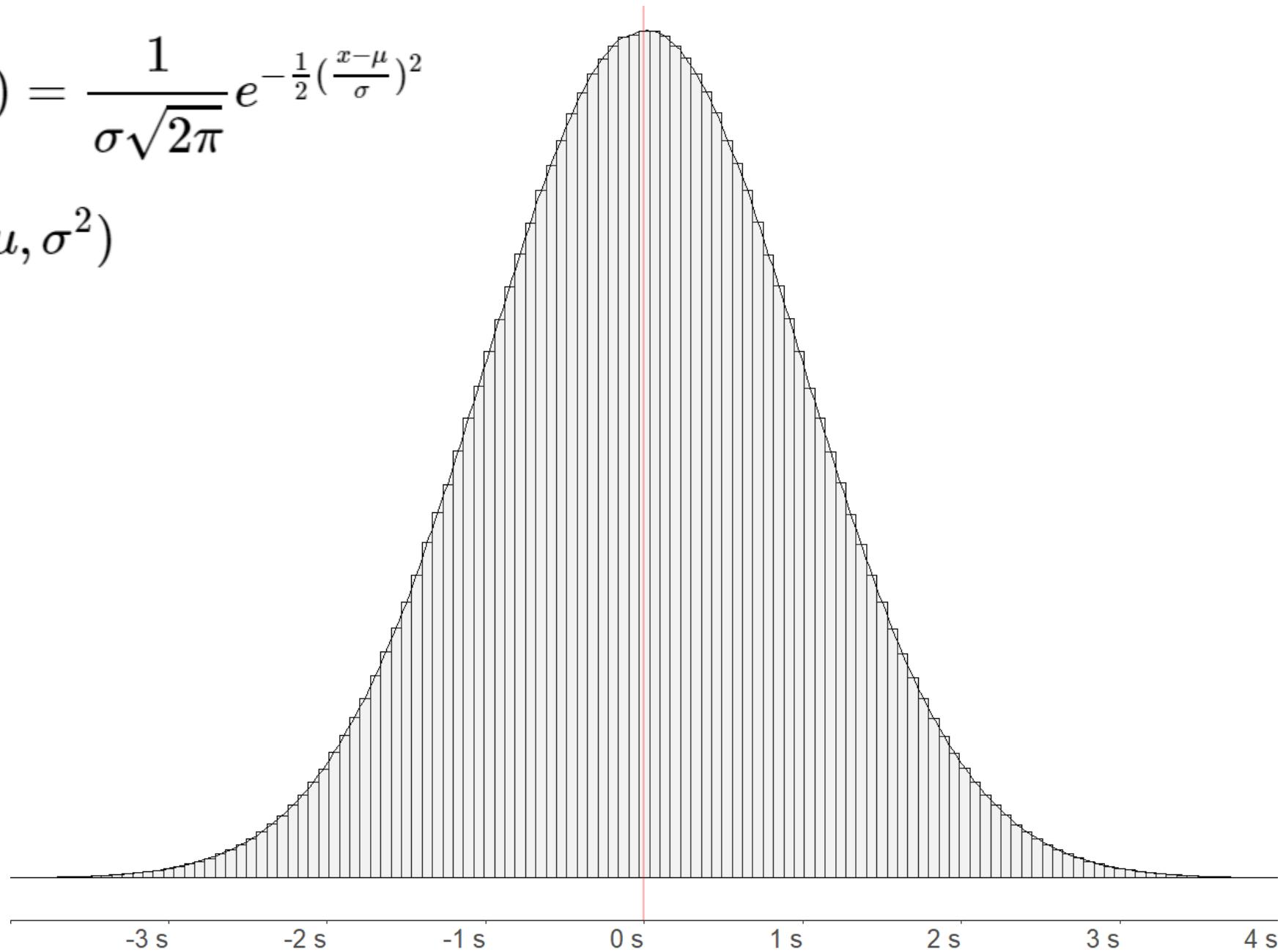
$$\mathcal{N}(\mu, \sigma^2)$$



Loi Normale

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

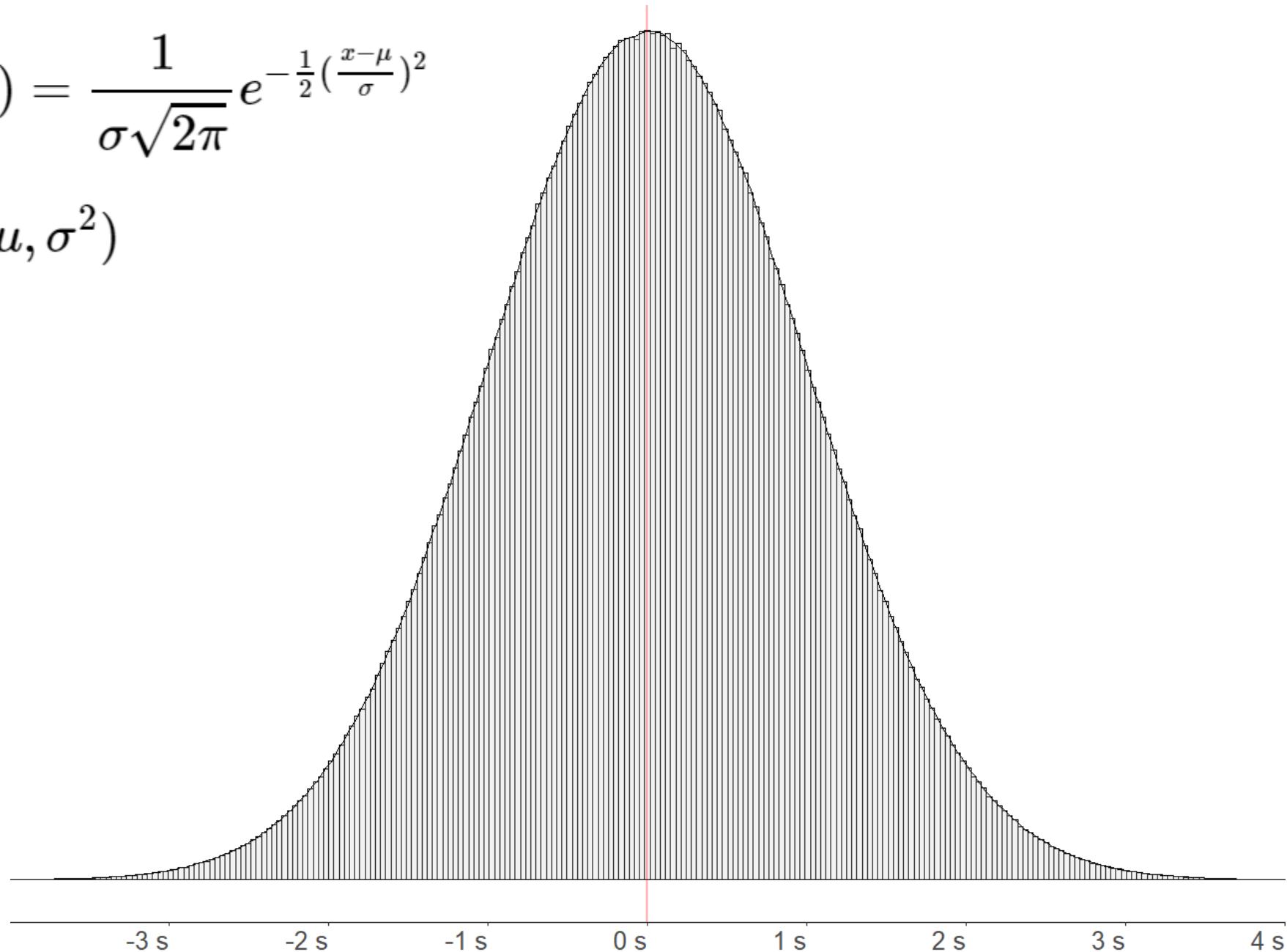
$$\mathcal{N}(\mu, \sigma^2)$$



Loi Normale

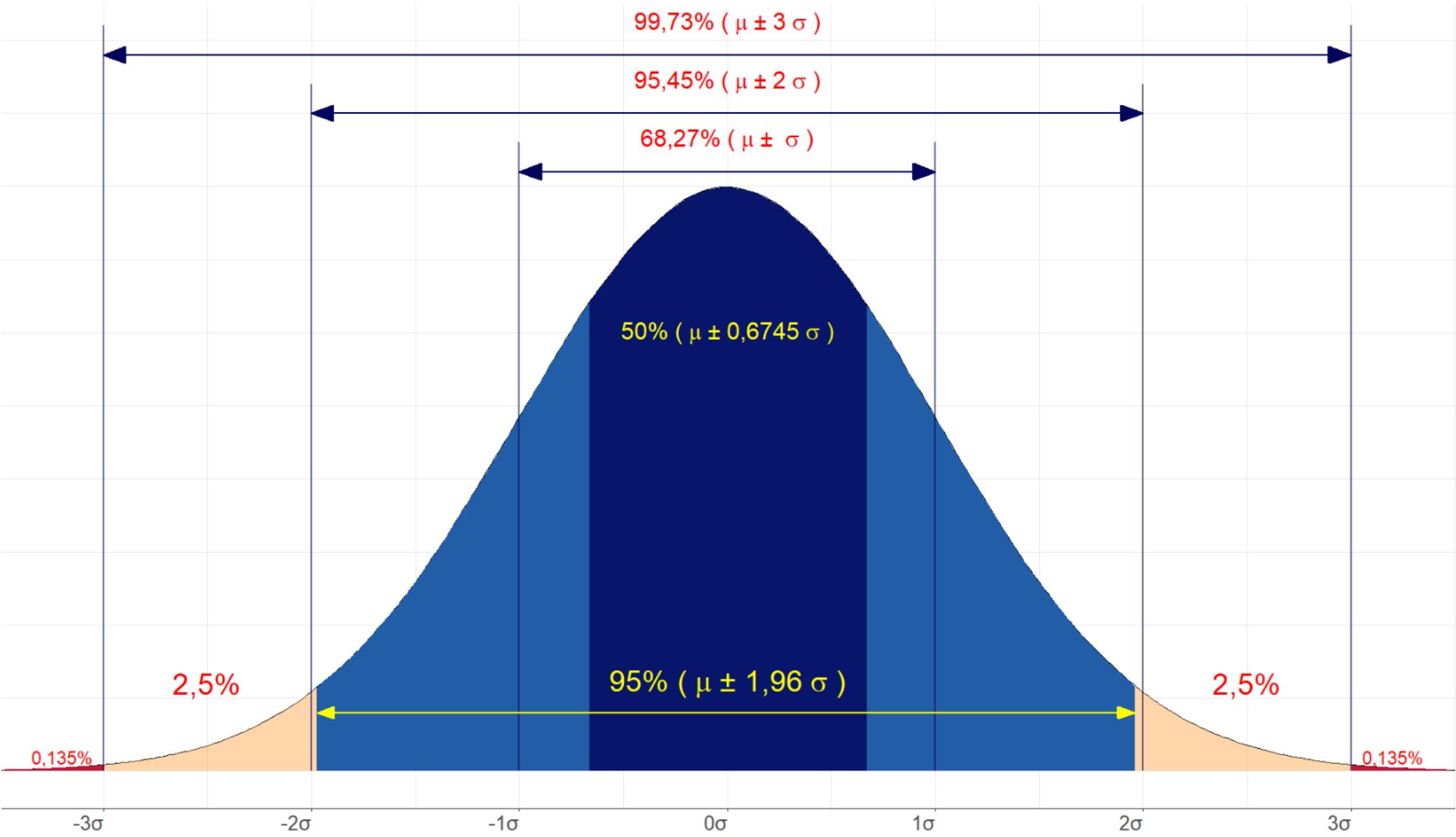
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$$

$$\mathcal{N}(\mu, \sigma^2)$$



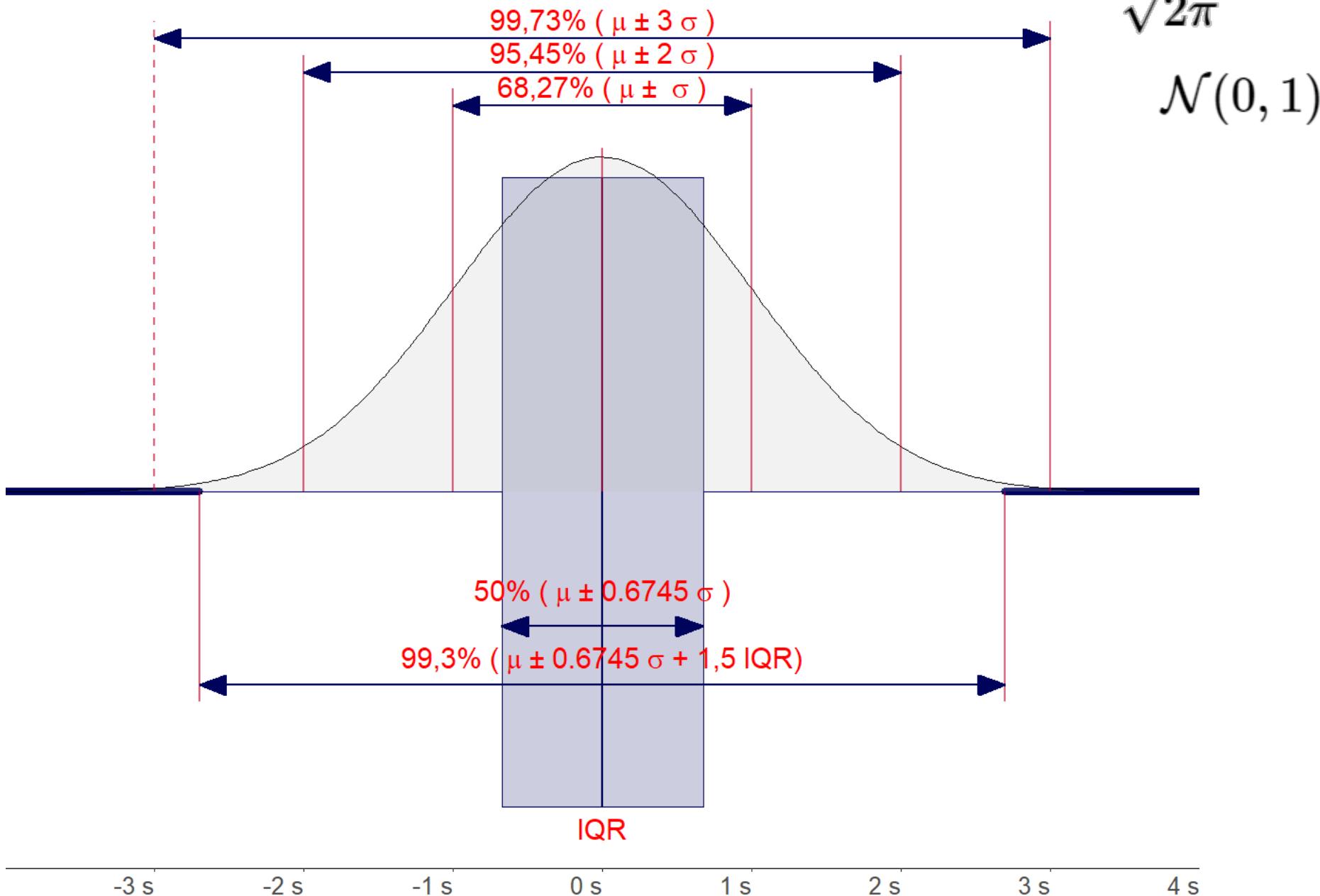
Loi Normale centrée réduite

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad \mathcal{N}(0, 1)$$

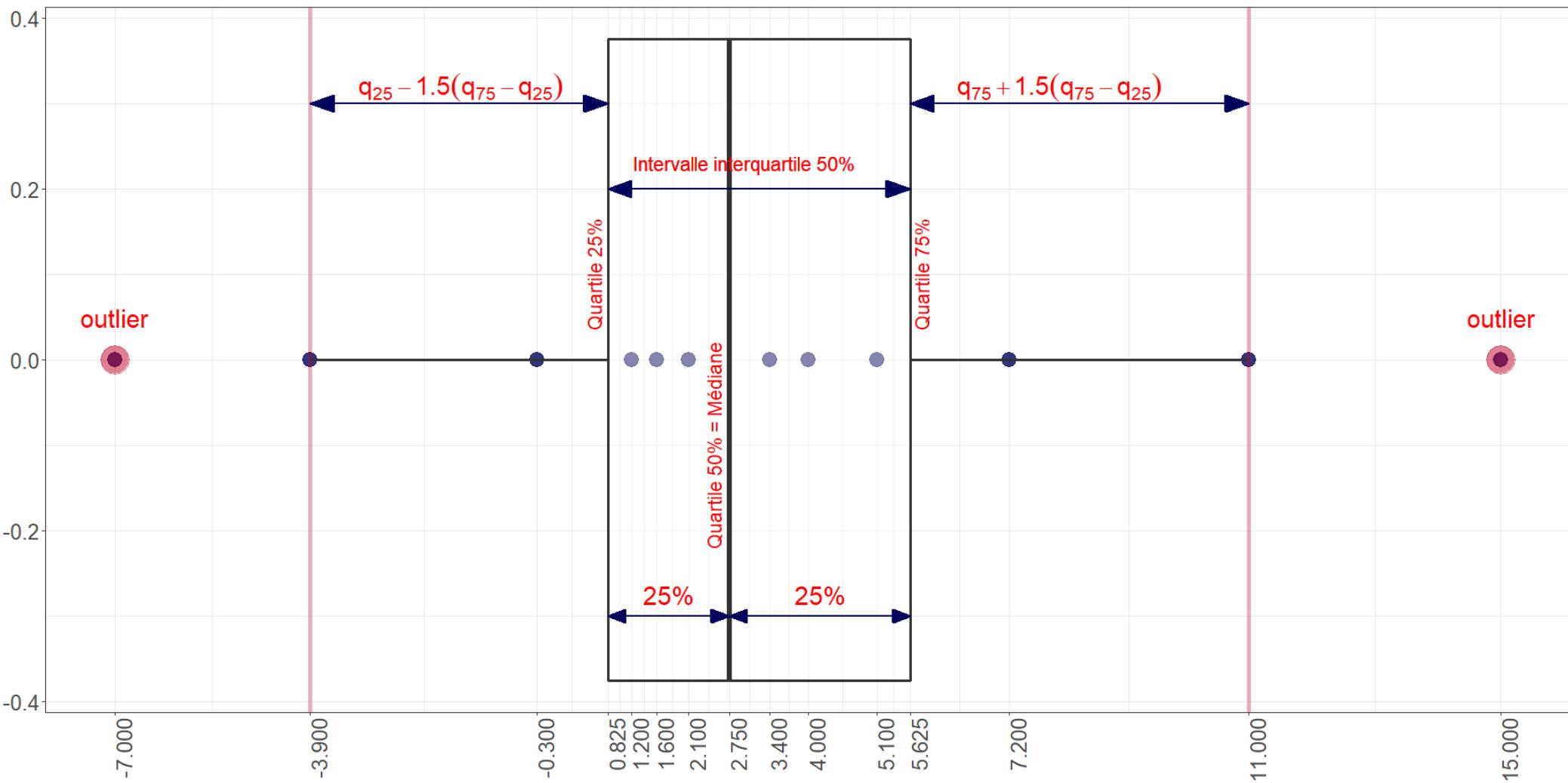


Loi Normale centrée réduite

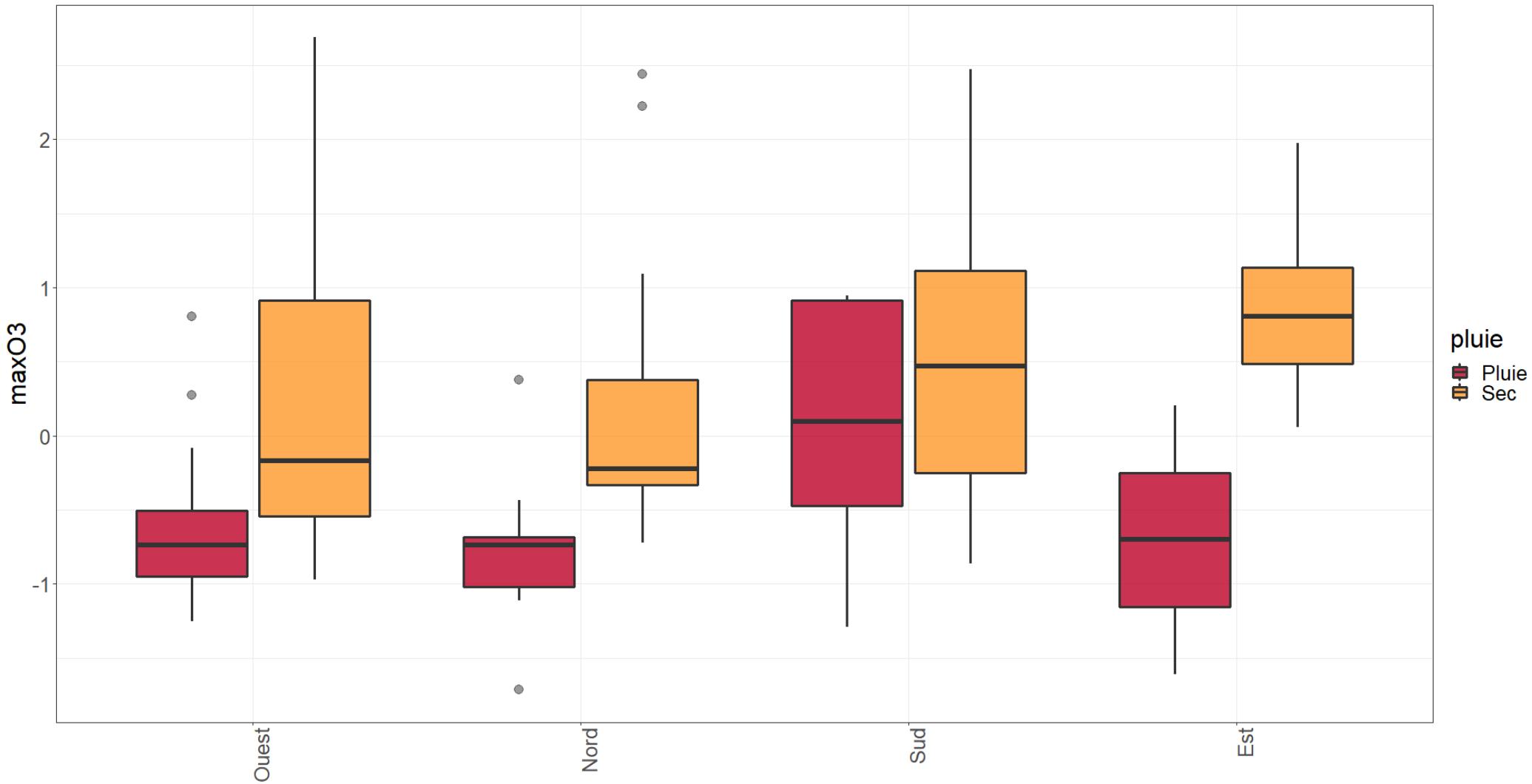
$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$
$$\mathcal{N}(0, 1)$$



La boîte à moustaches



La boîte à moustaches



Changement d'origine et d'unité

On appelle **changement d'origine** l'opération consistant à ajouter (ou soustraire) la même quantité $a \in \mathbb{R}$ à toutes les observations

$$y_i = a + x_i \quad i = 1, \dots, n$$

On appelle **changement d'unité** l'opération consistant à multiplier (ou diviser) la même quantité $b \in \mathbb{R}$ à toutes les observations

$$y_i = bx_i \quad i = 1, \dots, n$$

On appelle **changement d'origine et d'unité** l'opération consistant à multiplier toutes les observations par la même quantité $b \in \mathbb{R}$ puis à ajouter la même quantité $a \in \mathbb{R}$ à toutes les observations

$$y_i = a + bx_i \quad i = 1, \dots, n$$

Si on effectue un changement d'origine et d'unité sur une variable X , alors sa moyenne est affectée du même changement d'origine et d'unité.

$$y_i = a + bx_i \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n (a + bx_i) = a + b \frac{1}{n} \sum_{i=1}^n x_i = a + b\bar{x}$$

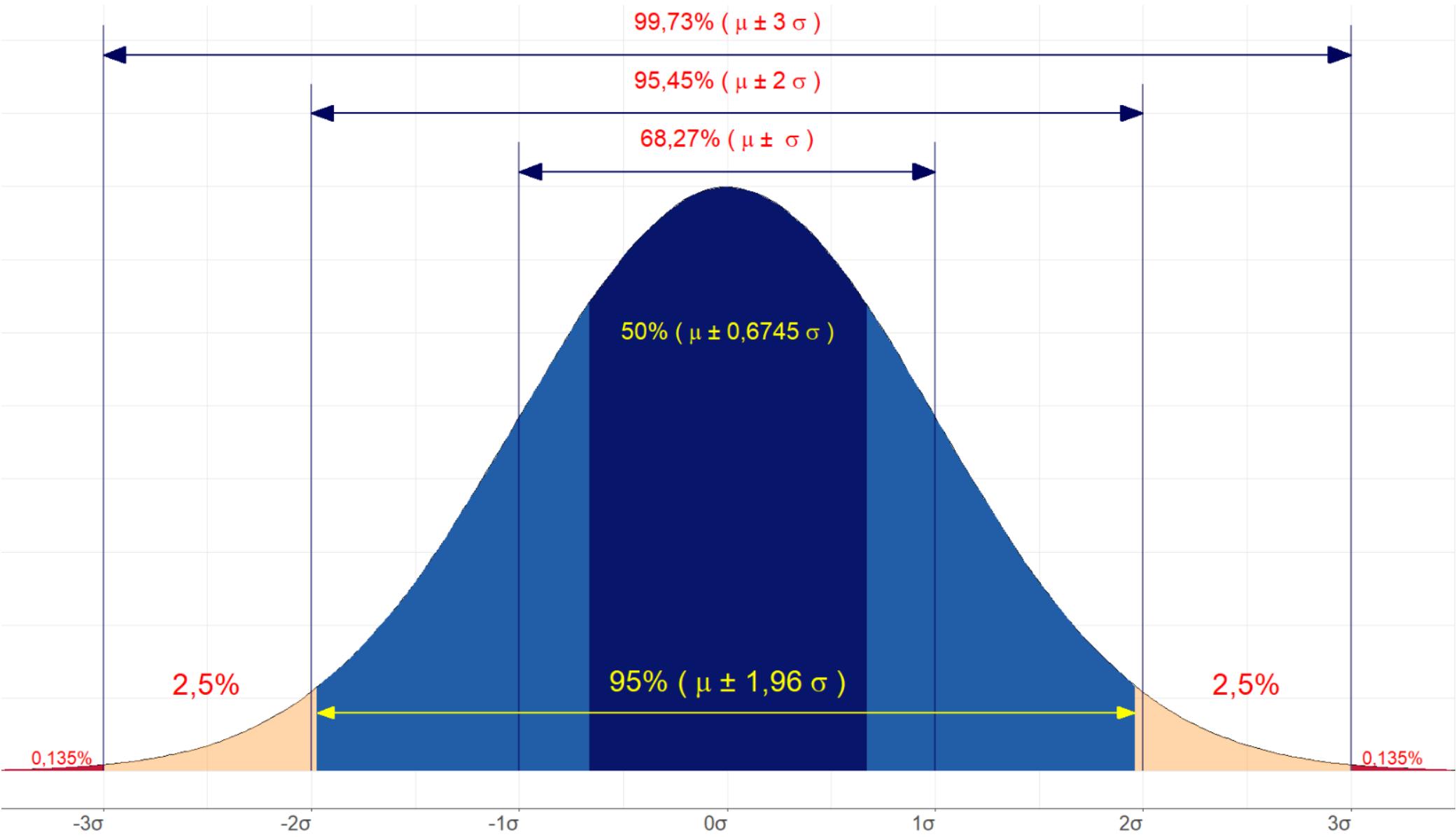
Centrage et réduction

$$si \quad a = -\bar{x} \quad et \quad b = \frac{1}{s_x} \Rightarrow y_i = \frac{x_i - \bar{x}}{s_x}$$

matière	mathématiques 1 ^{er} trimestre	philosophie 1 ^{er} trimestre
note	18	15
moyenne	10.64	8.99
écart à la moyenne	7.36	6.01
écart-type	2.85	2
note centrée réduite	$\frac{18-10.64}{2.85} = \frac{7.36}{2.85} = 2.58$	$\frac{15-8.99}{2} = \frac{6.01}{2} = 3.005$
probabilité associée	0.005	0.001
P[X > n.c.r.] ; L[X] = Normale		

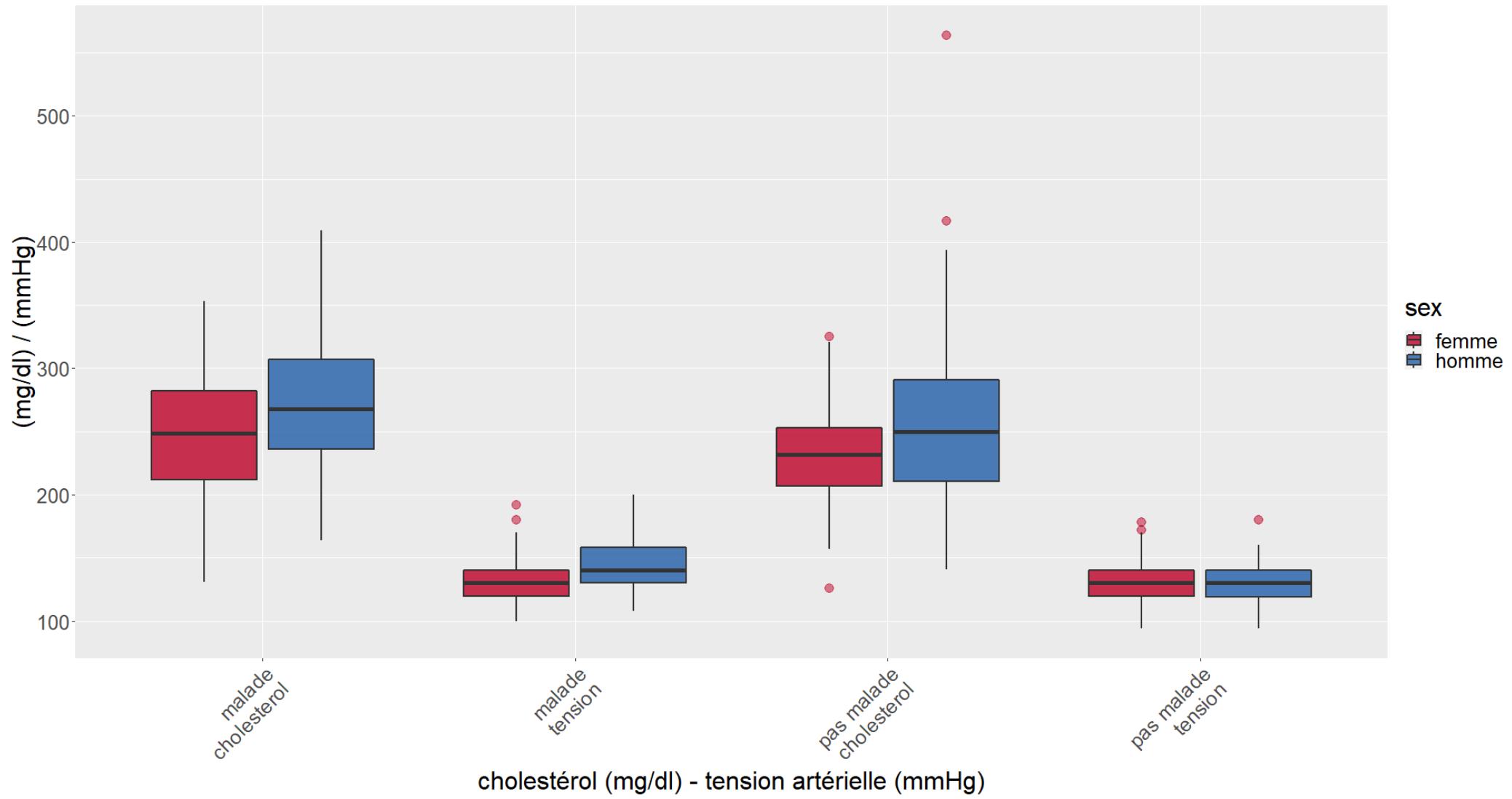
Loi Normale centrée réduite

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad \mathcal{N}(0, 1)$$



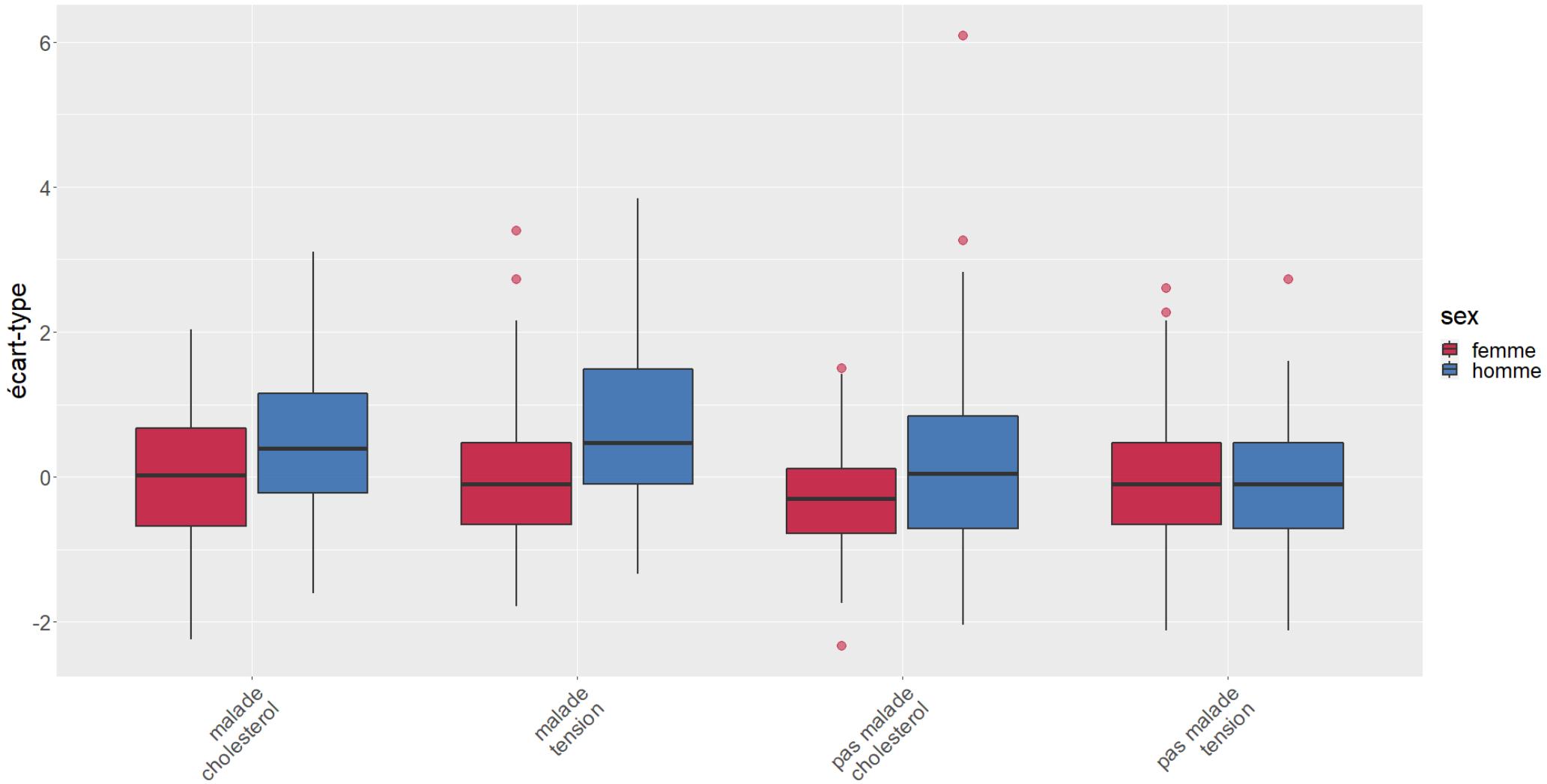
Centrage et réduction

$$si \quad a = -\bar{x} \quad et \quad b = \frac{1}{s_x} \Rightarrow y_i = \frac{x_i - \bar{x}}{s_x}$$



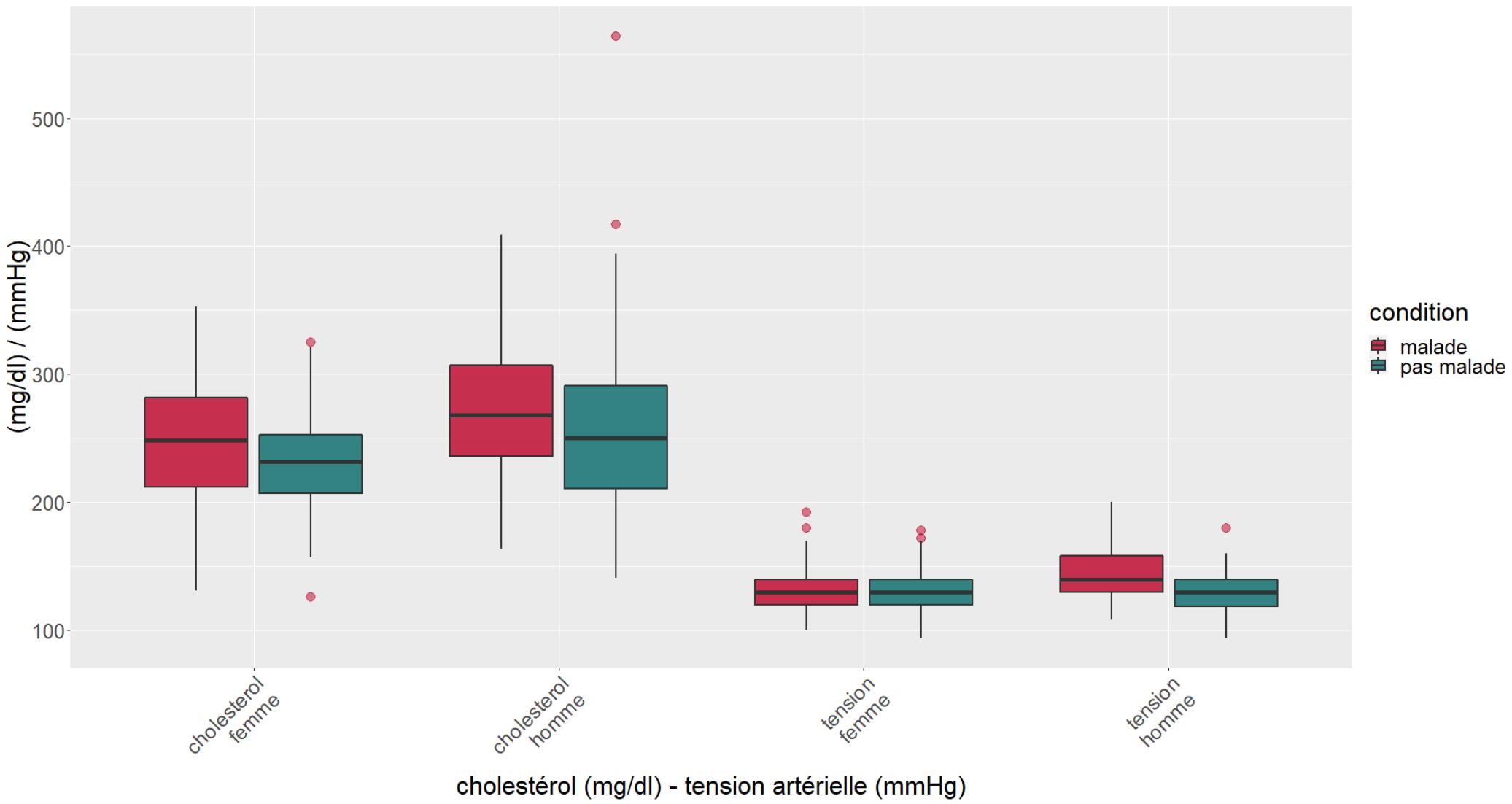
Centrage et réduction

$$si \quad a = -\bar{x} \quad et \quad b = \frac{1}{s_x} \Rightarrow y_i = \frac{x_i - \bar{x}}{s_x}$$



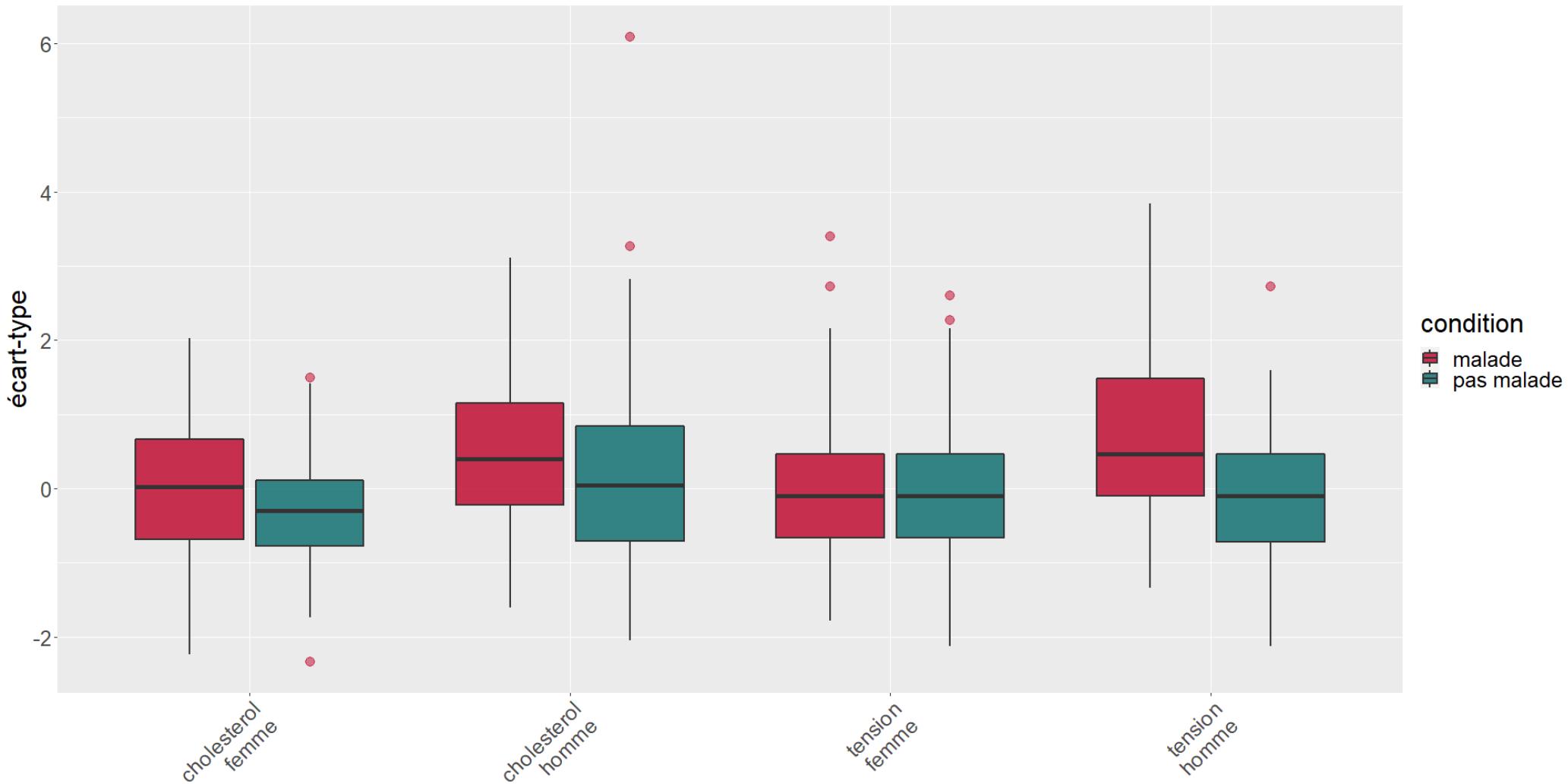
Centrage et réduction

$$si \quad a = -\bar{x} \quad et \quad b = \frac{1}{s_x} \Rightarrow y_i = \frac{x_i - \bar{x}}{s_x}$$



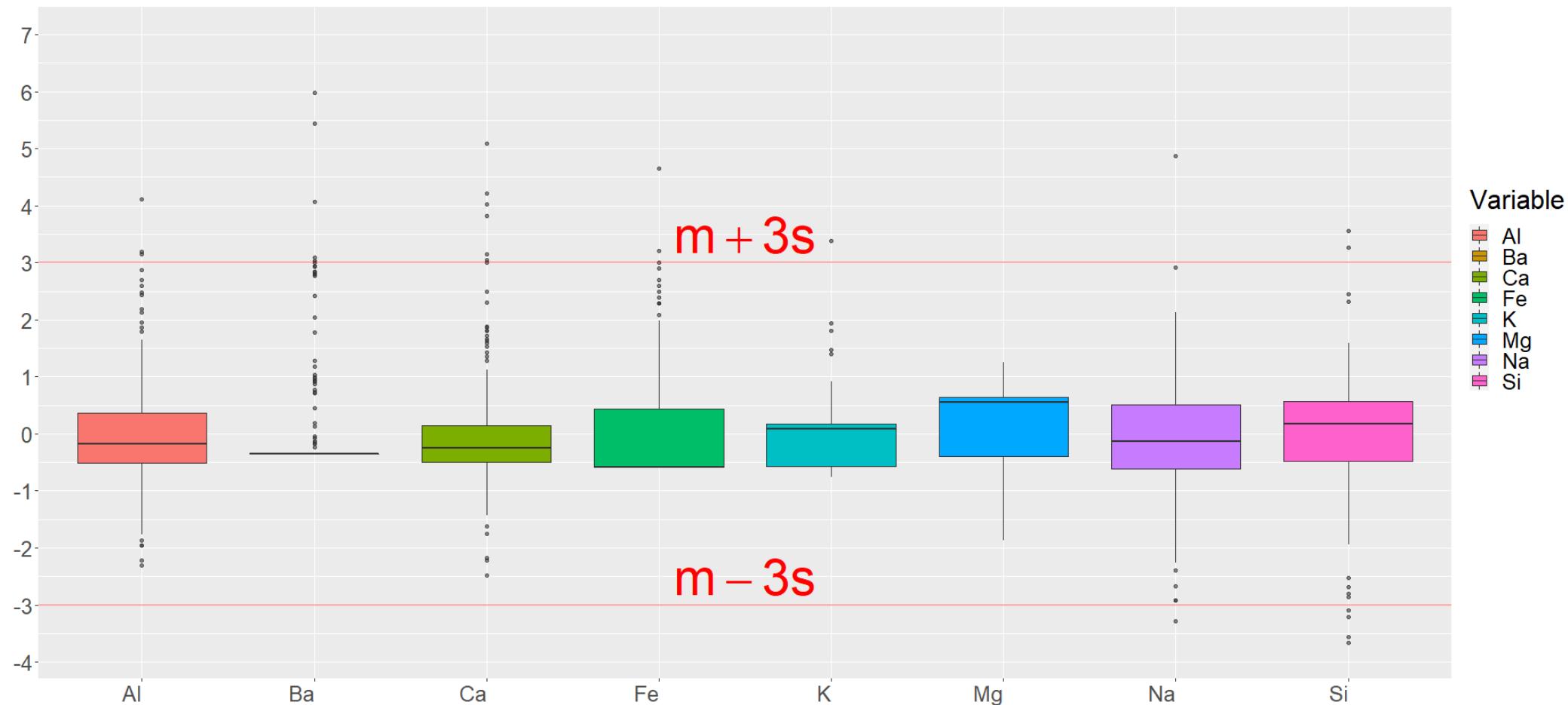
Centrage et réduction

$$si \quad a = -\bar{x} \quad et \quad b = \frac{1}{s_x} \Rightarrow y_i = \frac{x_i - \bar{x}}{s_x}$$



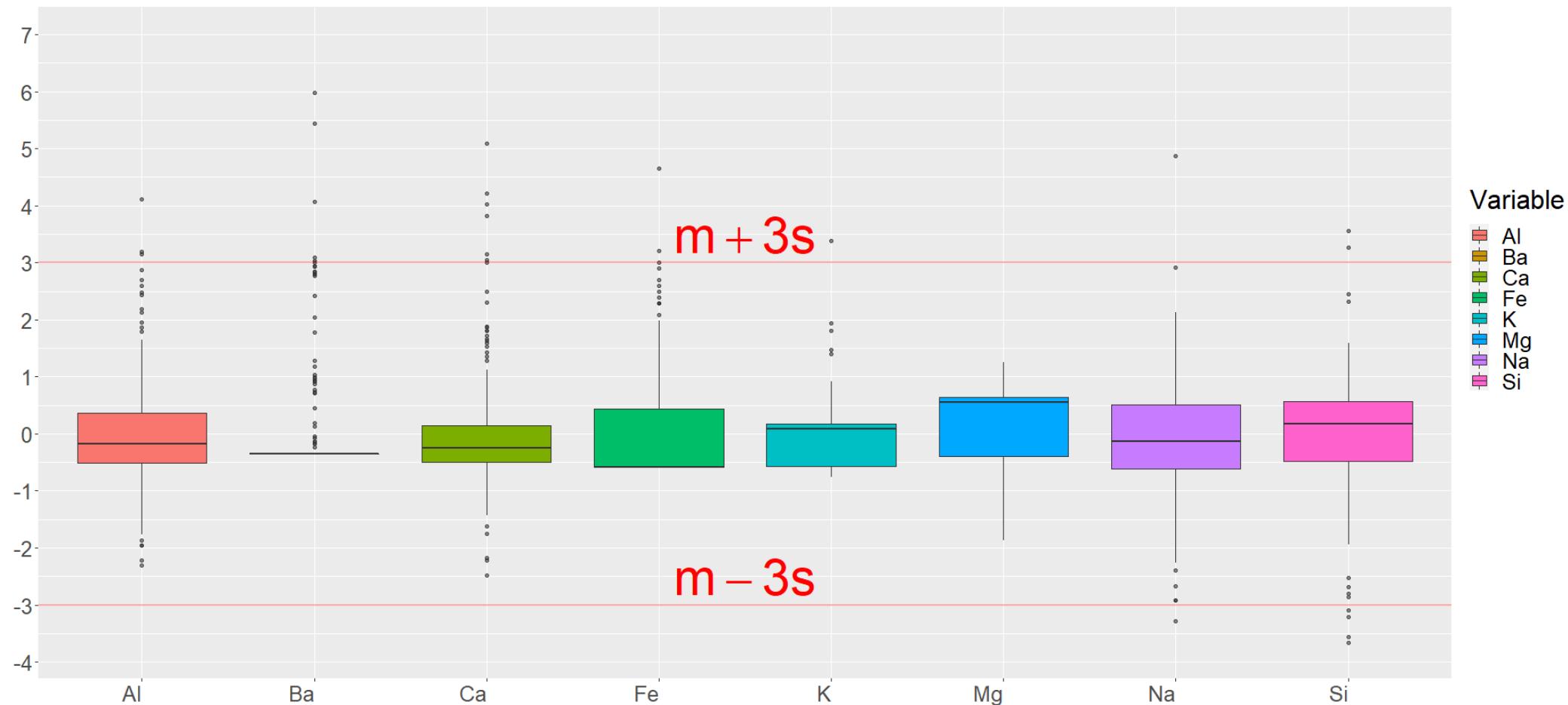
Centrage et réduction

$$si \quad a = -\bar{x} \quad et \quad b = \frac{1}{s_x} \Rightarrow y_i = \frac{x_i - \bar{x}}{s_x}$$



Centrage et réduction

$$si \quad a = -\bar{x} \quad et \quad b = \frac{1}{s_x} \Rightarrow y_i = \frac{x_i - \bar{x}}{s_x}$$



Module 2 : Les tableaux statistiques

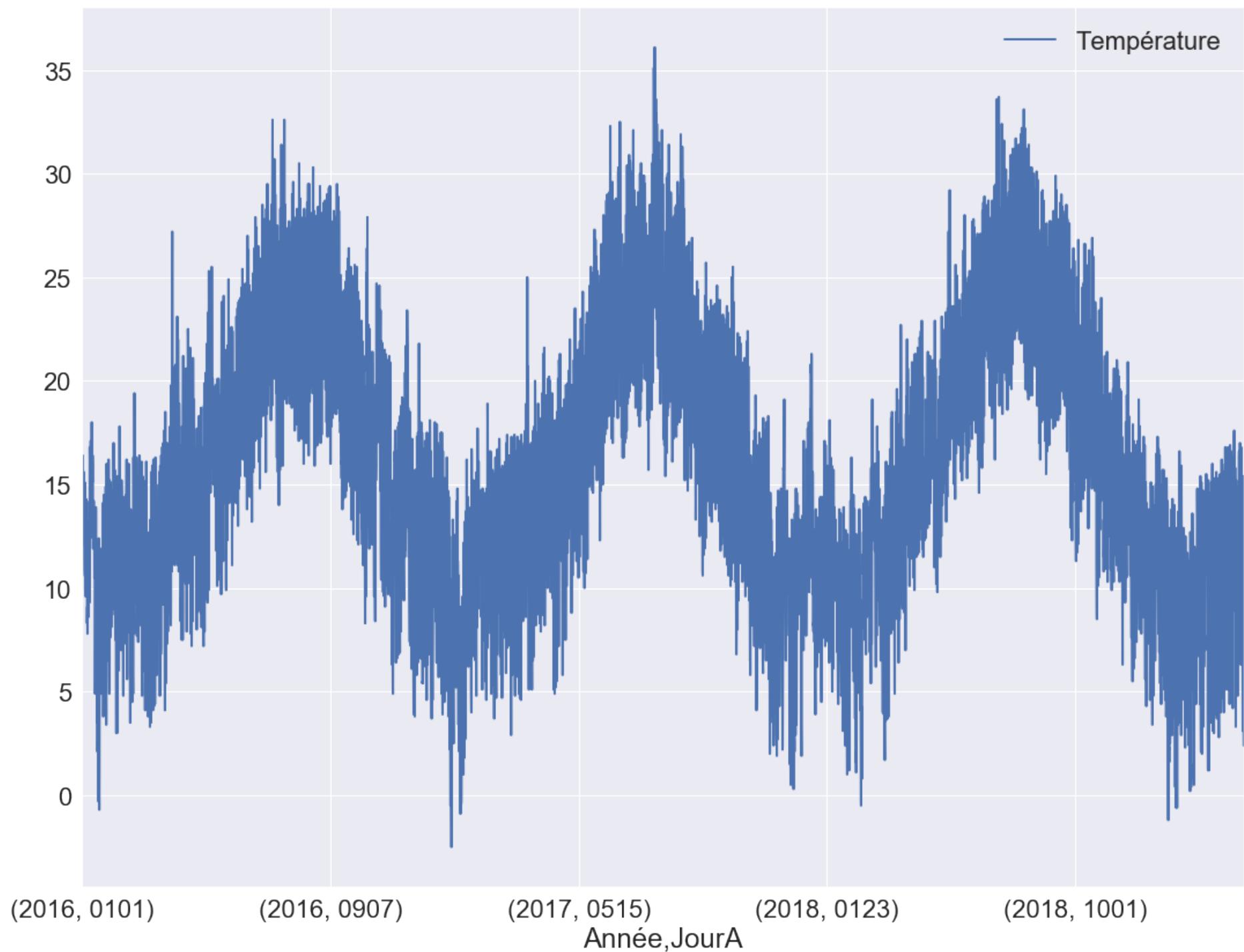
	Nom	Latitude	Longitude	Altitude	date	t	u	vv	pres
numer_sta									
7005	ABBEVILLE	50.136	1.834	69	19960101000000	276.04	99.0	NaN	99090.0
7005	ABBEVILLE	50.136	1.834	69	19960101030000	275.85	99.0	NaN	99070.0
7005	ABBEVILLE	50.136	1.834	69	19960101060000	276.04	99.0	100.0	99090.0
7005	ABBEVILLE	50.136	1.834	69	19960101090000	276.75	99.0	800.0	99110.0
7005	ABBEVILLE	50.136	1.834	69	19960101120000	277.54	99.0	1600.0	99040.0

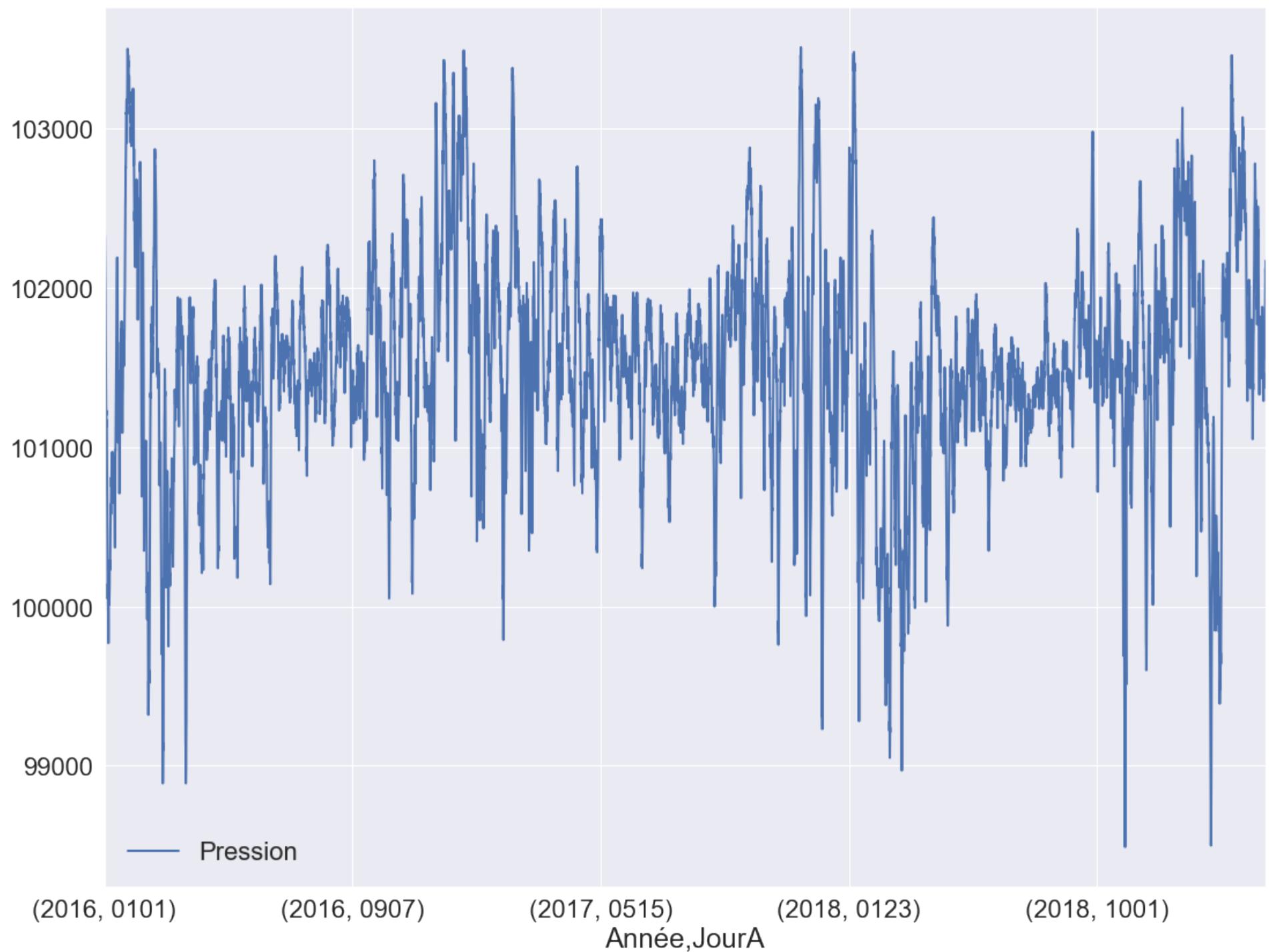
	Nom	Latitude	Longitude	Altitude	u	vv	pres	Année	Mois	Jour	JourA	Température
numer_sta												
7005	ABBEVILLE	50.136	1.834	69	99.0	NaN	99090.0	1996	1	1	0101	2.89
7005	ABBEVILLE	50.136	1.834	69	99.0	NaN	99070.0	1996	1	1	0101	2.70
7005	ABBEVILLE	50.136	1.834	69	99.0	100.0	99090.0	1996	1	1	0101	2.89
7005	ABBEVILLE	50.136	1.834	69	99.0	800.0	99110.0	1996	1	1	0101	3.60
7005	ABBEVILLE	50.136	1.834	69	99.0	1600.0	99040.0	1996	1	1	0101	4.39

Nom	Latitude	Longitude	Altitude	Température							
				Mois	1	2	3	4	5	6	7
ABBEVILLE	50.136000	1.834000	69	4.290521	4.815389	7.055397	9.711432	12.880101	15.673264	17.522640	17.6321
AJACCIO	41.918000	8.792667	5	8.917447	8.895447	10.900823	13.648195	17.345244	21.328785	23.563754	23.6919
ALENCON	48.445500	0.110167	143	4.390527	4.802228	7.267374	9.914059	13.367927	16.719701	18.431318	18.2775
BALE-MULHOUSE	47.614333	7.510000	263	1.939337	2.848941	6.497734	10.587650	14.958589	18.585189	19.884407	19.4428
BASTIA	42.540667	9.485167	10	9.090281	9.079022	11.181139	13.852164	17.806549	21.955024	24.620895	24.6610

Nom	Latitude	Longitude	Altitude	Température							
				JourA	0101	0102	0103	0104	0105	0106	0107
ABBEVILLE	50.136000	1.834000	69	4.232656	4.598698	4.117937	4.430677	4.424583	4.485625	5.031250	4.959115
AJACCIO	41.918000	8.792667	5	9.371875	9.843351	9.527842	9.307606	8.915313	9.334158	9.711667	9.727872
ALENCON	48.445500	0.110167	143	4.405842	4.638743	4.299583	4.402135	4.478125	4.601684	4.533125	4.952760
BALE-MULHOUSE	47.614333	7.510000	263	2.295521	2.436094	2.177552	2.027083	2.485625	2.191458	2.688115	2.722240
BASTIA	42.540667	9.485167	10	9.096563	9.810990	10.317566	9.345938	9.186094	9.071316	9.628351	9.824688

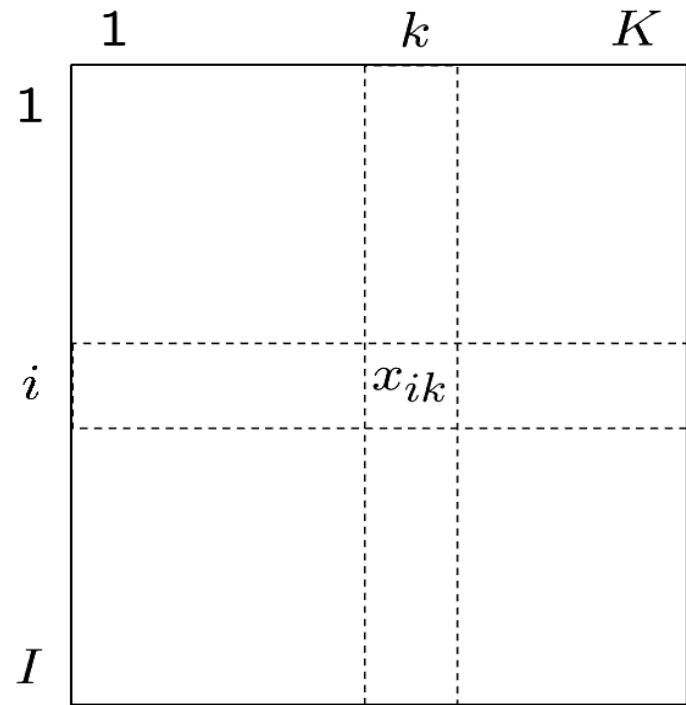
	janvier	février	mars	avril	mai	juin	juillet	août	septembre	octobre	novembre	décembre
Nom												
Abbeville	4.372309	4.945650	7.092692	9.841288	12.885313	15.764326	17.569472	17.764522	15.187675	11.868993	7.689304	5.042880
Ajaccio	8.965797	8.983649	10.928321	13.645100	17.333600	21.348559	23.678523	23.791242	20.658899	17.437811	13.049093	9.909476
Alencon	4.460286	4.931409	7.306194	10.051220	13.388578	16.758878	18.527981	18.375294	15.269251	11.773452	7.560239	4.879053
Bale	2.006641	3.041258	6.548861	10.649915	14.845584	18.637804	20.016305	19.554766	15.222541	10.951525	5.886351	2.770062
Bastia	9.135010	9.186098	11.185869	13.839347	17.752433	21.957800	24.680482	24.734411	21.067004	17.364209	13.101640	10.072712
Belle Ile	8.252692	8.025690	9.299535	11.197191	13.842194	16.570777	18.171541	18.294955	16.898853	14.577304	11.497191	9.195576
Bordeaux	6.822140	7.348053	10.317161	12.917454	16.327002	19.759232	21.383271	21.396970	18.520339	14.782295	9.808928	7.139346
Bourges	4.277366	4.900913	8.148068	11.195747	14.830750	18.706124	20.523184	20.321335	16.558331	12.537258	7.574953	4.754008
Brest	7.089676	7.083480	8.317183	9.916570	12.571513	15.169509	16.701156	16.666339	15.159649	12.784379	9.685450	7.700437
Caen	5.442578	5.844541	7.592919	9.832170	12.821443	15.876172	17.673384	17.841838	15.520532	12.444602	8.551714	6.031977
Cap Cépet	9.107509	9.016452	11.298338	13.727985	17.289659	21.405858	23.718360	23.852939	20.421422	16.889400	12.475549	9.966813
Clermont-Fd	4.137646	4.849099	8.006348	11.048576	14.822511	18.716563	20.547752	20.321455	16.367645	12.790320	7.635619	4.791446
Dijon	2.523281	3.582445	7.201618	10.823901	14.654741	18.740466	20.640479	20.175979	16.133918	11.550468	6.281968	3.196375
Embrun	1.419745	2.534277	6.557814	9.990218	13.817003	18.039846	20.313251	19.851152	15.680018	11.271405	5.485900	1.947364
Gourdon	5.220965	5.735507	9.004721	11.880580	15.263693	19.041794	20.787929	20.542360	17.112710	13.544116	8.282524	5.592242
Le Puy	1.131721	1.498085	4.670948	7.734377	11.594776	15.735622	17.660271	17.168941	13.164240	9.724798	4.703083	1.855387
Lille	3.897301	4.684211	7.186551	10.447000	13.815068	16.830526	18.652696	18.562877	15.582529	11.770374	7.389850	4.613175
Limoges	4.325551	4.607583	7.810112	10.559168	13.927378	17.621605	19.264866	19.297226	16.100861	12.461558	7.361151	4.939531
Lyon	3.726300	4.662335	8.524539	12.229397	16.077609	20.135900	21.938409	21.546680	17.436823	13.289958	7.622566	4.318790
Marignane	7.461079	7.825671	11.102441	14.383852	18.442467	22.802329	25.128455	24.641115	20.608042	16.805822	11.414476	8.036939
Millau	3.116205	3.450249	6.594085	9.442741	13.121010	17.526441	19.633983	19.467763	15.590353	11.801085	6.529040	3.942171
Mont-De-Marsan	6.050486	6.676959	9.833878	12.696449	16.102216	19.609619	21.269913	21.244089	18.077112	14.222011	9.043137	6.313867





Module 3 : Analyse en composantes principales

L'ACP s'intéresse à des tableaux de données rectangulaires avec des **individus** en lignes et des **variables quantitatives** en colonnes



Pour la variable k , on note :

$$\text{la moyenne : } \bar{x}_k = \frac{1}{I} \sum_{i=1}^I x_{ik}$$

$$\text{l'écart-type : } s_k = \sqrt{\frac{1}{I} \sum_{i=1}^I (x_{ik} - \bar{x}_k)^2}$$

Descriptif - exploratoire : visualisation de données par graphiques simples
Synthèse - résumé de grands tableaux individus \times variables

Les données température

15 individus (lignes) : villes de France

14 variables (colonnes) :

12 températures mensuelles moyennes (sur 30 ans)

2 variables géographiques (latitude, longitude)

	Janv	Févr	Mars	Avri	Mai	Juin	juil	Août	Sept	Octo	Nove	Déce	Lati	Long
Bordeaux	5.6	6.6	10.3	12.8	15.8	19.3	20.9	21	18.6	13.8	9.1	6.2	44.5	-0.34
Brest	6.1	5.8	7.8	9.2	11.6	14.4	15.6	16	14.7	12	9	7	48.24	-4.29
Clermont	2.6	3.7	7.5	10.3	13.8	17.3	19.4	19.1	16.2	11.2	6.6	3.6	45.47	3.05
Grenoble	1.5	3.2	7.7	10.6	14.5	17.8	20.1	19.5	16.7	11.4	6.5	2.3	45.1	5.43
Lille	2.4	2.9	6	8.9	12.4	15.3	17.1	17.1	14.7	10.4	6.1	3.5	50.38	3.04
Lyon	2.1	3.3	7.7	10.9	14.9	18.5	20.7	20.1	16.9	11.4	6.7	3.1	45.45	4.51
Marseille	5.5	6.6	10	13	16.8	20.8	23.3	22.8	19.9	15	10.2	6.9	43.18	5.24
Montpellier	5.6	6.7	9.9	12.8	16.2	20.1	22.7	22.3	19.3	14.6	10	6.5	43.36	3.53
Nantes	5	5.3	8.4	10.8	13.9	17.2	18.8	18.6	16.4	12.2	8.2	5.5	47.13	-1.33
Nice	7.5	8.5	10.8	13.3	16.7	20.1	22.7	22.5	20.3	16	11.5	8.2	43.42	7.15
Paris	3.4	4.1	7.6	10.7	14.3	17.5	19.1	18.7	16	11.4	7.1	4.3	48.52	2.2
Rennes	4.8	5.3	7.9	10.1	13.1	16.2	17.9	17.8	15.7	11.6	7.8	5.4	48.05	-1.41
Strasbourg	0.4	1.5	5.6	9.8	14	17.2	19	18.3	15.1	9.5	4.9	1.3	48.35	7.45
Toulouse	4.7	5.6	9.2	11.6	14.9	18.7	20.9	20.9	18.3	13.3	8.6	5.5	43.36	1.26
Vichy	2.4	3.4	7.1	9.9	13.6	17.1	19.3	18.8	16	11	6.6	3.4	46.08	3.26

$$\det(\text{mat} - \lambda \text{In}) = 0$$

$$R = \begin{pmatrix} 1 & r_{1,2} & L & r_{1,p} \\ r_{1,2} & 1 & L & r_{2,p} \\ L & L & L & L \\ r_{1,p} & r_{2,p} & L & 1 \end{pmatrix}$$

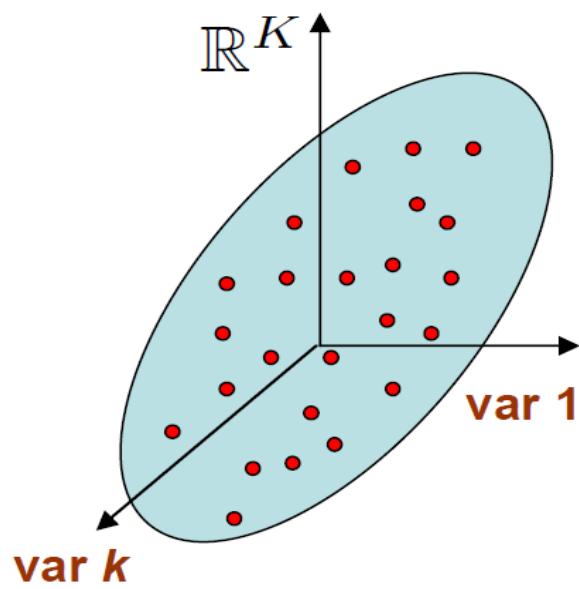
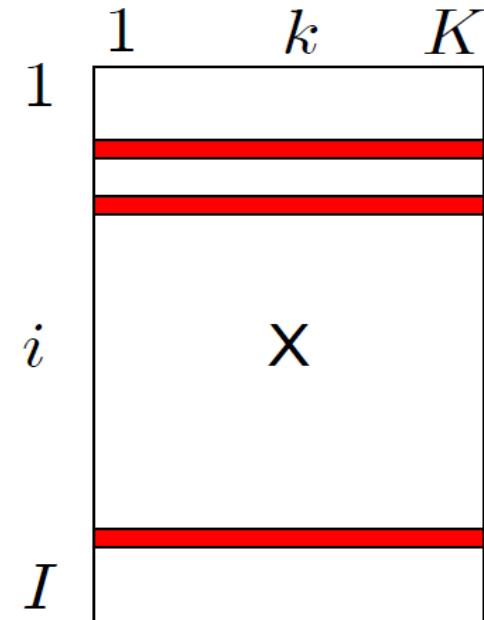
*matrice des
vecteurs propres*

*Matrice des valeurs
propres (diagonale)*

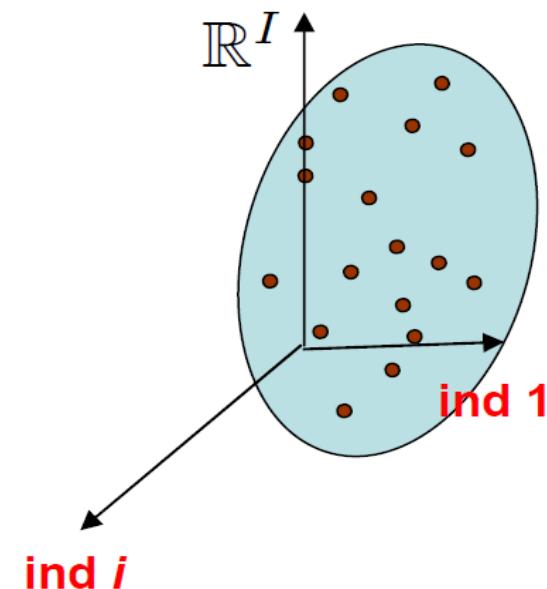
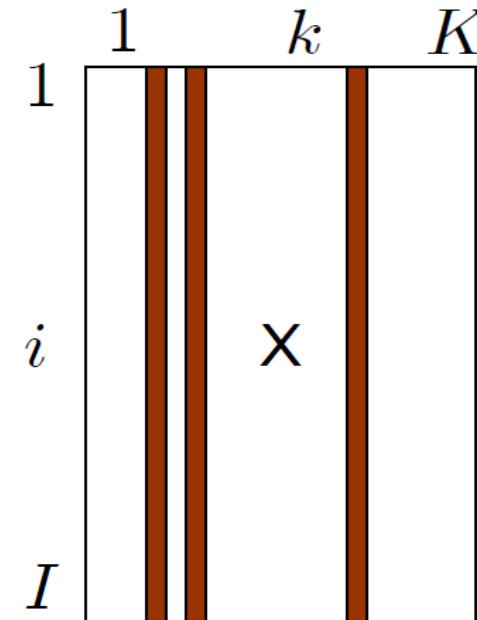
$$\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_p \end{pmatrix}$$

Deux nuages de points

Etude des individus



Etude des variables



L'étude du nuage des individus

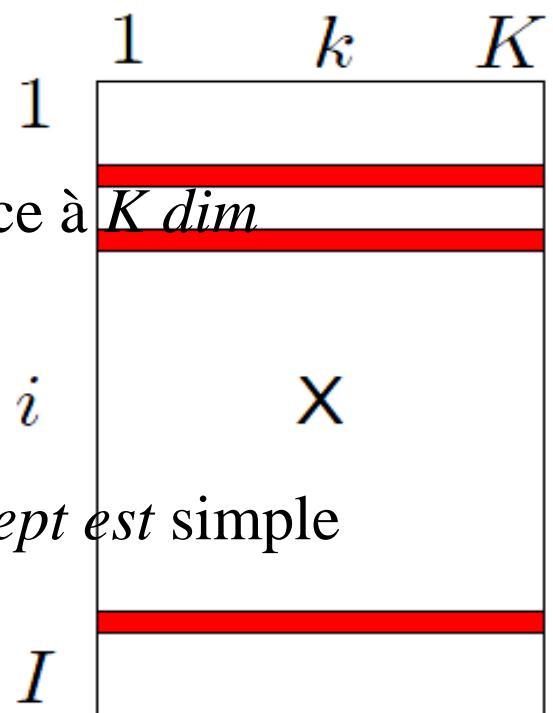
1 individu = 1 ligne du tableau) 1 point dans un espace à $K \dim$

Si $K = 1$: *Représentation axiale*

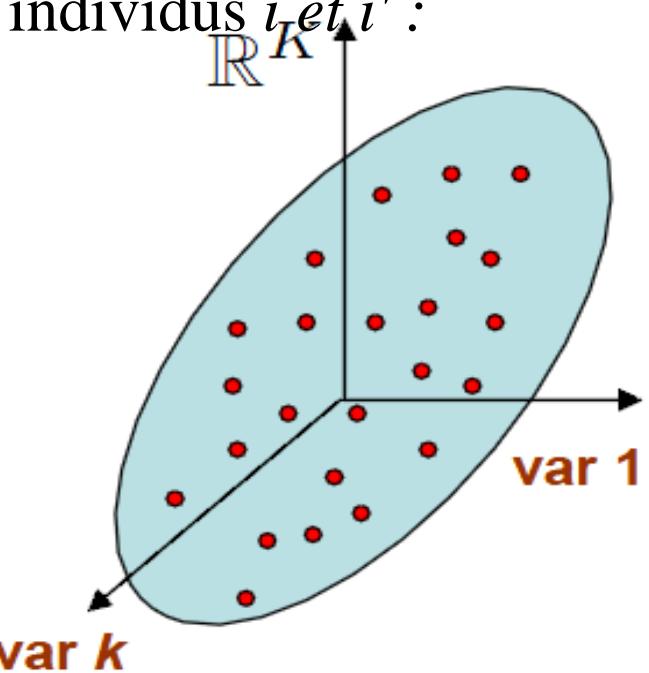
Si $K = 2$: *Nuage de points*

Si $K = 3$: *Représentation + difficile en 3D*

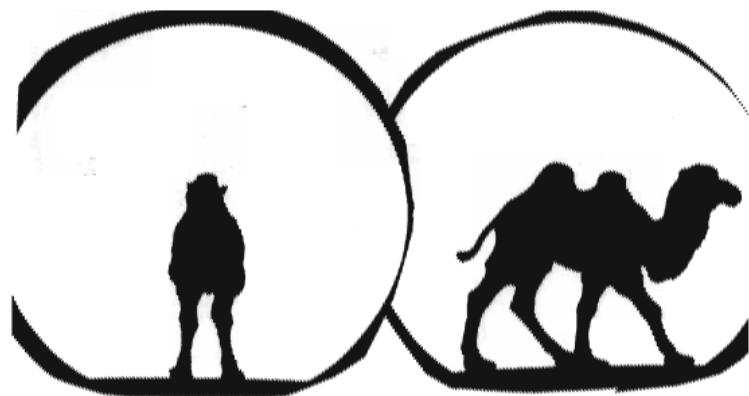
Si $K = 4$: *Impossible à représenter MAIS le concept est simple*



entre individus i et i' :



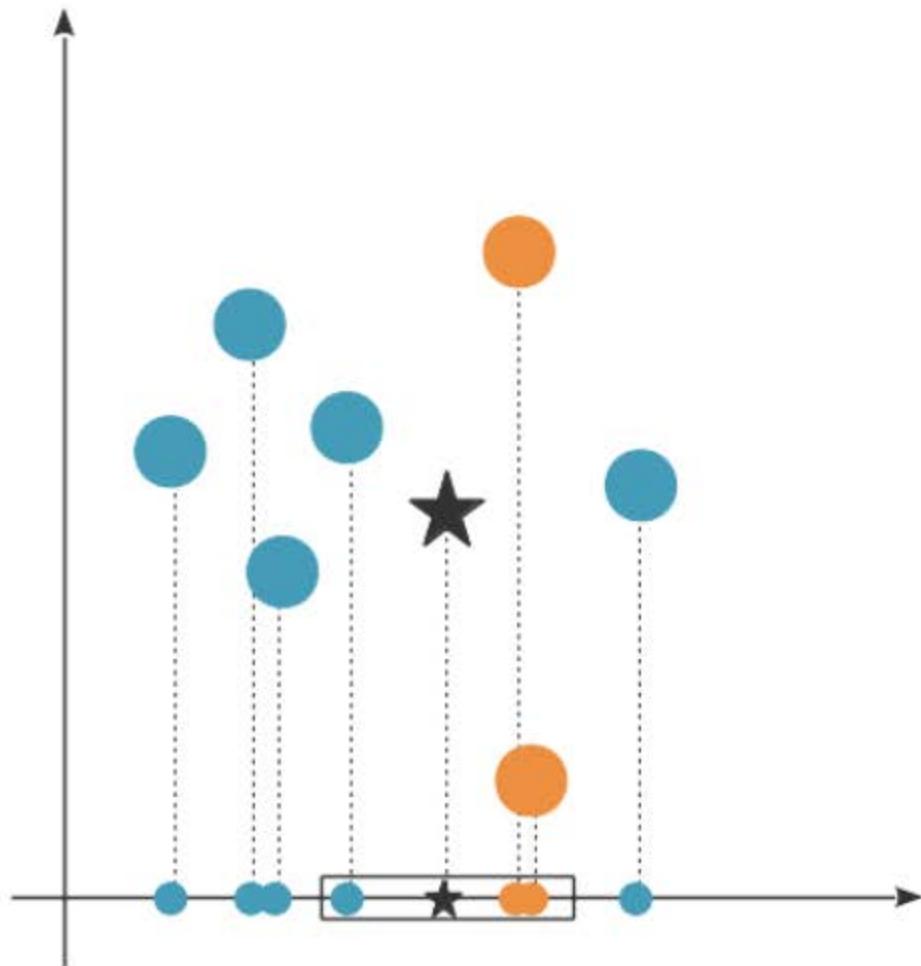
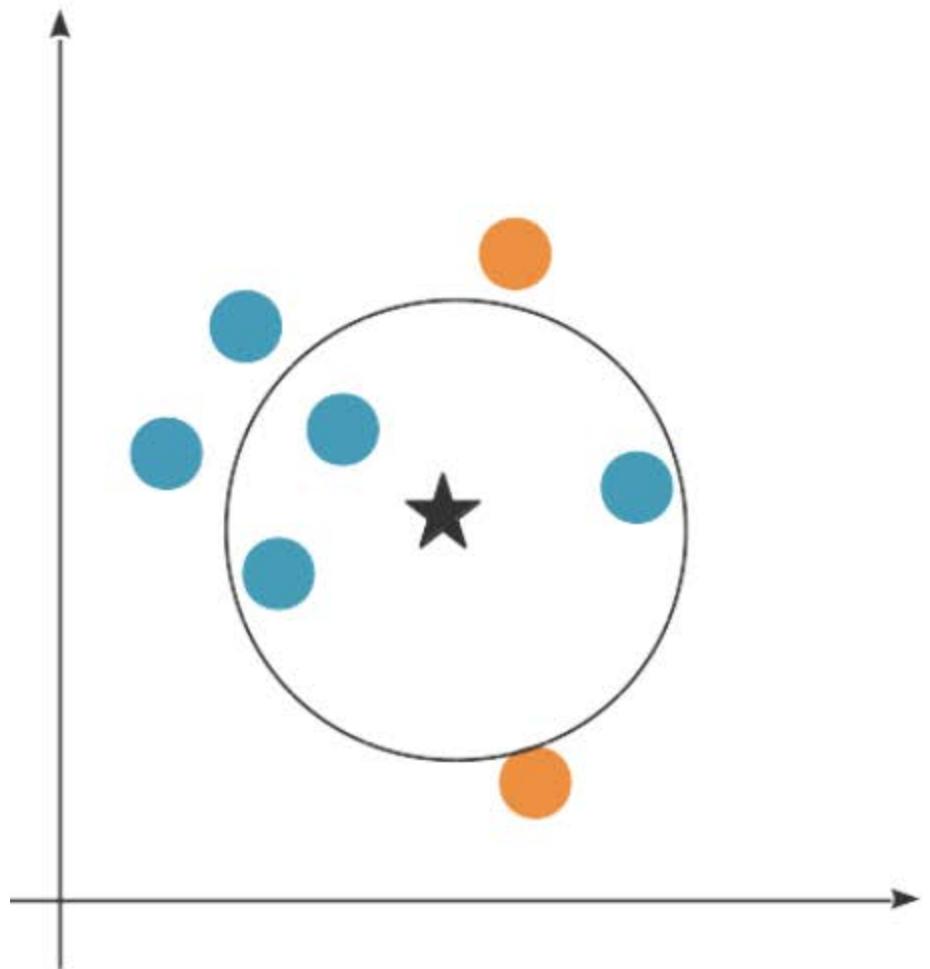
$$d^2(i, i') = \sum_{k=1}^K (x_{ik} - x_{i'k})^2$$

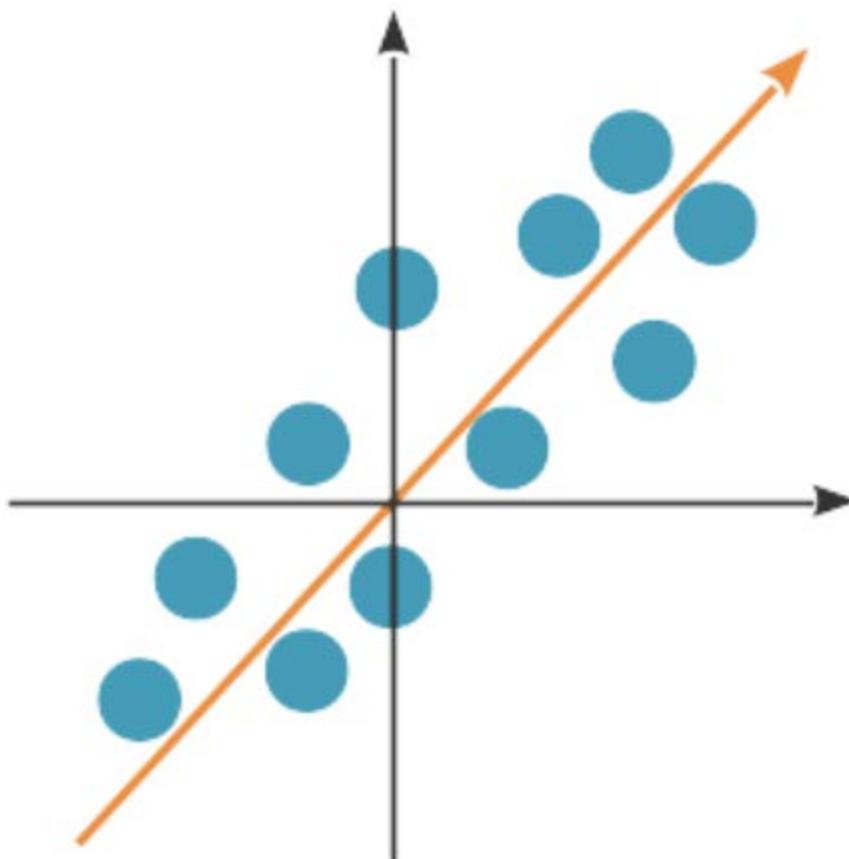


Le nuage des individus

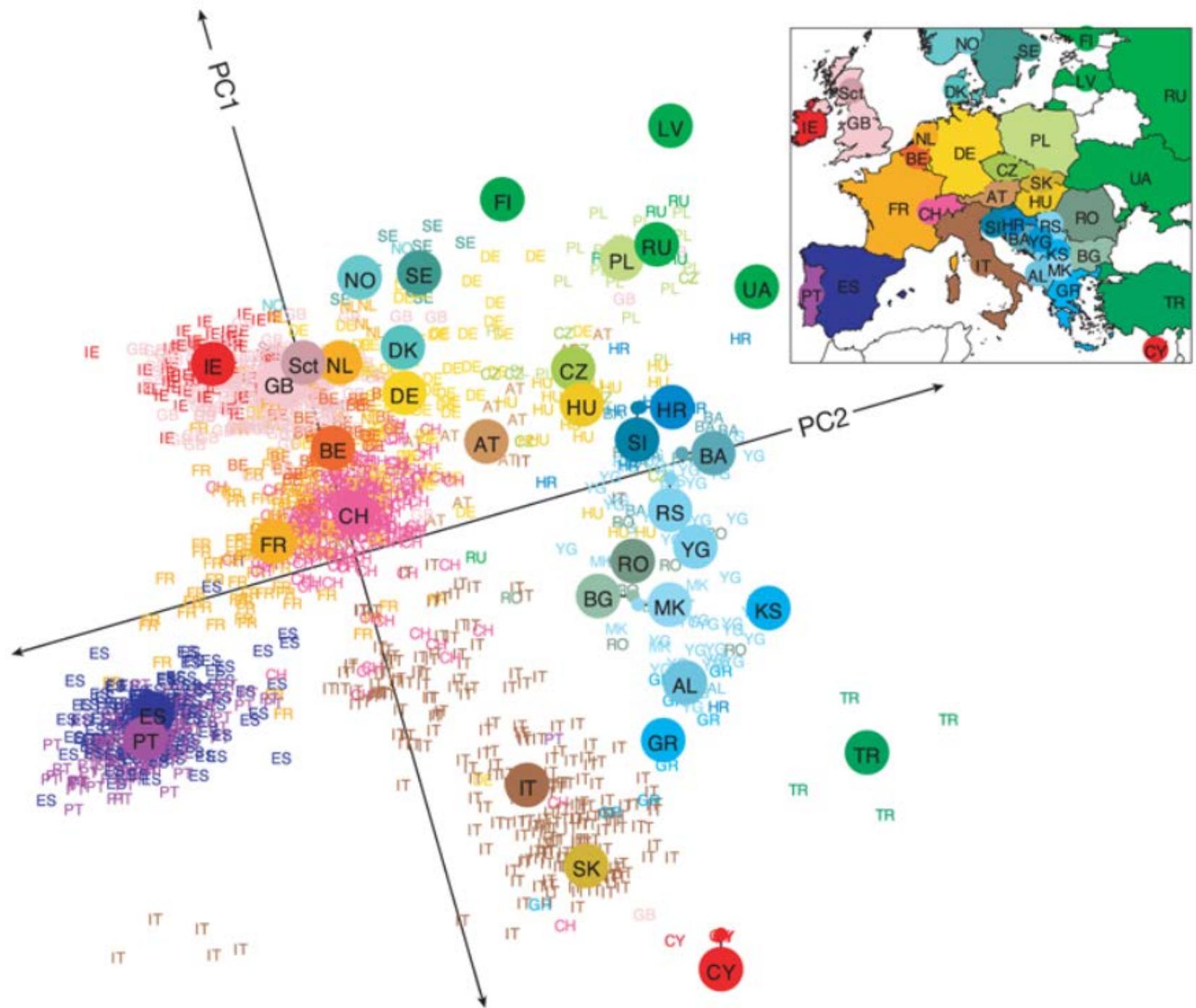
Etude des individus = Etude de la forme du nuage







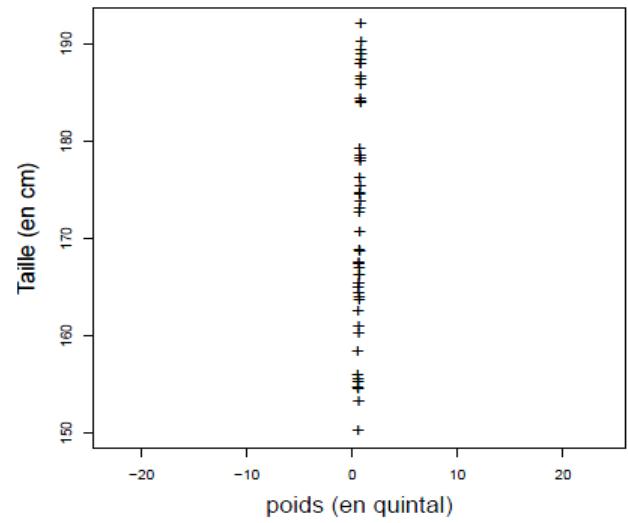
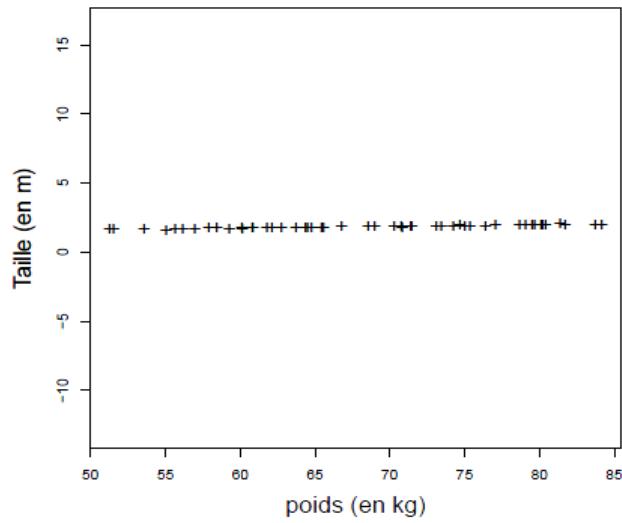
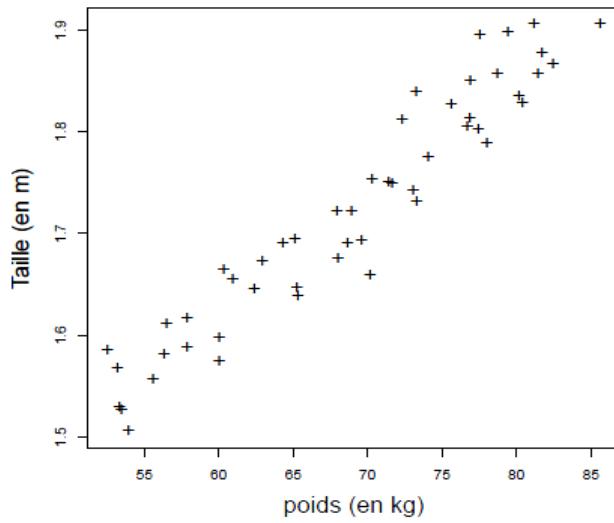
Variance des données



Centrage – Réduction des données

Centrer les données ne modifie pas la forme du nuage (toujours centrer)

$$x_{ik} \hookrightarrow x_{ik} - \bar{x}_k$$



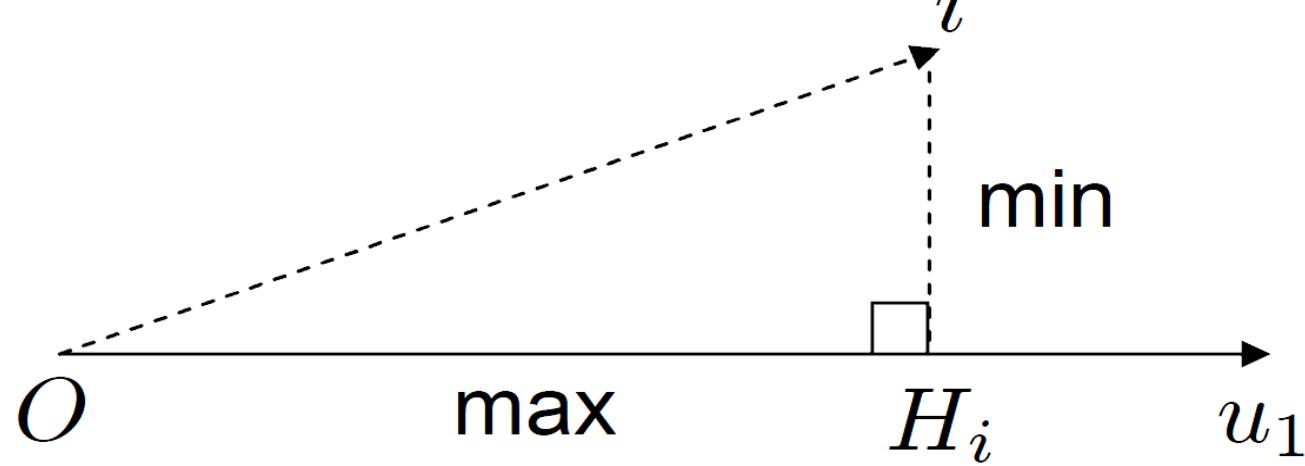
Réduire les données est indispensable si les unités de mesure sont différentes d'une variable à l'autre

$$x_{ik} \hookrightarrow \frac{x_{ik}}{s_k}$$

$$x_{ik} \hookrightarrow \frac{x_{ik} - \bar{x}_k}{s_k}$$

Ajustement du nuage des individus

Trouver l'axe (facteur) qui déforme le moins possible le nuage



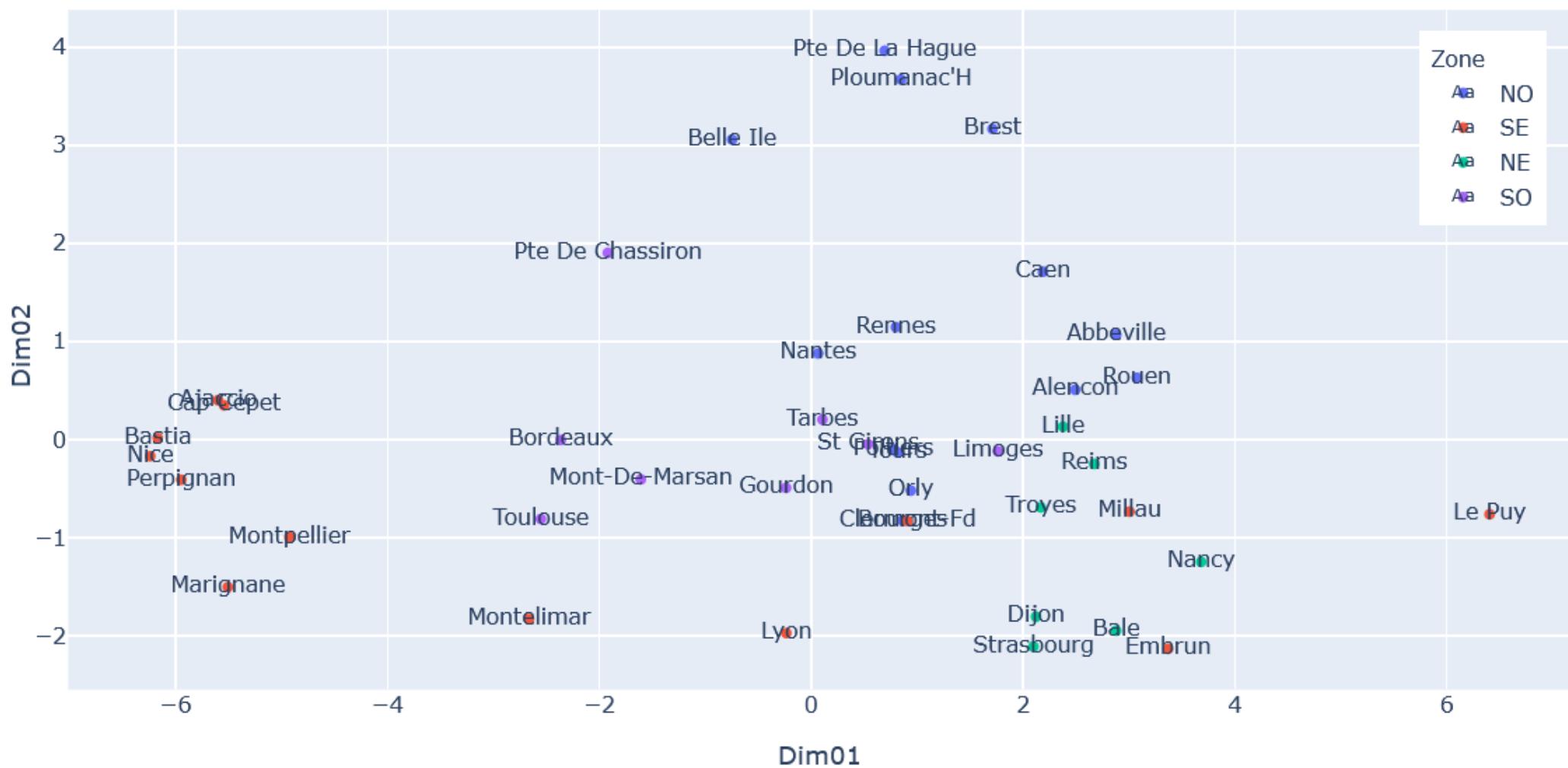
Trouver le meilleur plan : maximiser

$$\sum_i (OH_i)^2 \quad \text{avec } H_i \in \text{plan}$$

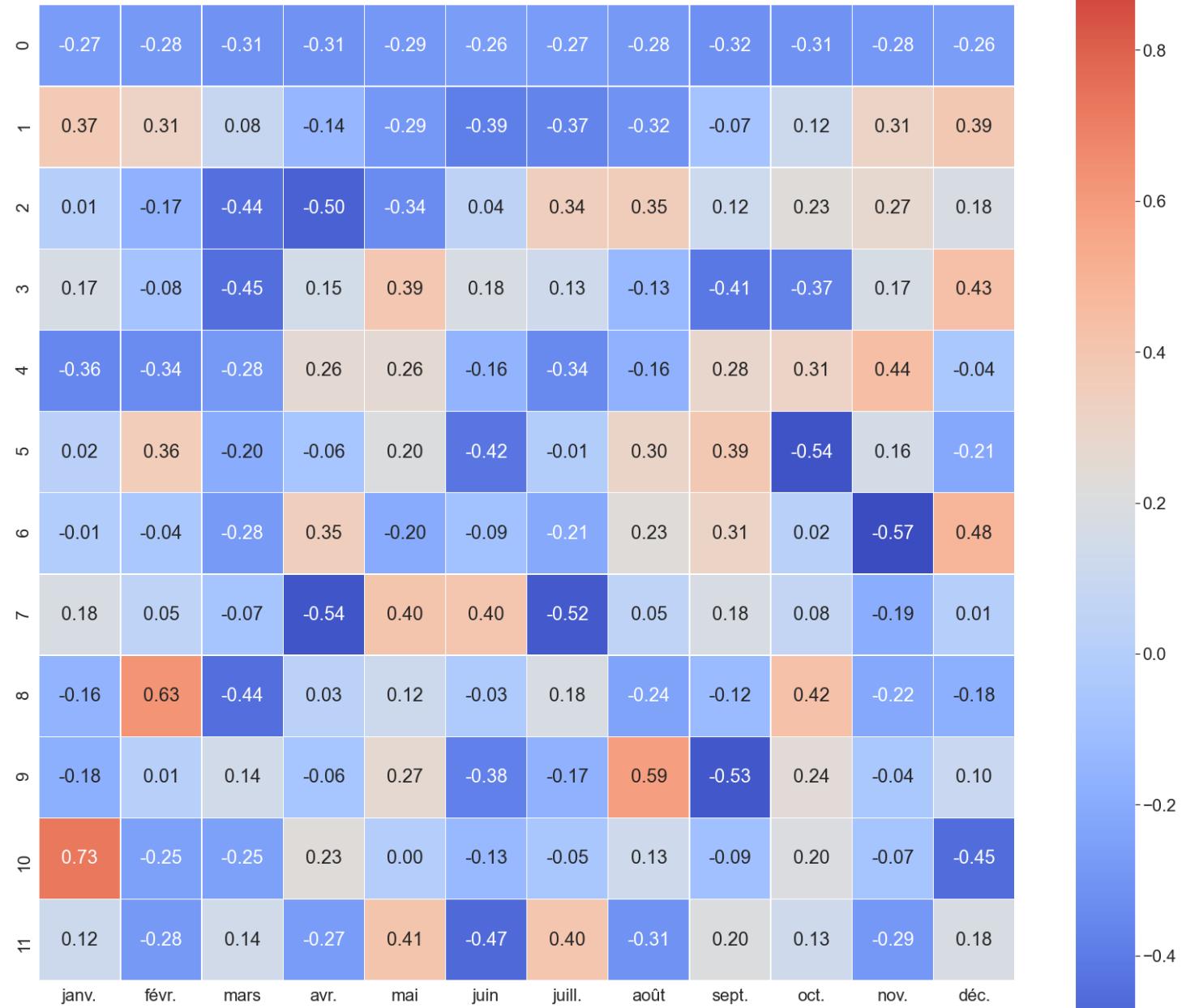
Meilleur plan contient le meilleur axe : on cherche $u_2 \perp u_1$ et maximisant

on peut chercher un 3ème axe, etc. d'inertie maximum

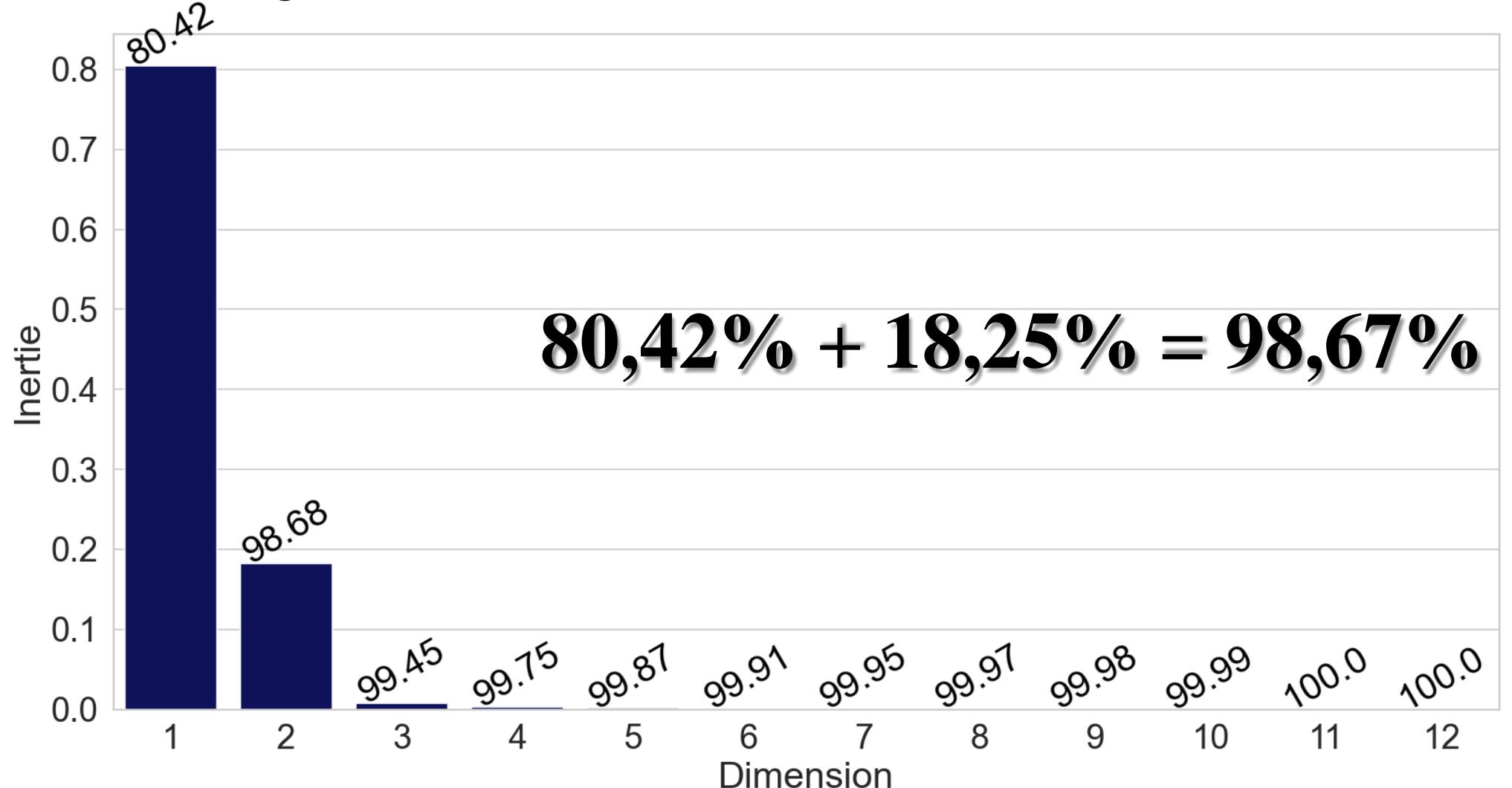
Graphe des individus



L'influence des variables dans les nouvelles dimensions



Pourcentage d'inertie

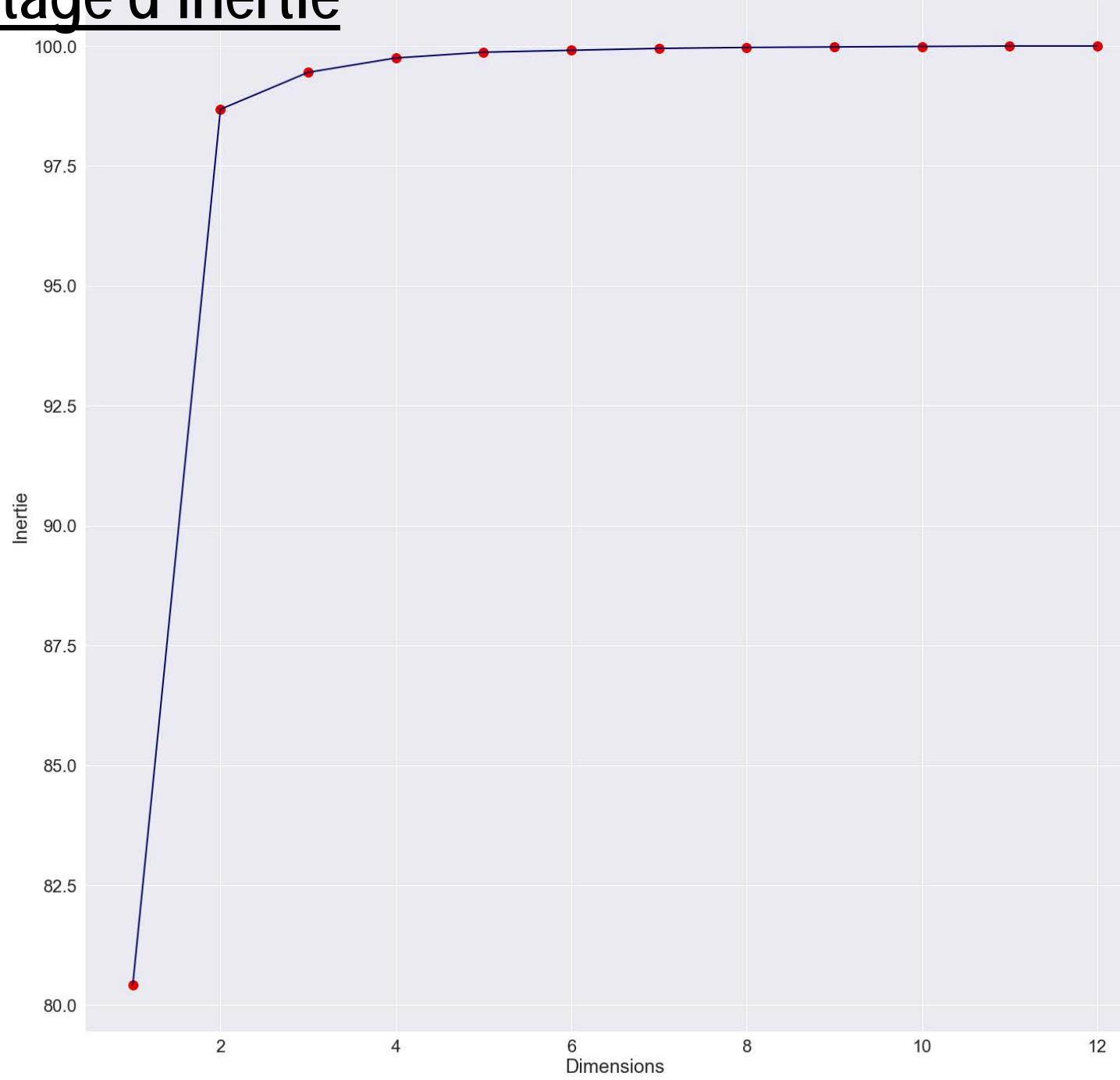


Comment quantifier la qualité d'une image ?

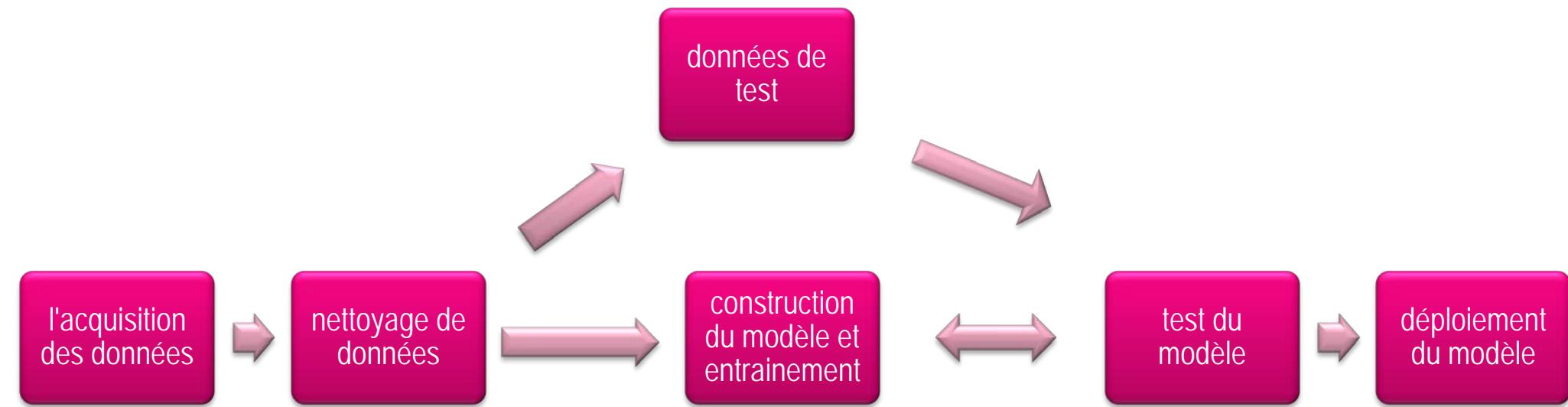
A l'aide de la notion de dispersion ou variabilité appelée **Inertie**

Inertie variance généralisée à plusieurs dimensions

Pourcentage d'inertie

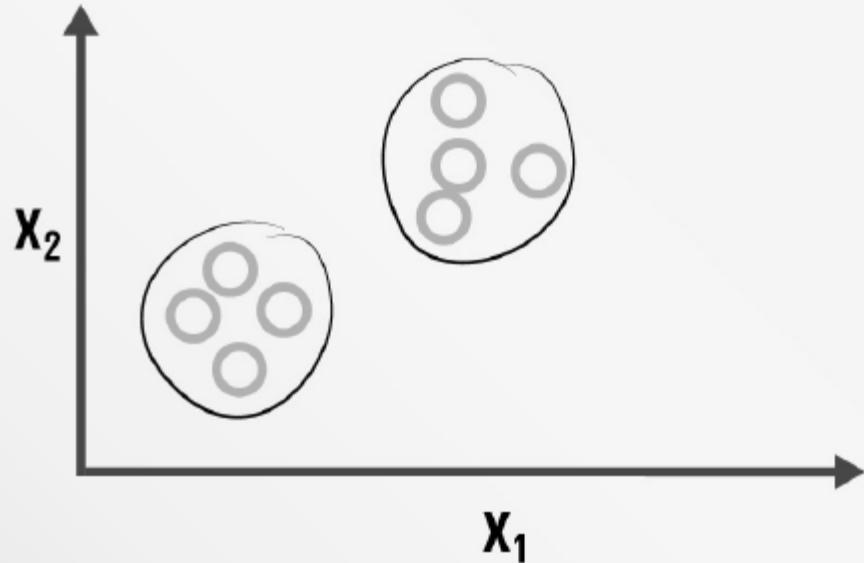


Module 4 : Vue d'ensemble du Machine Learning

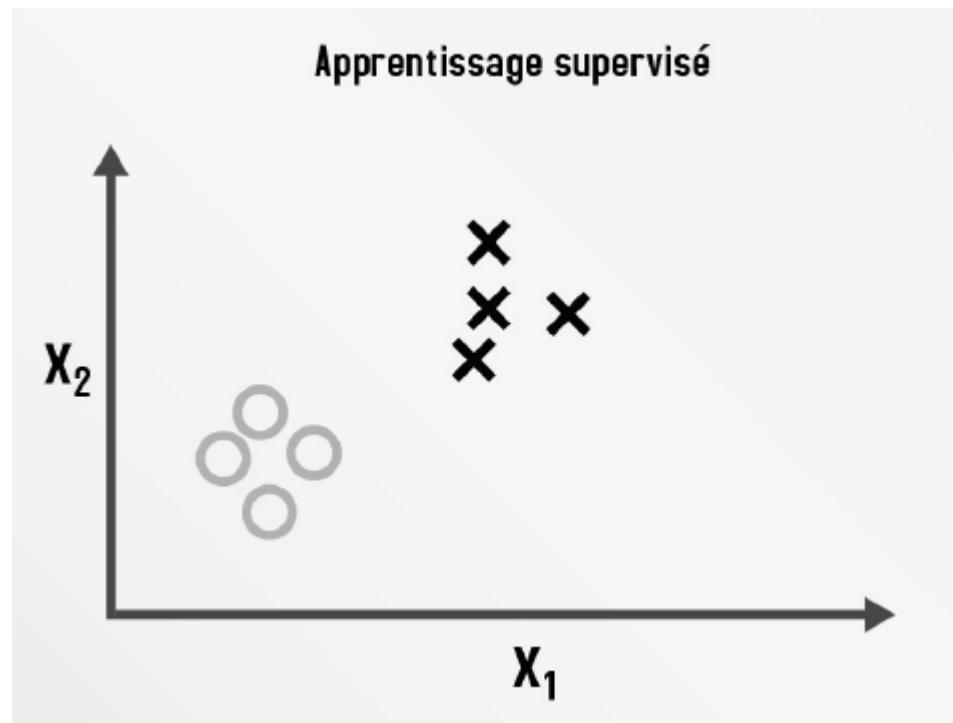


Algorithmes supervisés et non supervisés

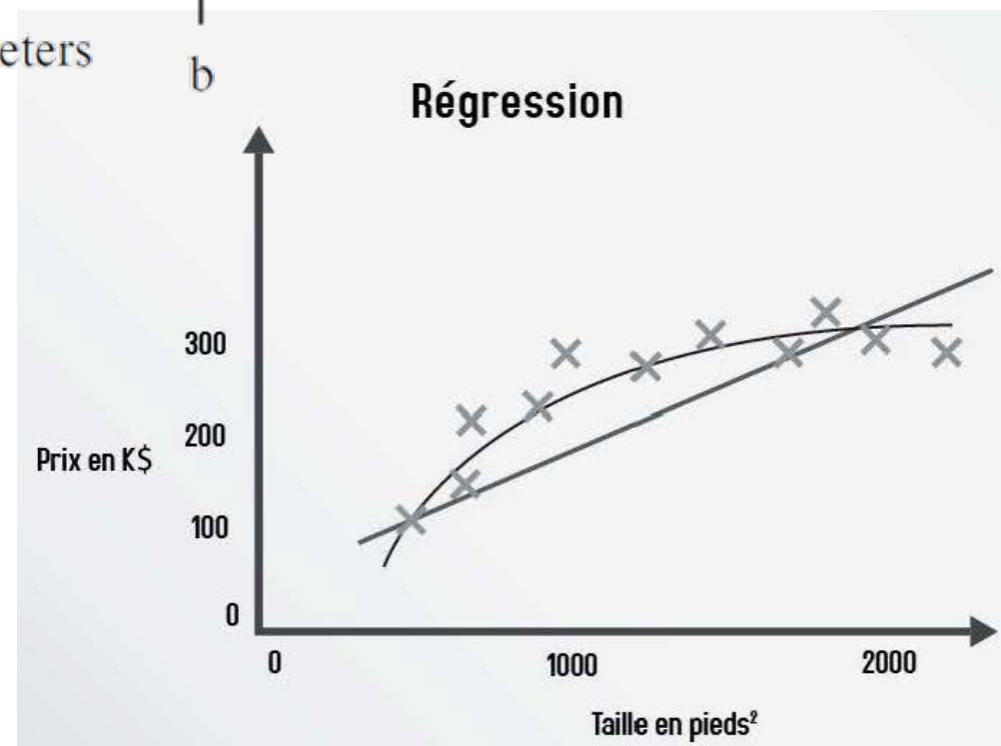
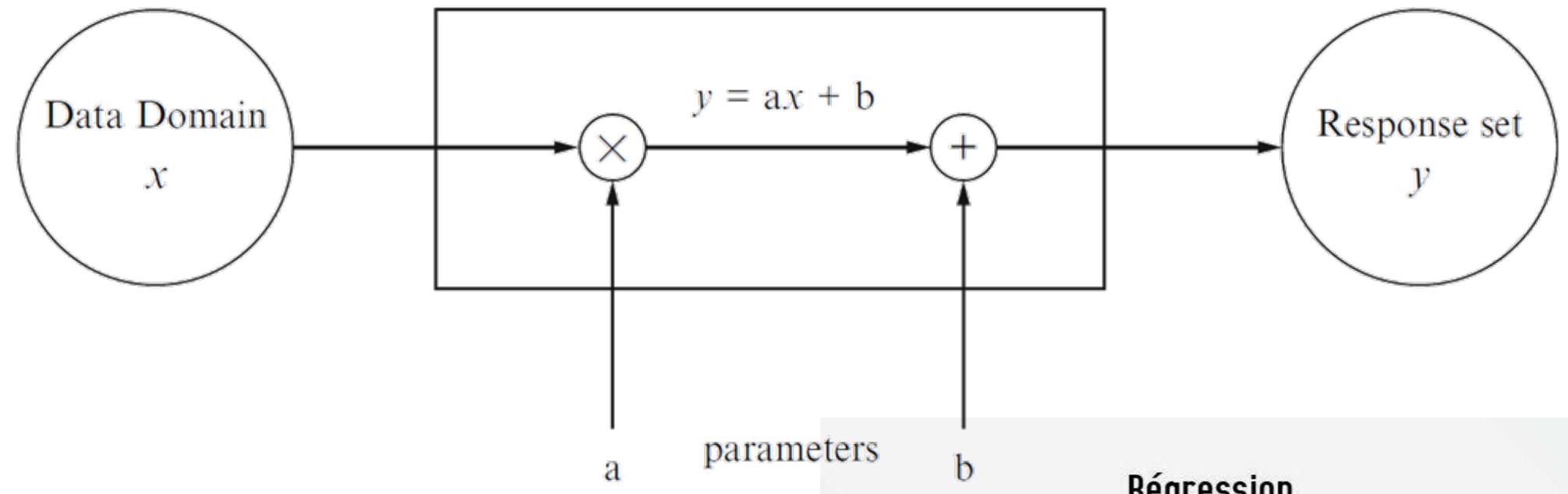
Apprentissage non supervisé



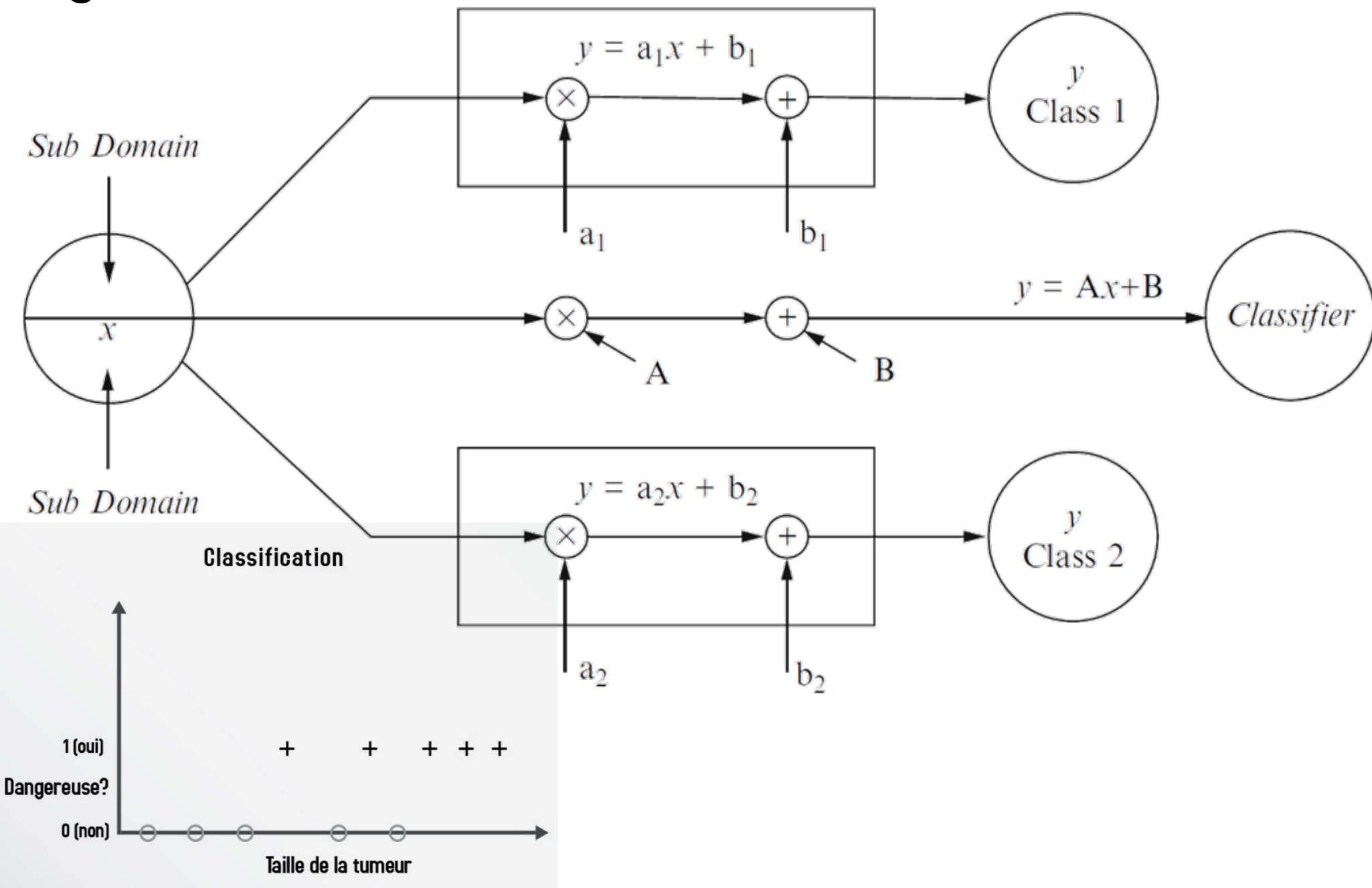
Apprentissage supervisé

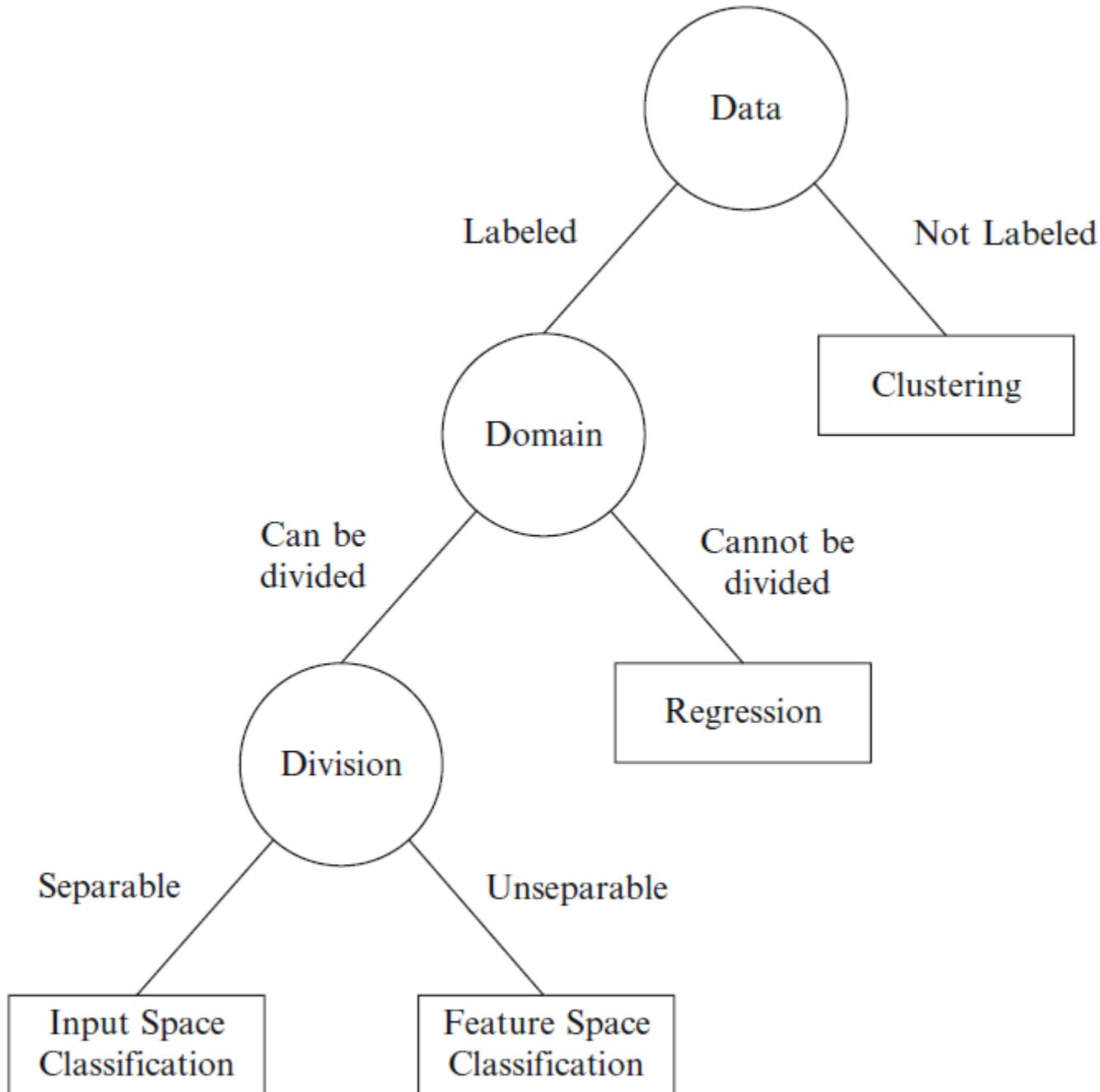


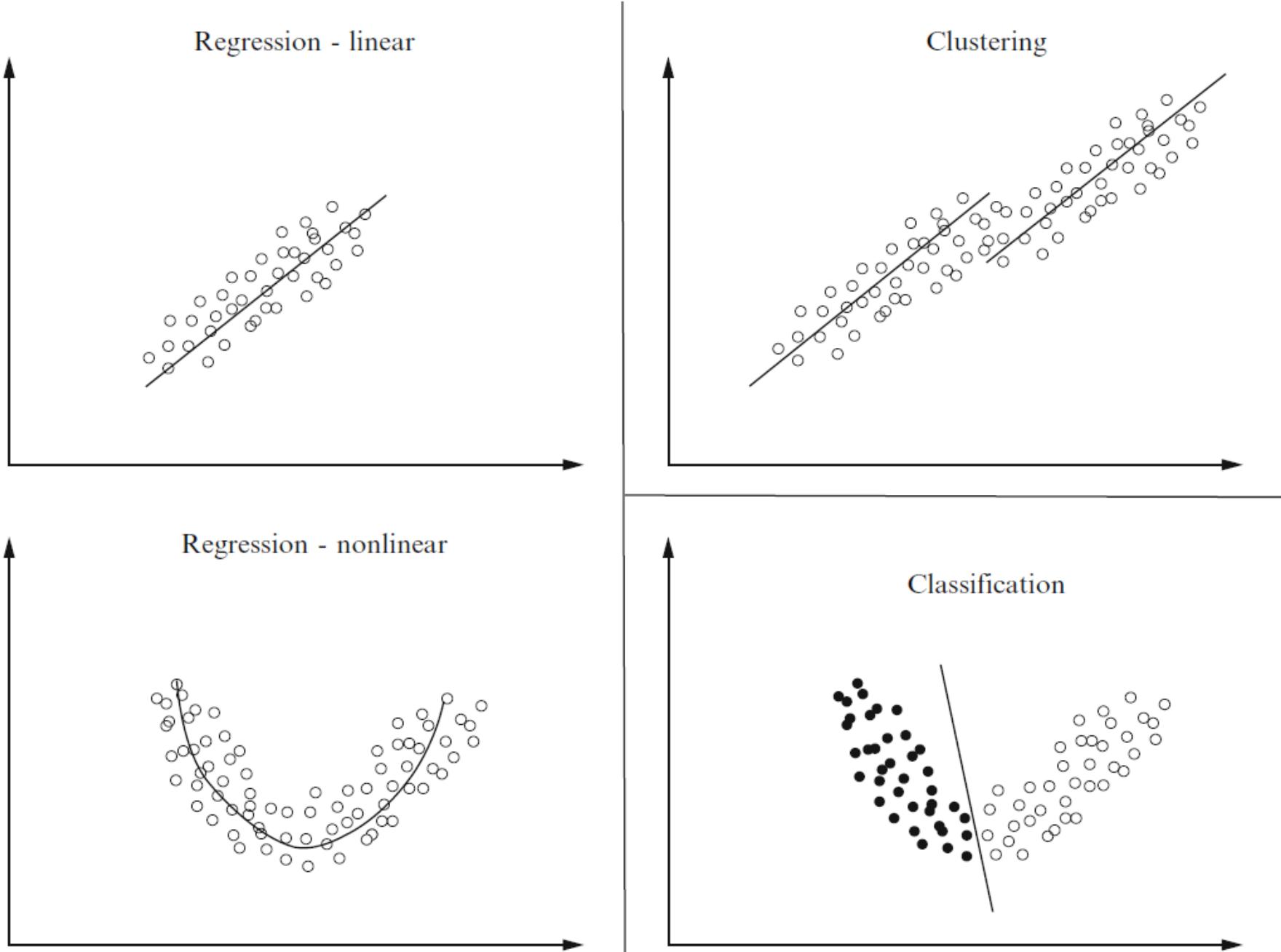
Algorithmes de régression



Algorithmes de classification

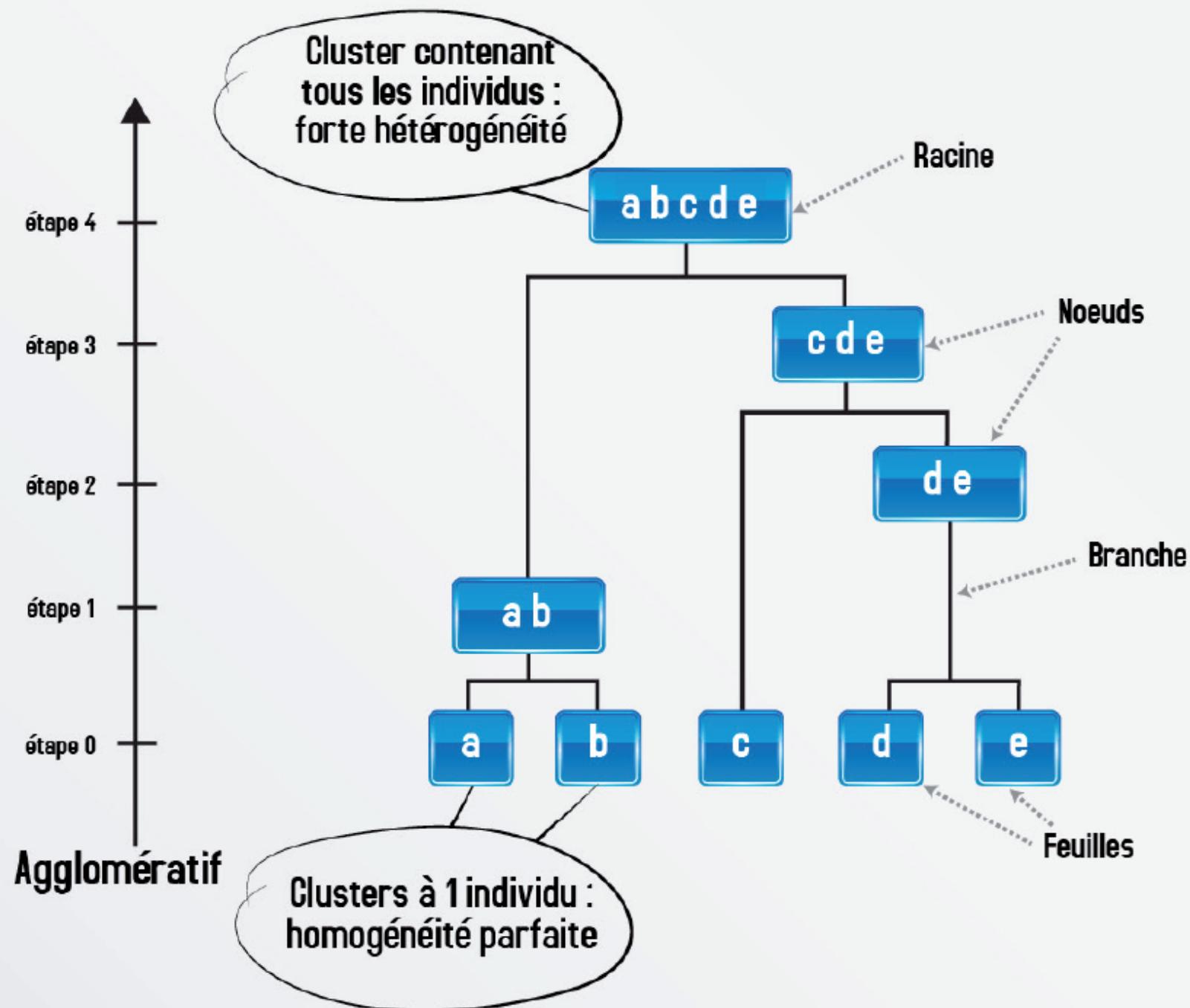






Module 5 : Classification ascendante hiérarchique

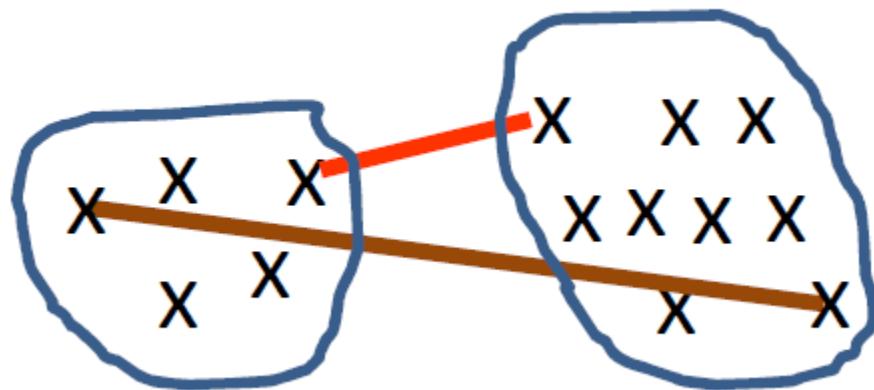
Divisif



Ressemblance entre individus

la distance euclidienne

$$d(x, y) = \sqrt{\sum_{i=1}^m (x_i - y_i)^2}.$$



la distance de Manhattan

$$d(x, y) = \sum_{i=1}^m |x_i - y_i|.$$

la distance de Minkowski

$$d(x, y) = \left(\sum_{i=1}^m |x_i - y_i|^p \right)^{\frac{1}{p}}.$$

la distance de Canberra

$$d(x, y) = \sum_{i=1}^m \frac{|x_i - y_i|}{|x_i + y_i|}.$$

la distance maximum

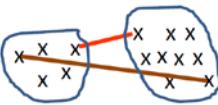
$$d(x, y) = \sup_{i \in \{1, \dots, m\}} |x_i - y_i|.$$

la distance de Jaccard

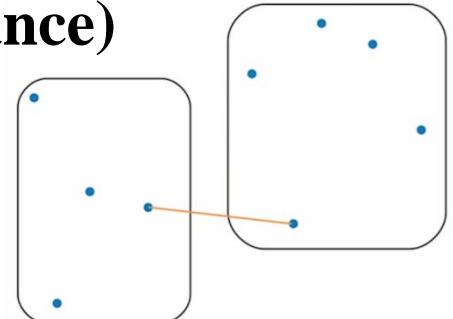
$$d(x_1, x_2) = 1 - J$$

$$J = \frac{M_{11}}{M_{01} + M_{10} + M_{11}} = \frac{M_{11}}{n - M_{00}}$$

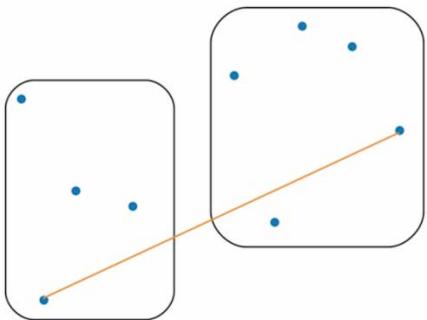
Ressemblance entre groupes d'individus



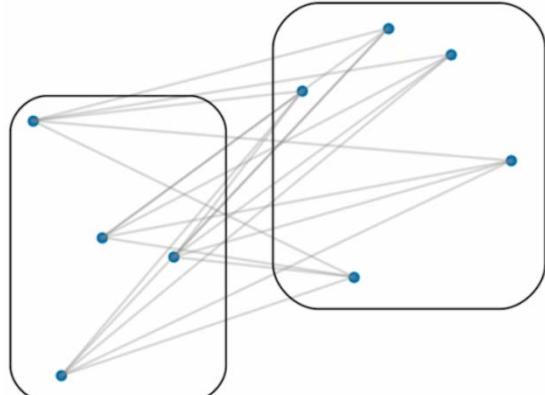
single - saut minimum ou lien simple (plus petite distance)



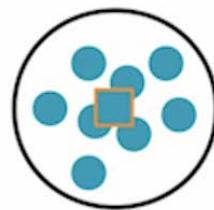
complete - lien complet (plus grande distance)



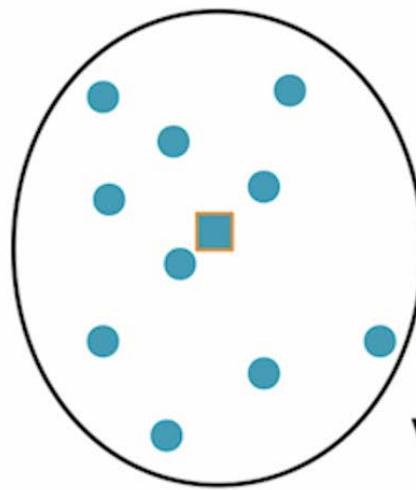
average - lien moyen



ward - critère de Ward



variance faible



variance élevée

$$\text{Var}_{\text{in}}(\mathcal{C}) = \frac{1}{|\mathcal{C}|} \sum_{x \in \mathcal{C}} \|x - \mu_{\mathcal{C}}\|^2$$

Algorithme

7^e regroupement

ABC	4.07
DEFGH	

4

{ABCDEFGH}

6^e regroupement

ABC	DE
DE	4.72
FGH	4.07 1.81

3

5^e regroupement

ABC	DE	FG
DE	4.72	
FG	4.23 1.81	
H	4.07 2.90	1.12

2

4^e regroupement

ABC	D	E	FG
D	4.72		
E	5.55 1.00		
FG	4.07 2.01 1.81		
H	4.75 3.16 2.90 1.12		

1

3^e regroupement

ABC	D	E	F	G
D	4.72			
E	5.55 1.00			
F	4.07 2.01 2.06			
G	4.68 2.06 1.81 0.61			
H	4.75 3.16 2.90 1.28 1.12			

0

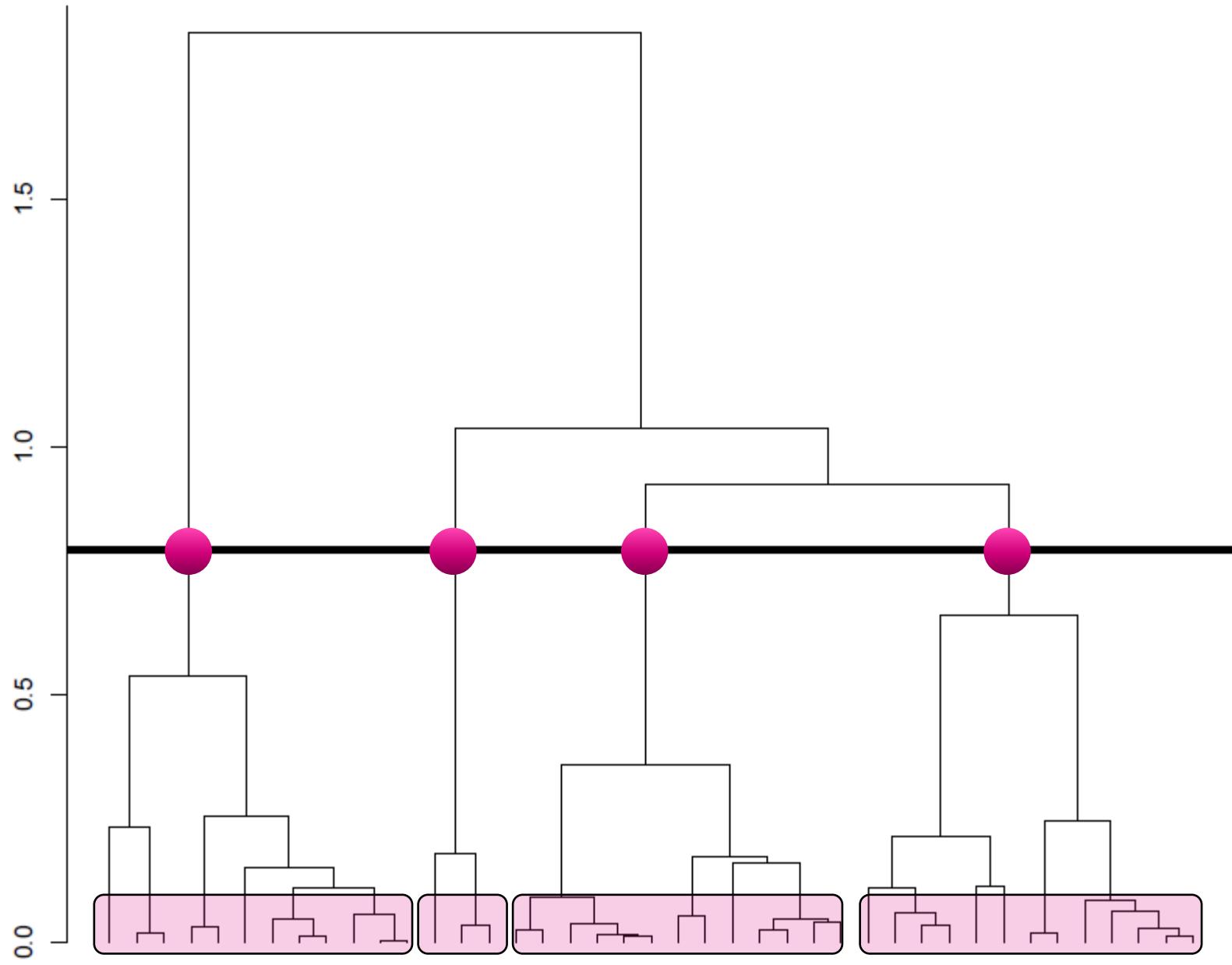
2^e regroupement

AC	B	D	E	F	G
B	0.50				
D	4.80 4.72				
E	5.57 5.55 1.00				
F	4.07 4.23 2.01 2.06				
G	4.68 4.84 2.06 1.81 0.61				
H	4.75 5.02 3.16 2.90 1.28 1.12				

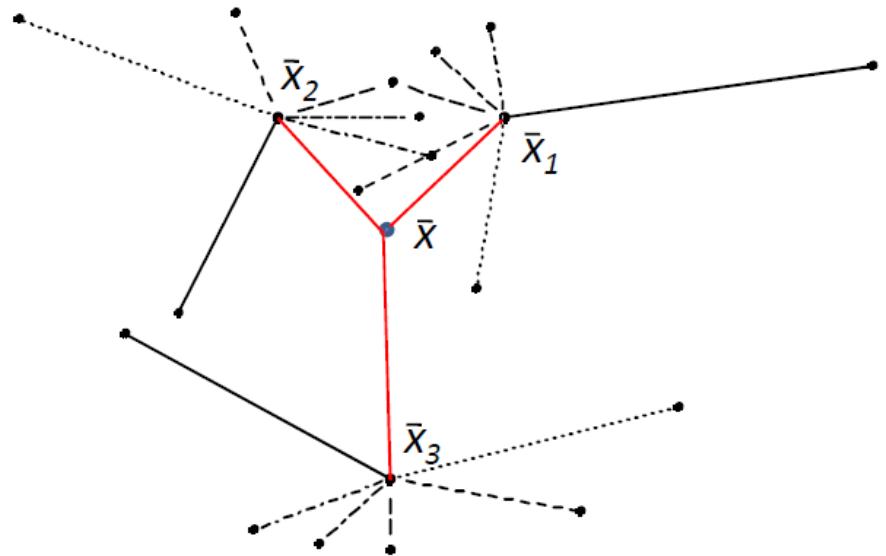
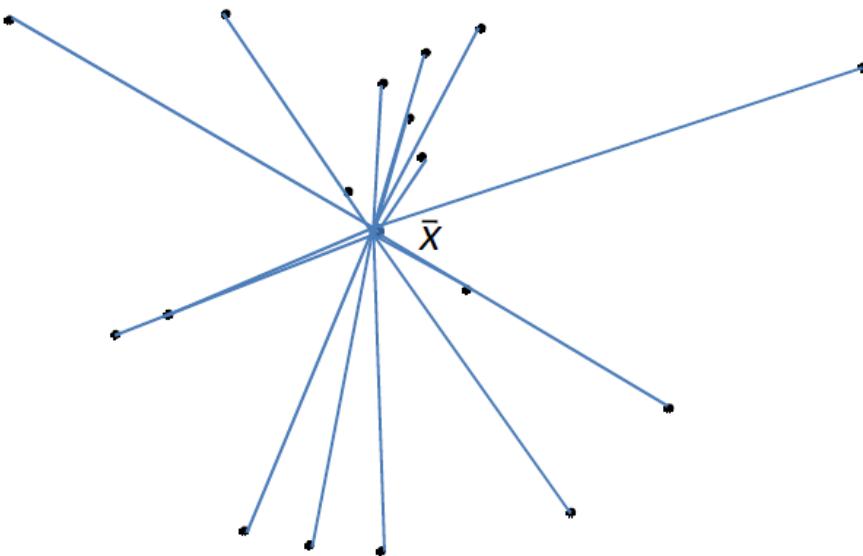
1

A	B	C	D	E	F	G
B	0.50					
C	0.25 0.56					
D	5.00 4.72 4.80					
E	5.78 5.55 5.57 1.00					
F	4.32 4.23 4.07 2.01 2.06					
G	4.92 4.84 4.68 2.06 1.81 0.61					
H	5.00 5.02 4.75 3.16 2.90 1.28 1.12					

1^{er} regroupement



$$\underbrace{\sum_{k=1}^K \sum_{q=1}^Q \sum_{i=1}^I (x_{iqk} - \bar{x}_k)^2}_{\text{Inertie totale}} = \underbrace{\sum_{k=1}^K \sum_{q=1}^Q \sum_{i=1}^I (x_{iqk} - \bar{x}_{qk})^2}_{\text{Inertie intra}} + \underbrace{\sum_{k=1}^K \sum_{q=1}^Q \sum_{i=1}^I (\bar{x}_{qk} - \bar{x}_k)^2}_{\text{Inertie inter}}$$

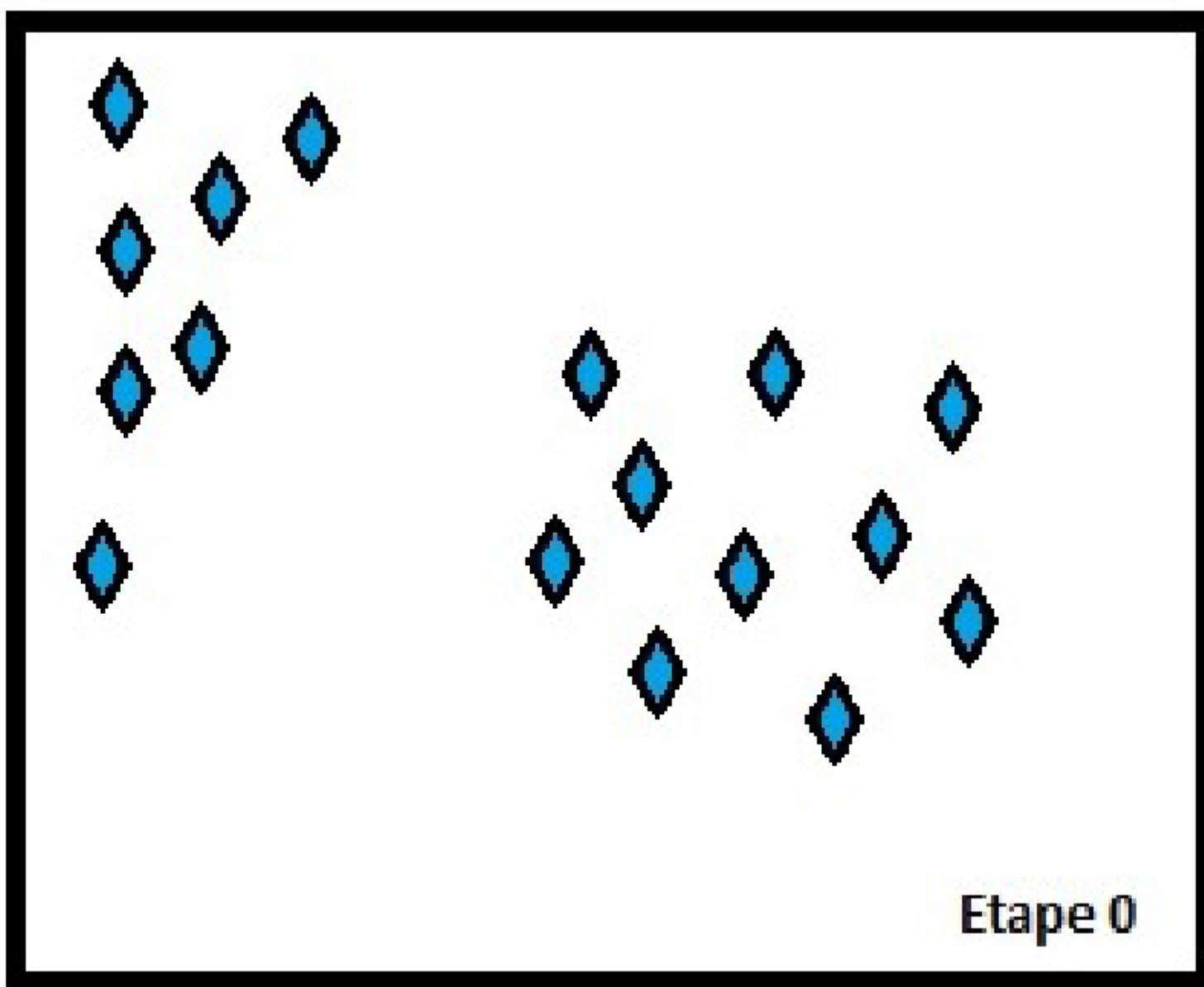


Module 6 : Algorithme des centres mobiles

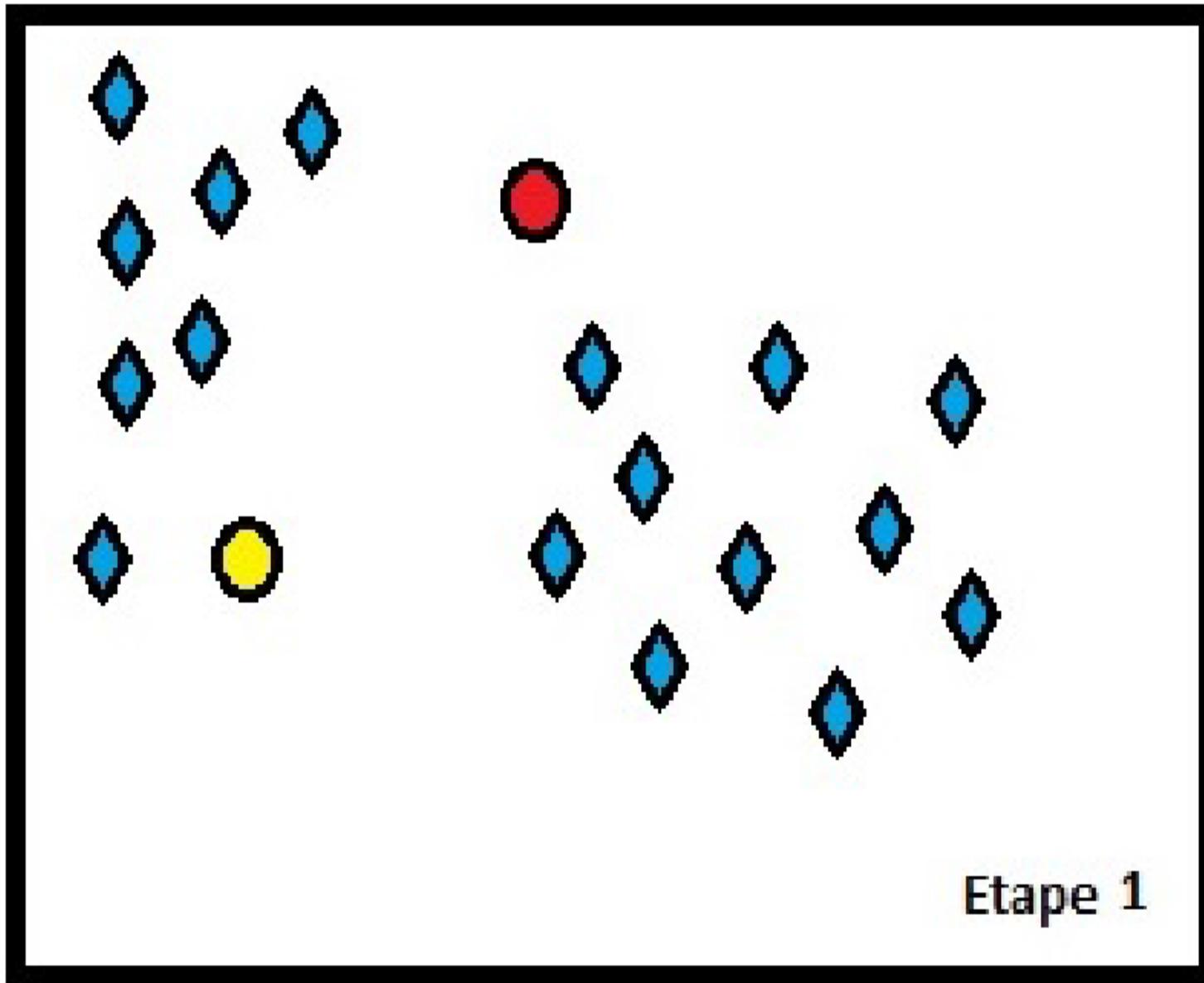
Algorithme des centres mobiles (k means)

- On choisit q points au hasard.
- On calcule le tableau de distances entre tous les individus et les q centres.
- On forme alors q partitions des groupes des individus les plus proches de ce centre que d'un autre.
- On calcule le centre de gravité de chacun des q groupes.
- Les q centres de gravité sont nos nouveaux q centres.
- On itère la procédure précédente jusqu'à ce que deux itérations conduisent à la même partition.

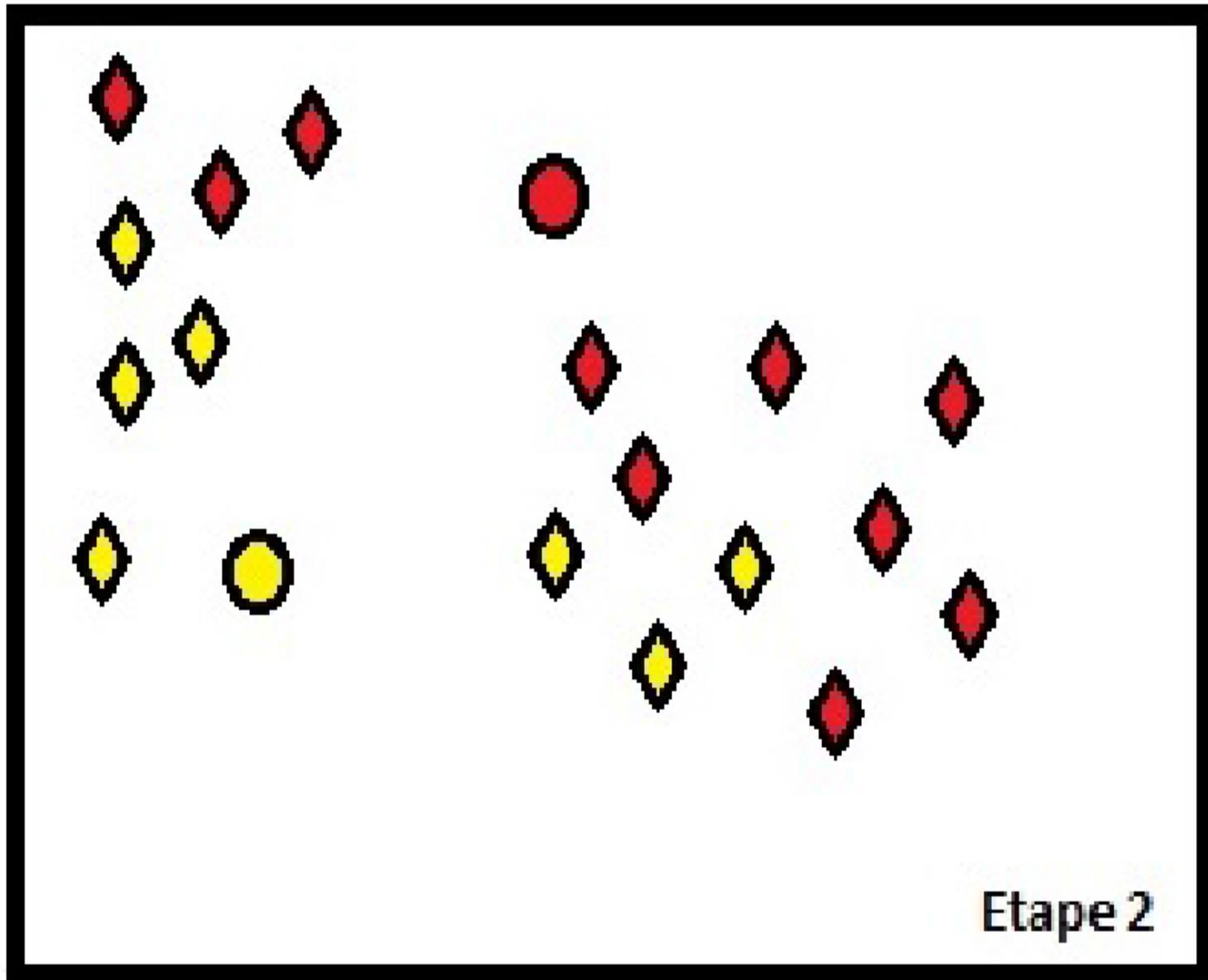
Algorithme de Lloyd



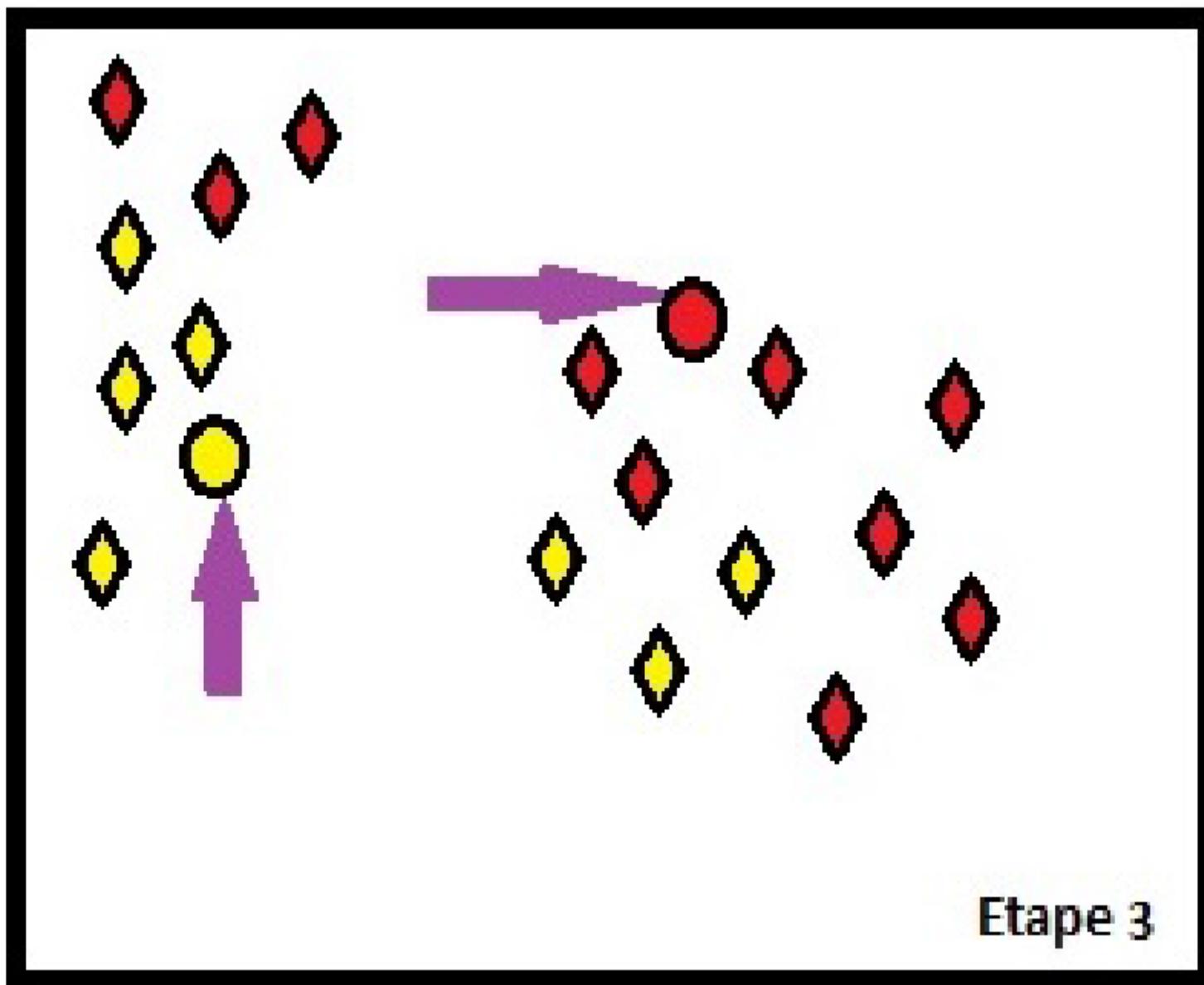
Algorithme de Lloyd



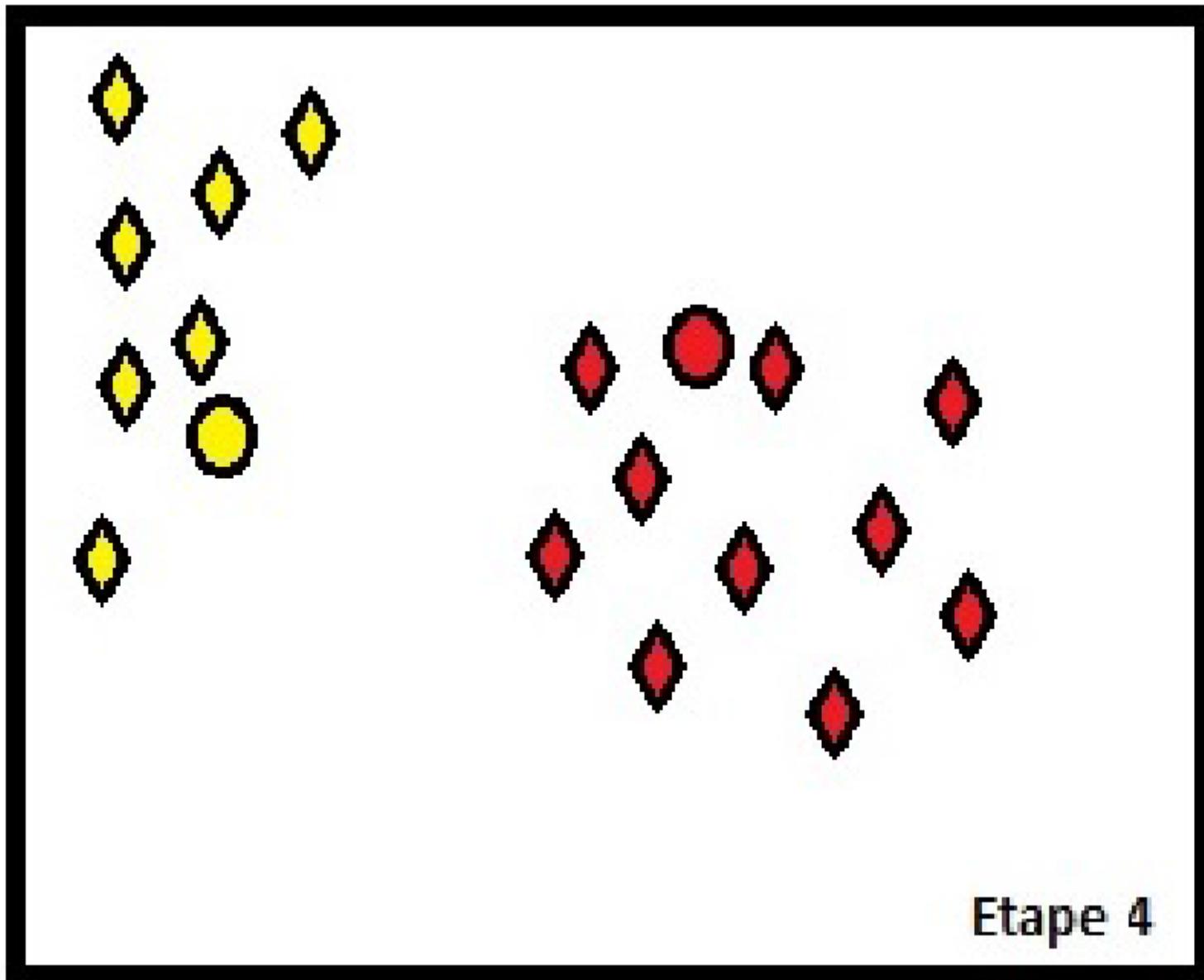
Algorithme de Lloyd



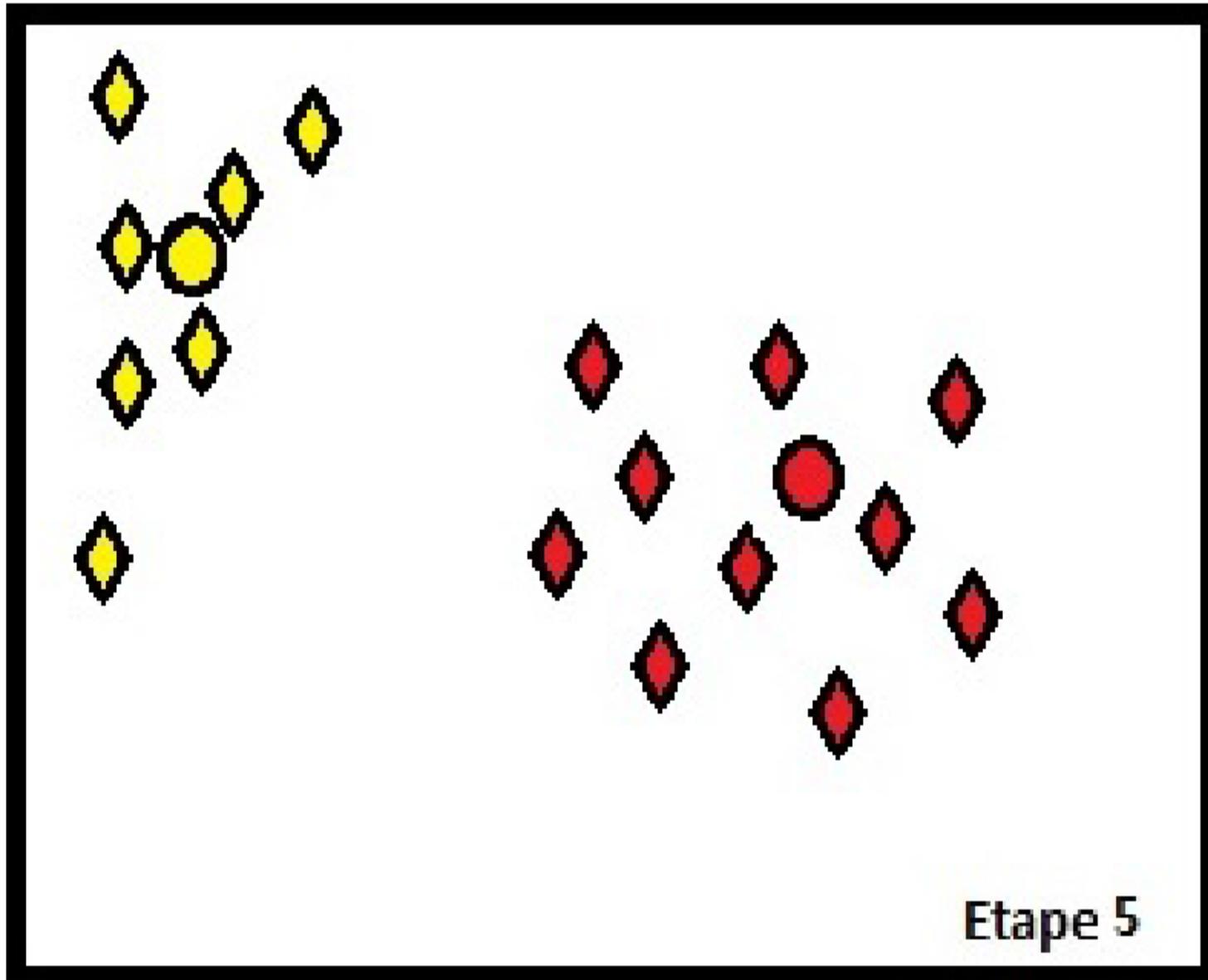
Algorithme de Lloyd

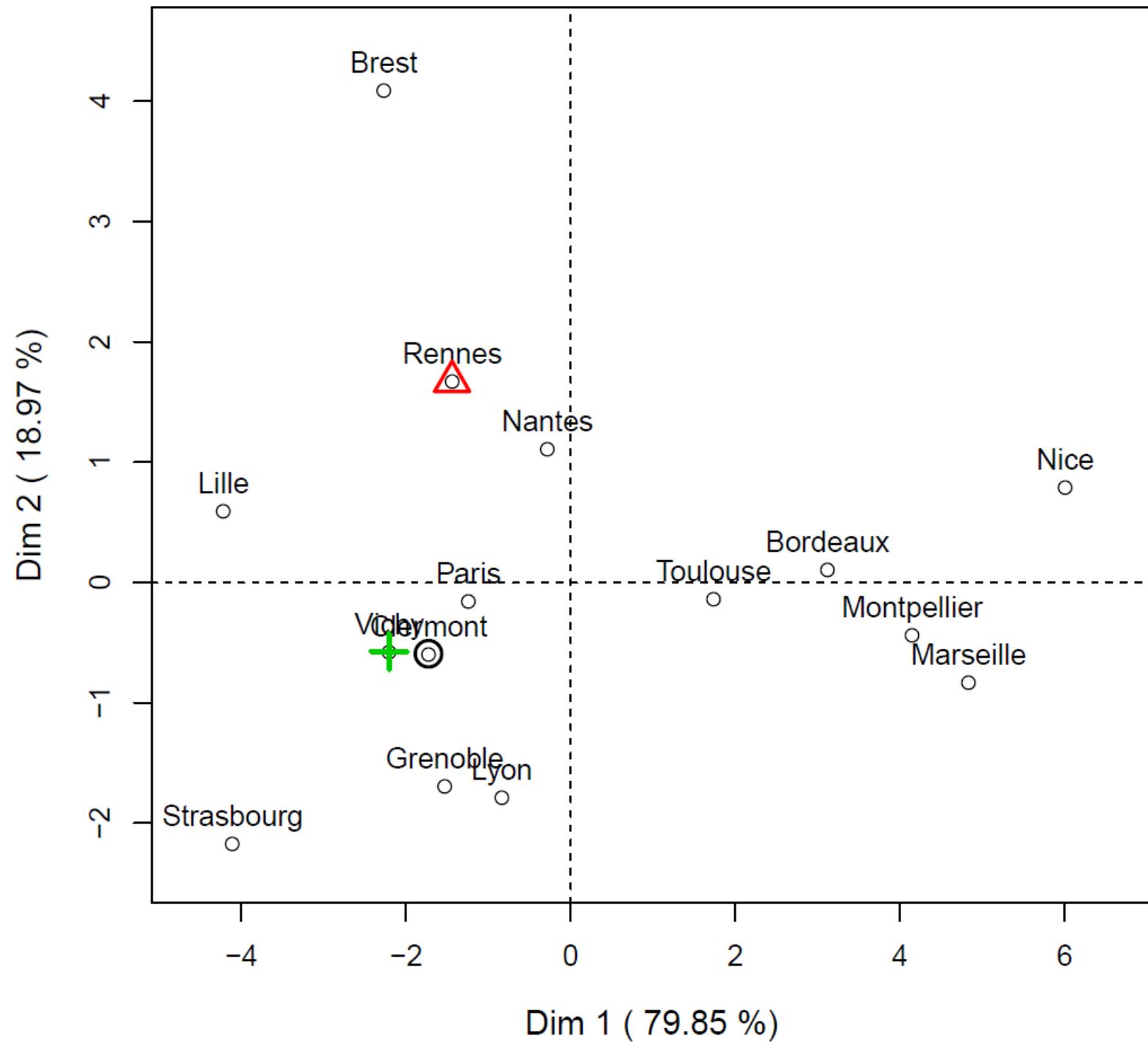


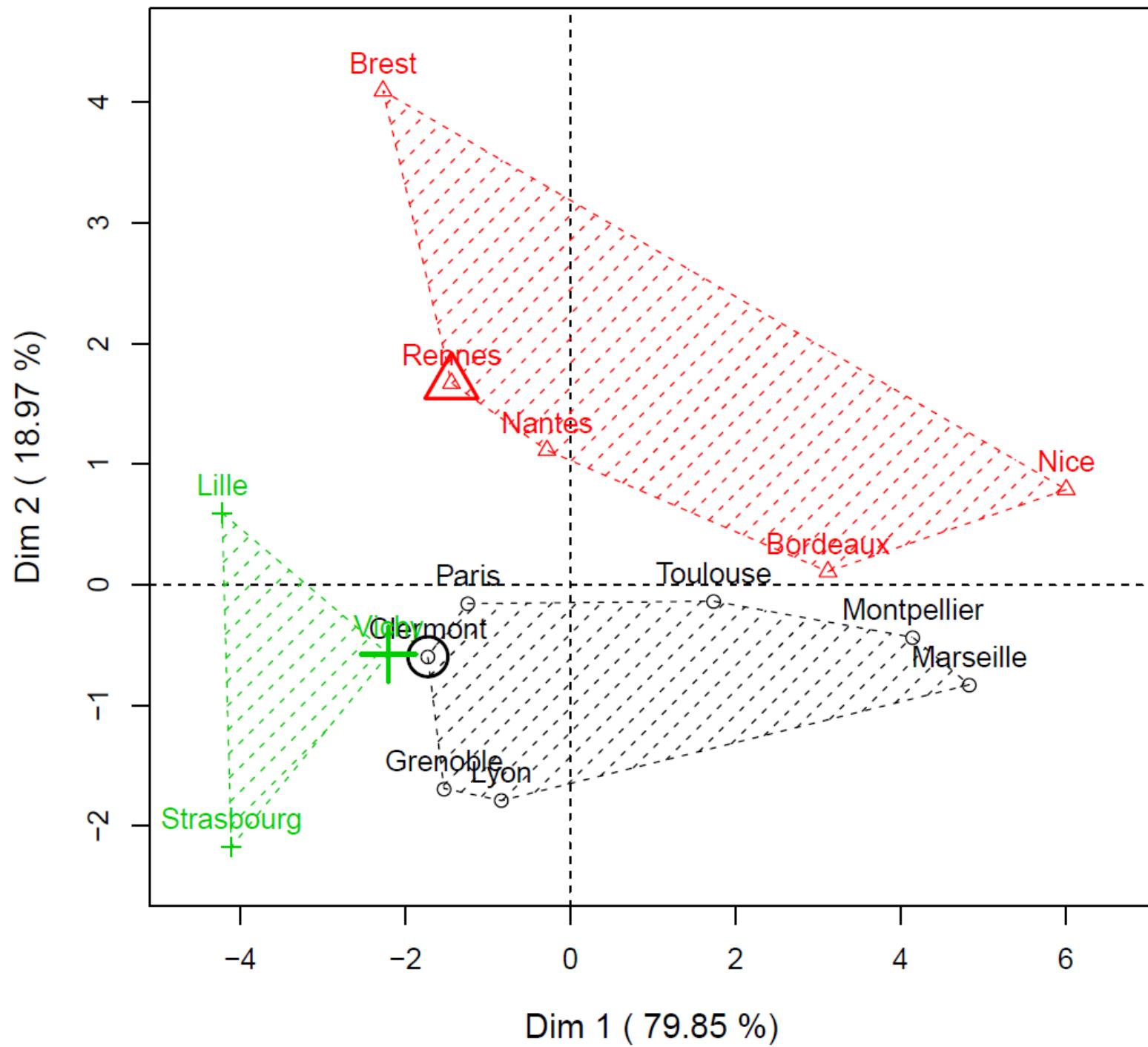
Algorithme de Lloyd

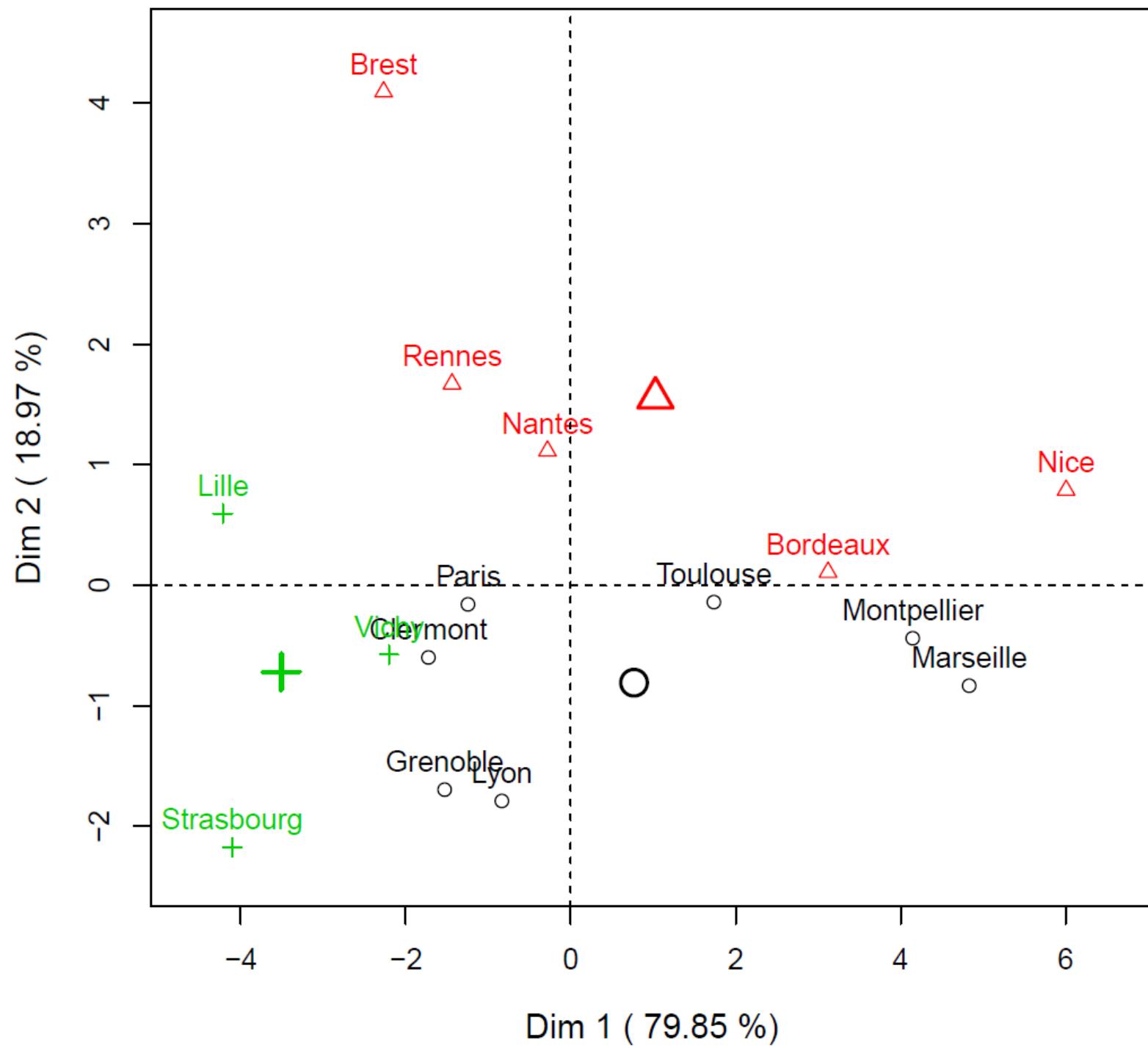


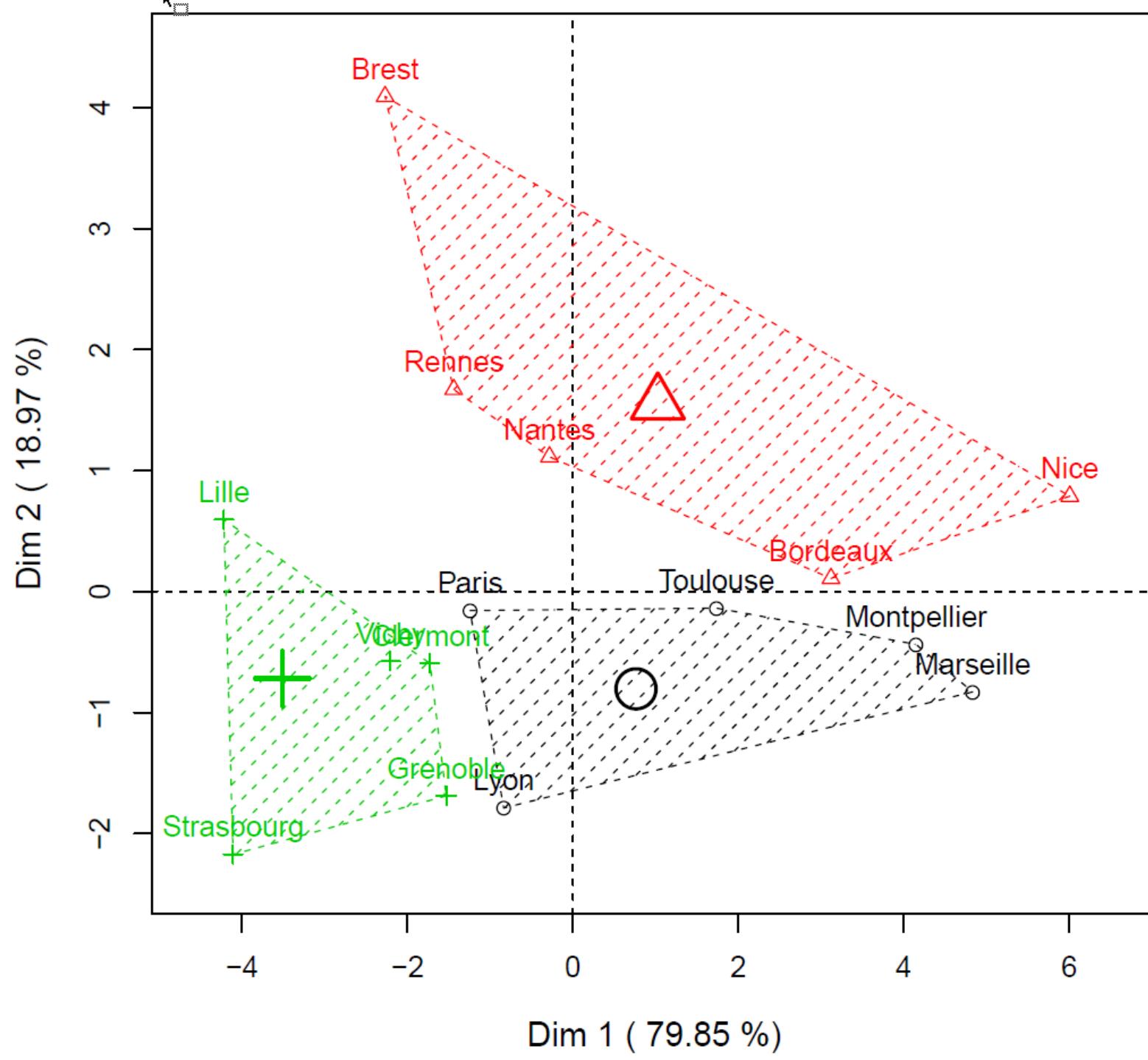
Algorithme de Lloyd

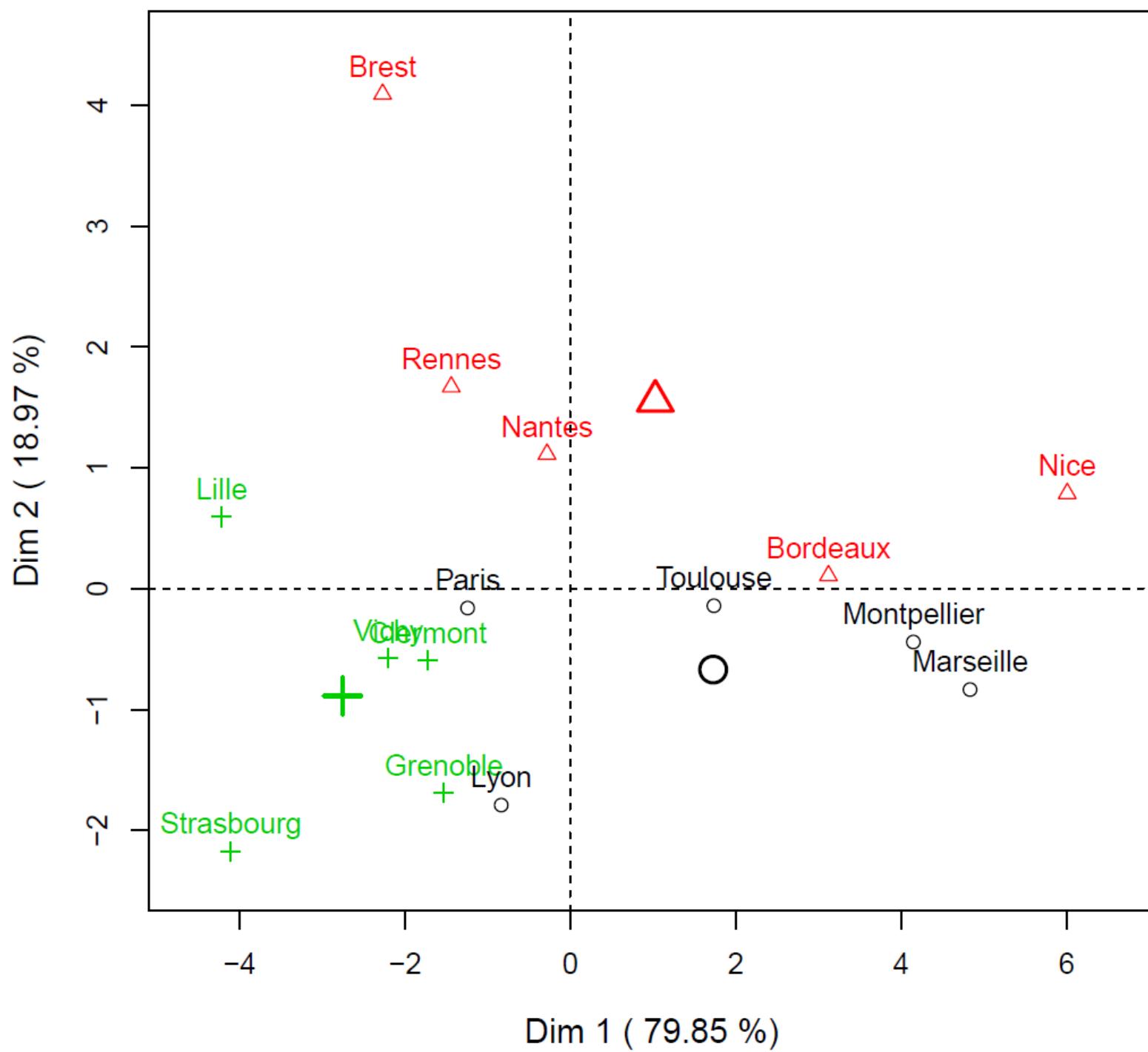


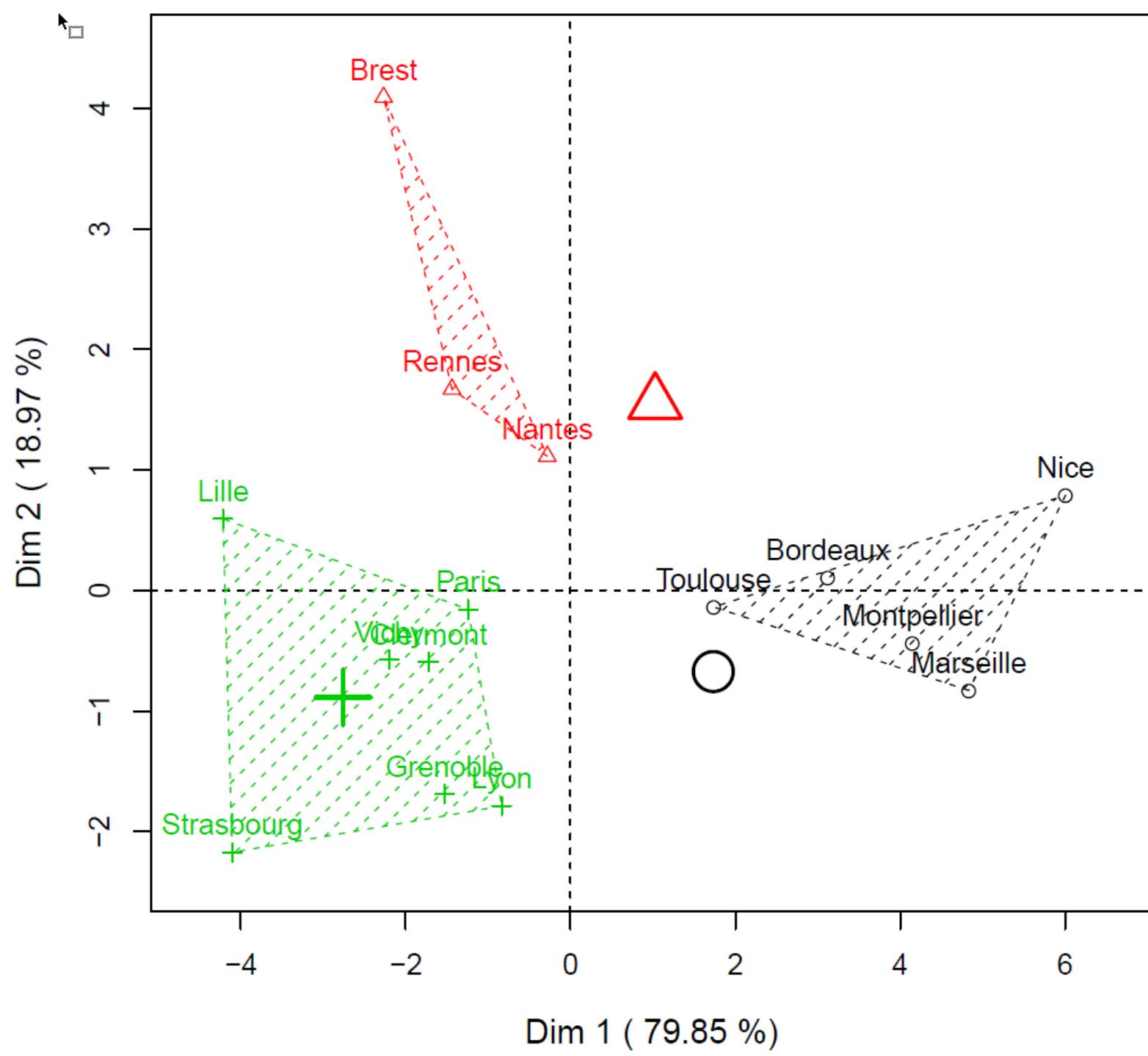


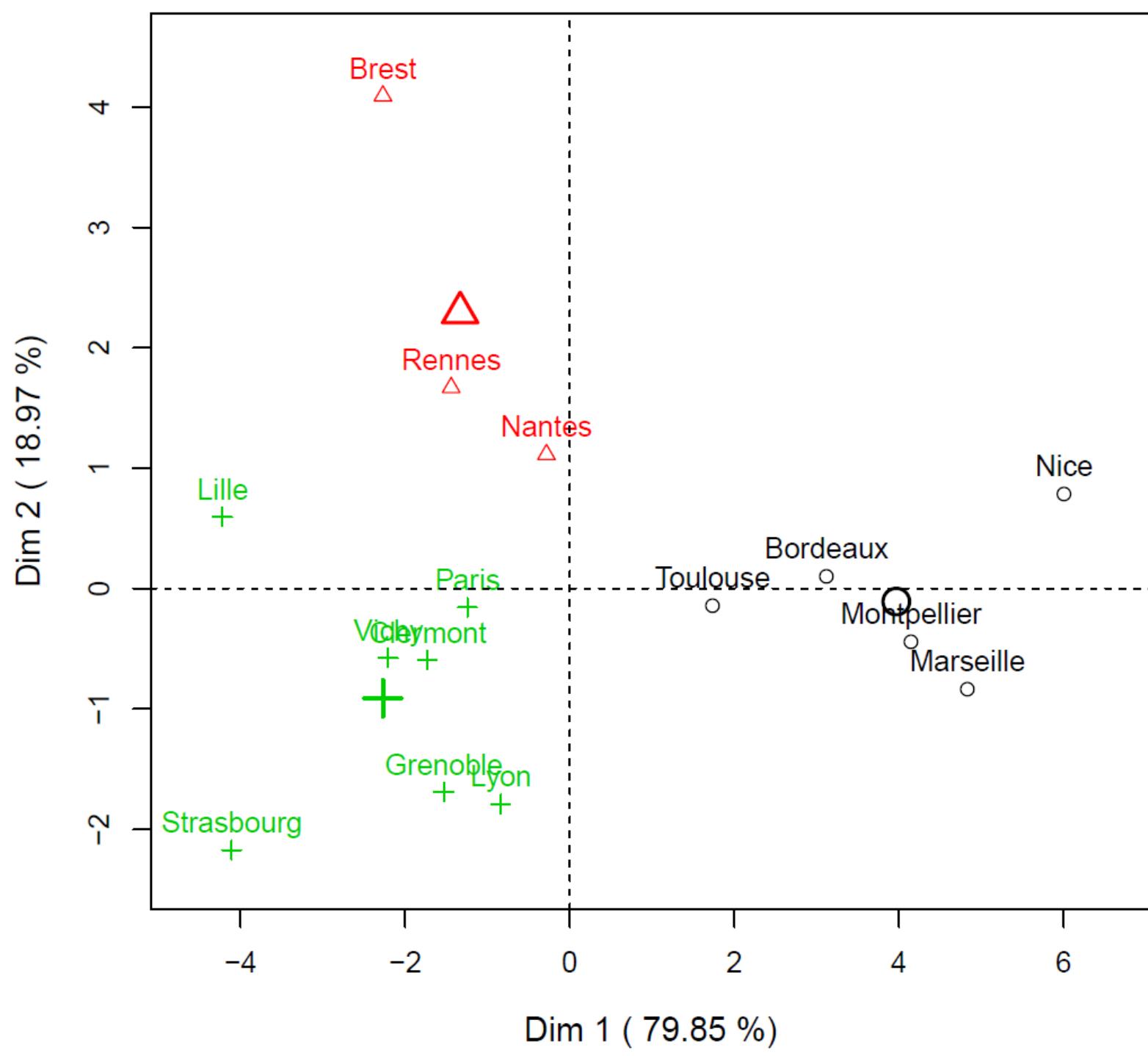


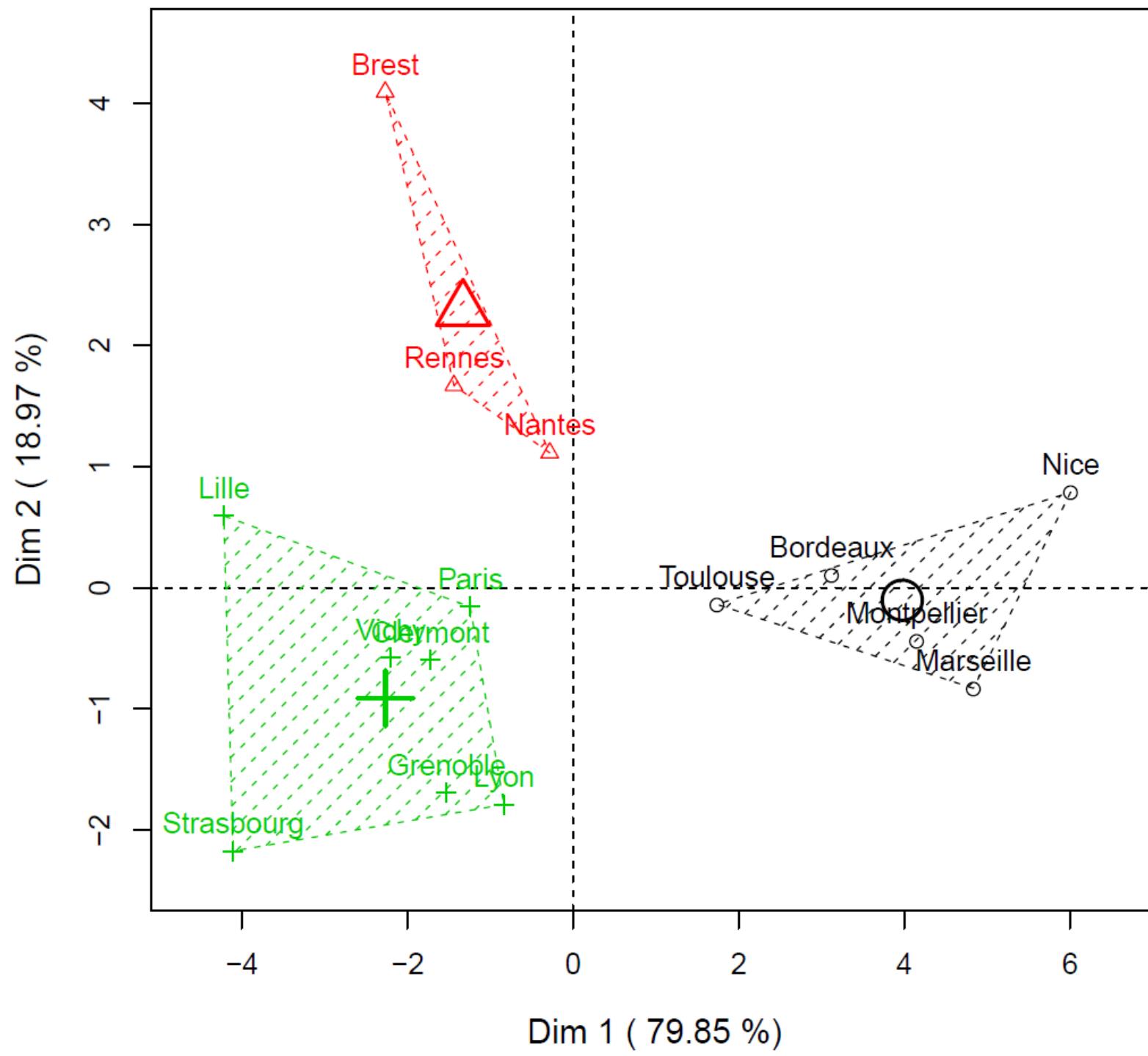


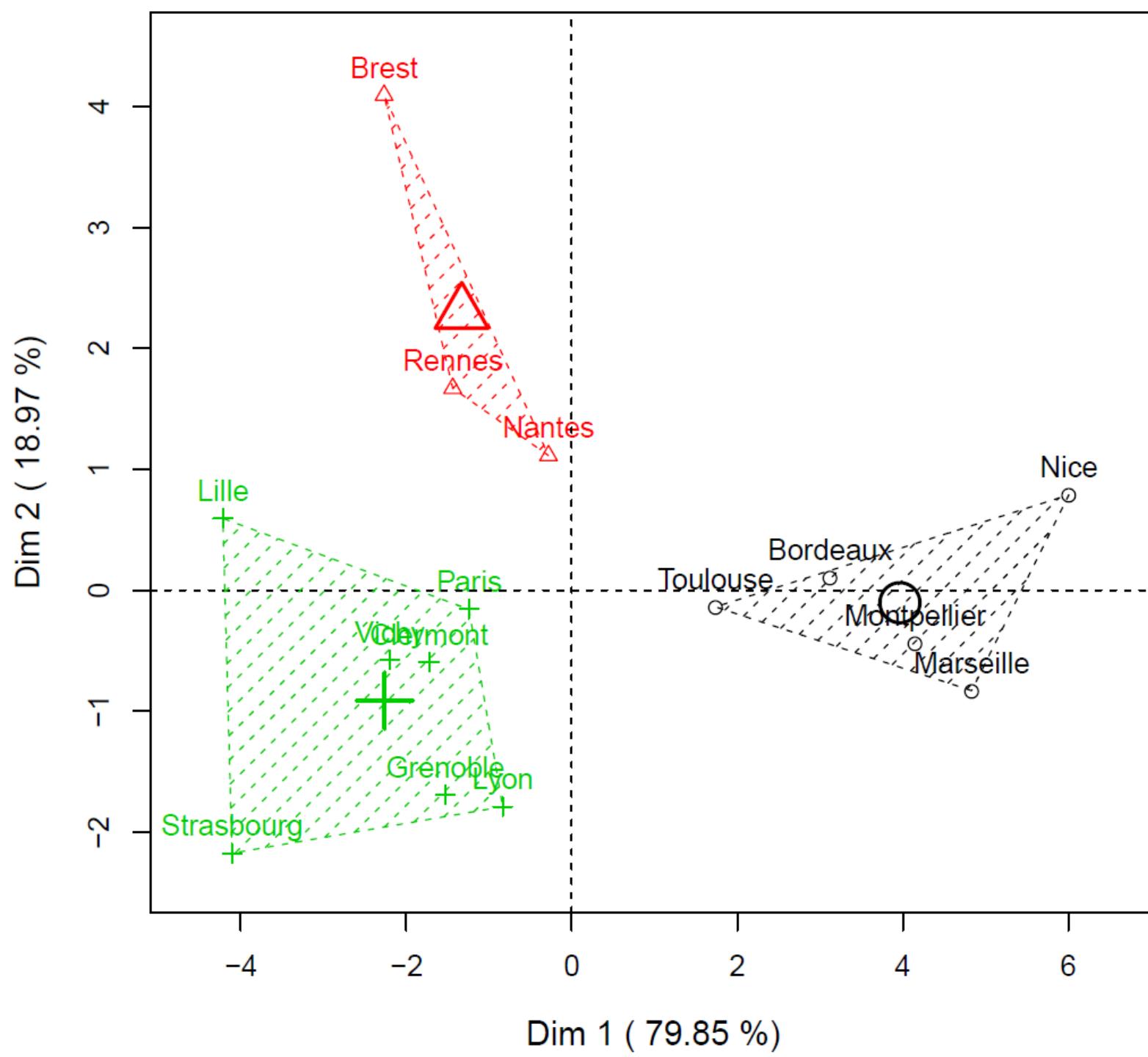


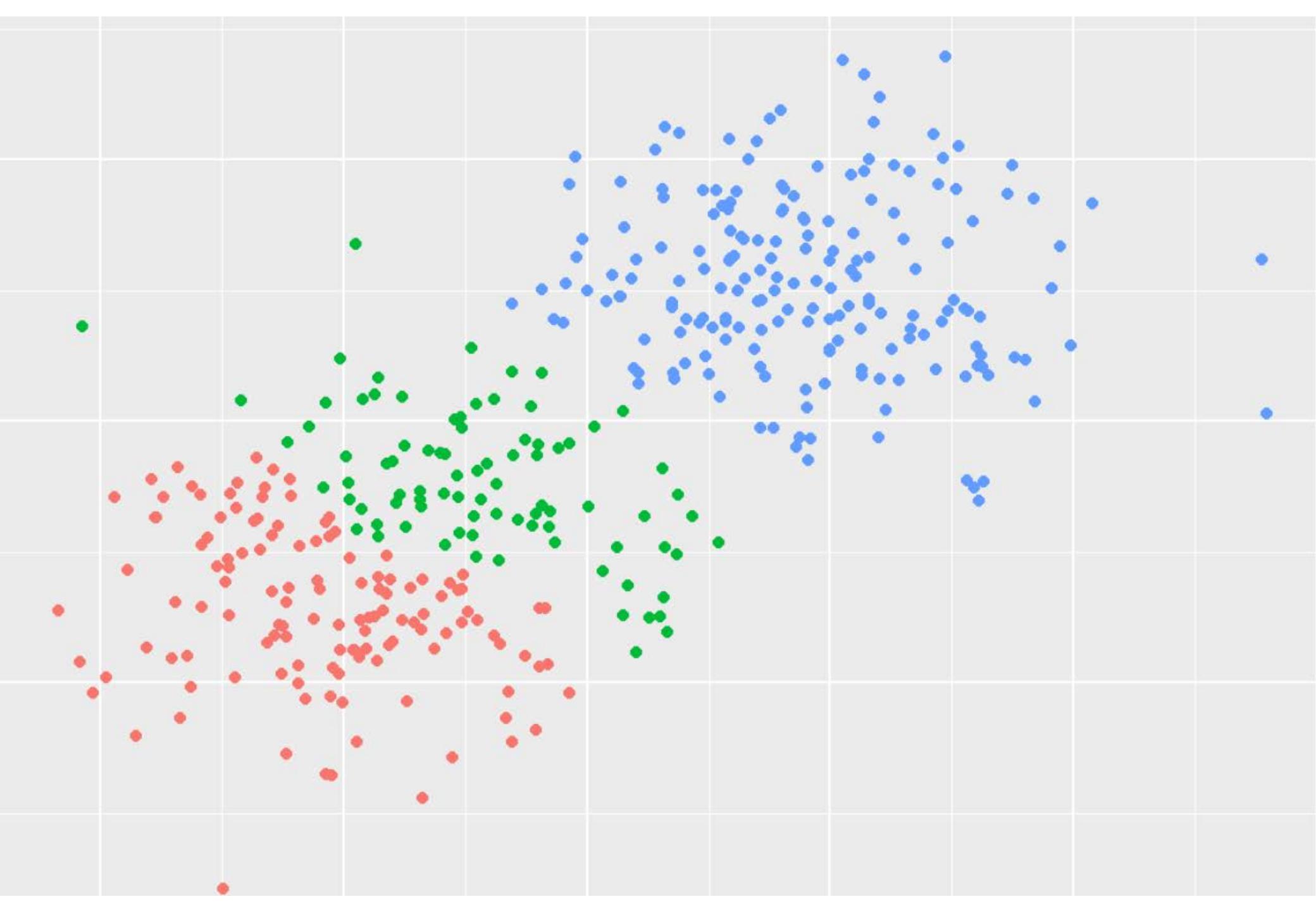




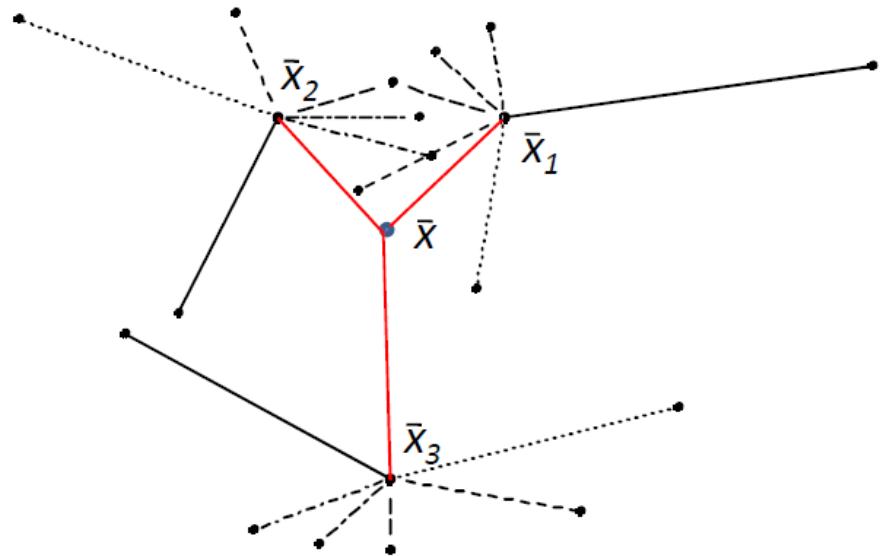
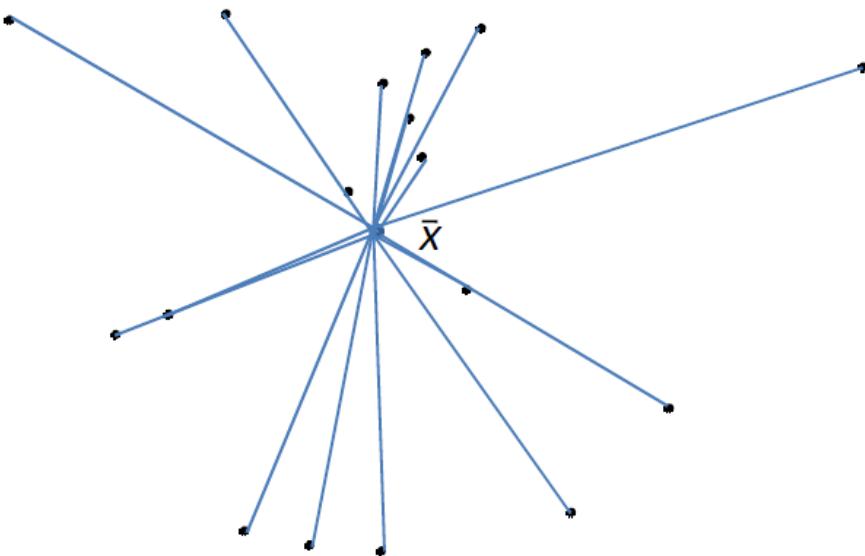




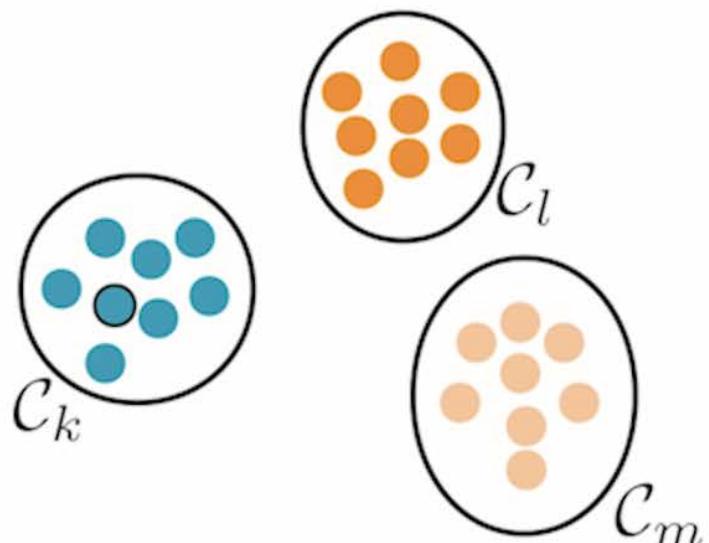




$$\underbrace{\sum_{k=1}^K \sum_{q=1}^Q \sum_{i=1}^I (x_{iqk} - \bar{x}_k)^2}_{\text{Inertie totale}} = \underbrace{\sum_{k=1}^K \sum_{q=1}^Q \sum_{i=1}^I (x_{iqk} - \bar{x}_{qk})^2}_{\text{Inertie intra}} + \underbrace{\sum_{k=1}^K \sum_{q=1}^Q \sum_{i=1}^I (\bar{x}_{qk} - \bar{x}_k)^2}_{\text{Inertie inter}}$$



Silhouette

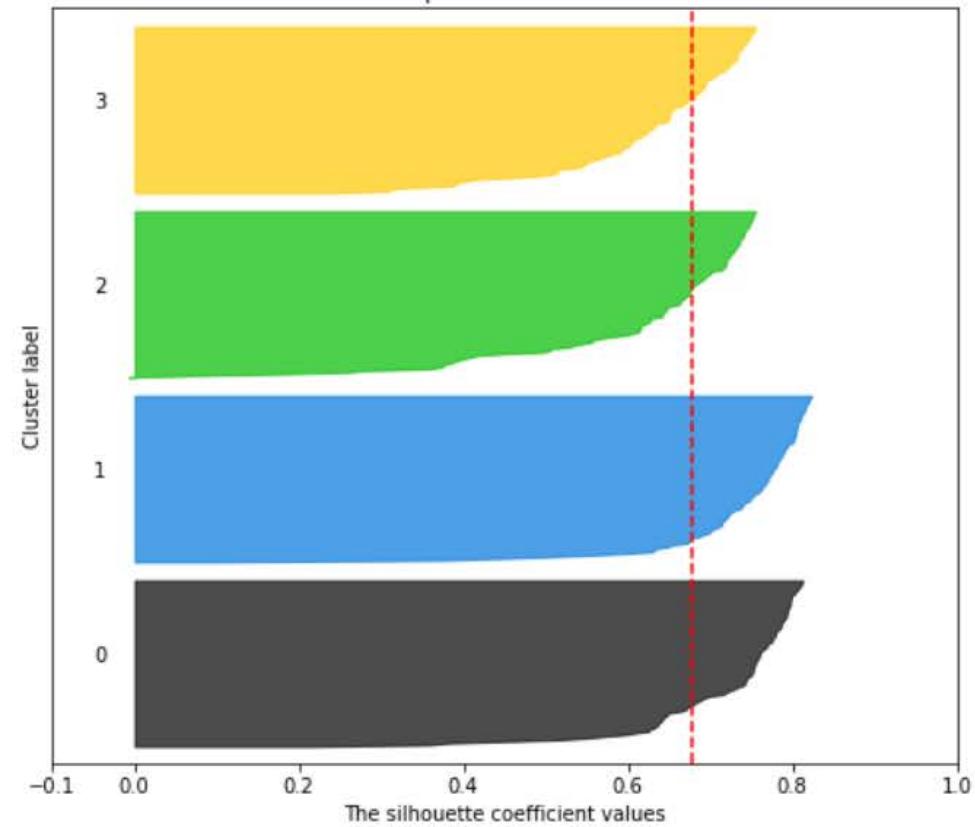


$$a(\mathbf{x}) = \frac{1}{n_k - 1} \sum_{\mathbf{u} \in \mathcal{C}_k, \mathbf{u} \neq \mathbf{x}} d(\mathbf{x}, \mathbf{u})$$

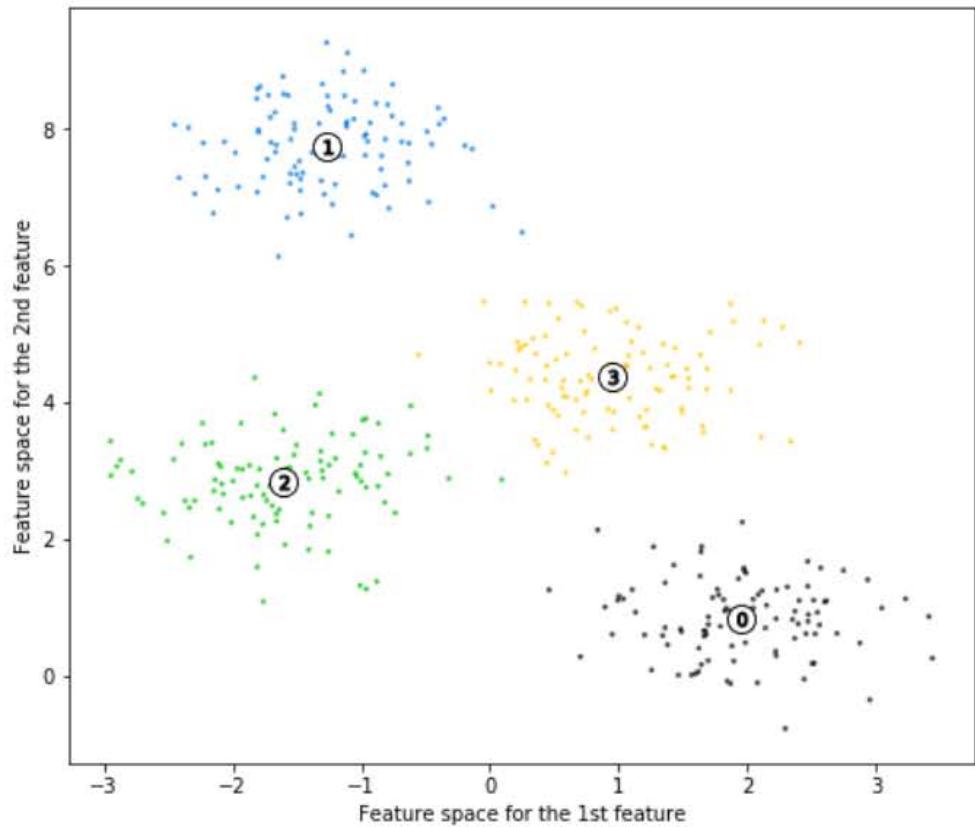
$$b(\mathbf{x}) = \min_{l \neq k} \frac{1}{n_l - 1} \sum_{\mathbf{u} \in \mathcal{C}_l} d(\mathbf{x}, \mathbf{u})$$

$$s(\mathbf{x}) = \frac{b(\mathbf{x}) - a(\mathbf{x})}{\max(a(\mathbf{x}), b(\mathbf{x}))}$$

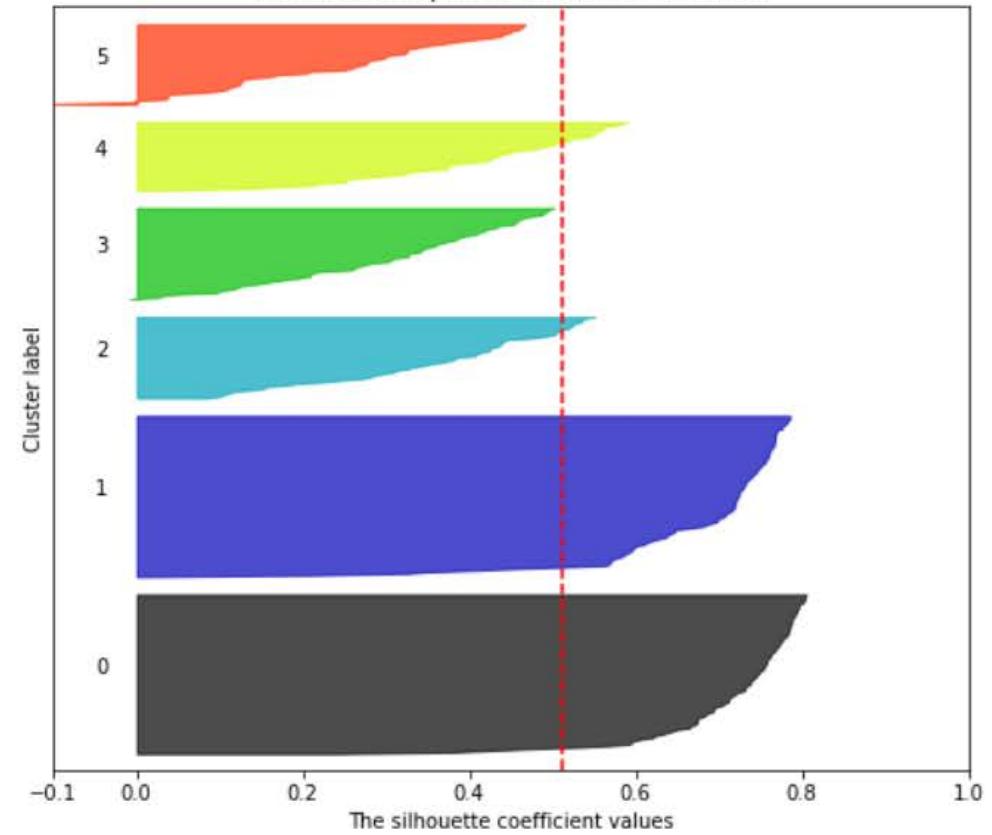
The silhouette plot for the various clusters.



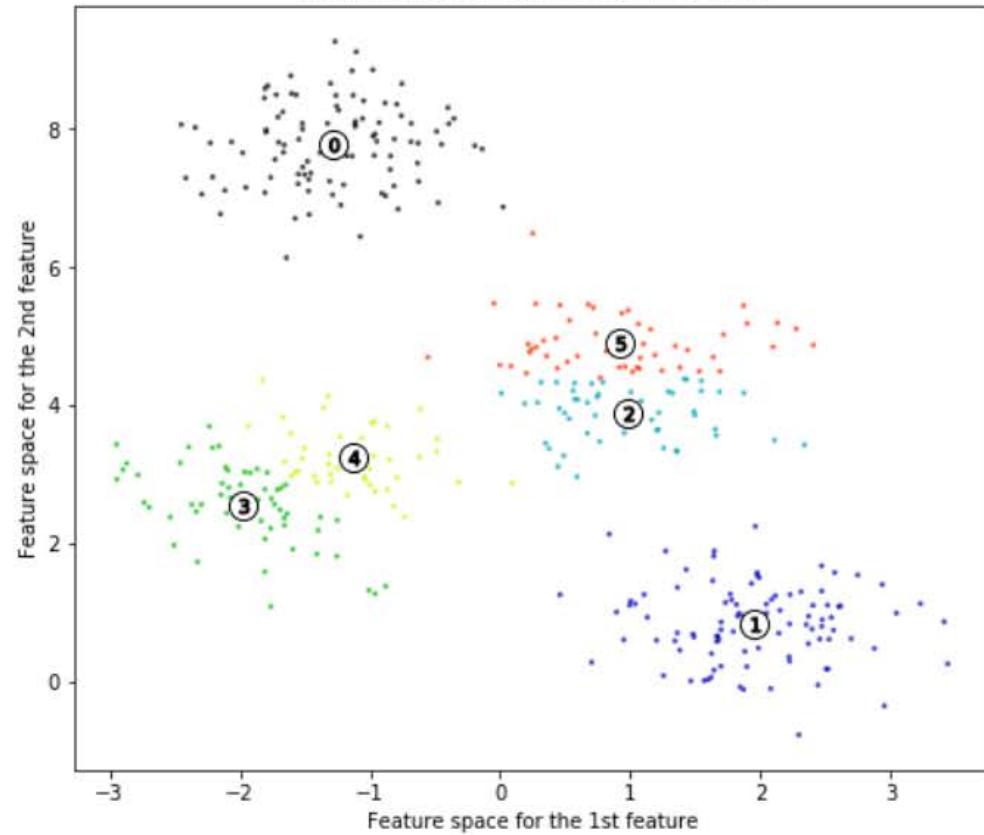
The visualization of the clustered data.

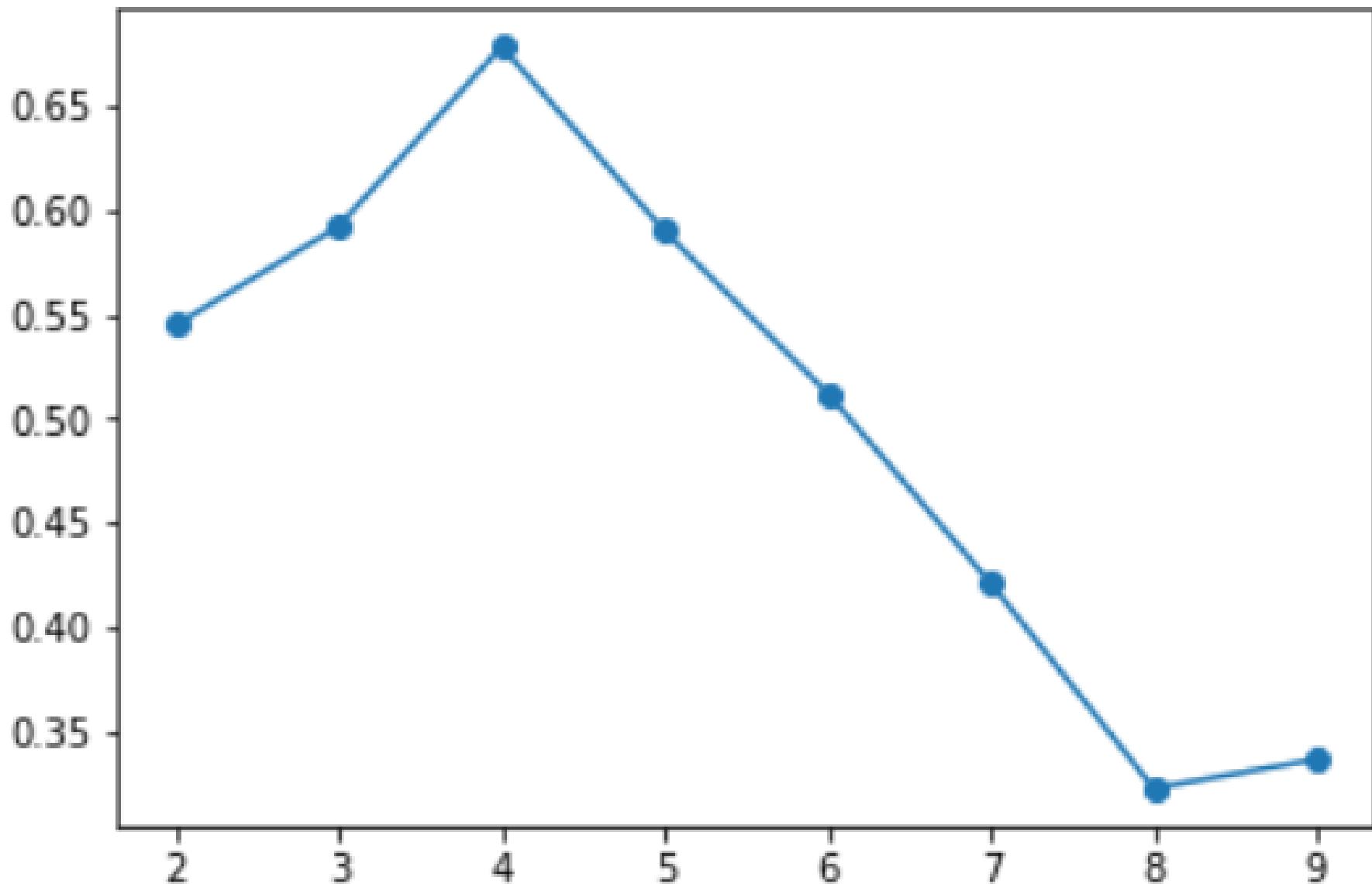


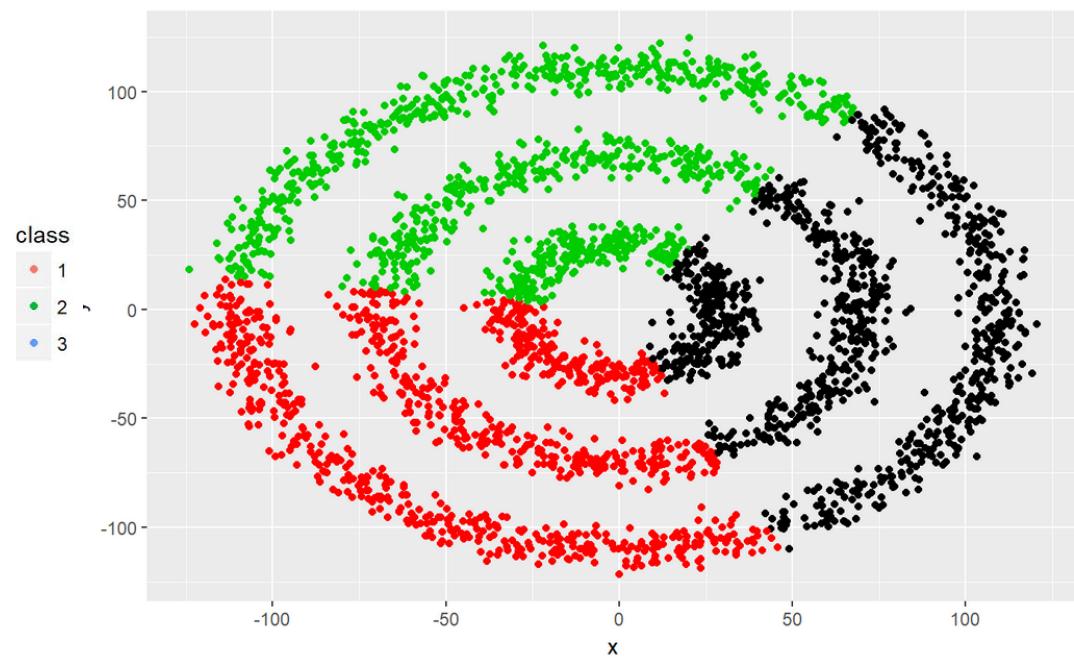
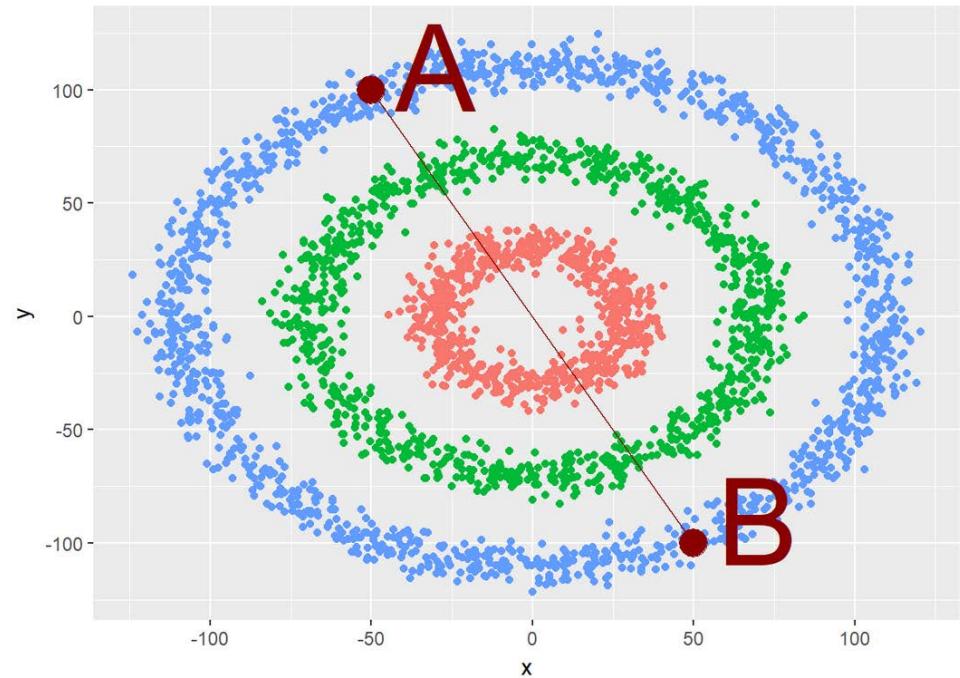
The silhouette plot for the various clusters.



The visualization of the clustered data.

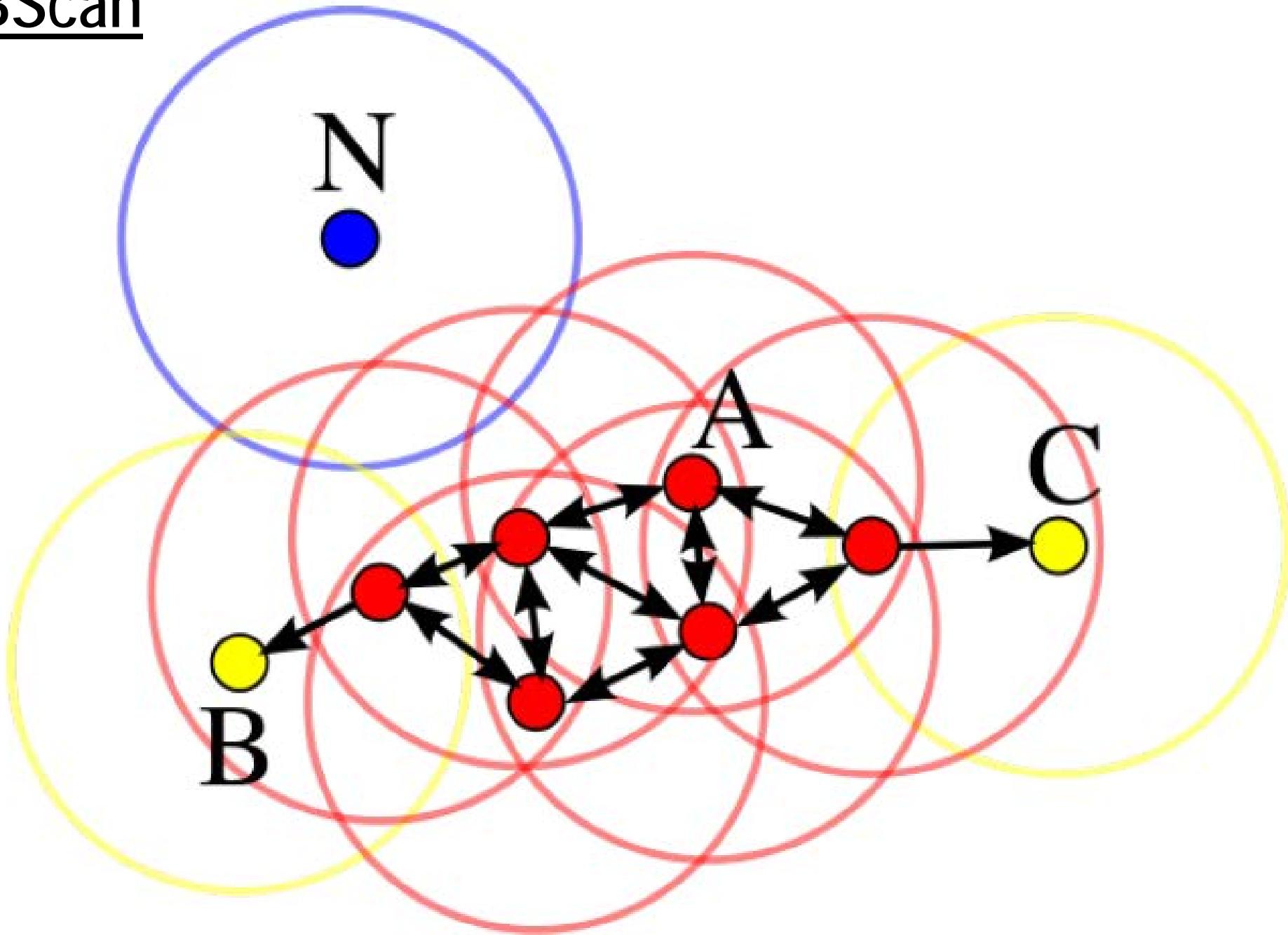


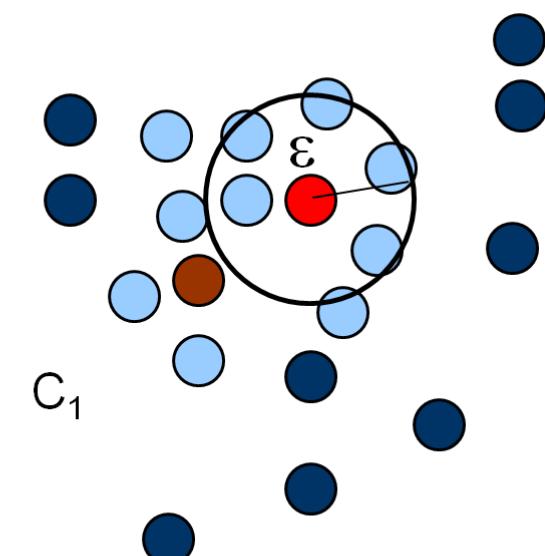
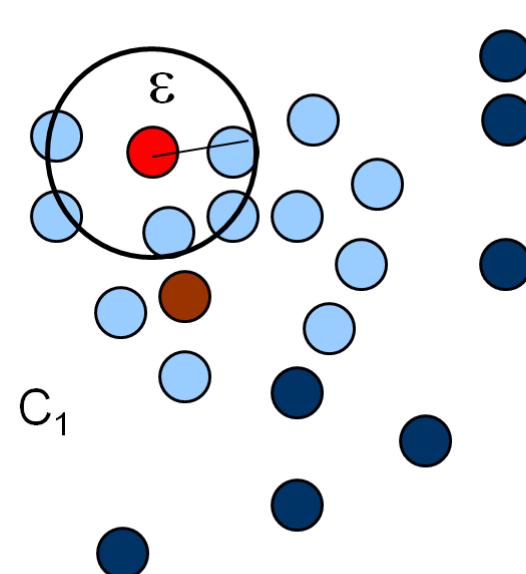
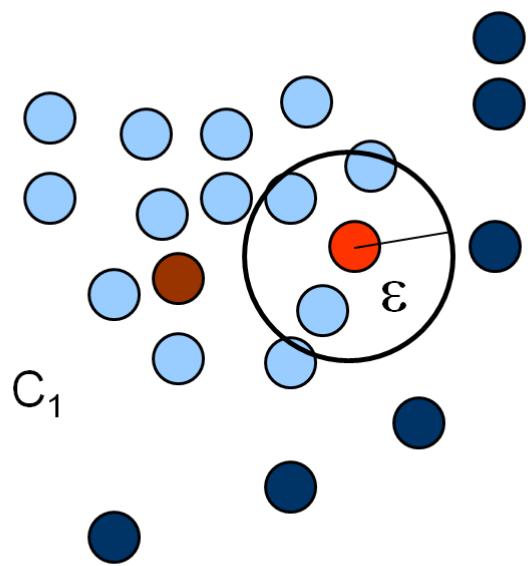
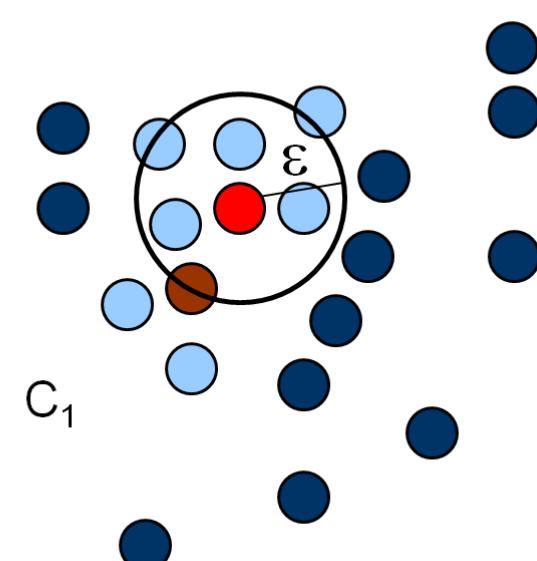
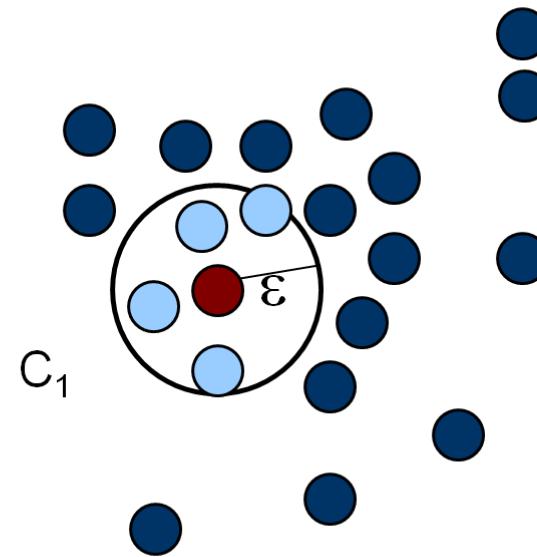
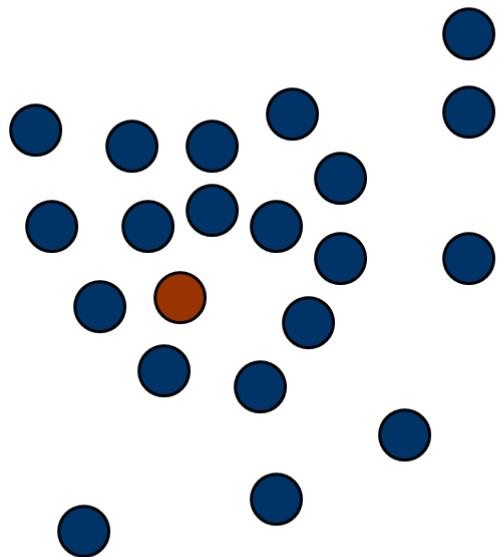


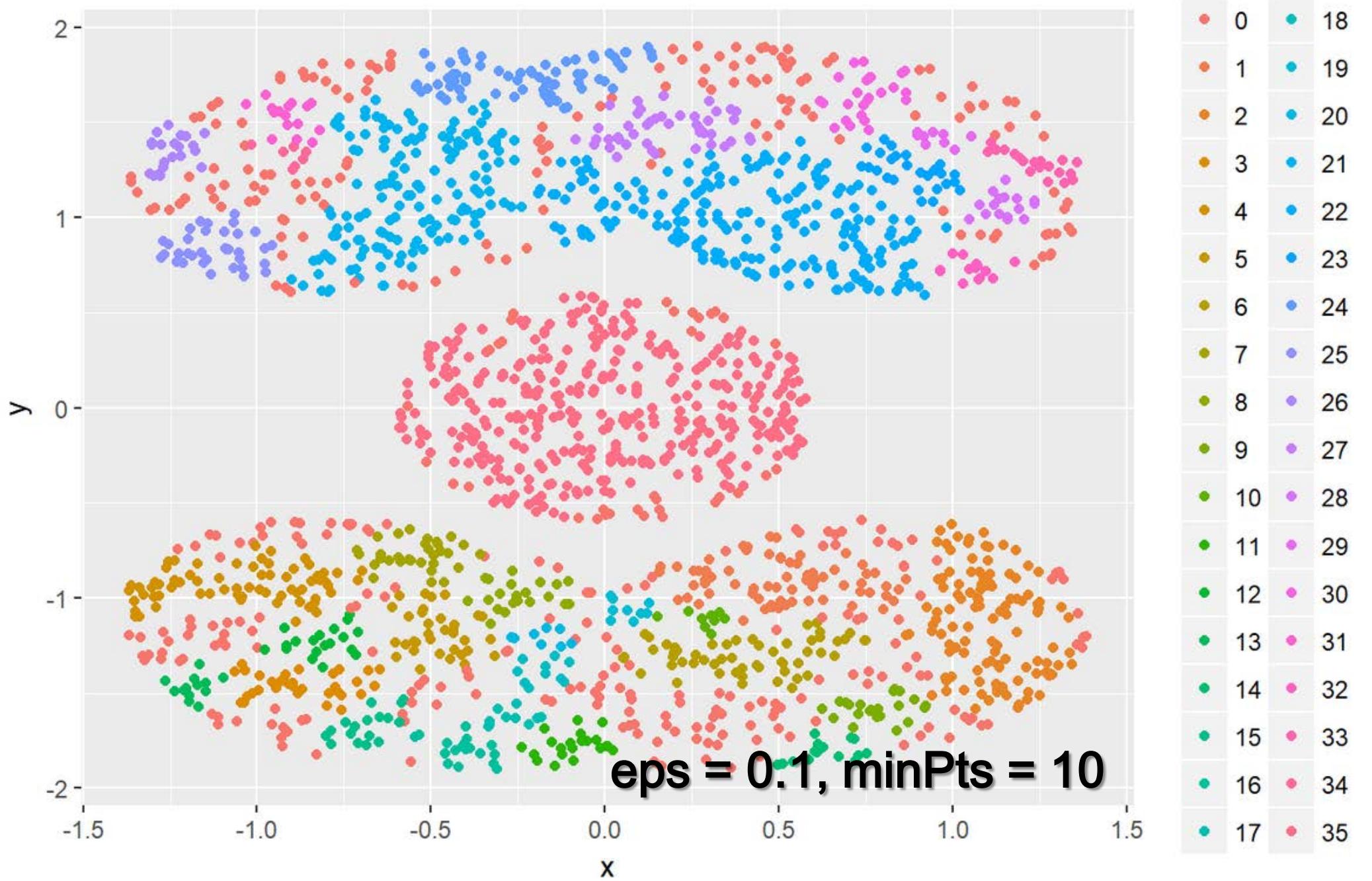


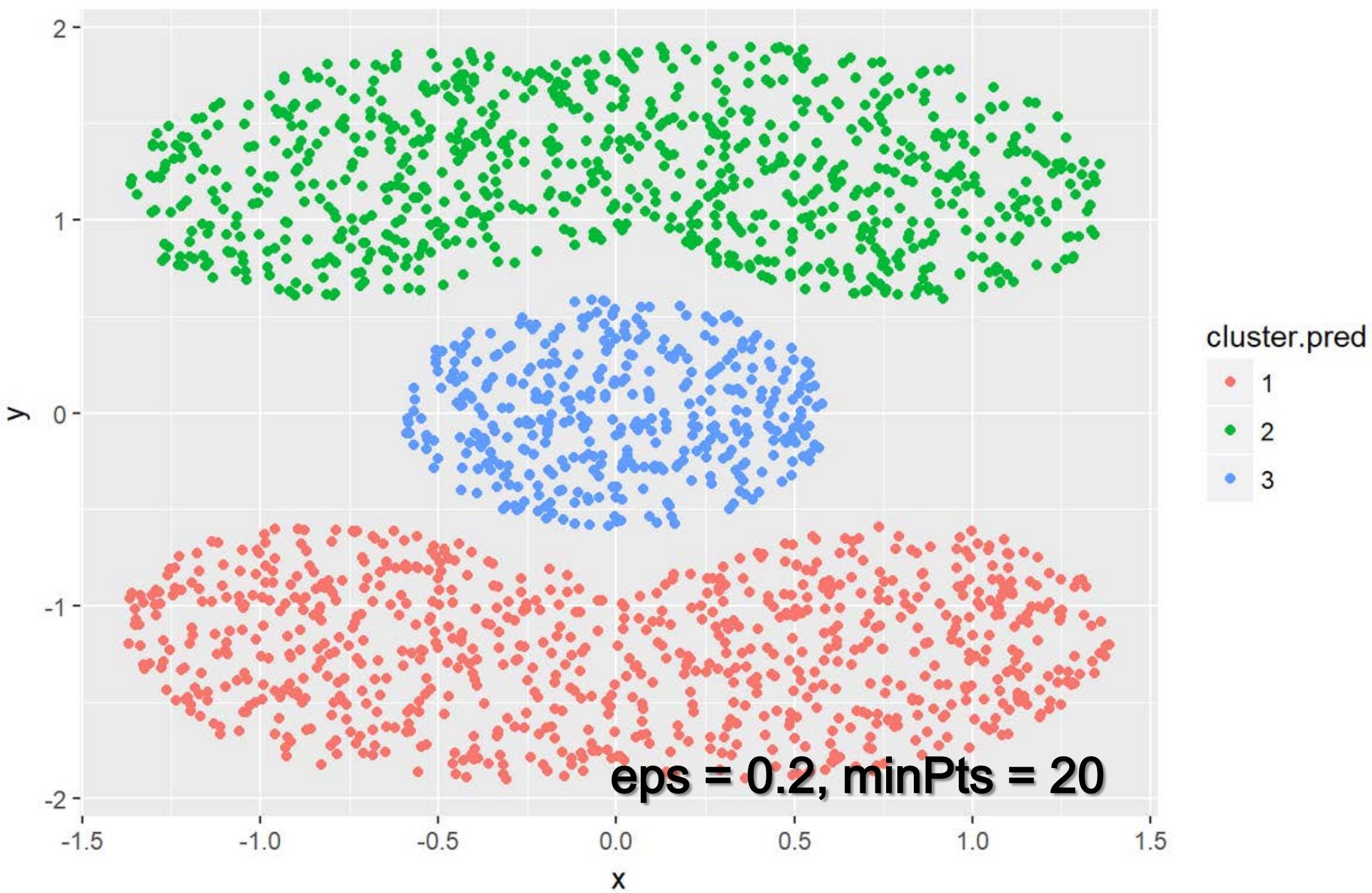
Module 7 : Density-Based Spatial Clustering of Applications with Noise

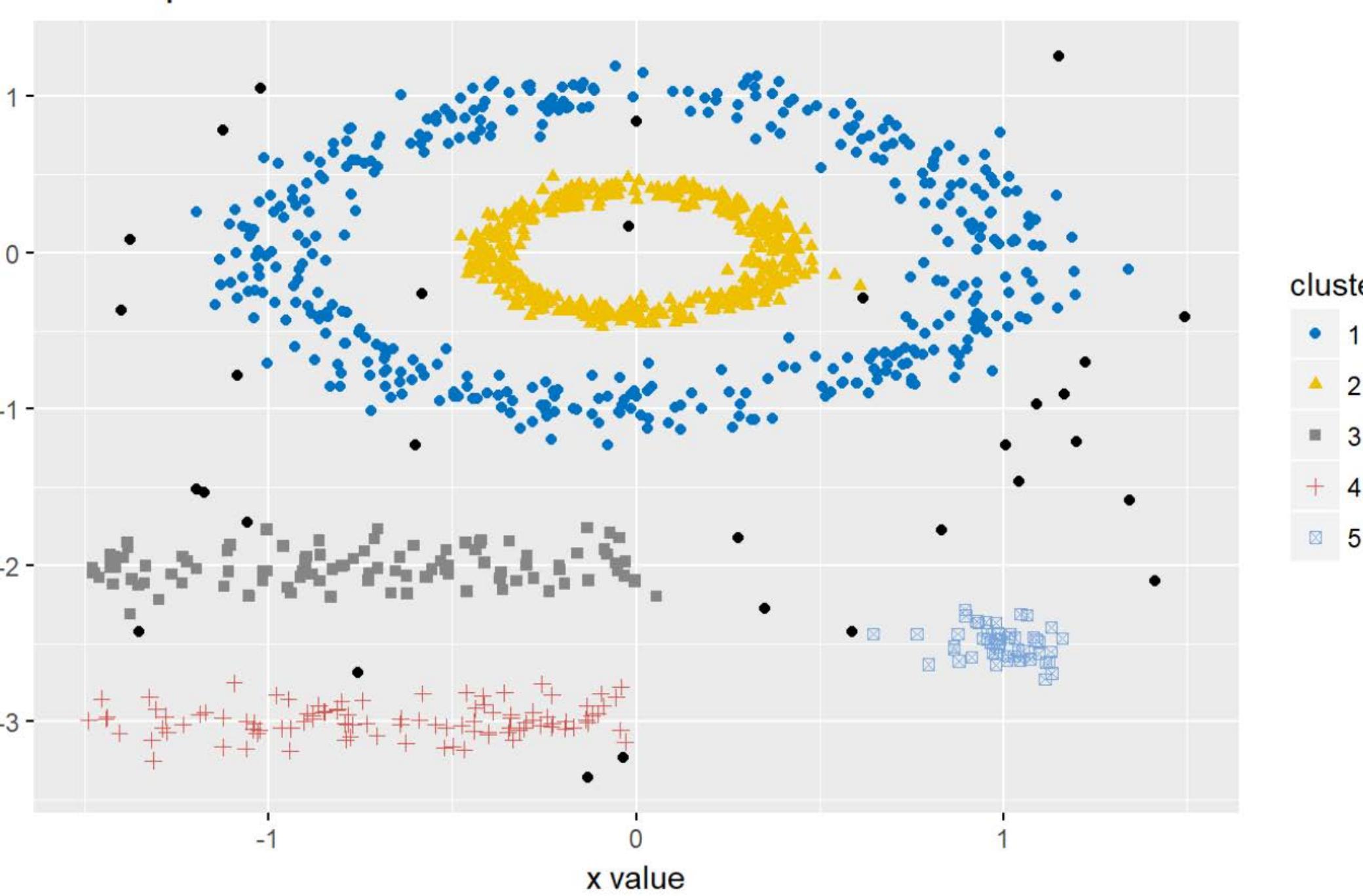
DBScan











Module 8 : Démystifier les algorithmes

3



9



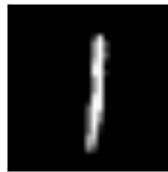
5



2



1



3



1



3



6



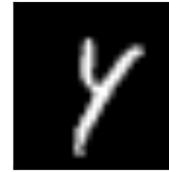
5



7



4



2



2



6



3



2



6



5



4



8



9

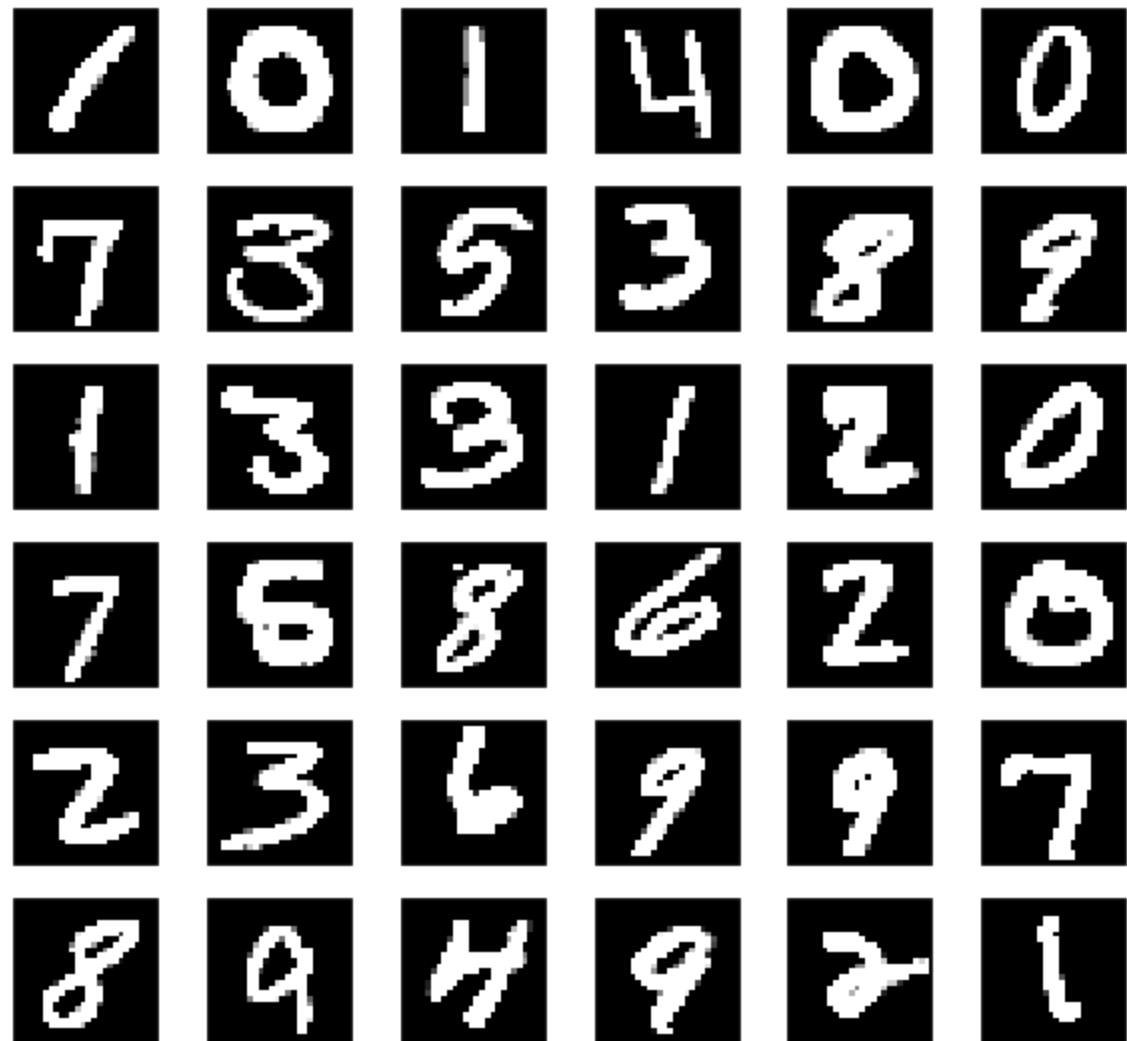


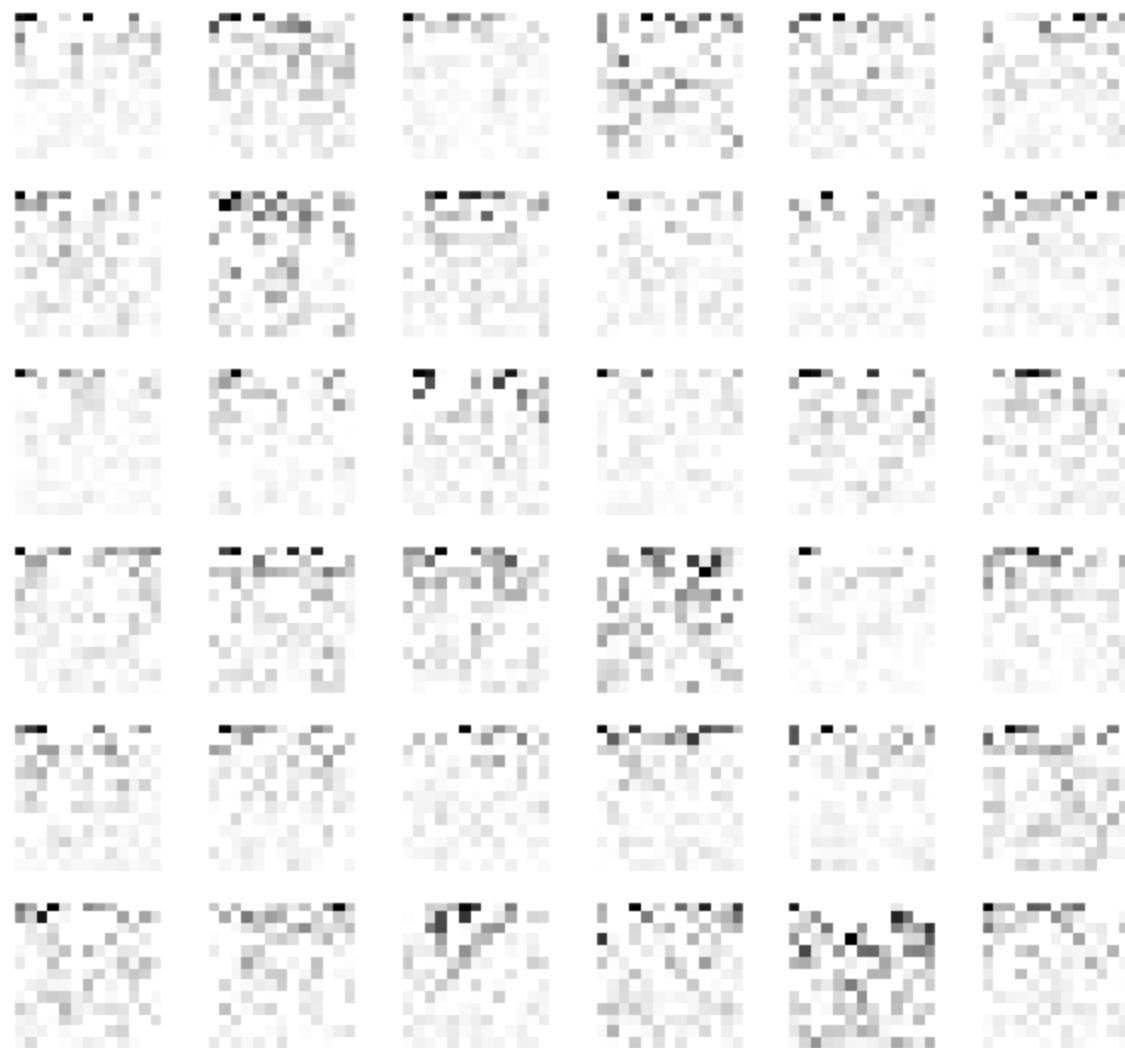
7



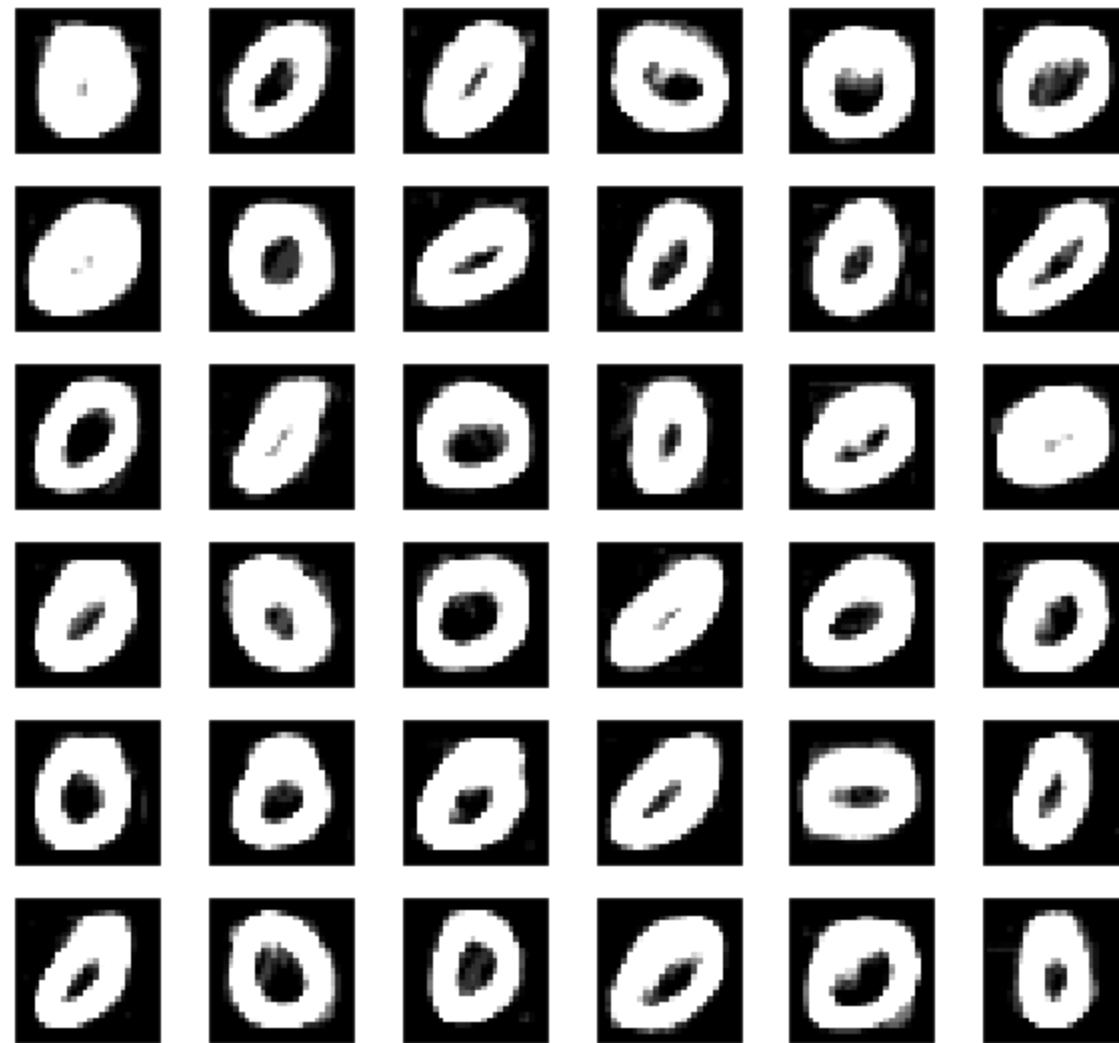
1







K-Means

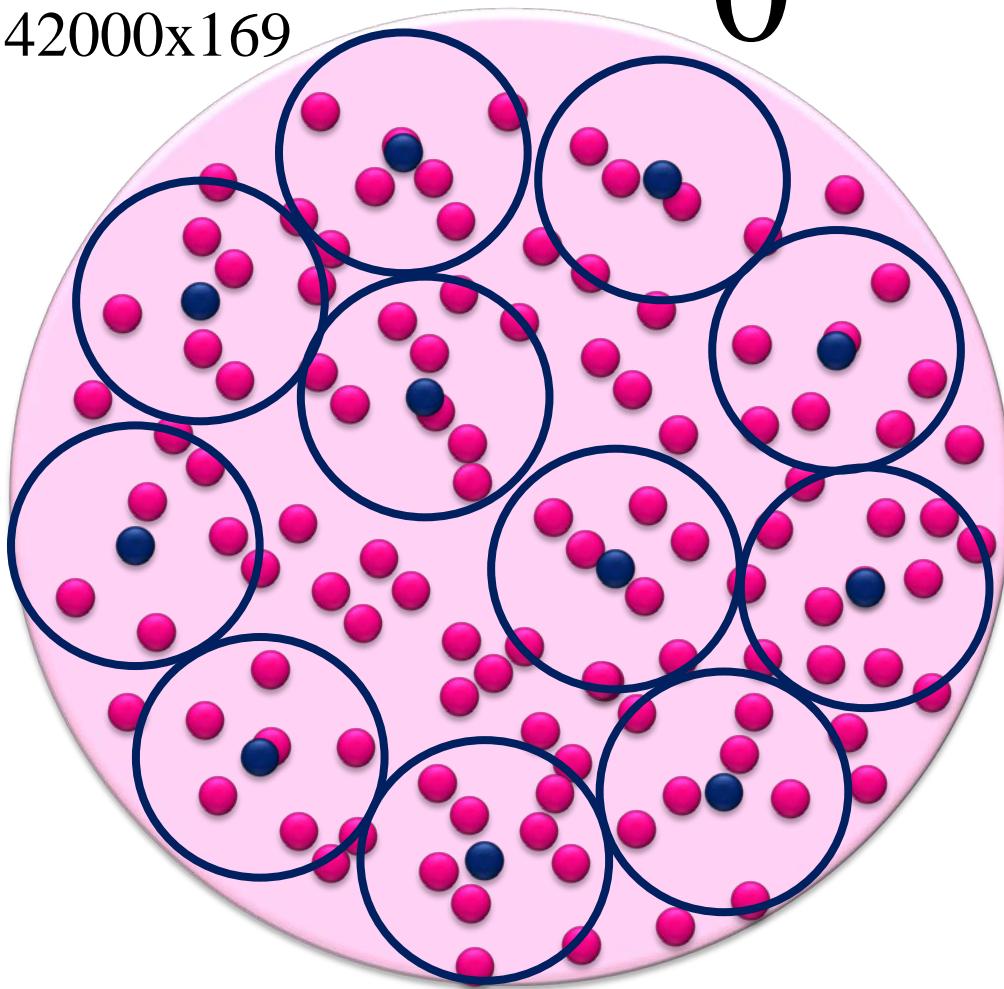


K-Means

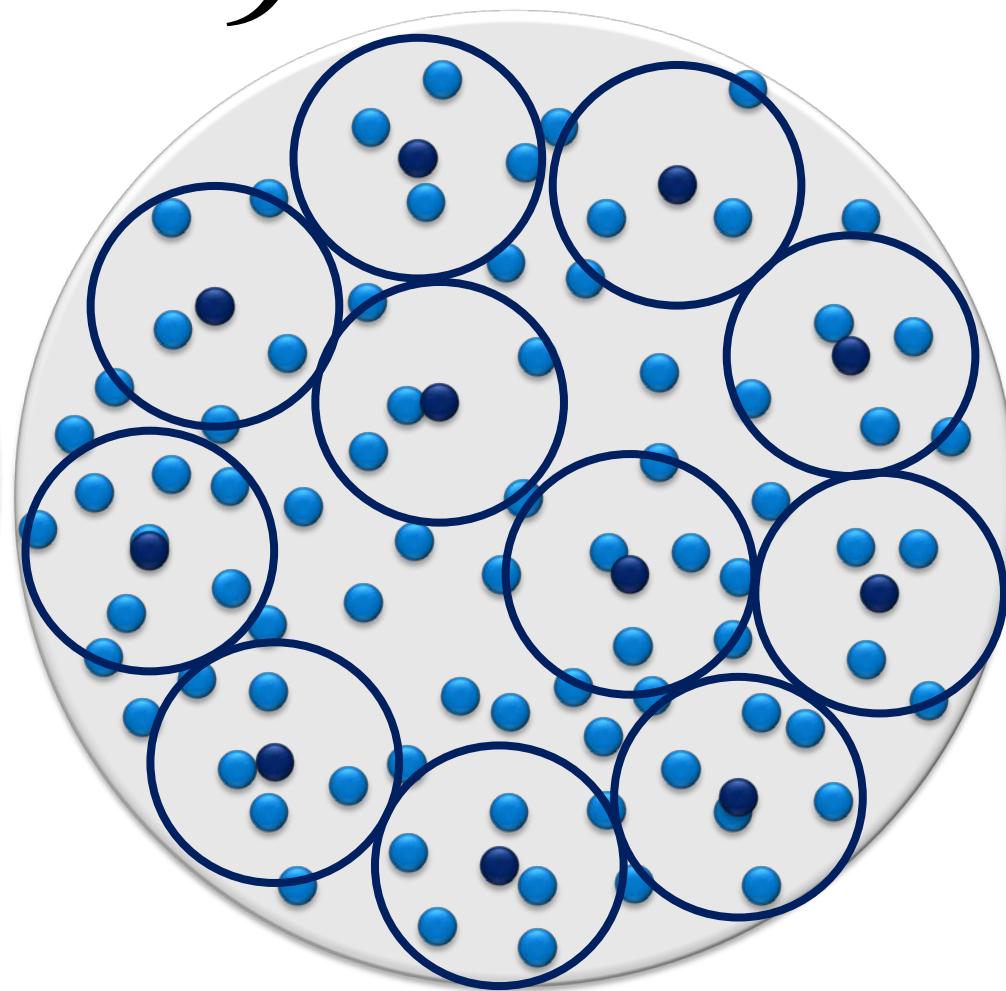
42000x784

42000x169

0



9



10x50x784

10x100x784

10x150x784

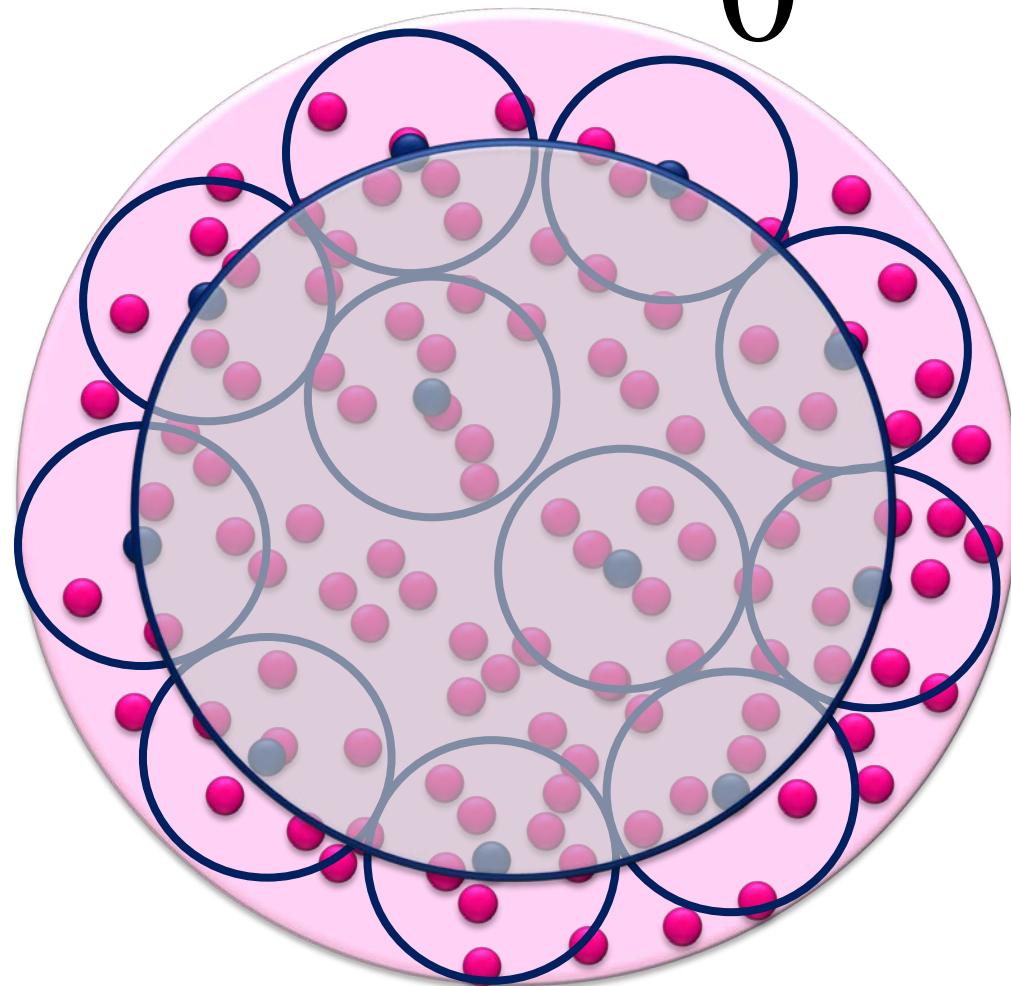
10x50x169

10x100x169

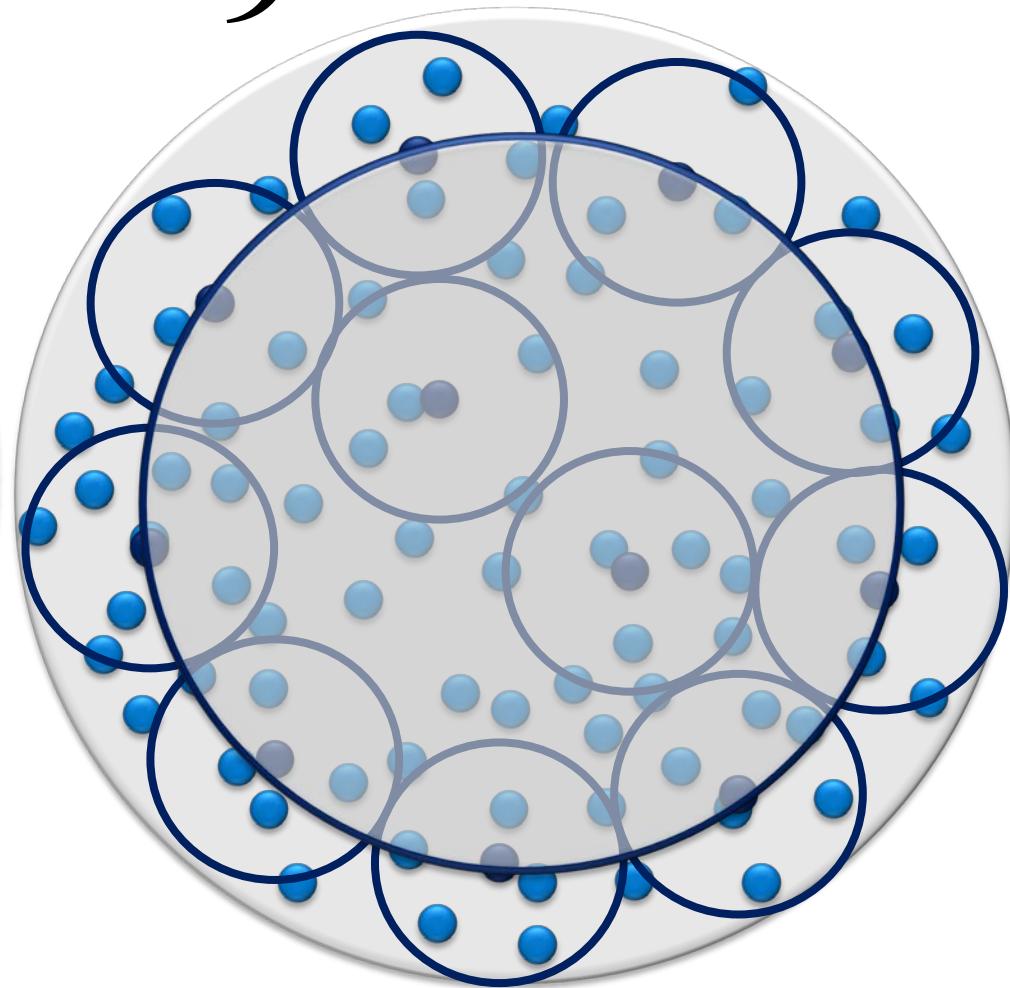
10x150x169

K-Means

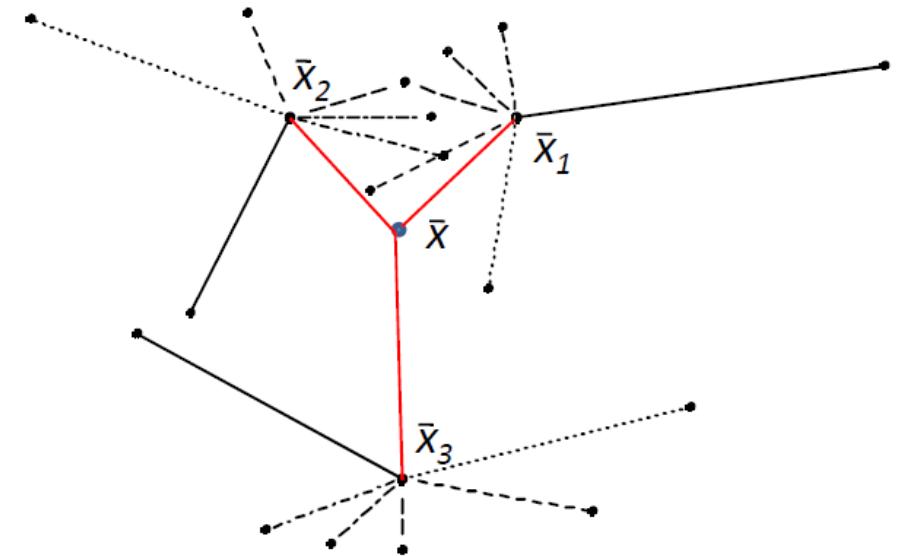
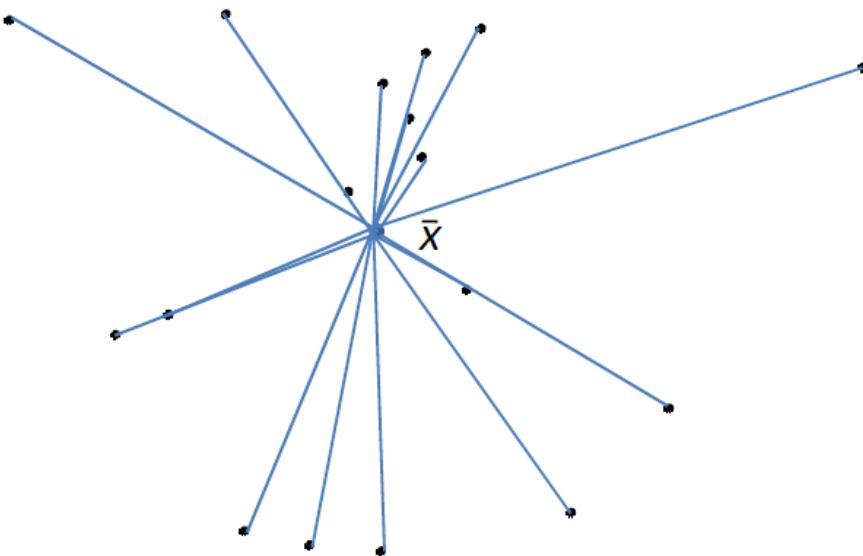
0



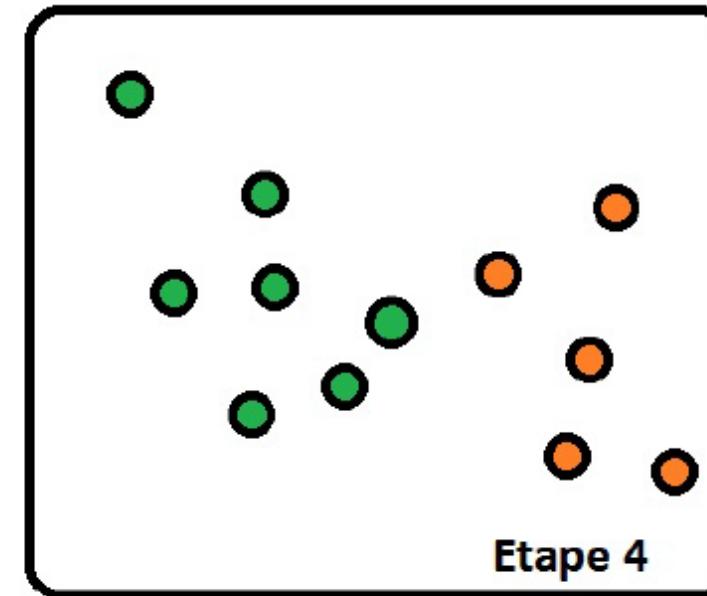
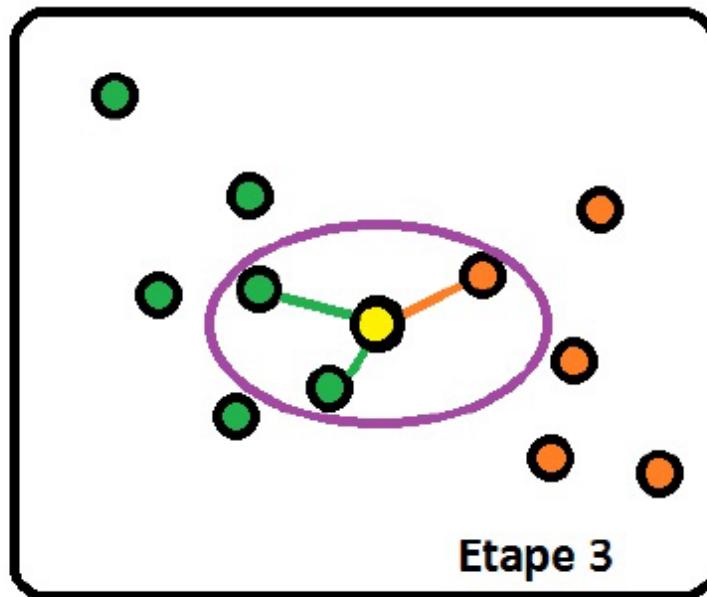
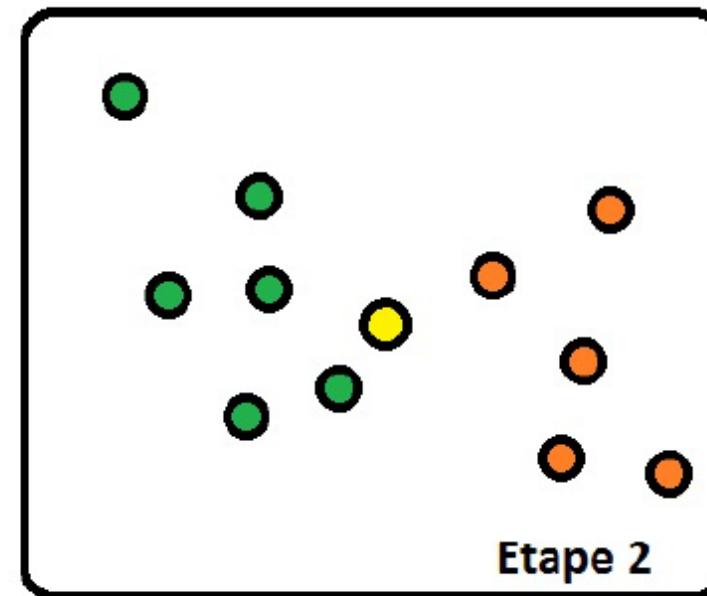
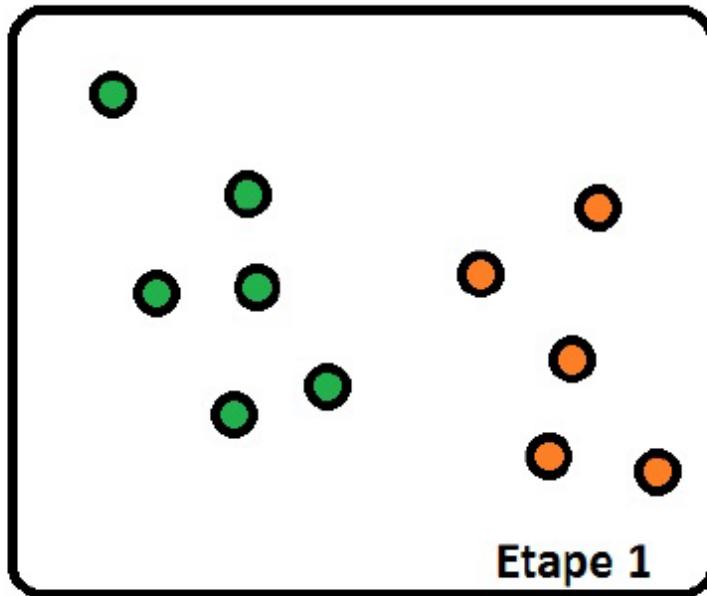
9



$$\underbrace{\sum_{k=1}^K \sum_{q=1}^Q \sum_{i=1}^I (x_{iqk} - \bar{x}_k)^2}_{\text{Inertie totale}} = \underbrace{\sum_{k=1}^K \sum_{q=1}^Q \sum_{i=1}^I (x_{iqk} - \bar{x}_{qk})^2}_{\text{Inertie intra}} + \underbrace{\sum_{k=1}^K \sum_{q=1}^Q \sum_{i=1}^I (\bar{x}_{qk} - \bar{x}_k)^2}_{\text{Inertie inter}}$$



KNN avec k=3



Performance du modèle

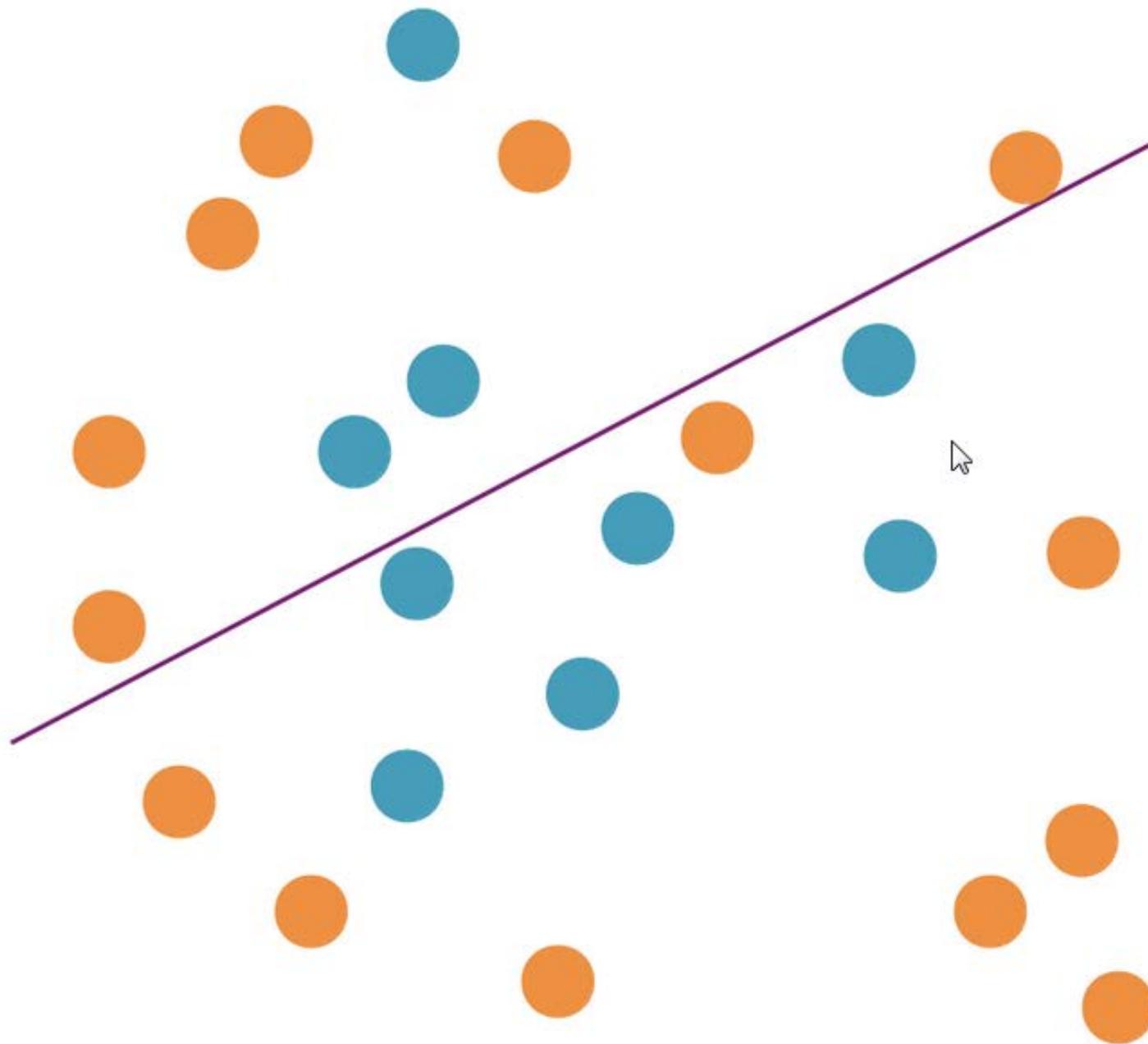
96.19 %

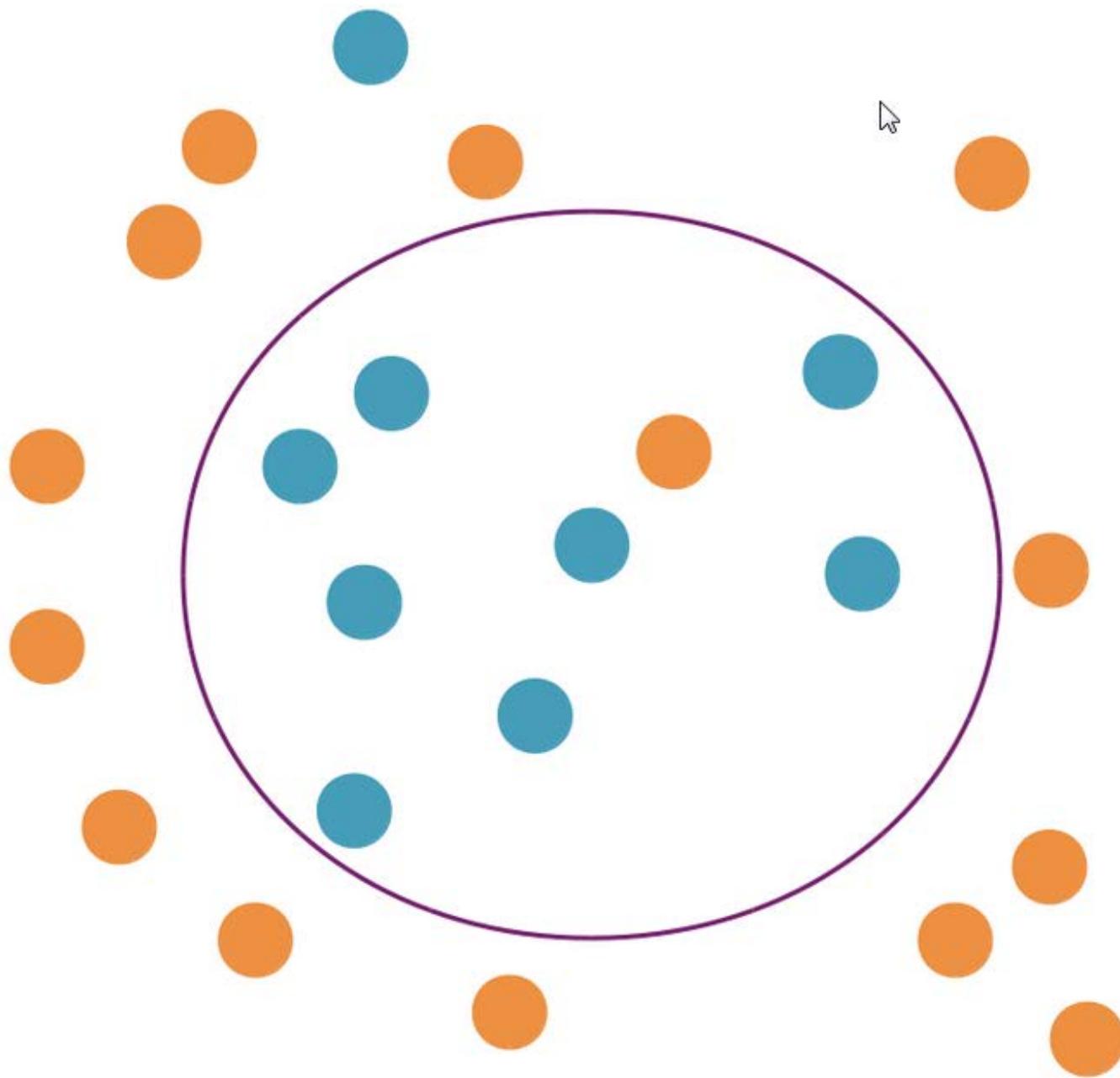
label	0	1	2	3	4	5	6	7	8	9
predict	4067	0	17	3	2	12	16	3	7	8
0	1	4645	26	11	26	6	7	28	24	4
1	9	6	4015	35	5	2	3	25	10	11
2	4	4	17	4085	1	44	1	2	47	26
3	2	8	5	3	3849	3	3	29	13	74
4	12	0	1	86	1	3654	22	2	62	11
5	28	2	11	4	29	25	4076	2	20	6
6	2	11	43	29	7	3	1	4208	11	73
7	4	2	32	61	3	26	8	3	3838	12
8	3	6	10	34	149	20	0	99	31	3963

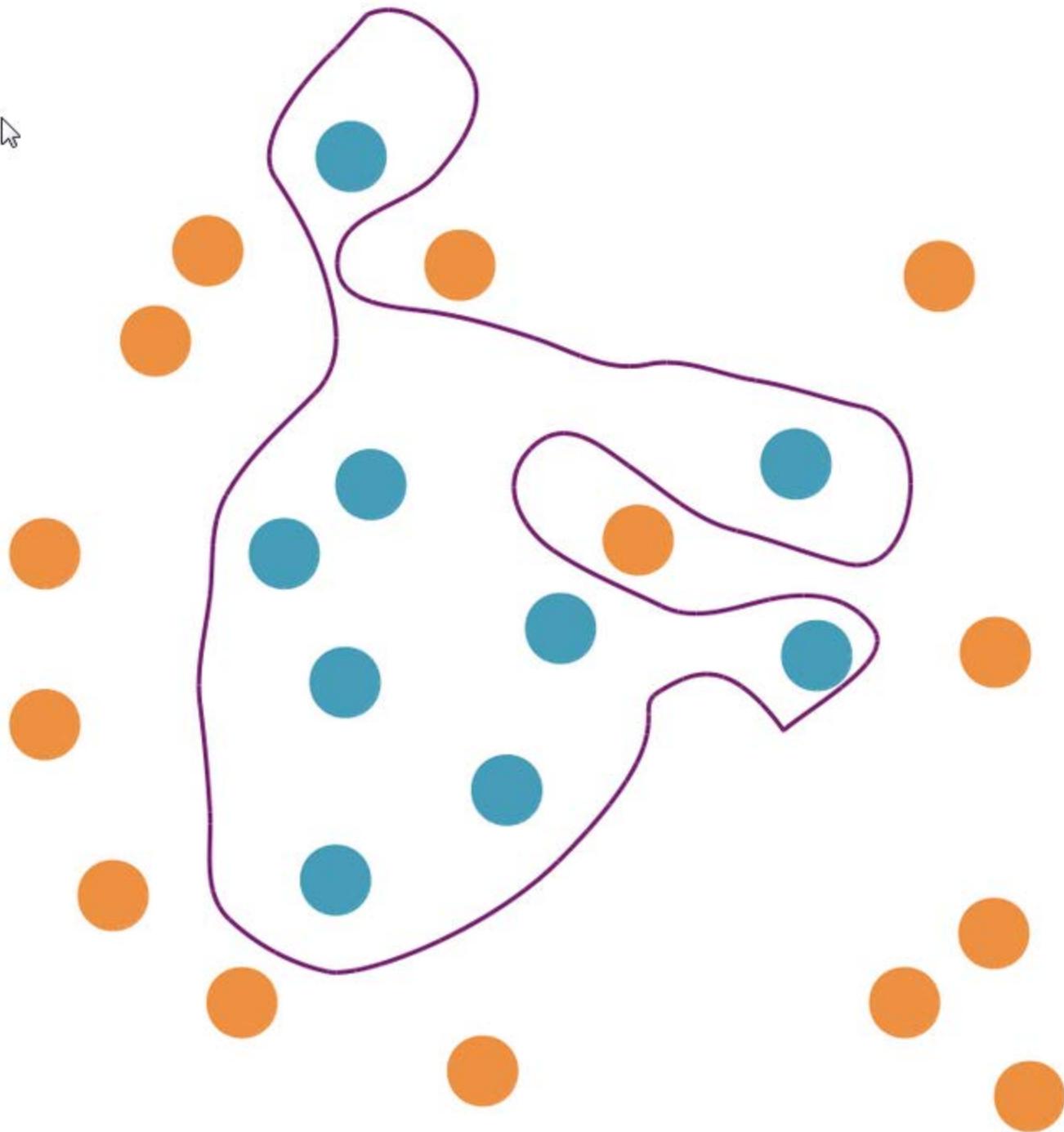
Module 9 : Évaluer un modèle



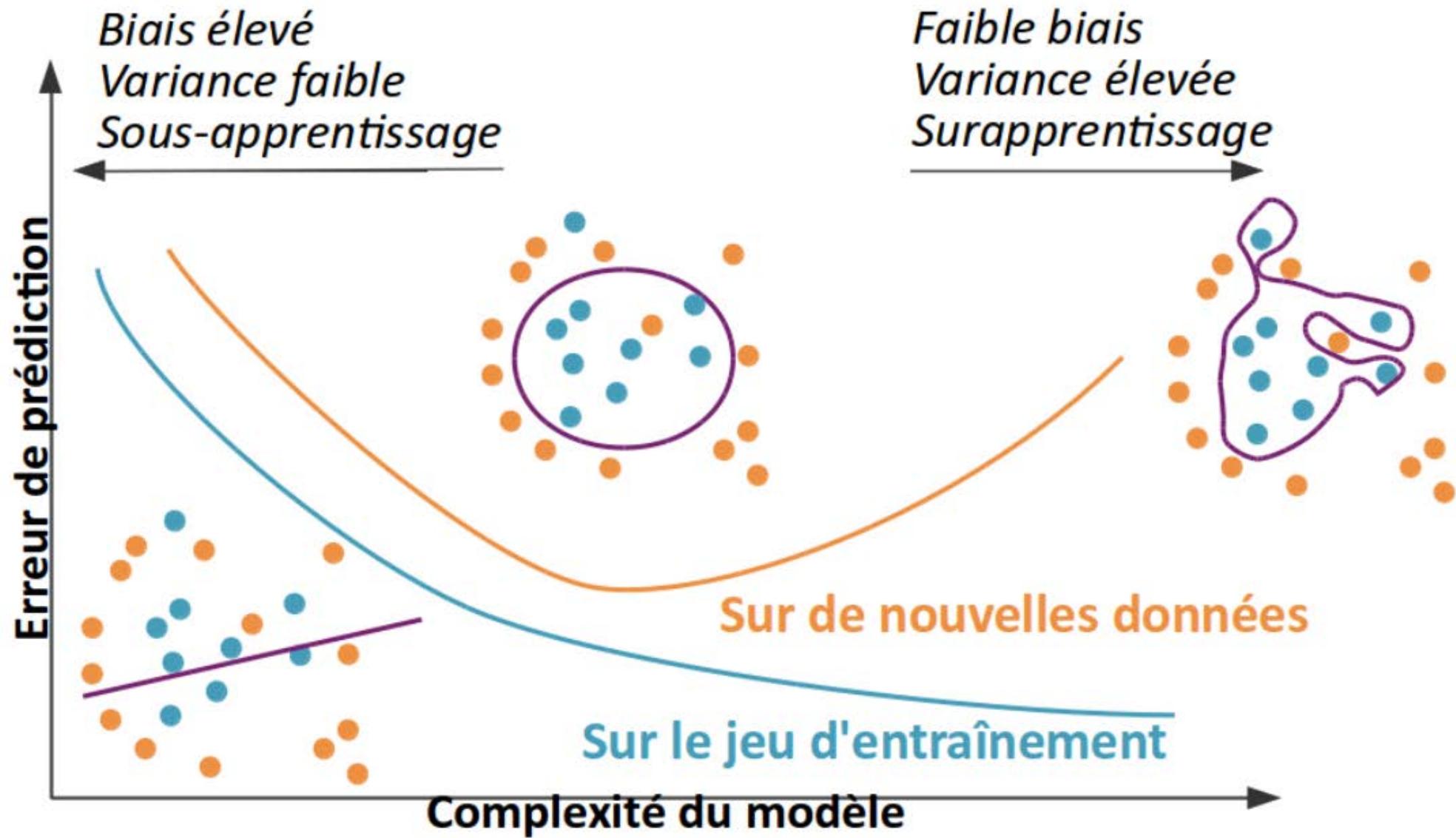
- 70% **apprentissage**
- 30% **test**







↗



1

L'échantillon initial



2

L'échantillon traitement



L'échantillon validation



3

L'échantillon apprentissage



L'échantillon test



4

La validation croisée



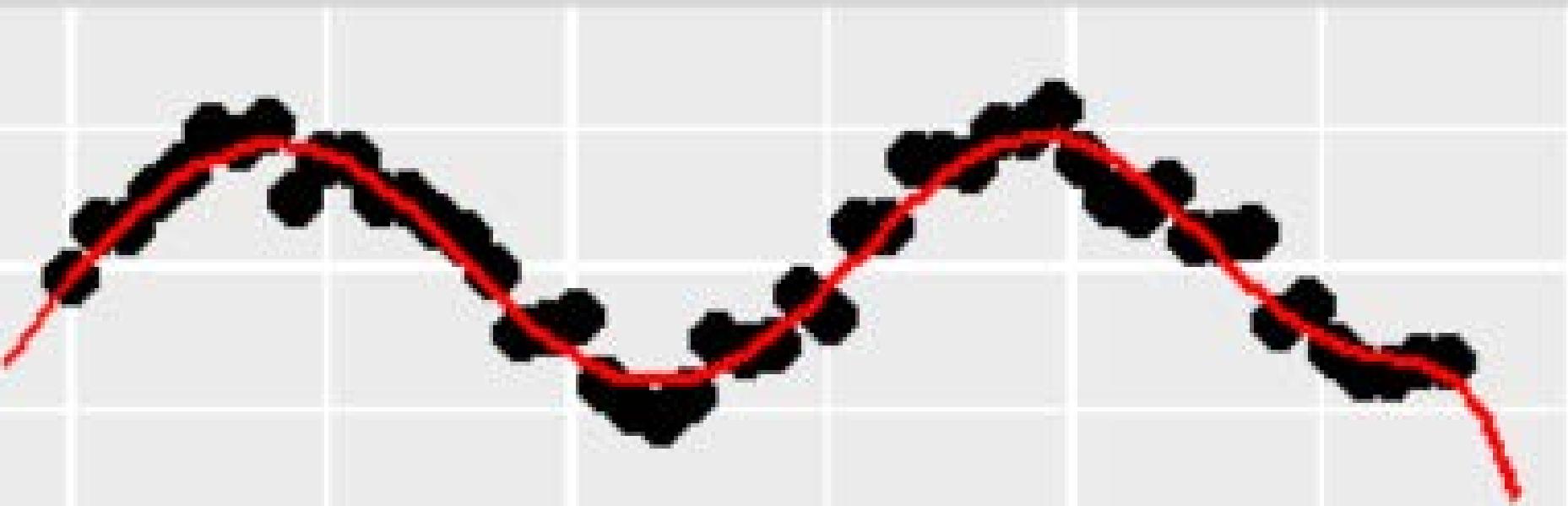
L'échantillon validation

5

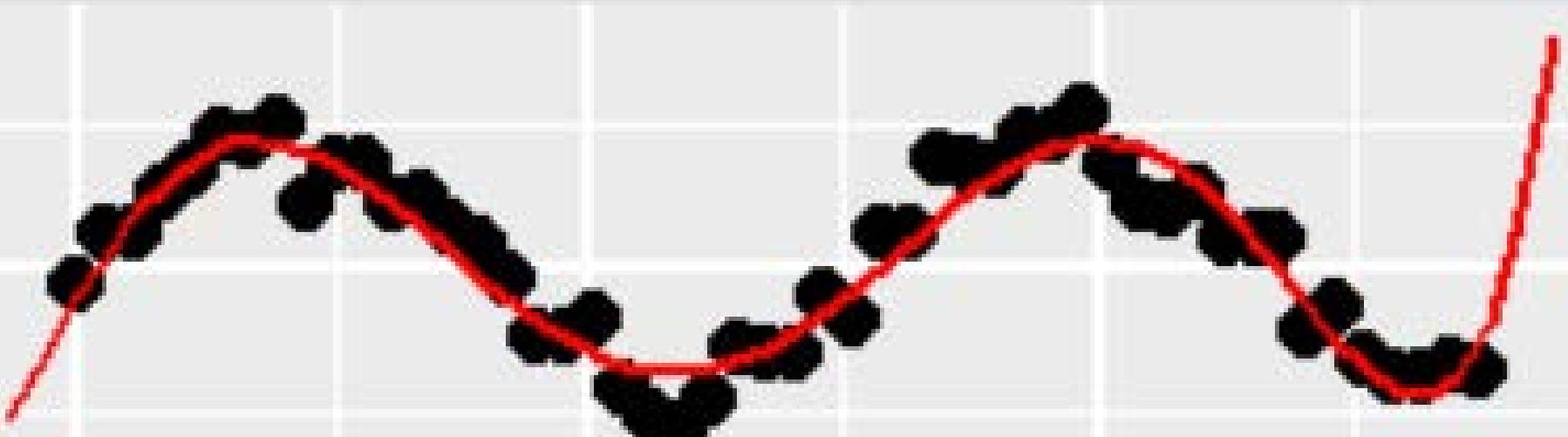
La validation finale du modèle



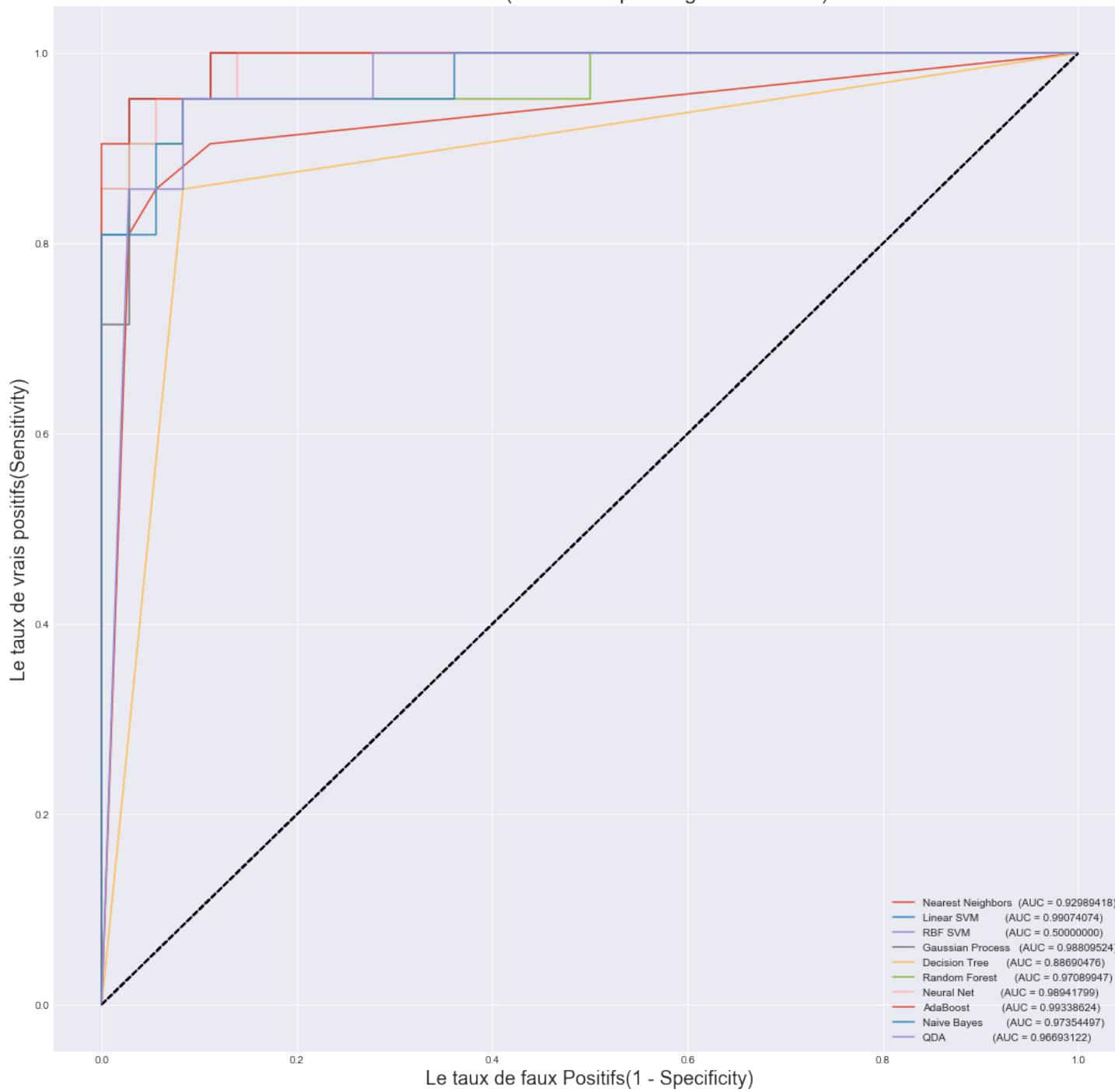
mod13



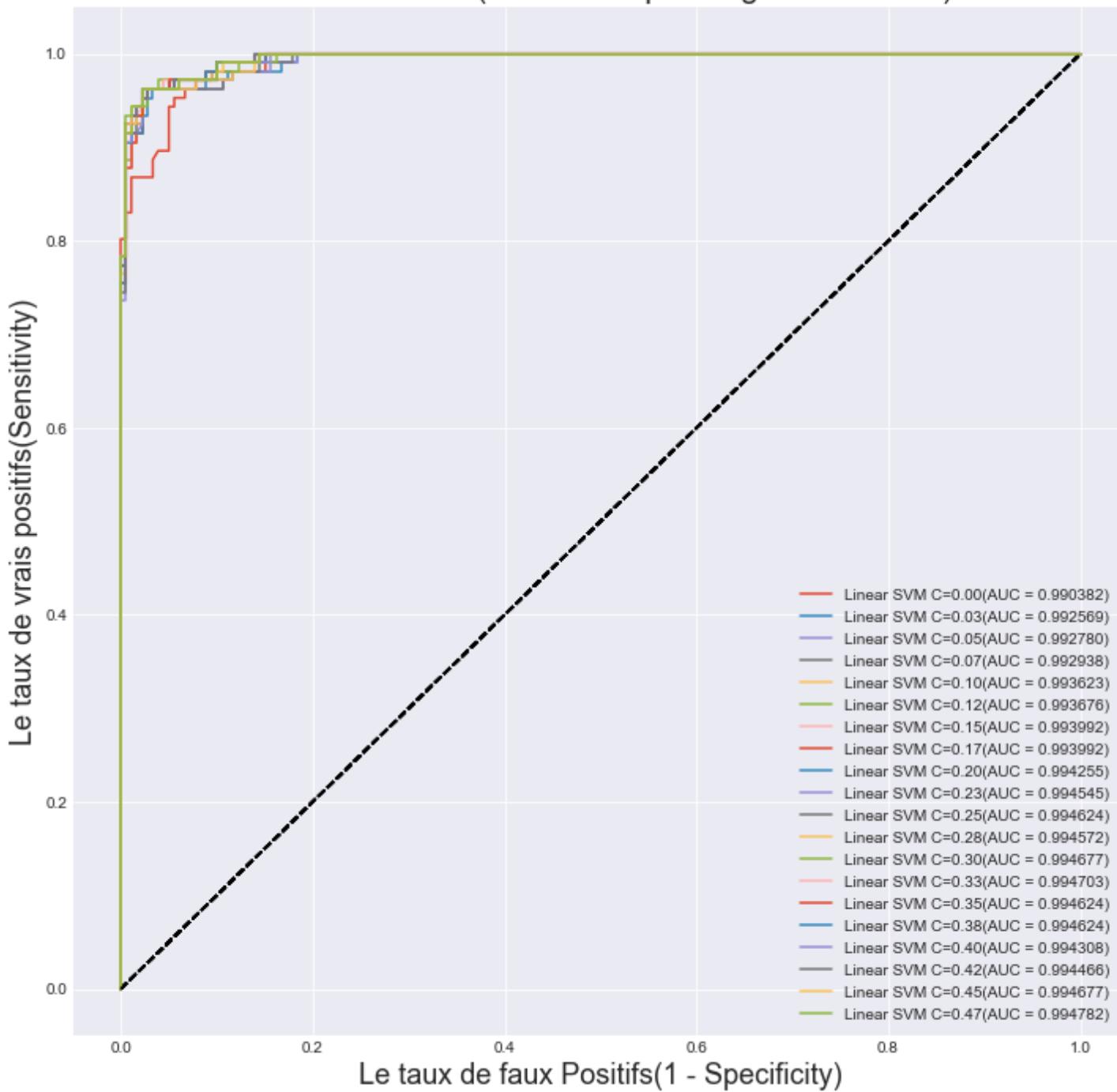
mod11



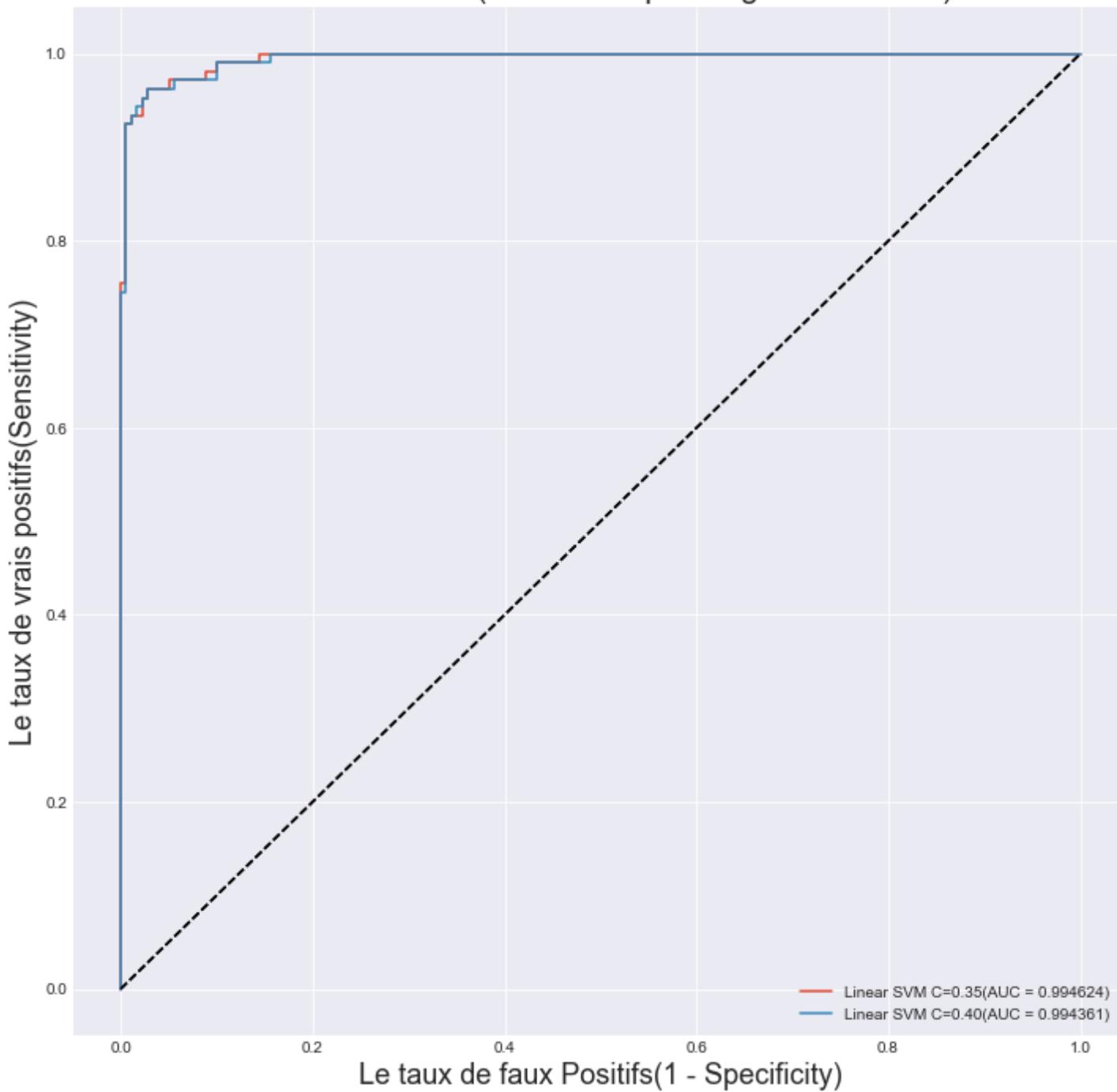
La courbe ROC (Receiver Operating Characteristic)



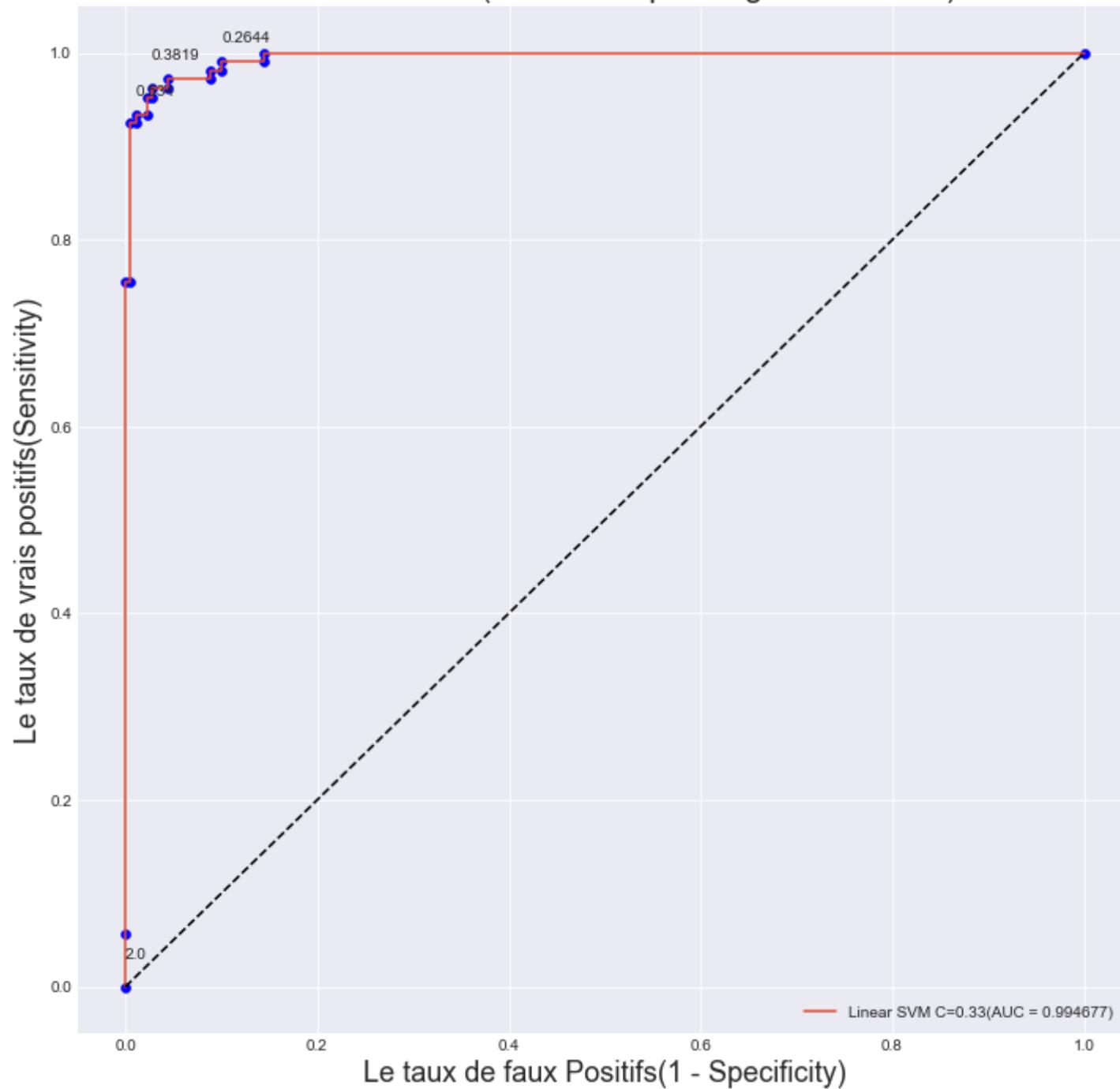
La courbe ROC (Receiver Operating Characteristic)



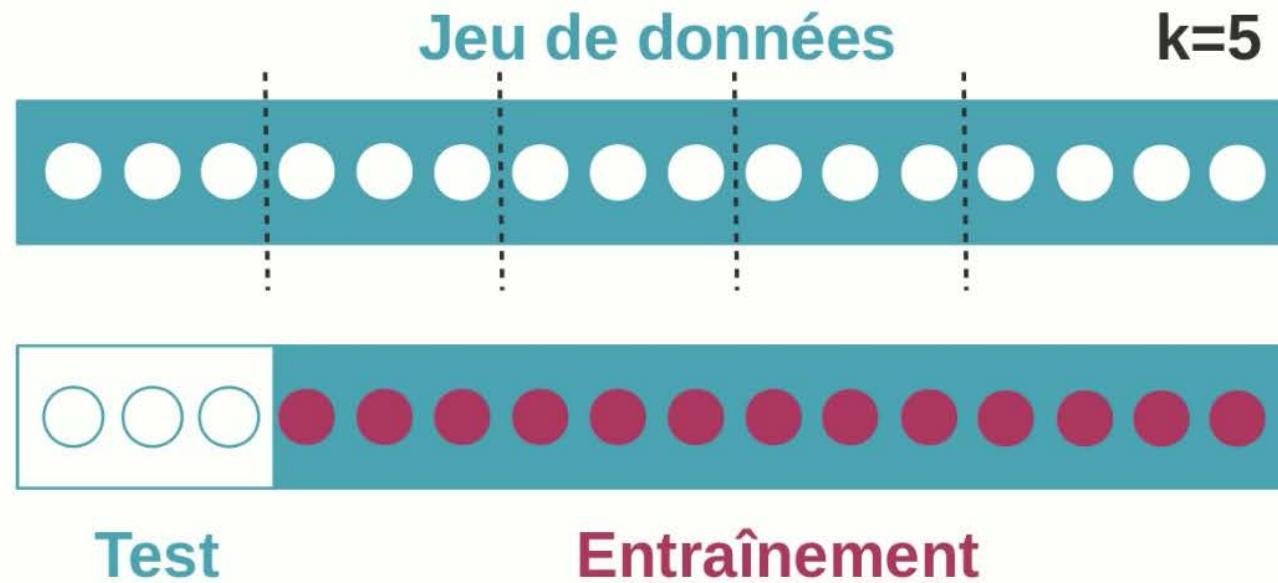
La courbe ROC (Receiver Operating Characteristic)



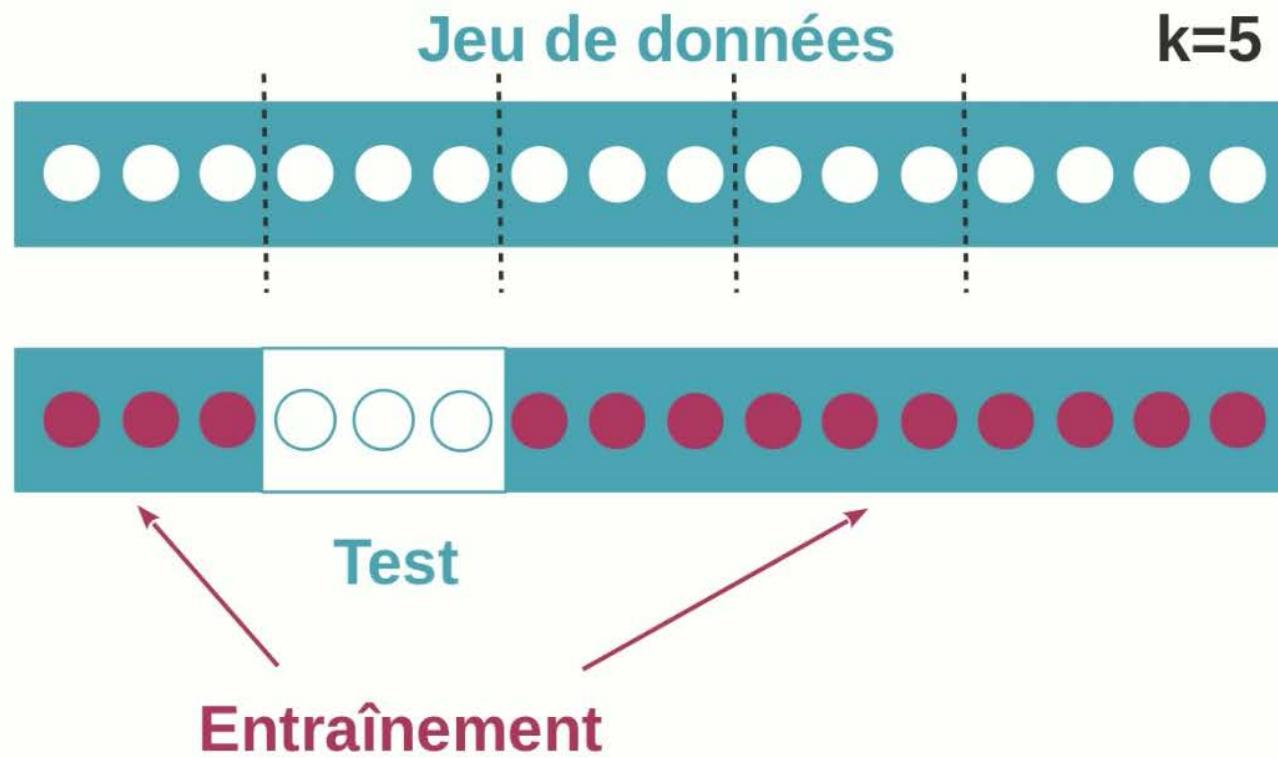
La courbe ROC (Receiver Operating Characteristic)



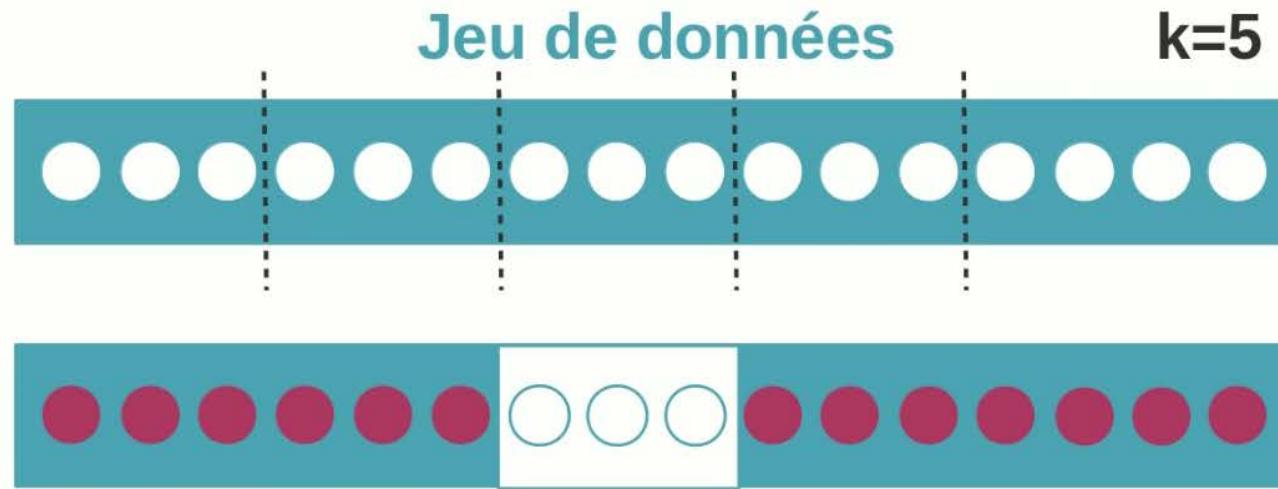
La validation croisée



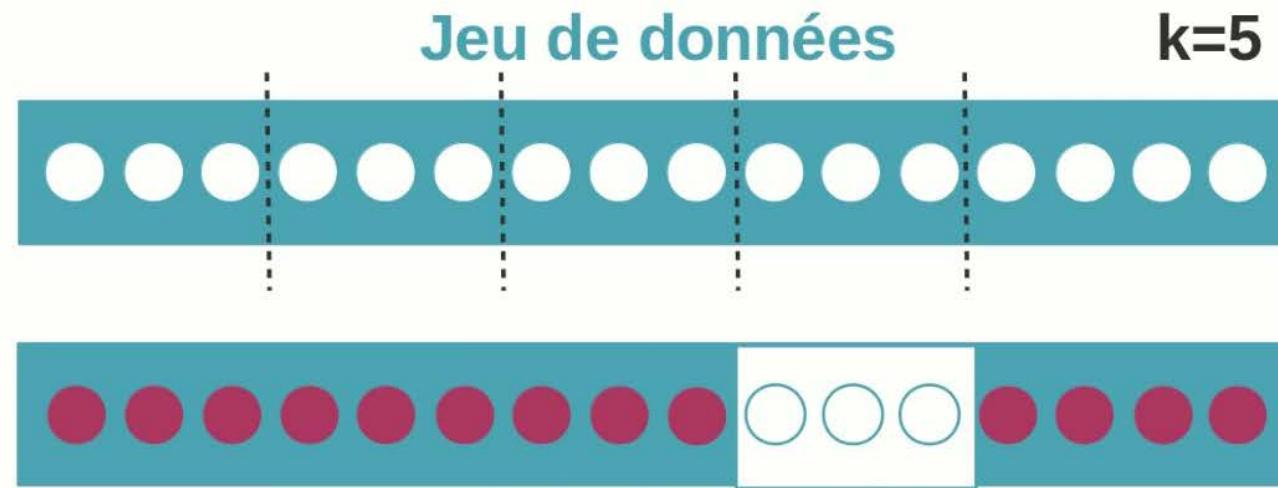
La validation croisée



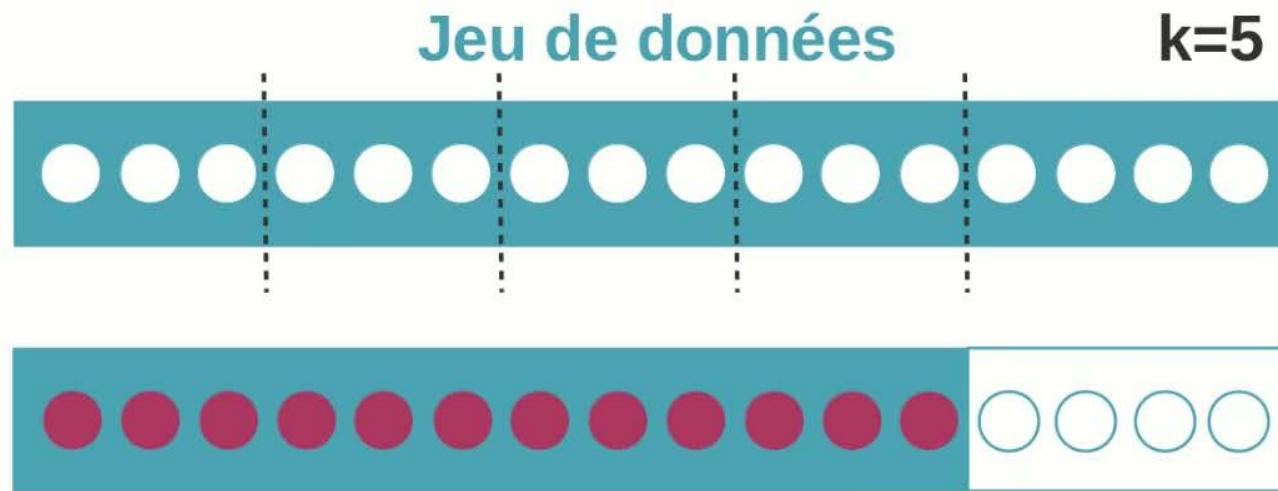
La validation croisée



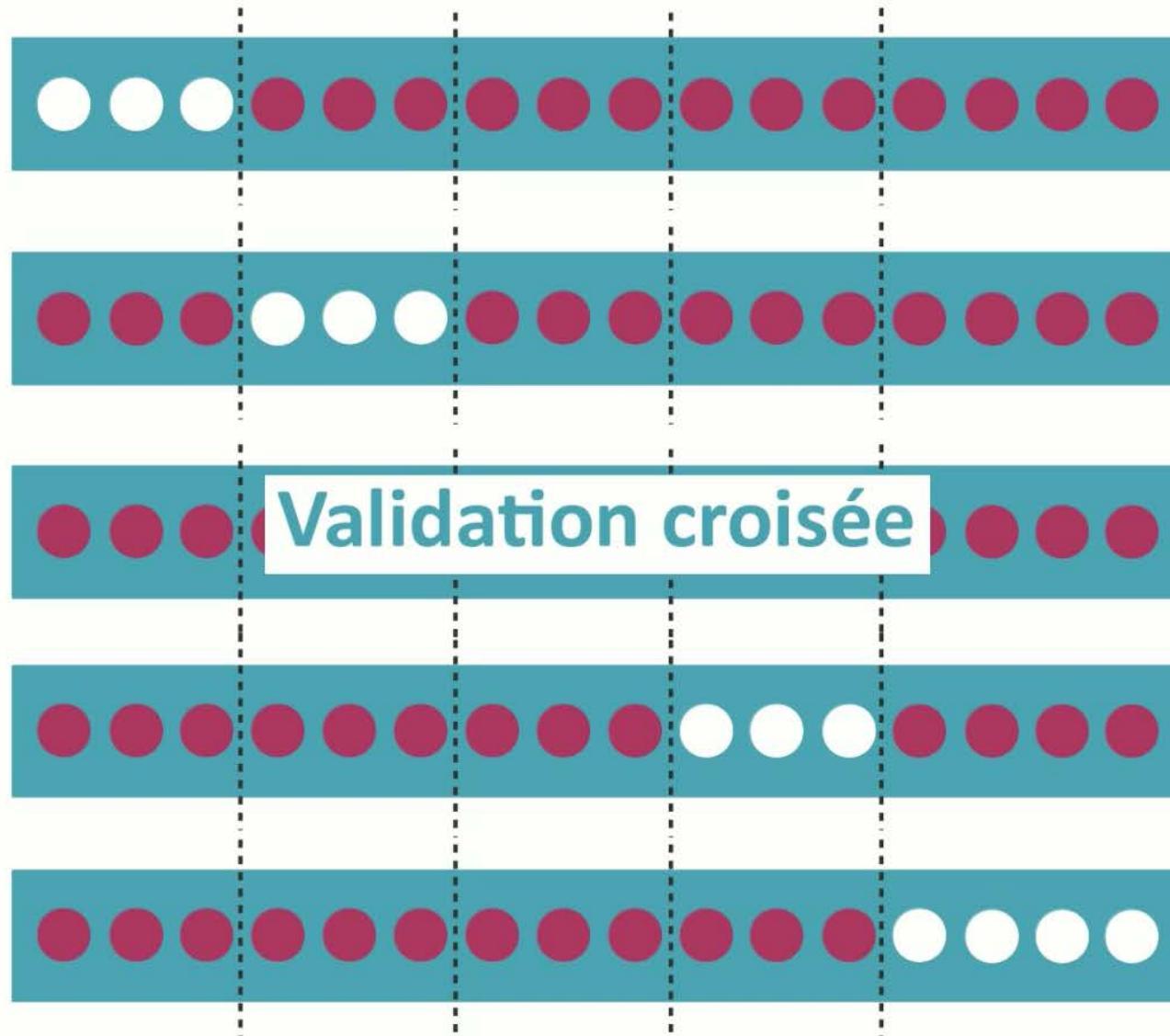
La validation croisée



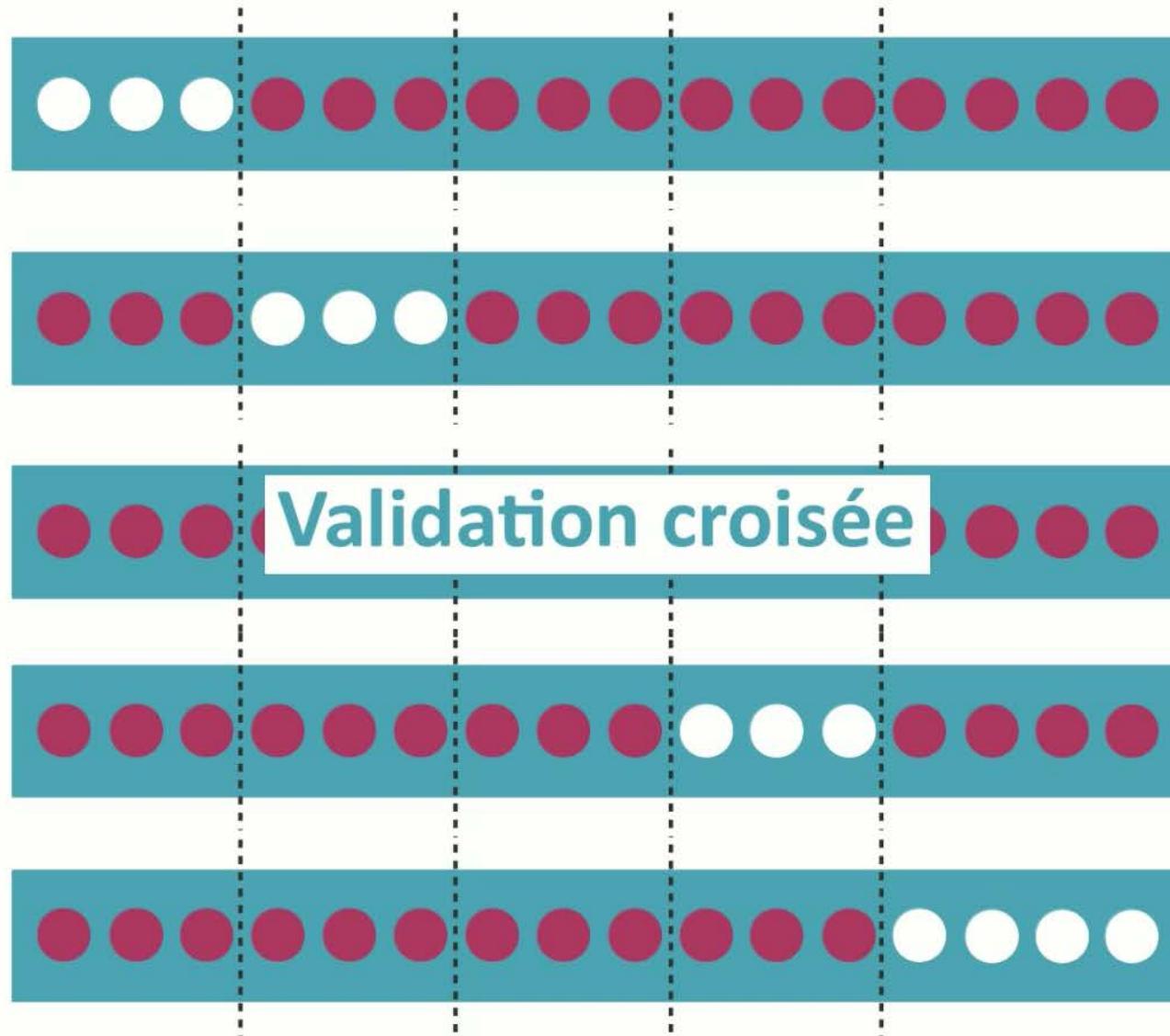
La validation croisée



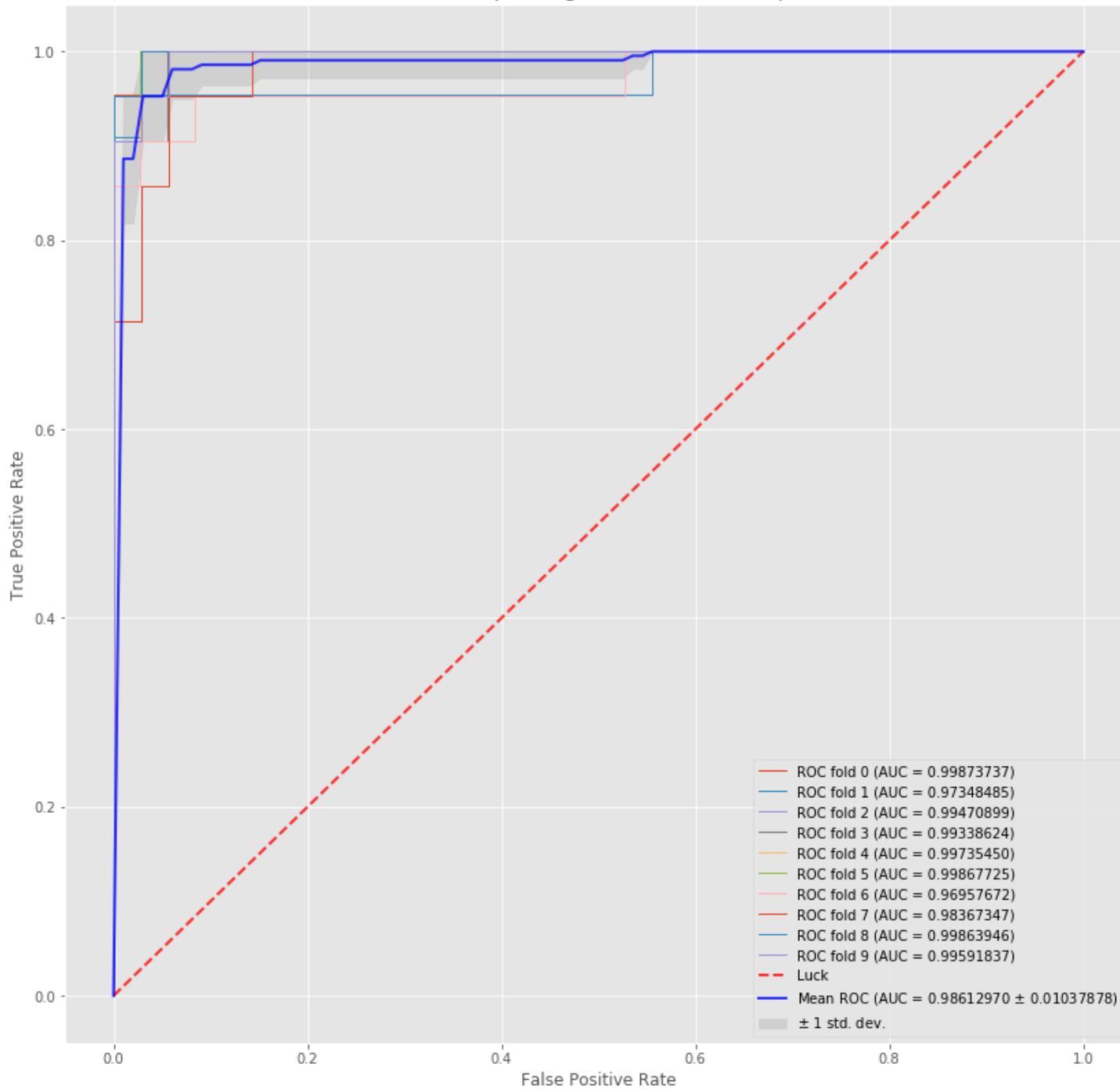
La validation croisée



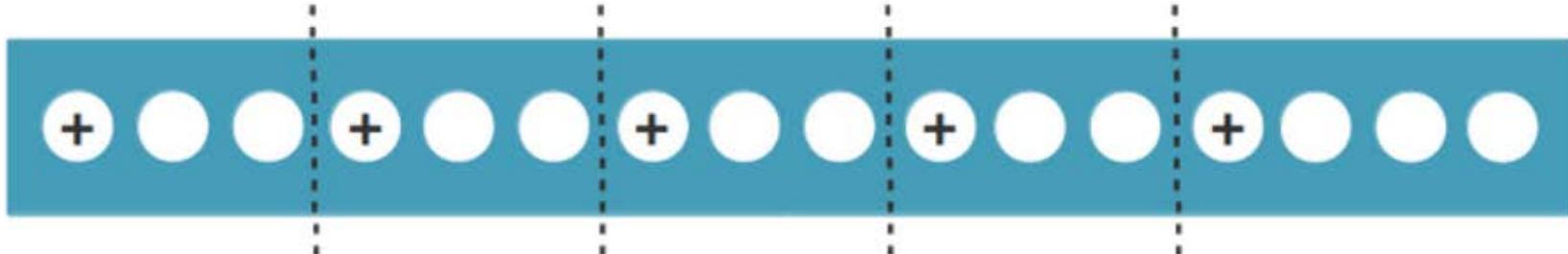
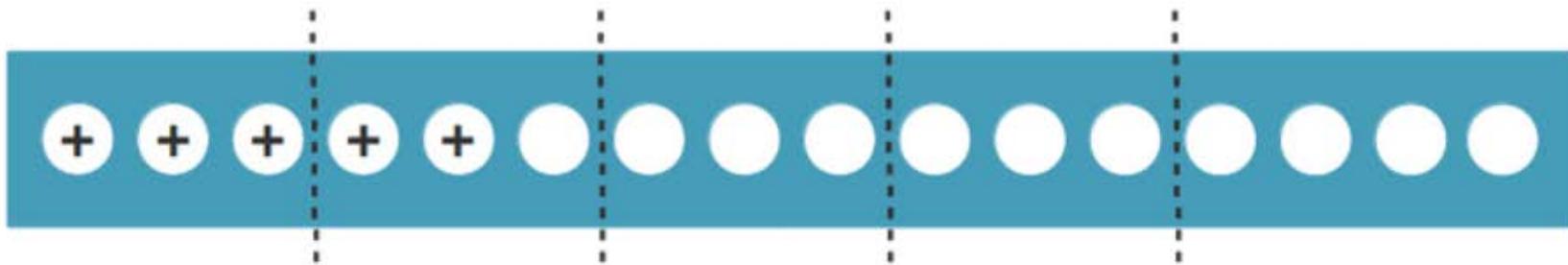
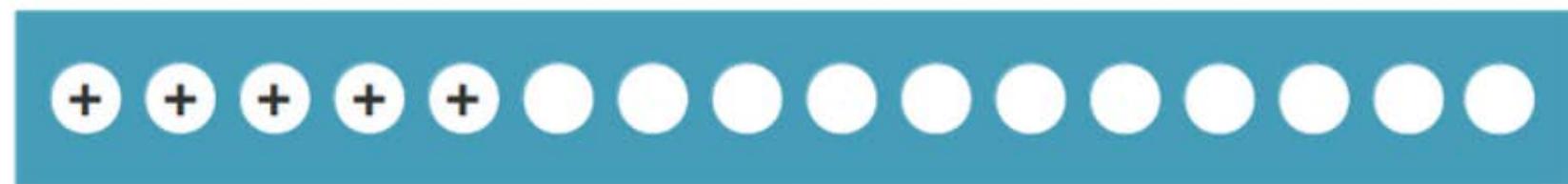
La validation croisée



Receiver operating characteristic example



Stratification



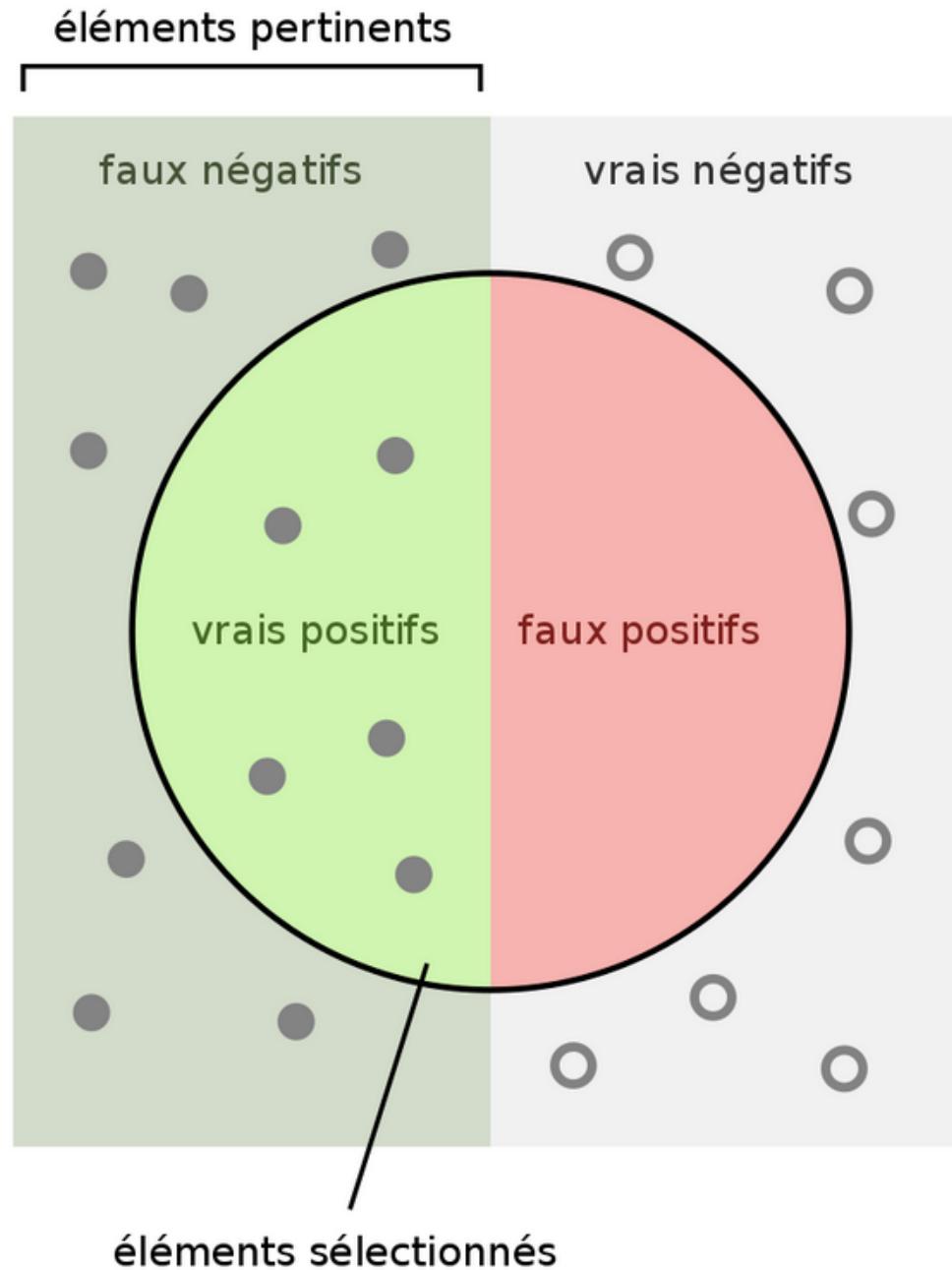
Leave-one-out: k=n



... etc...

Module 10 : Choix des métrique pour la classification

La matrice de confusion



La matrice de confusion

		Observations		Total
Prédictions	+	+	-	
	+	Vrais positifs (VP)	Faux positifs (FP)	Total des positifs prédits (VP + FP)
	-	Faux négatifs (FN)	Vrais négatifs (VN)	Total des négatifs prédits (FN + VN)
Total		Total des vrais positifs observés (VP + FN)	Total des vrais négatifs observés (FP + VN)	Taille totale de l'échantillon (N)

Les termes définis par la matrice de confusion

		Observations		Total
Prédictions	+	+	-	
	+	Vrais positifs (VP)	Faux positifs (FP)	Total des positifs prédits (VP + FP)
	-	Faux négatifs (FN)	Vrais négatifs (VN)	Total des négatifs prédits (FN + VN)
Total		Total des vrais positifs observés (VP + FN)	Total des vrais négatifs observés (FP + VN)	Taille totale de l'échantillon (N)

$$\text{Sensibilité} \text{ (Sensitivity)} = \frac{\text{VP}}{\text{VP} + \text{FN}}$$

$$\text{Précision} = \frac{\text{VP}}{\text{VP} + \text{FP}}$$

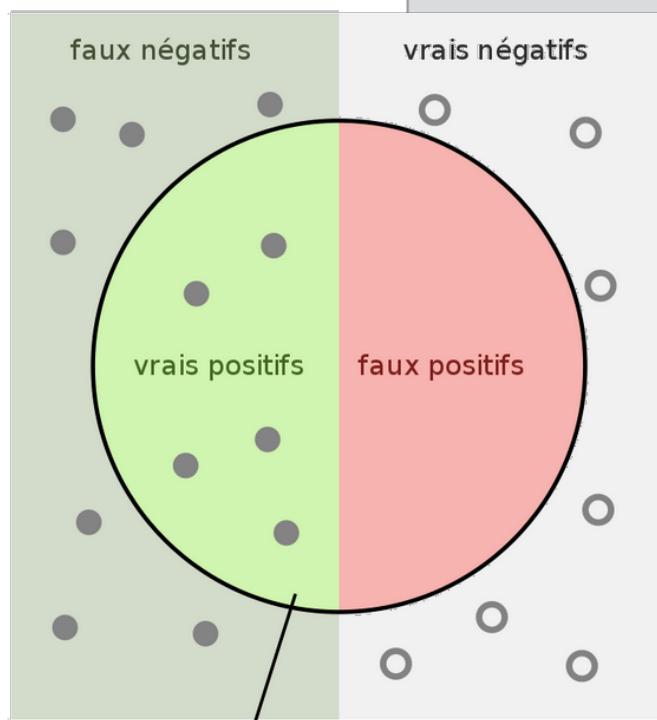
$$\text{Rappel} \text{ (Recall)} = \frac{\text{VP}}{\text{VP} + \text{FN}}$$

$$\text{F1-score} = \frac{2 \times \text{P} \times \text{R}}{\text{P} + \text{R}}$$

$$\text{Spécificité} \text{ (Specificity)} = \frac{\text{VN}}{\text{FP} + \text{VN}}$$

La matrice de confusion

		Observations		Total
Prédictions	+	+	-	
		Vrais positifs (VP)	Faux positifs (FP)	
éléments pertinents	+	Faux négatifs (FN)	Vrais négatifs (VN)	Total des positifs prédits (VP + FP)
	-	Total des vrais positifs observés (VP + FN)	Total des vrais négatifs observés (FP + VN)	Total des négatifs prédits (FN + VN)
Total		Taille totale de l'échantillon (N)		



Combien de candidats sélectionnés sont pertinents ?

$$\text{Précision} = \frac{\text{vrais positifs}}{\text{éléments sélectionnés}}$$



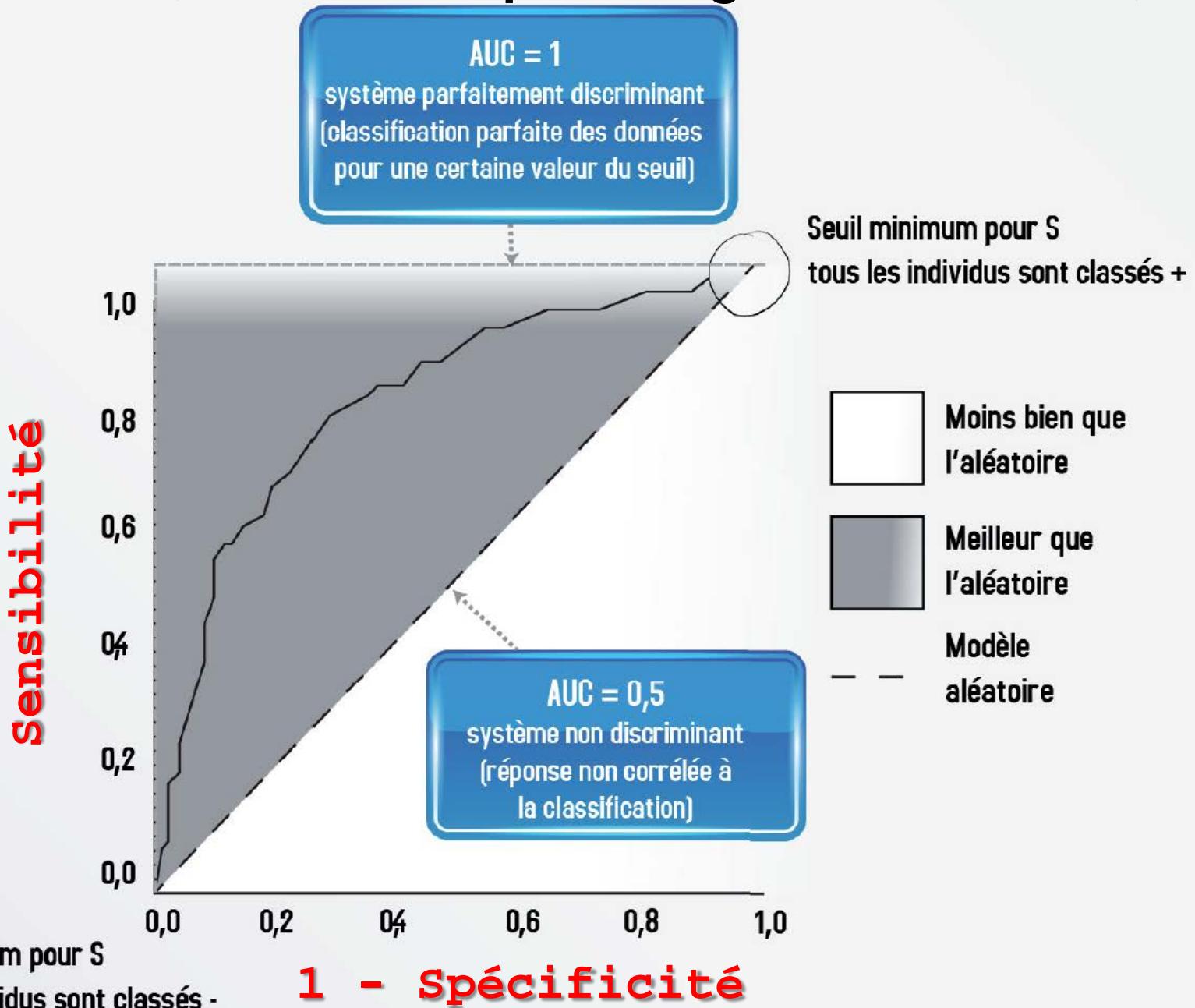
Combien d'éléments pertinents sont sélectionnés ?

$$\text{Rappel} = \frac{\text{vrais positifs}}{\text{éléments pertinents}}$$

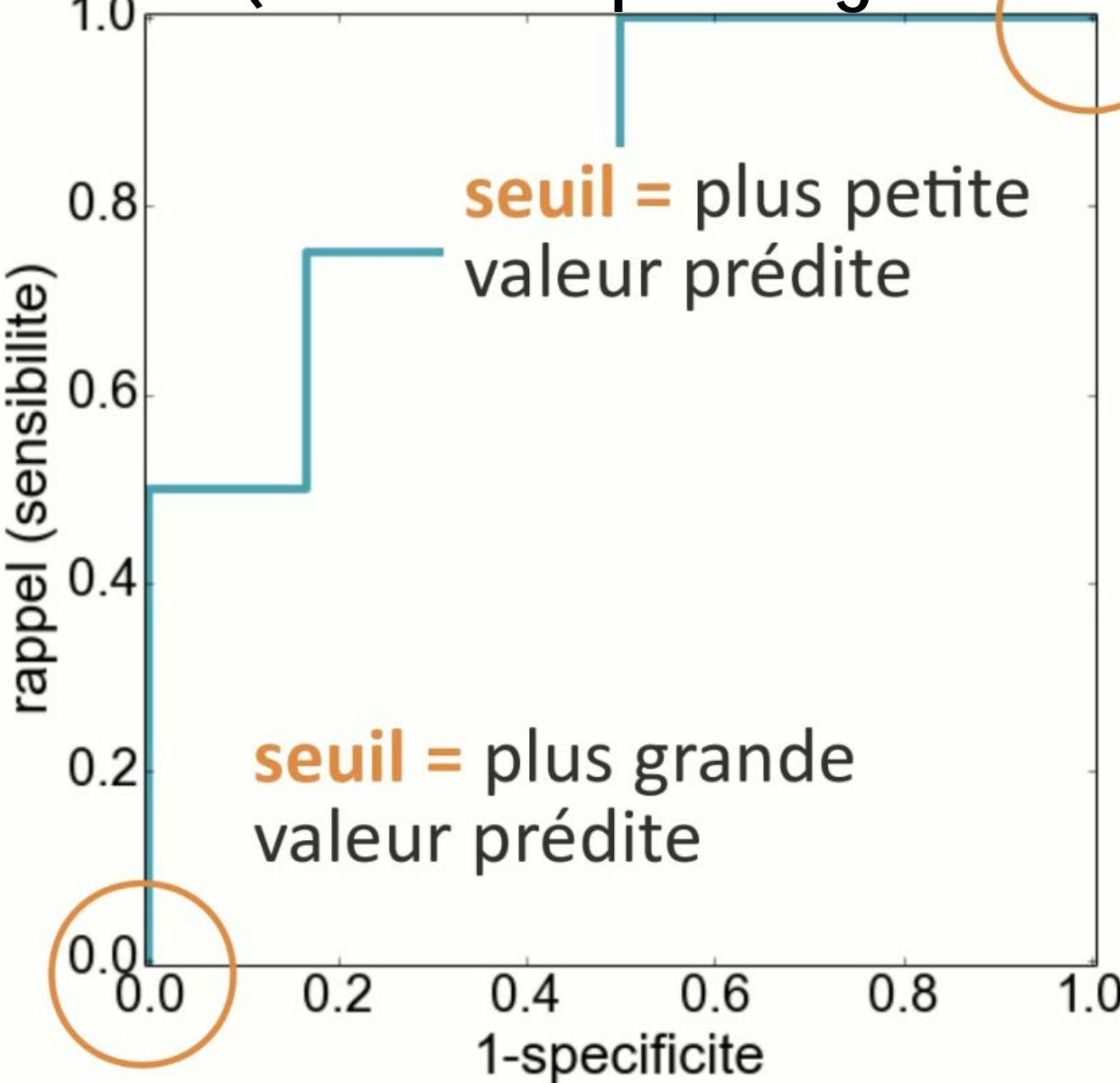


Sensibilité

La courbe ROC (Receiver Operating Characteristic)

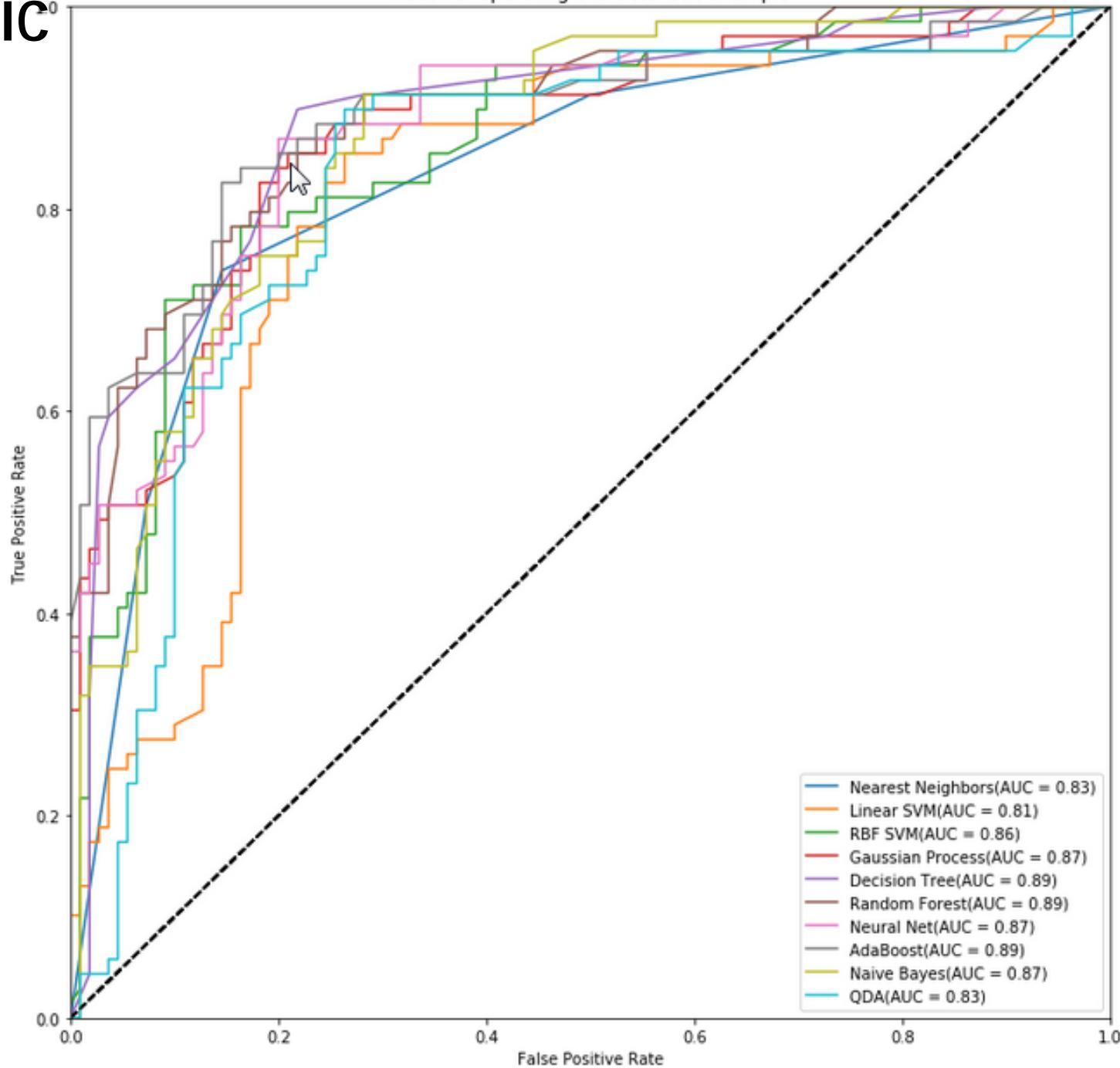


La courbe ROC (Receiver Operating Characteristic)



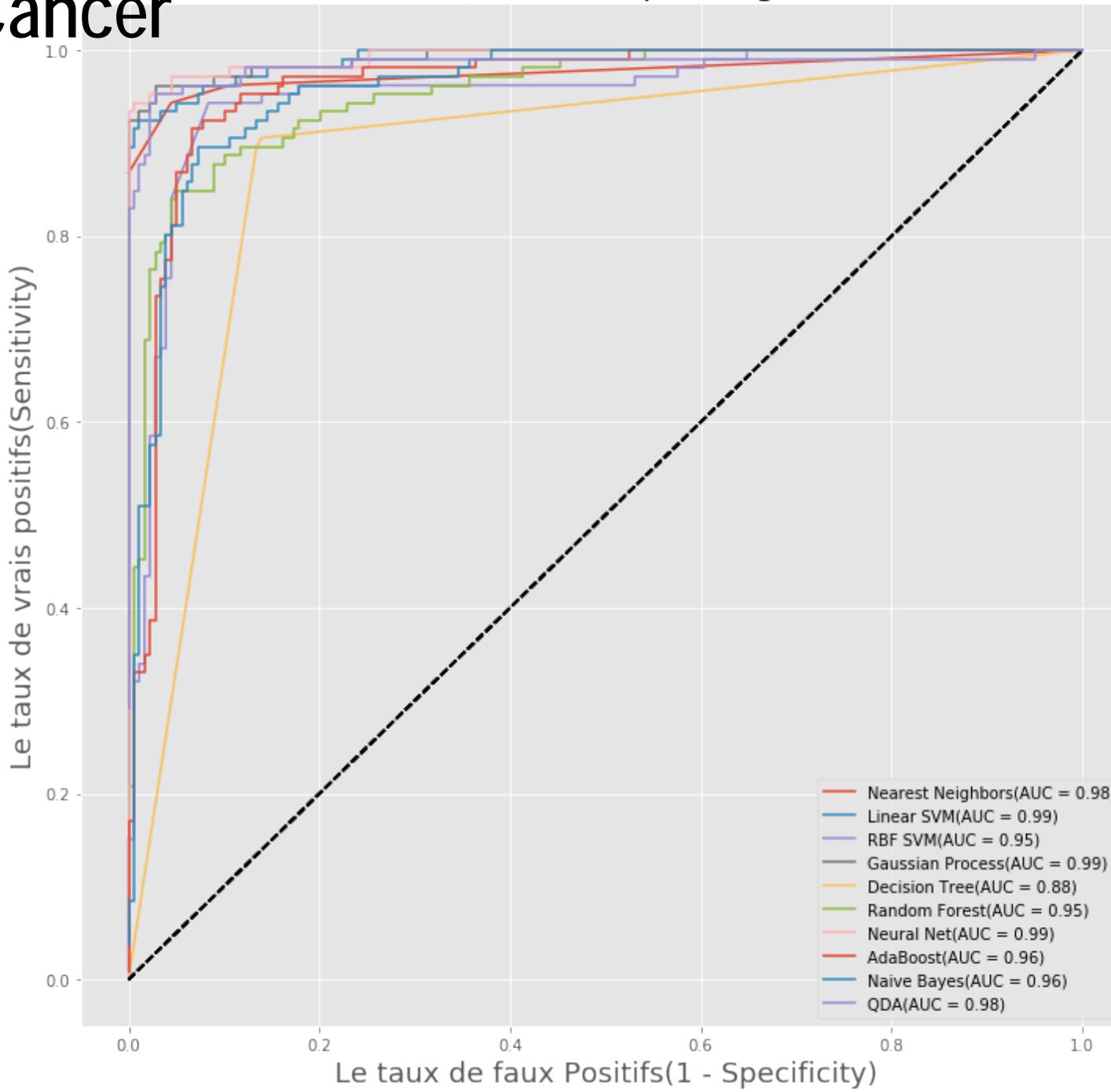
Titanic⁻⁰

Receiver operating characteristic example



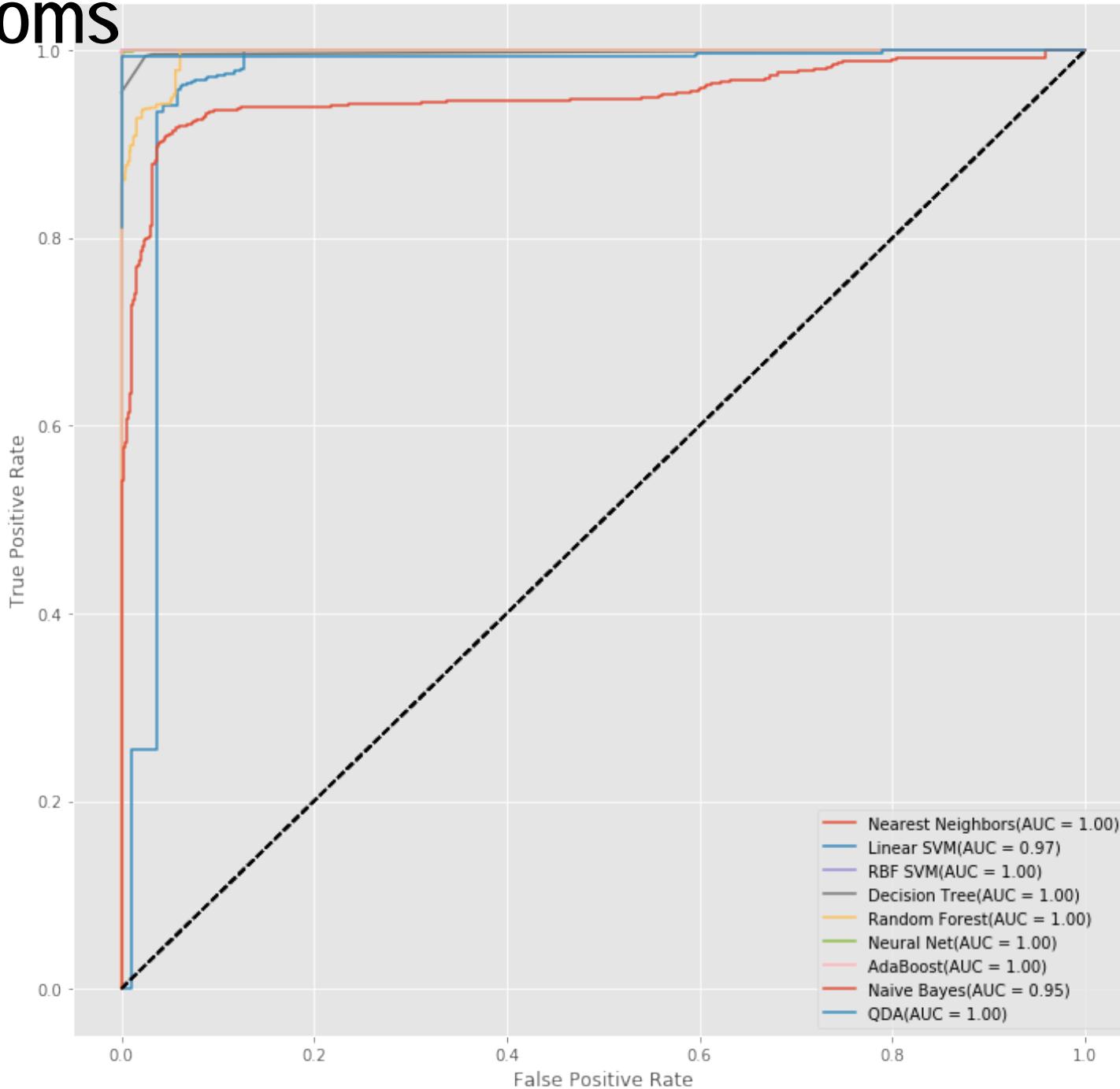
BreastCancer

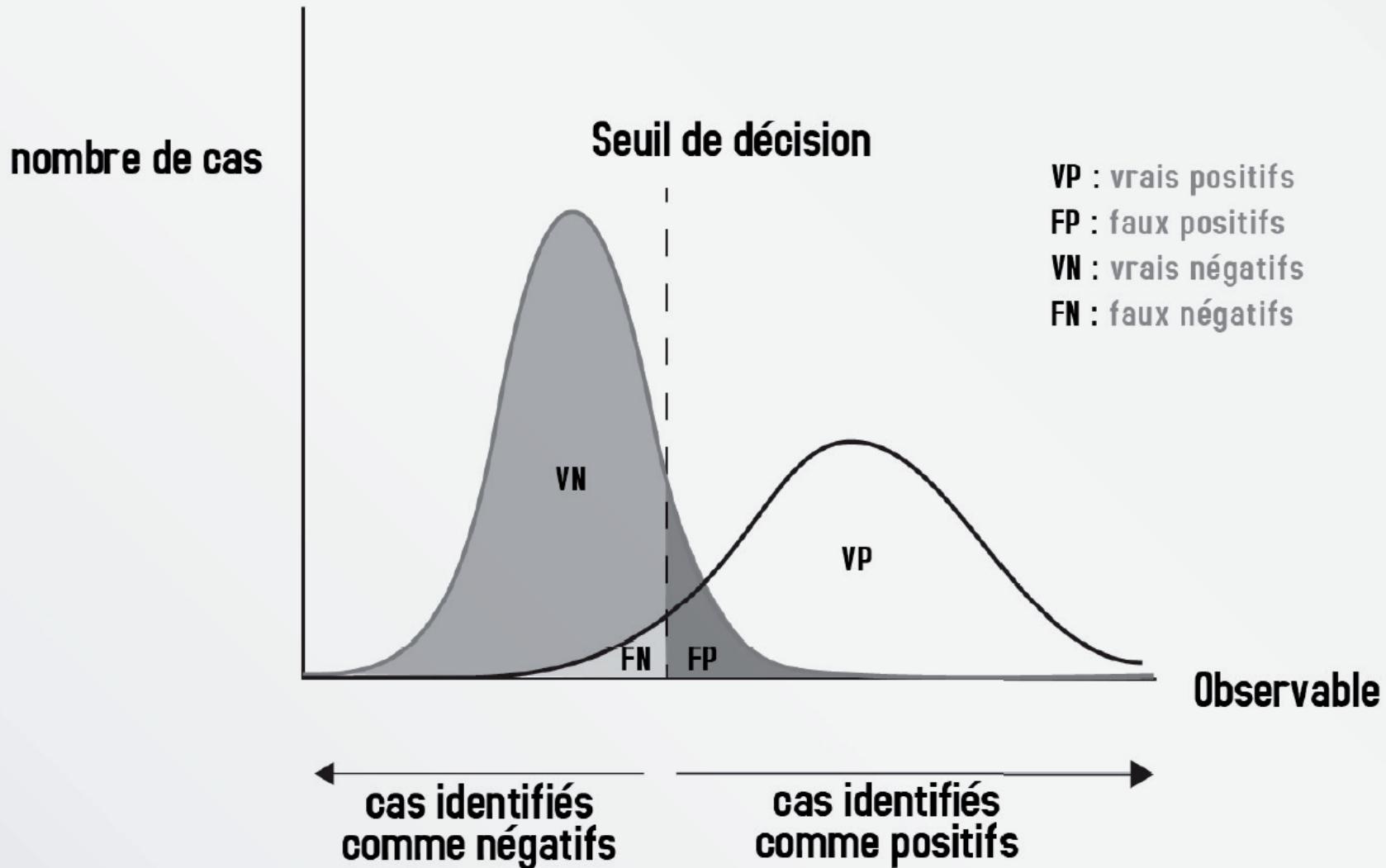
La courbe ROC (Receiver Operating Characteristic)



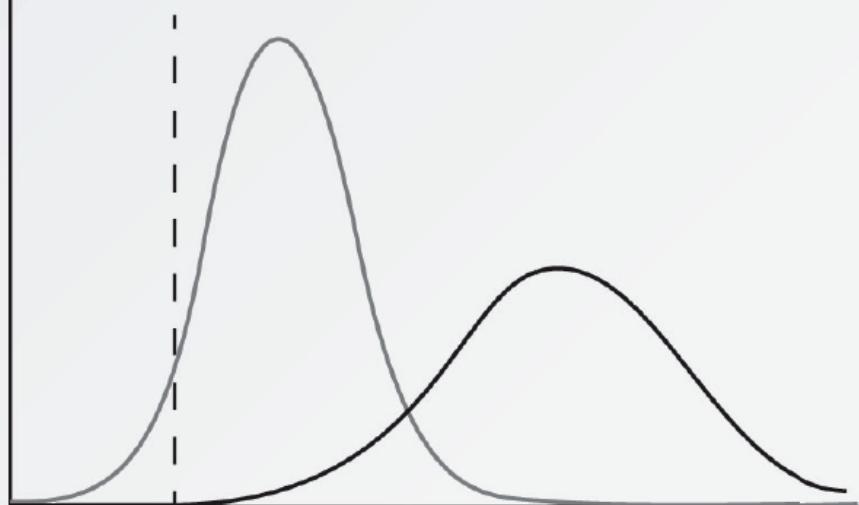
Mushrooms

Receiver operating characteristic example



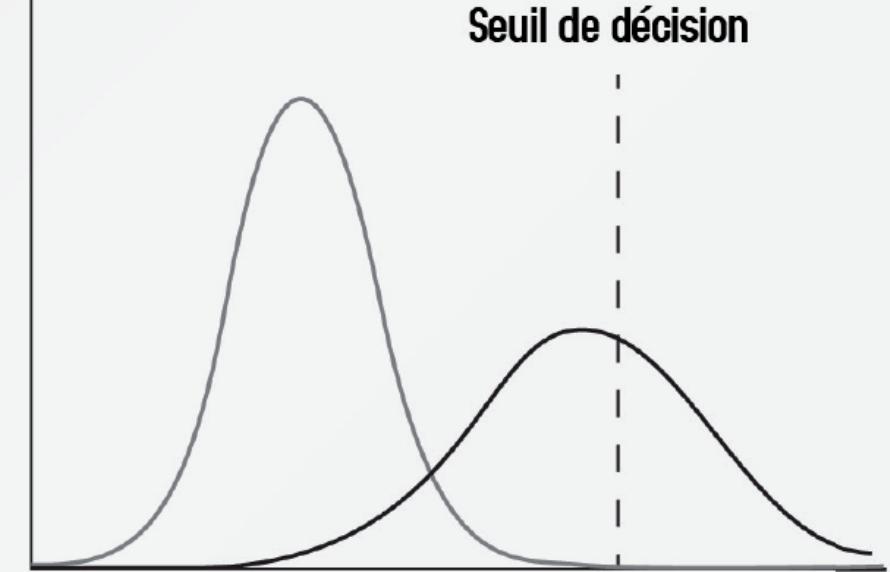


Seuil de décision

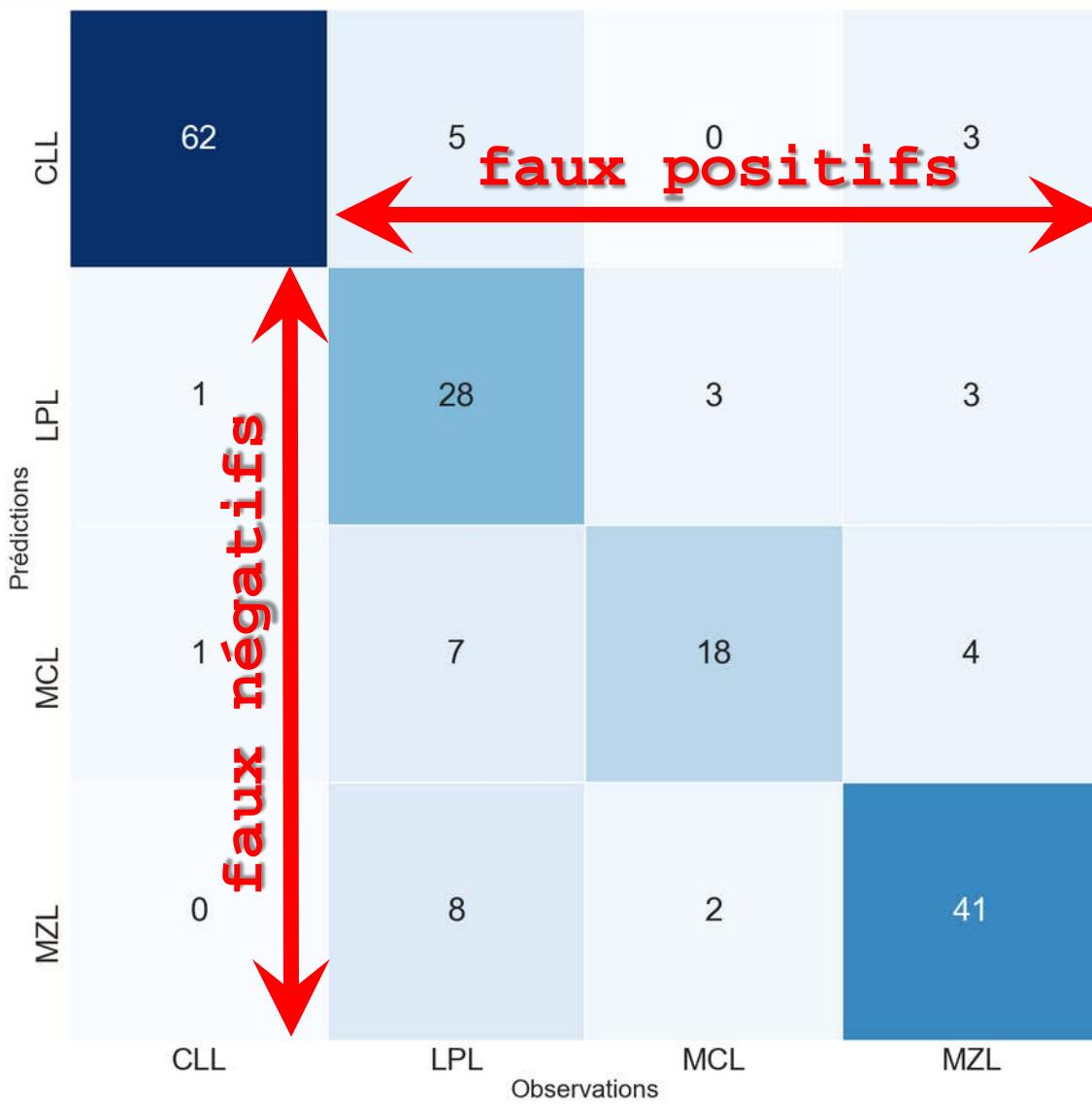


forte sensibilité, faible spécificité

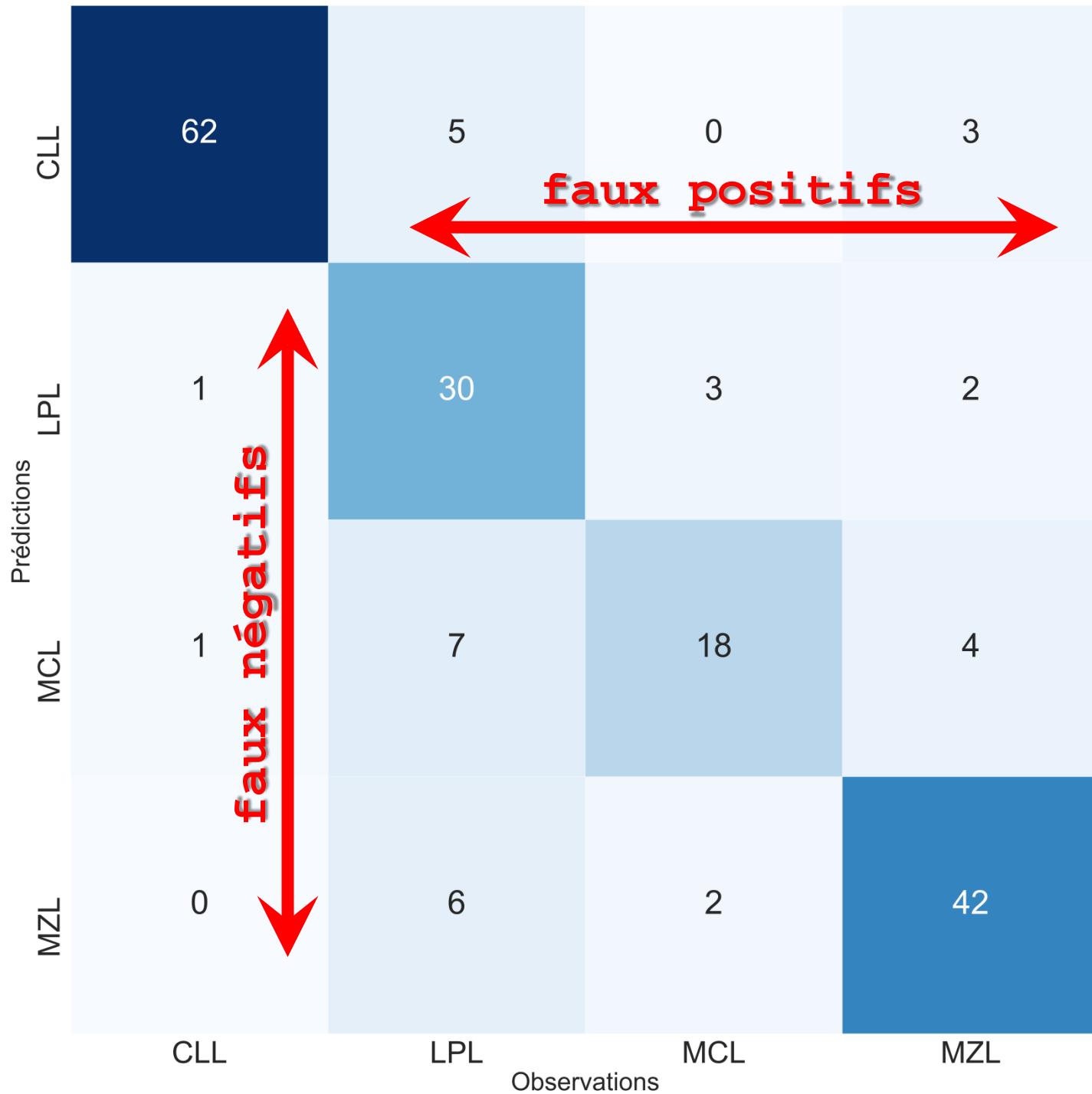
Seuil de décision



faible sensibilité, forte spécificité

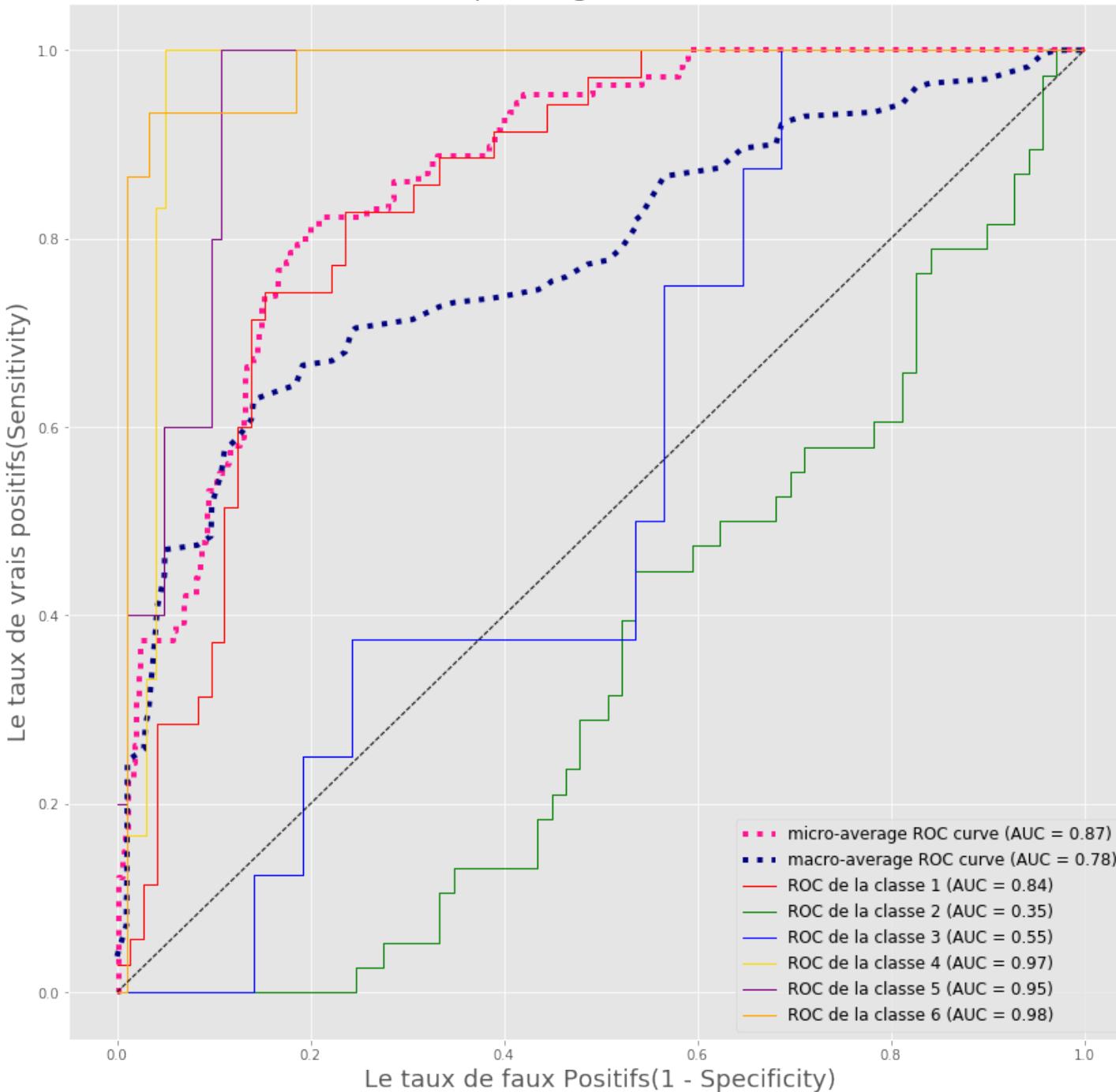


autre	114	2	autre	131	20	autre	151	5	autre	125	10
CLL	8	62	LPL	7	28	MCL	12	18	MZL	10	41
	autre	CLL		autre	LPL		autre	MCL		autre	MZL



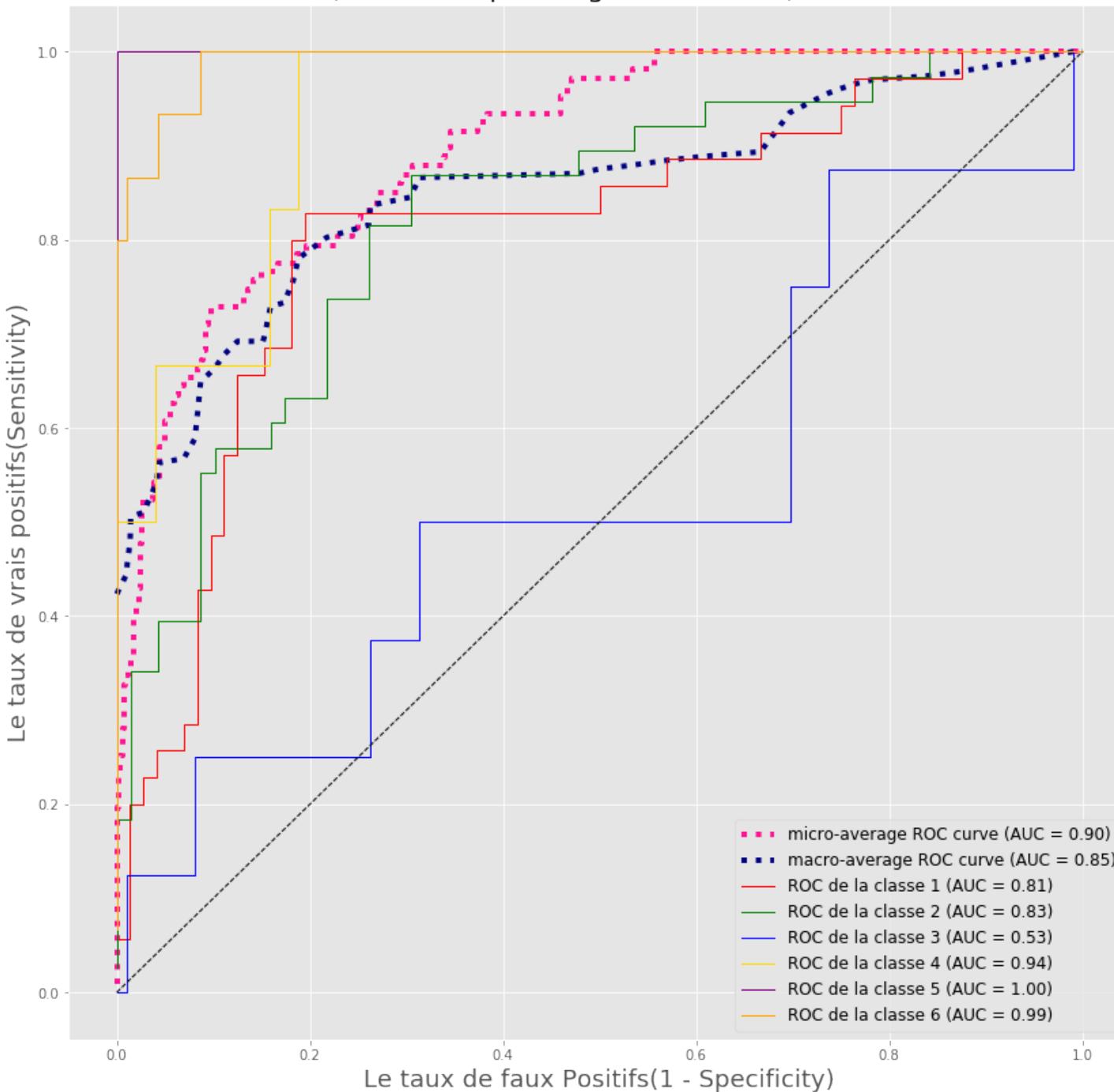
La courbe ROC (Receiver Operating Characteristic) multi-class Linear SVM

Glass

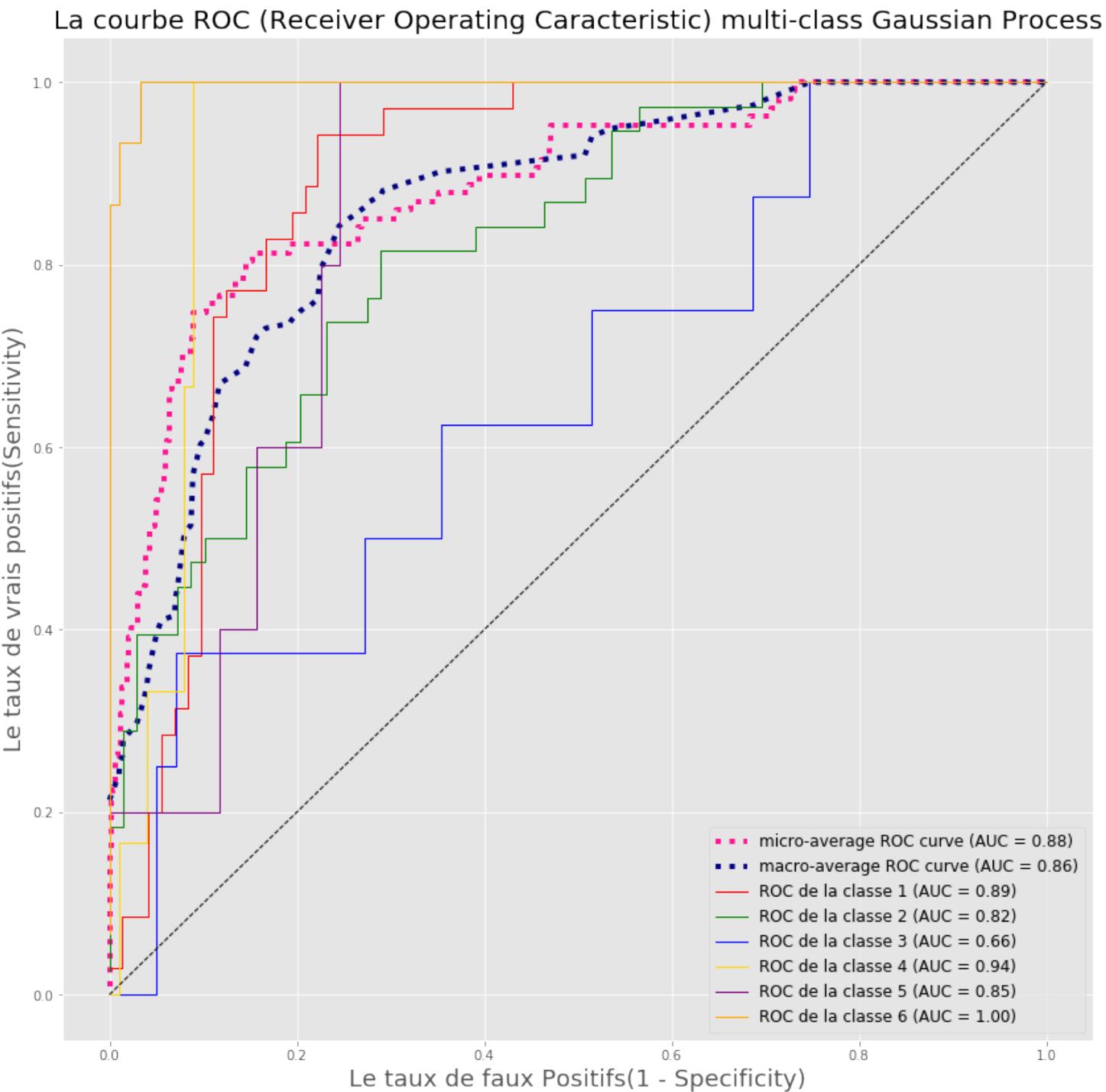


La courbe ROC (Receiver Operating Characteristic) multi-class RBF SVM

Glass

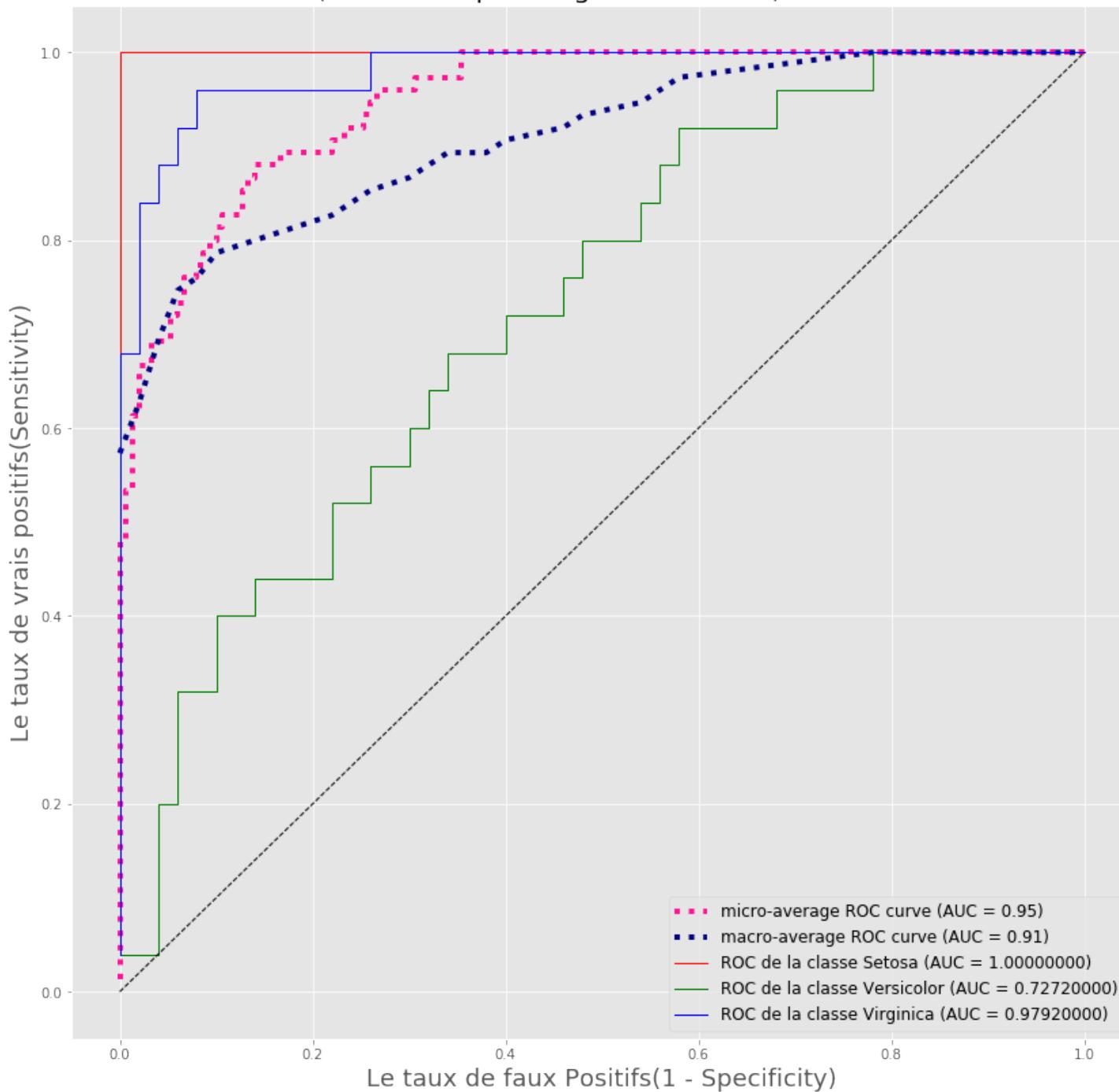


Glass



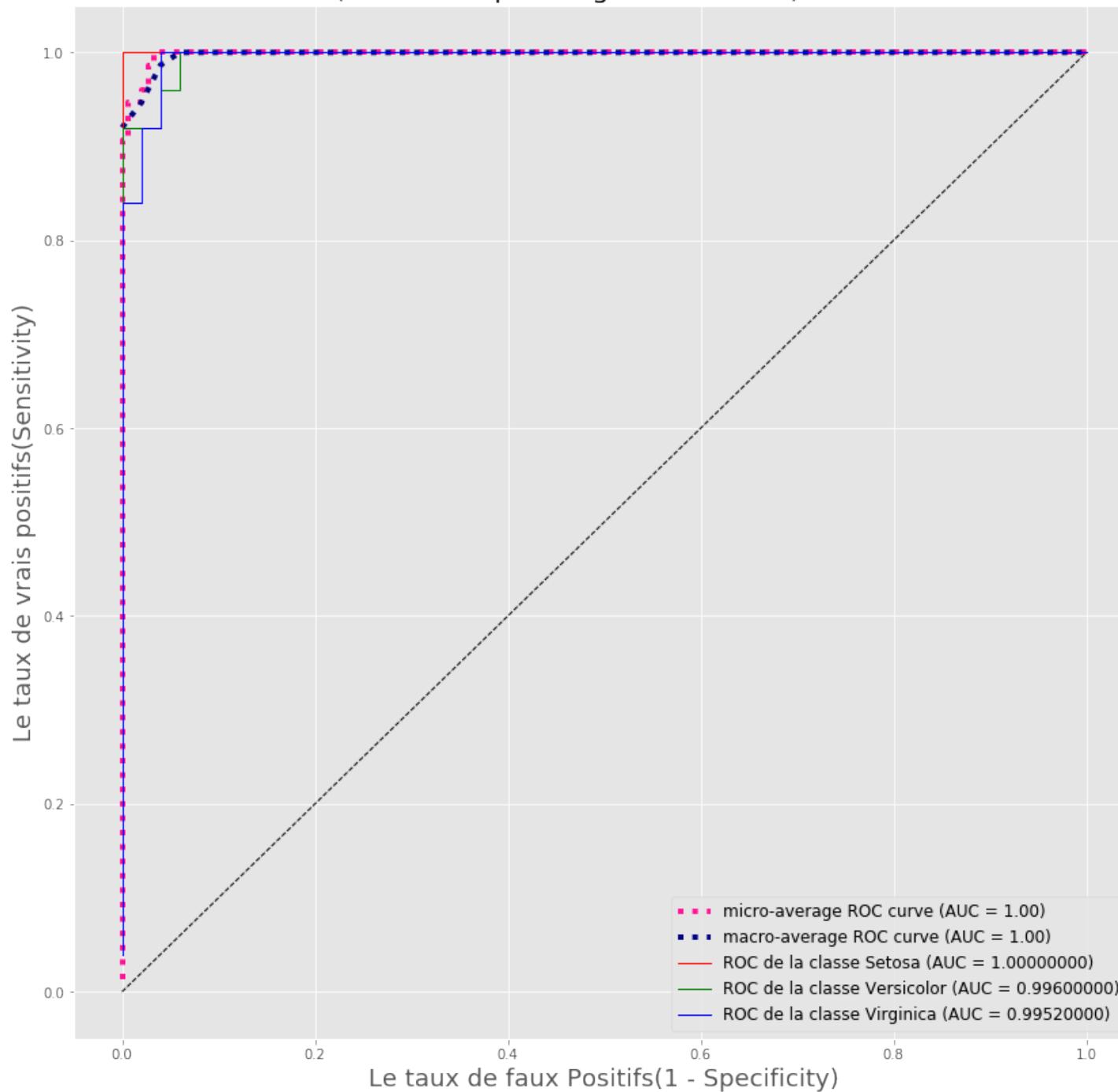
Iris

La courbe ROC (Receiver Operating Characteristic) multi-class Linear SVM



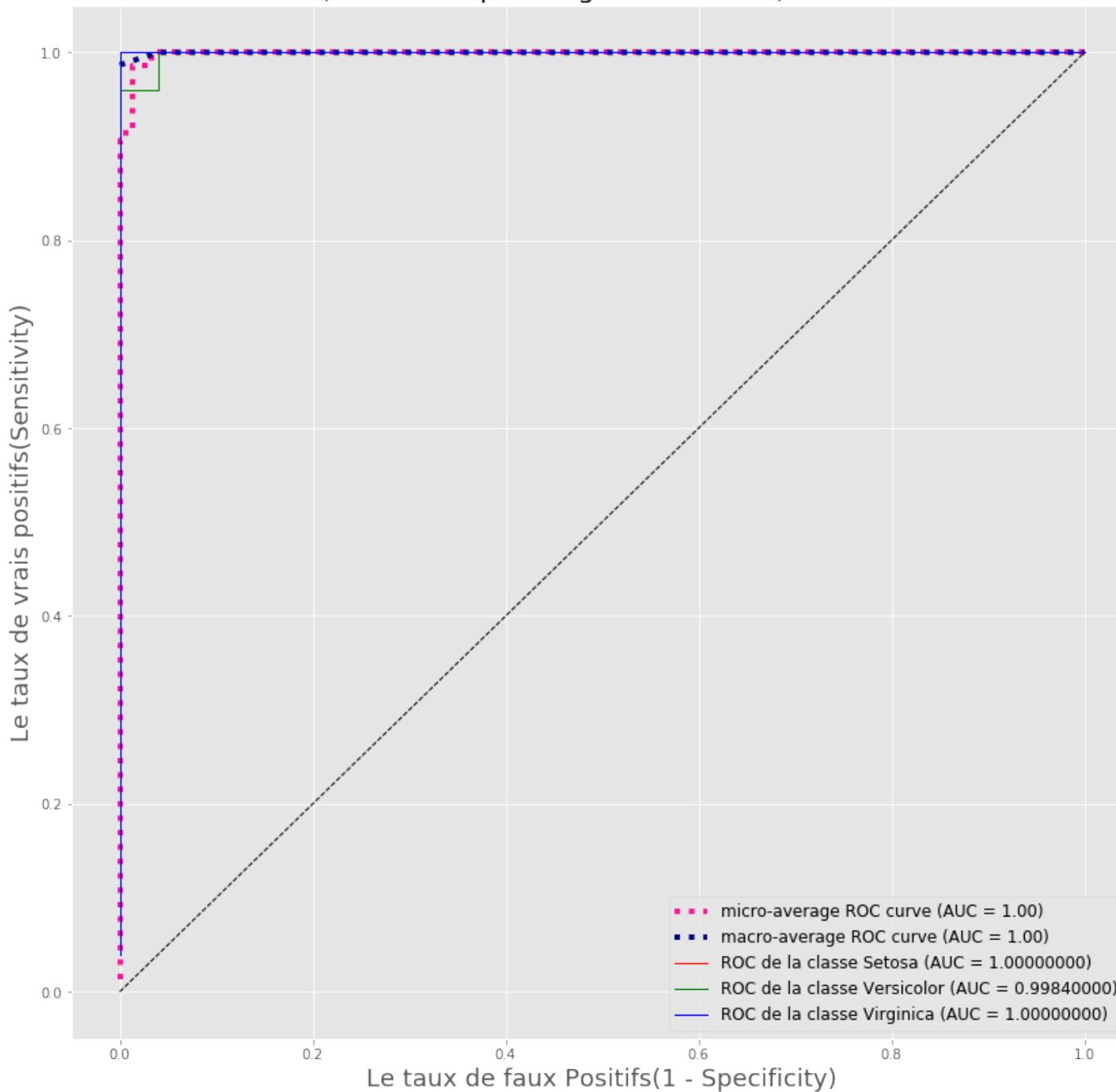
Iris

La courbe ROC (Receiver Operating Characteristic) multi-class RBF SVM

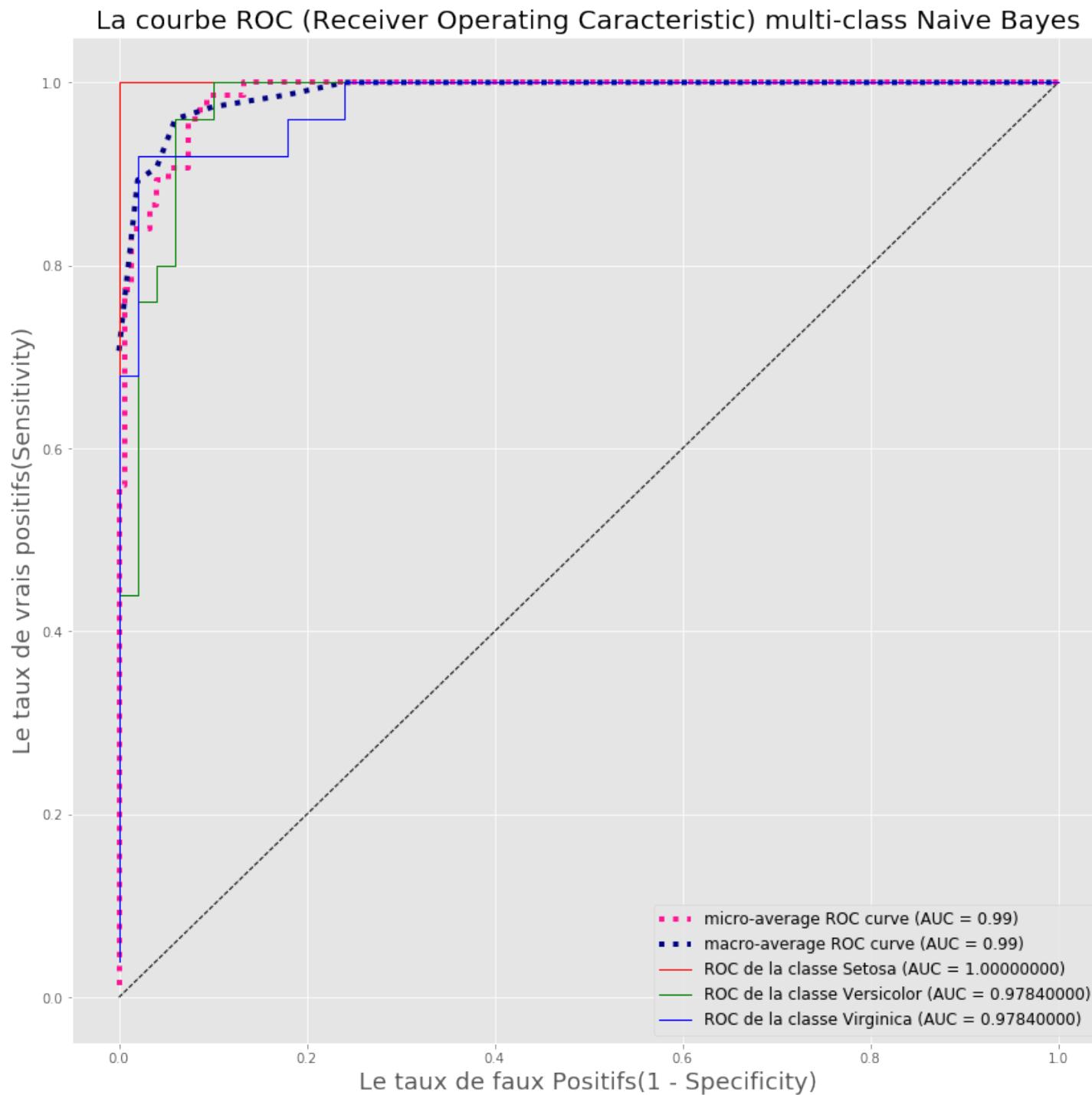


Iris

La courbe ROC (Receiver Operating Characteristic) multi-class Neural Net



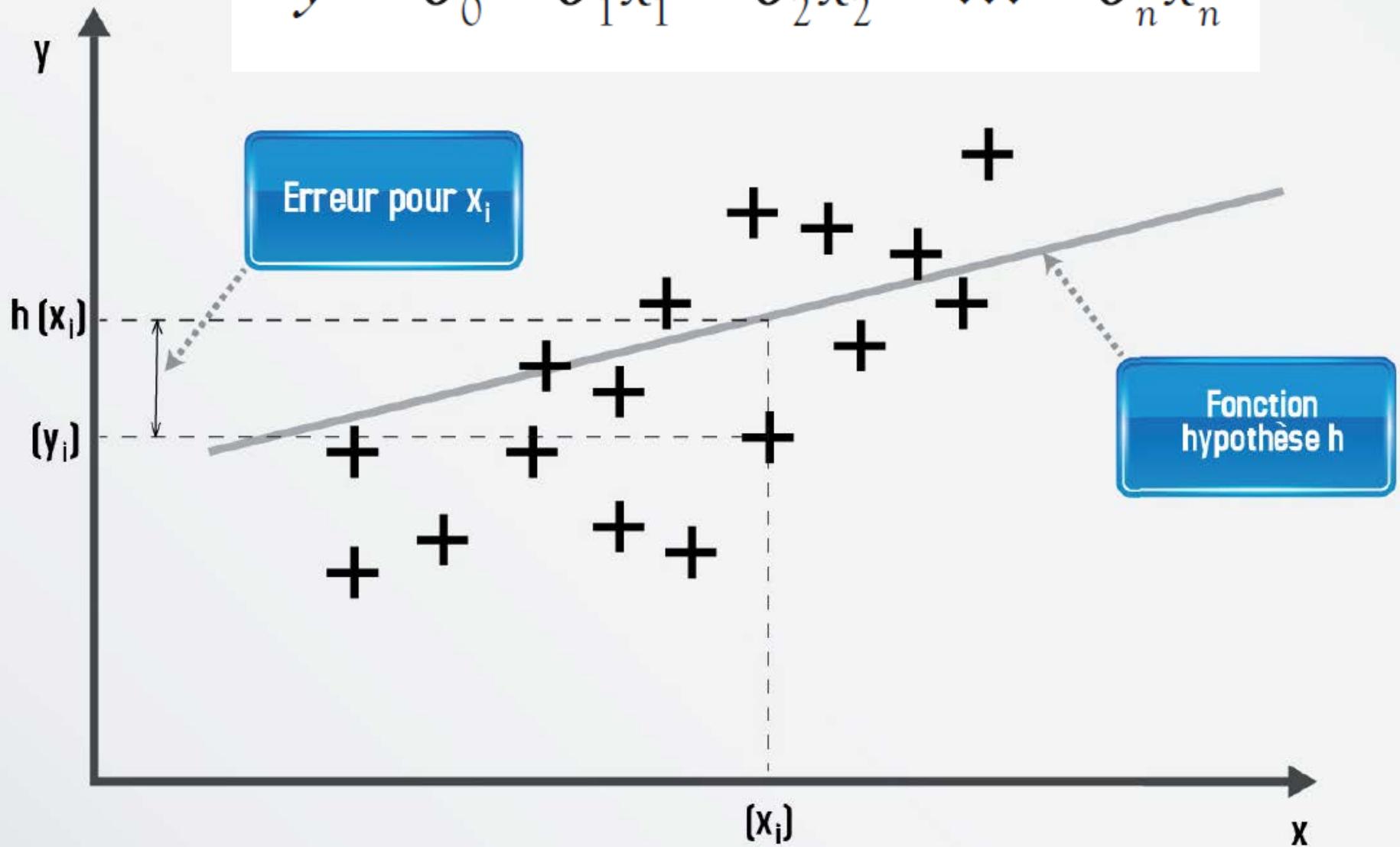
Iris



Algorithme	Mode d'apprentissage	Type de problème à traiter
Régression linéaire univariée	Supervisé	Régression
Régression linéaire multivariée	Supervisé	Régression
Régression polynomiale	Supervisé	Régression
Régression régularisée	Supervisé	Régression
Naive Bayes	Supervisé	Classification
Régression logistique	Supervisé	Classification
Clustering hiérarchique	Non supervisé	-
Clustering non hiérarchique	Non supervisé	-
Arbres de décision	Supervisé	Régression ou classification
Random forest	Supervisé	Régression ou classification
Gradient boosting	Supervisé	Régression ou classification
Support Vector Machine	Supervisé	Régression ou classification
PCA -- Analyse en composantes principales		-

Module 11 : Choix des métrique pour la régression

La régression linéaire





$$y = f_{\beta}(x) + \epsilon$$

y

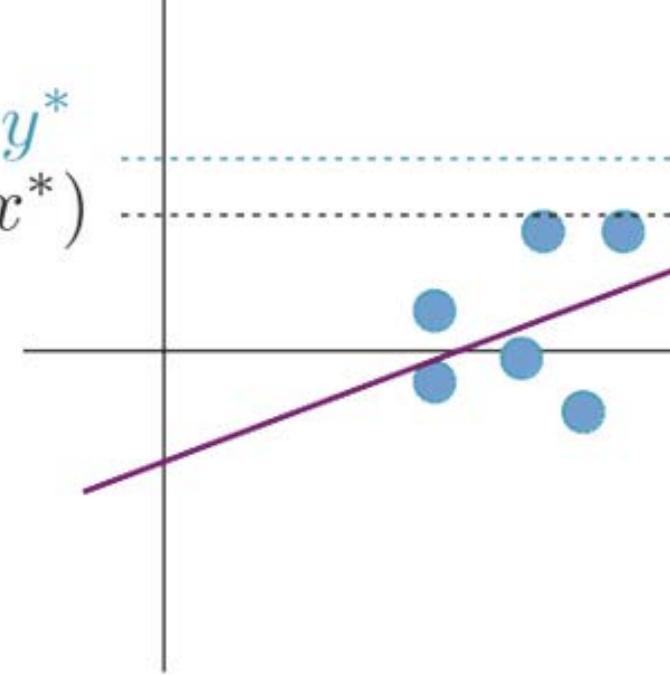
$$f_{\beta}(x^*)$$

$$f_{\beta}(x) = \beta_1 x + \beta_0$$

$$p(y|x^*) \sim \mathcal{N}(f_{\beta}(x^*), \sigma^2)$$

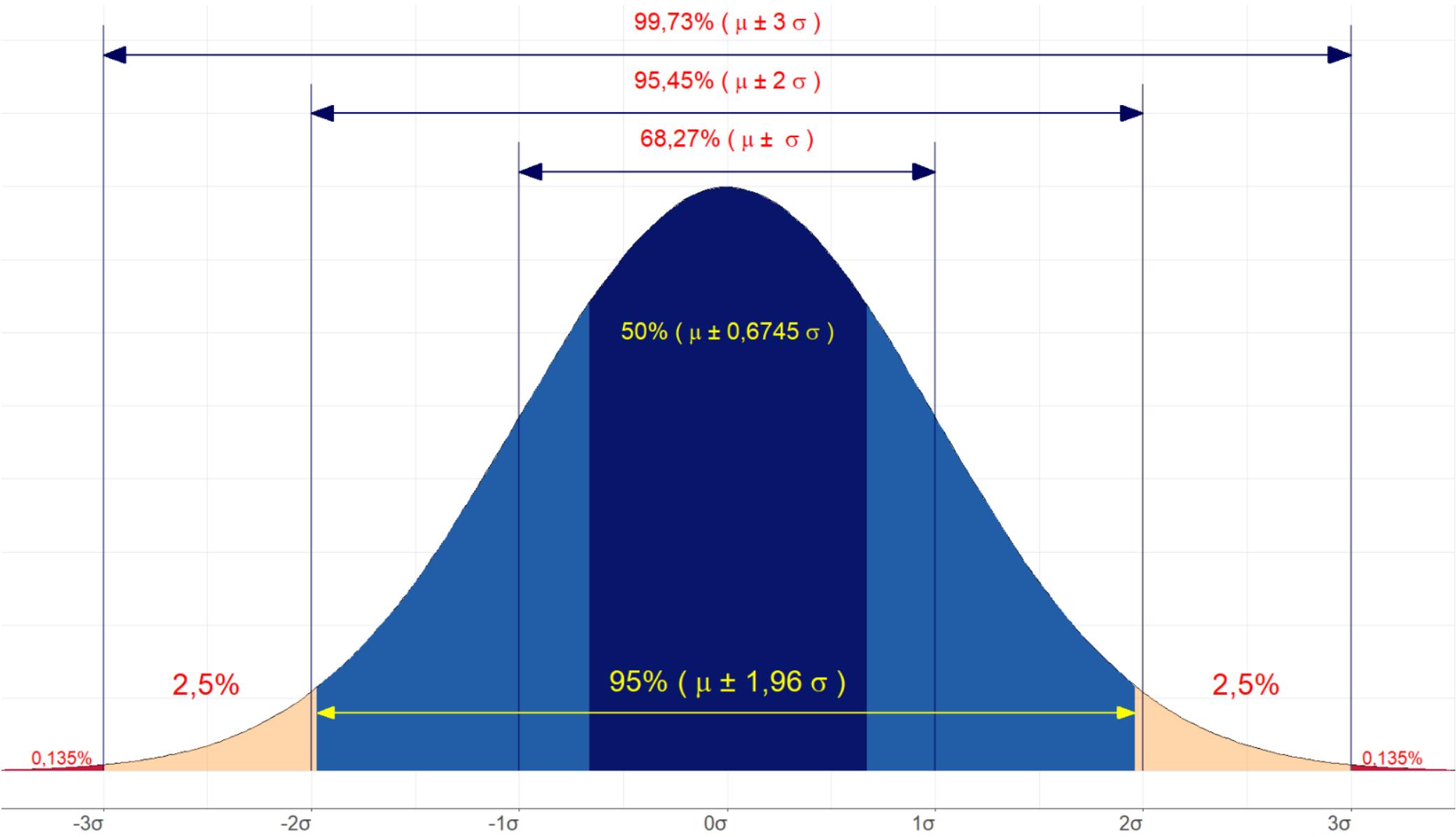
x

x^{*}



Loi Normale centrée réduite

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad \mathcal{N}(0, 1)$$



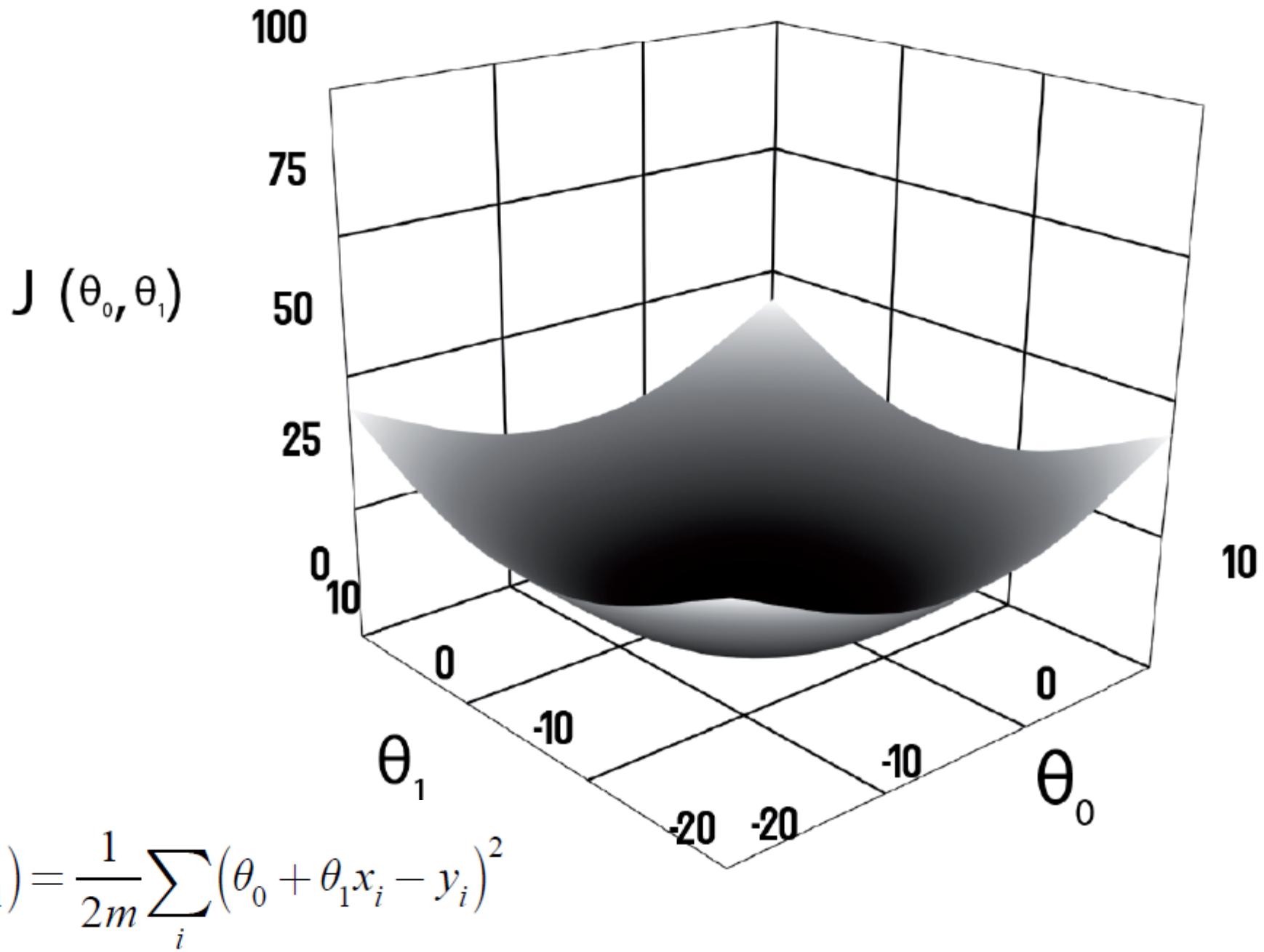
Box-Cox

$$y_i^{(\lambda)} = \begin{cases} \frac{y_i^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0, \\ \ln(y_i) & \text{if } \lambda = 0, \end{cases}$$

Yeo-Johnson

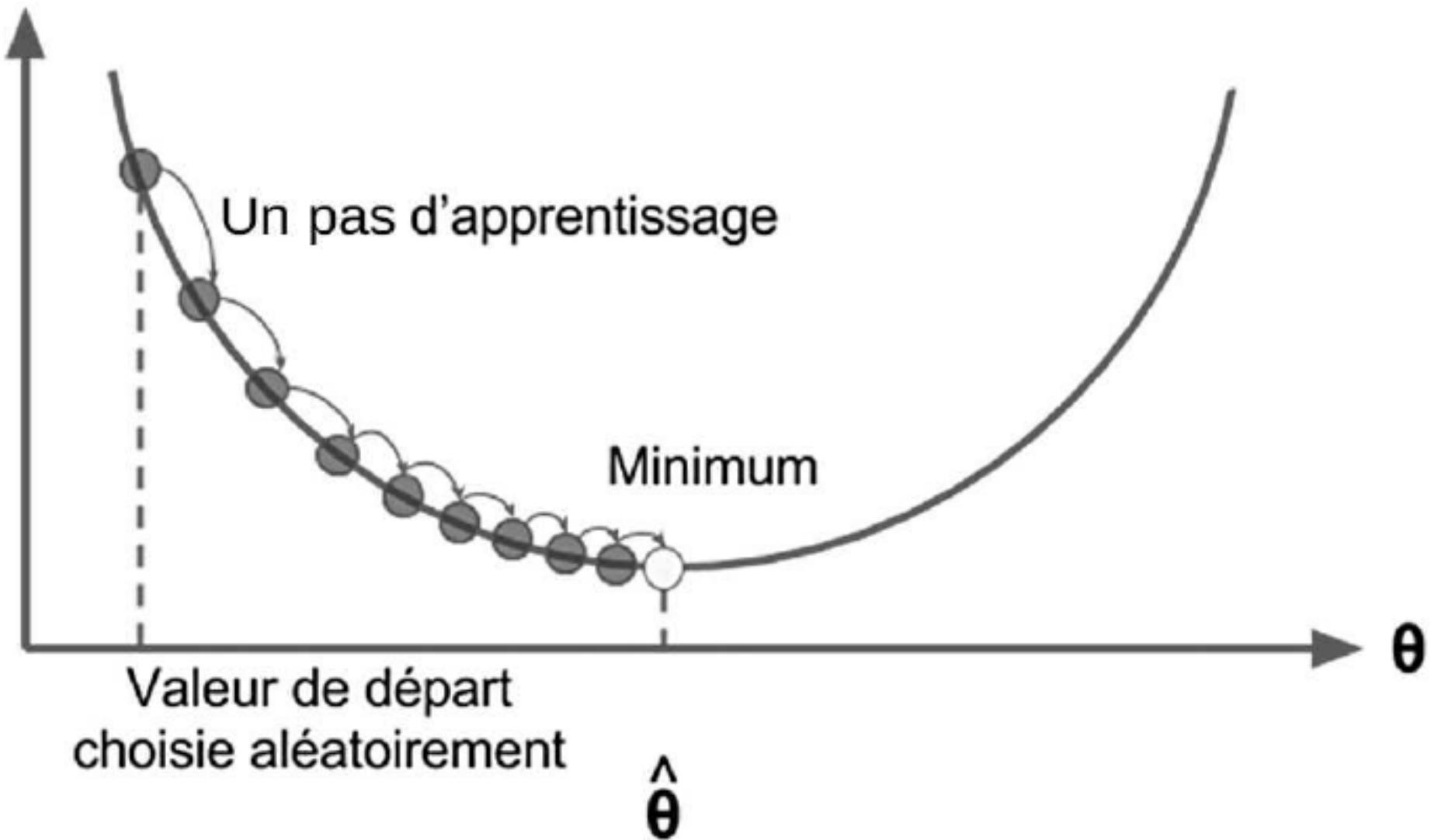
$$y_i^{(\lambda)} = \begin{cases} ((y_i + 1)^\lambda - 1)/\lambda & \text{if } \lambda \neq 0, y \geq 0 \\ \log(y_i + 1) & \text{if } \lambda = 0, y \geq 0 \\ -[(-y_i + 1)^{(2-\lambda)} - 1]/(2 - \lambda) & \text{if } \lambda \neq 2, y < 0 \\ -\log(-y_i + 1) & \text{if } \lambda = 2, y < 0 \end{cases}$$

Minimiser la fonction de coût



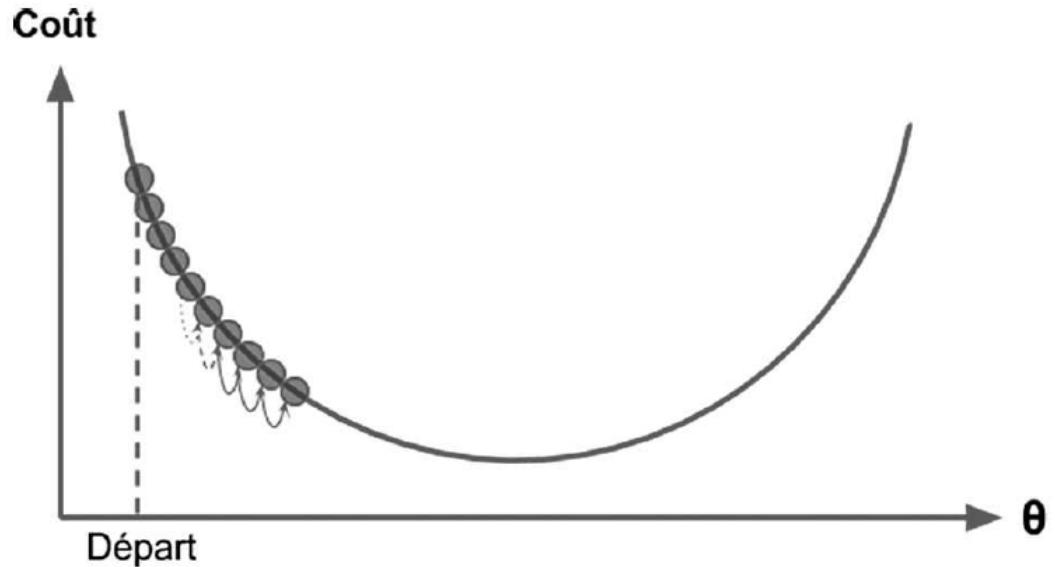
La descente de gradient

Coût

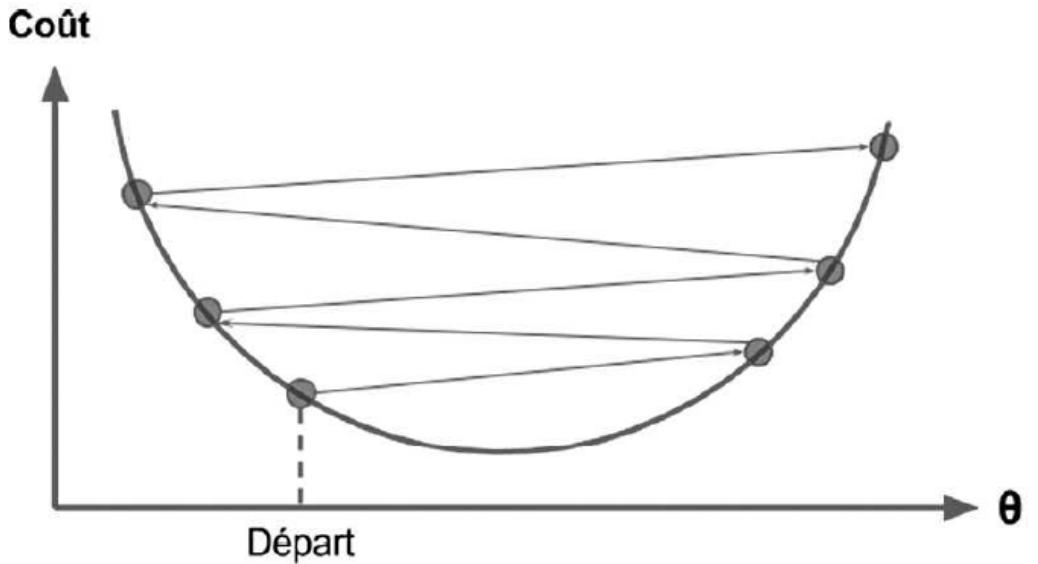


La descente de gradient

Taux d'apprentissage trop petit

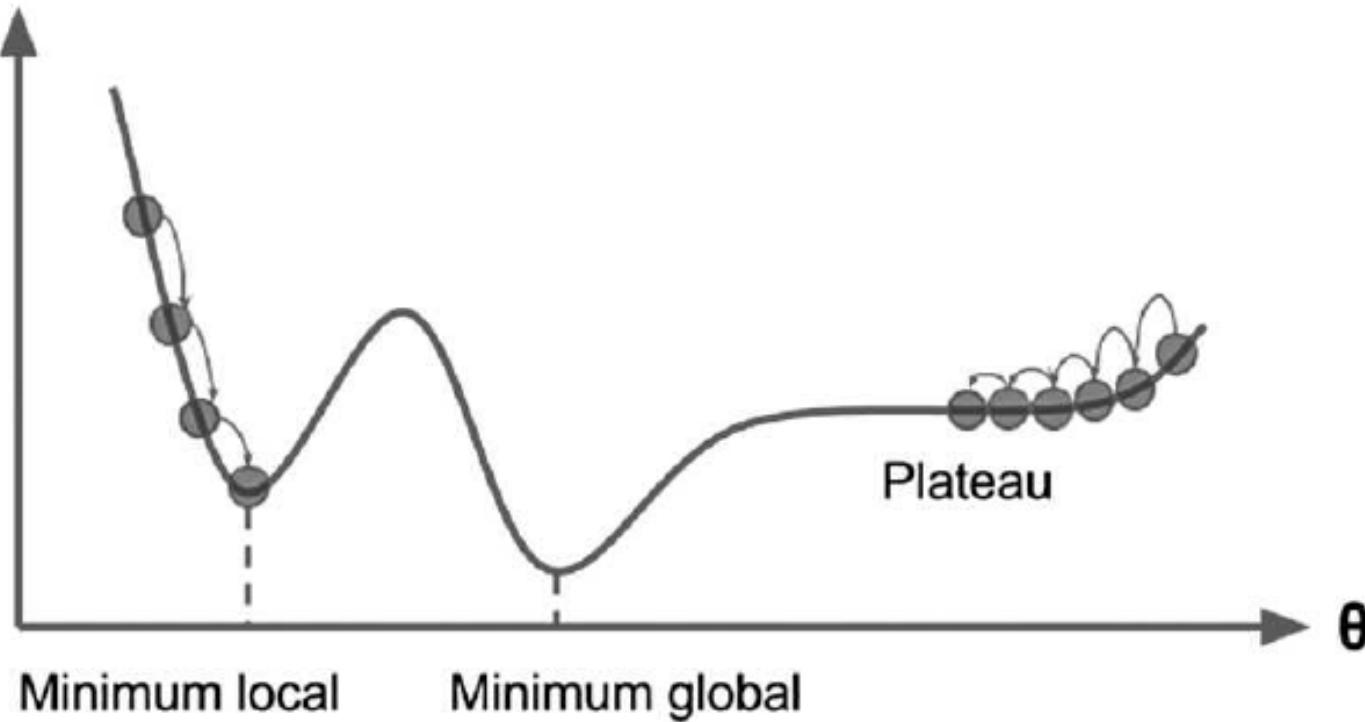


Taux d'apprentissage trop élevé



Pièges de la descente de gradient

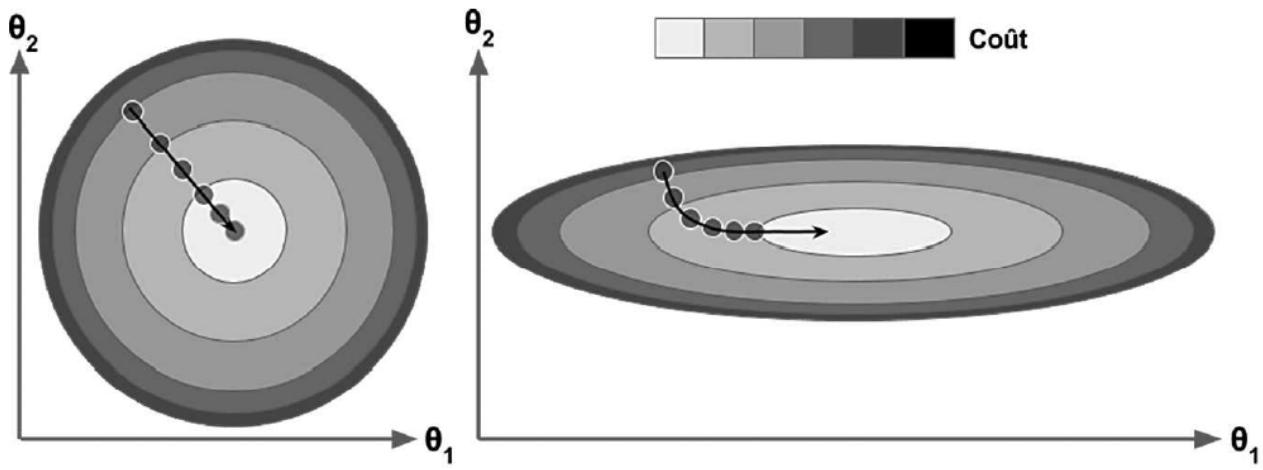
Coût



Minimum local

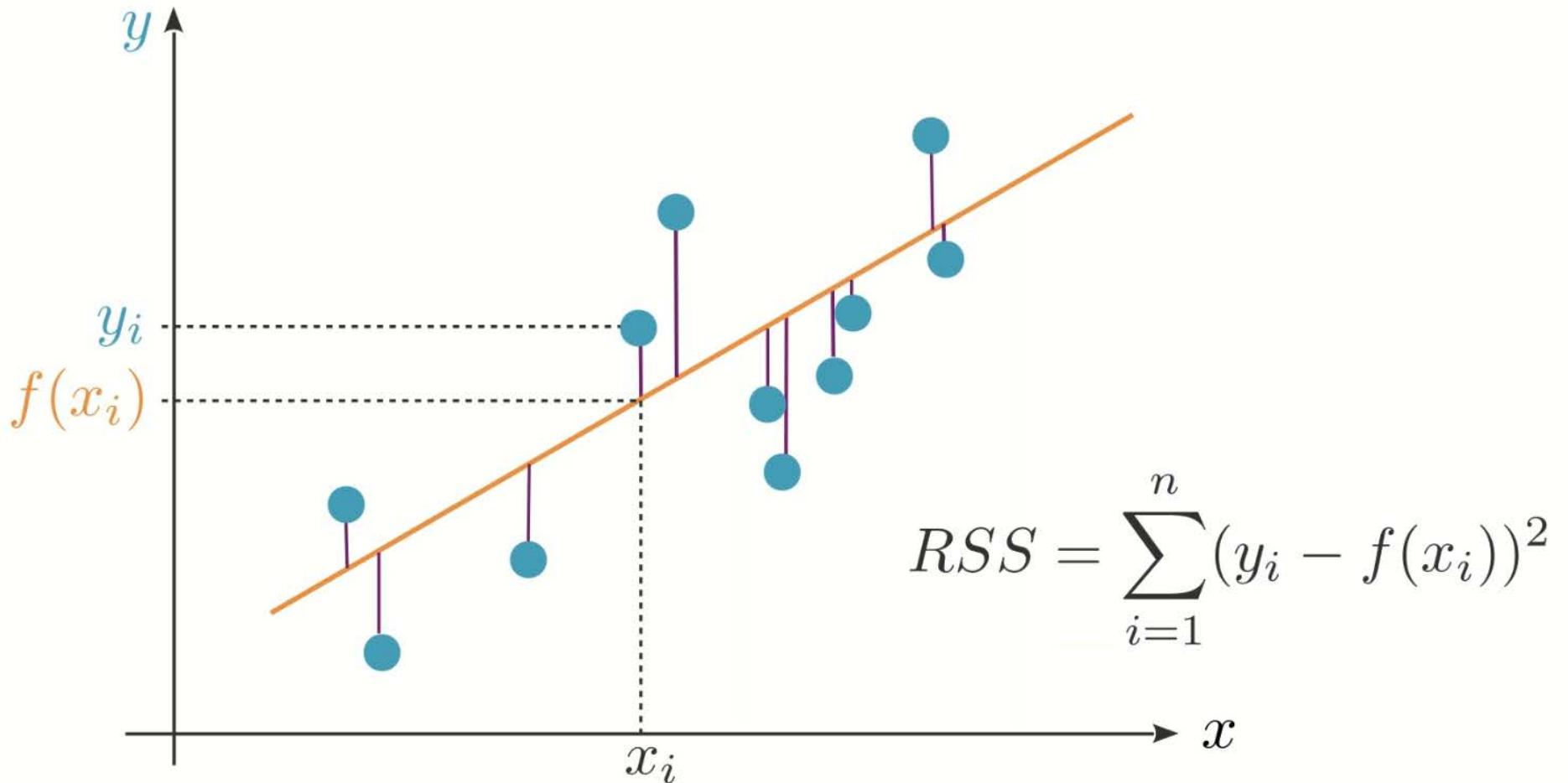
Minimum global

Plateau



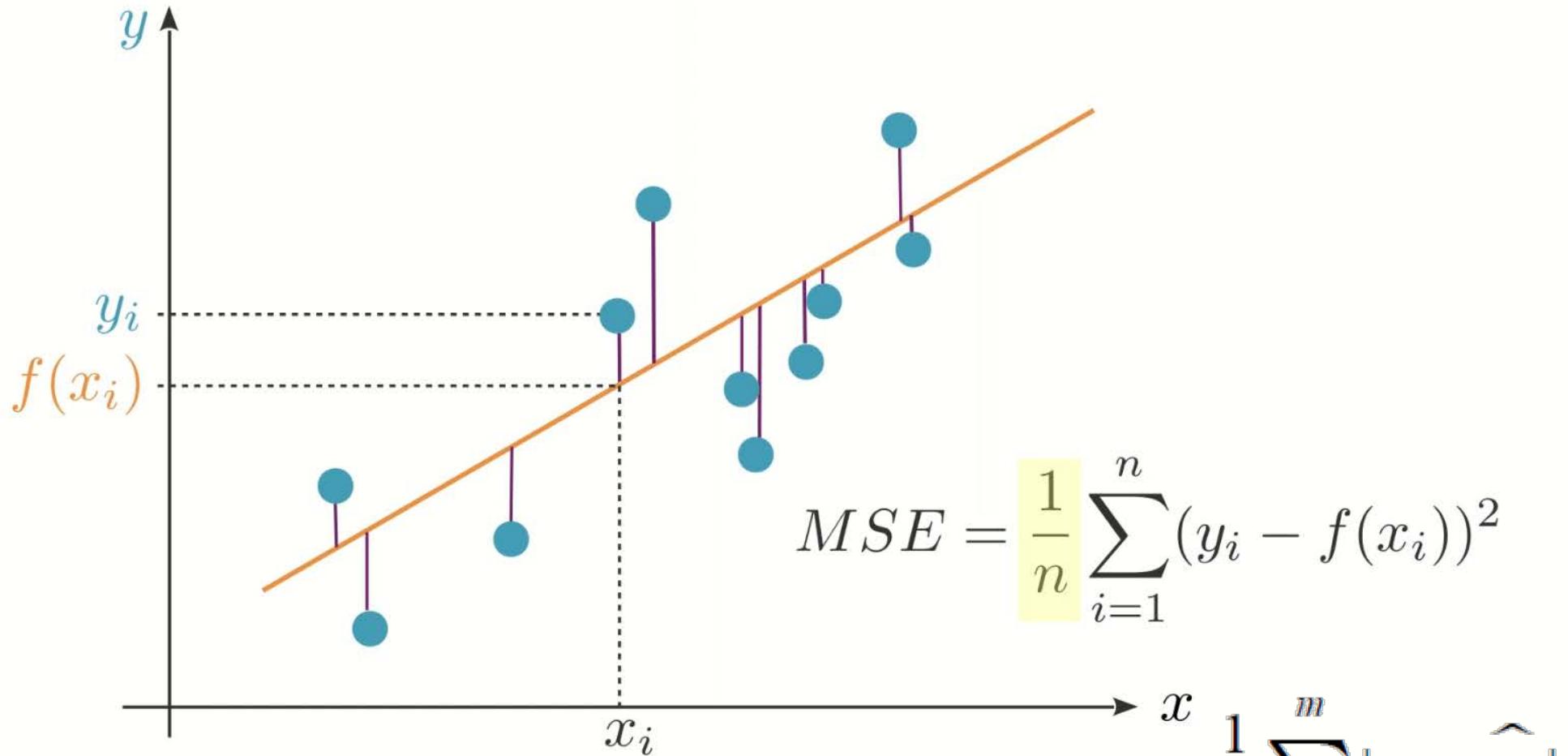
Choix de la métrique de performance

Residual Sum of Squares (RSS)
(somme des carrés des résidus)



Choix de la métrique de performance

Mean squared error (MSE) (erreur quadratique moyenne)

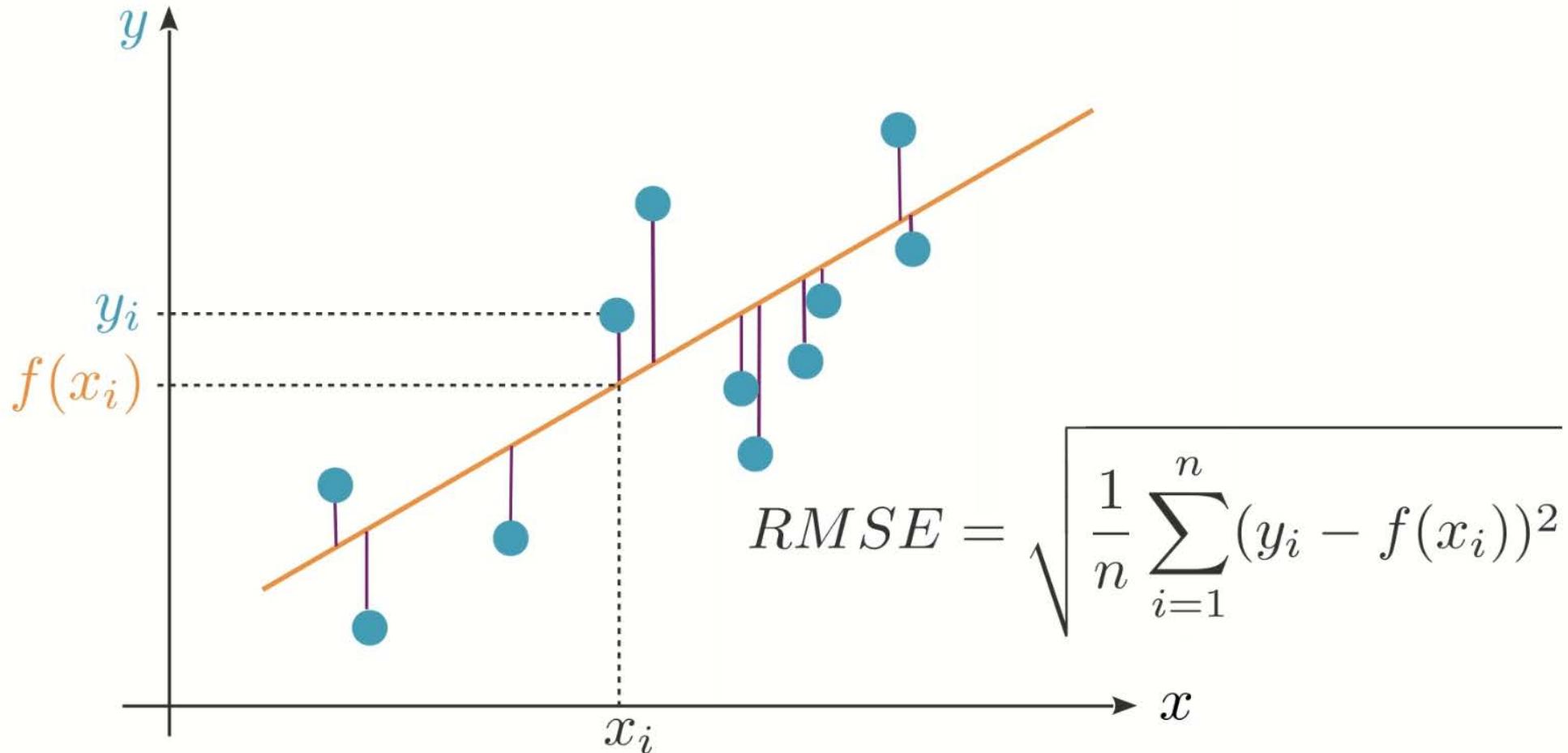


l'erreur moyenne absolue (MAE, Mean Absolute Error)

$$\frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i|$$

Choix de la métrique de performance

Root mean squared error (RMSE)
(racine carrée de la MSE)



Choix de la métrique de performance

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - f(x_i))^2}$$

$$(4 - 104)^2 = 10\,000$$

$$(8\,000 - 8\,100)^2 = 10\,000$$

↓ log

$$(\log(4 + 1) - \log(104 + 1))^2 \approx 1.75$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\log(y_i + 1) - \log(f(x_i) + 1))^2}$$

$$(\log(8\,000 + 1) - \log(8\,100 + 1))^2 \approx 0.000029$$

Choix de la métrique de performance

Relative Squared Error (RSE) *(erreur carrée relative)*

$$RSE = \frac{\sum_{i=1}^n (y_i - f(x_i))^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

Choix de la métrique de performance

Coefficient de détermination

$$R^2 = 1 - RSE$$
$$= (\text{corrélation de Pearson})^2$$

$$RSE = \frac{\sum_{i=1}^n (y_i - f(x_i))^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

ρ = corrélation de Pearson

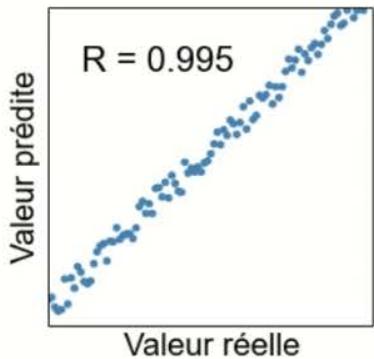
$$R = \frac{\sum_{i=1}^n (y_i - \bar{y})(f(x_i) - \bar{f(x)})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2} \sqrt{\sum_{i=1}^n (f(x_i) - \bar{f(x)})^2}}$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

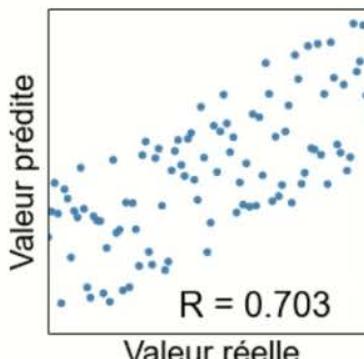
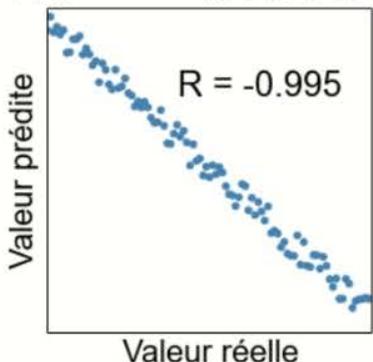
Choix de la métrique de performance

$\rho = \text{corrélation de Pearson}$

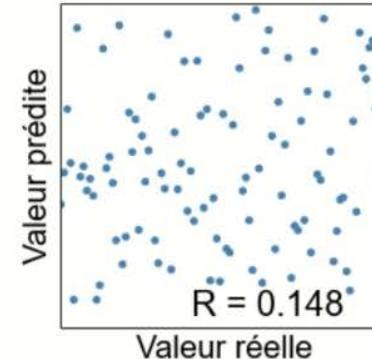
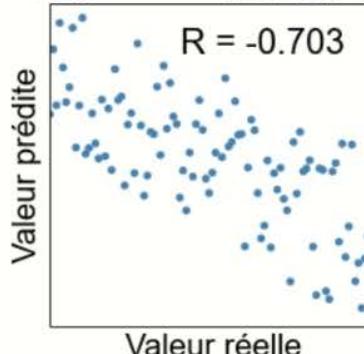
$$R = \frac{\sum_{i=1}^n (y_i - \bar{y})(f(x_i) - \overline{f(x)})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2} \sqrt{\sum_{i=1}^n (f(x_i) - \overline{f(x)})^2}}$$



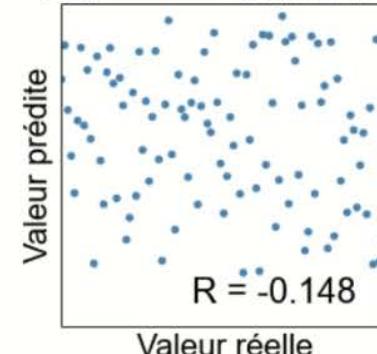
$$R^2 = 0.990$$



$$R^2 = 0.494$$



$$R^2 = 0.022$$

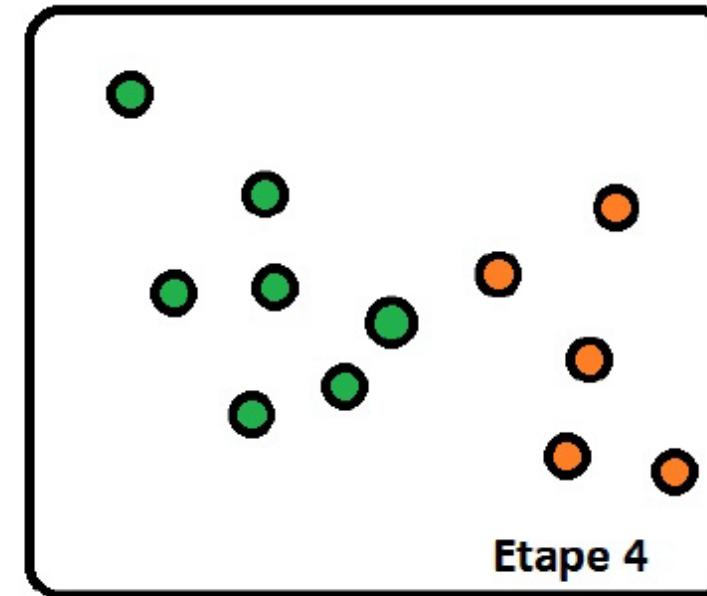
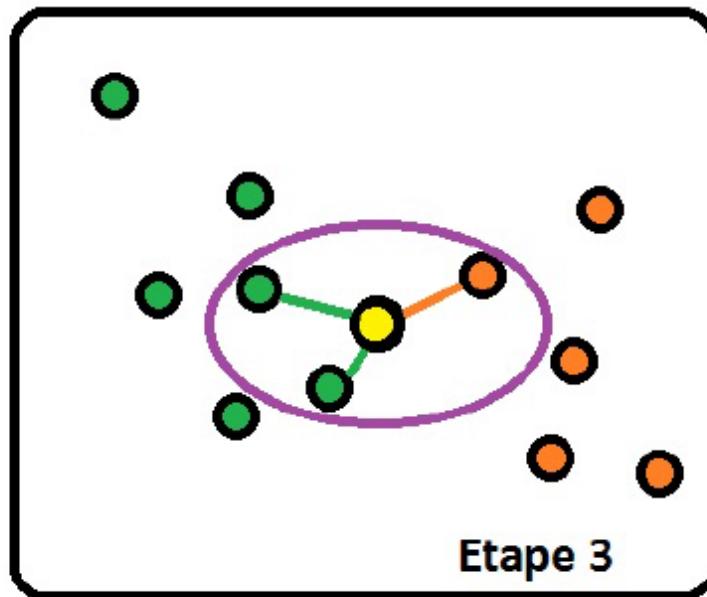
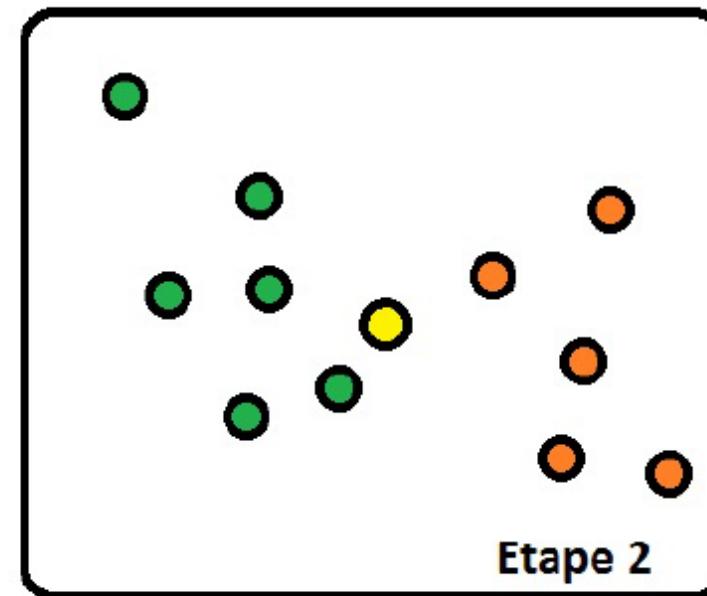
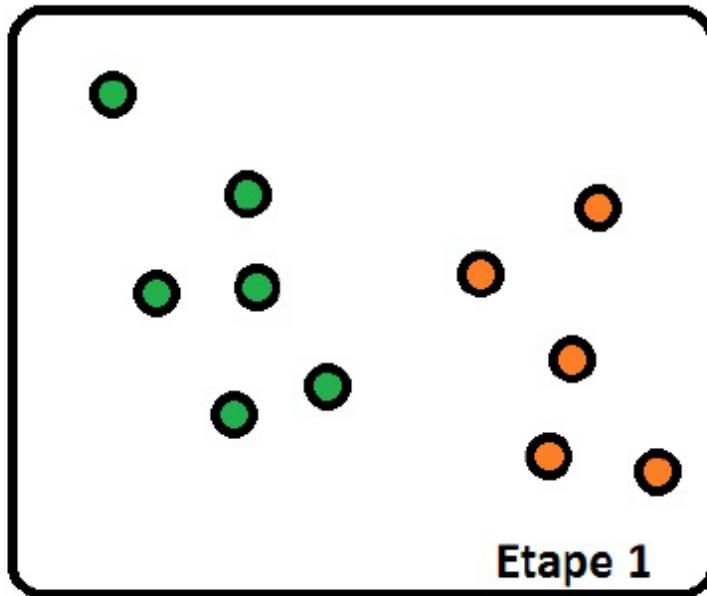


Module 12 : Algorithmes de la classification

Algorithme	Mode d'apprentissage	Type de problème à traiter
Régression linéaire univariée	Supervisé	Régression
Régression linéaire multivariée	Supervisé	Régression
Régression polynomiale	Supervisé	Régression
Régression régularisée	Supervisé	Régression
Naive Bayes	Supervisé	Classification
Régression logistique	Supervisé	Classification
Clustering hiérarchique	Non supervisé	-
Clustering non hiérarchique	Non supervisé	-
Arbres de décision	Supervisé	Régression ou classification
Random forest	Supervisé	Régression ou classification
Gradient boosting	Supervisé	Régression ou classification
Support Vector Machine	Supervisé	Régression ou classification
Réseau de neurones	Supervisé	Régression ou classification

K Nearest Neighbors

KNN avec k=3



Naïve Bayes

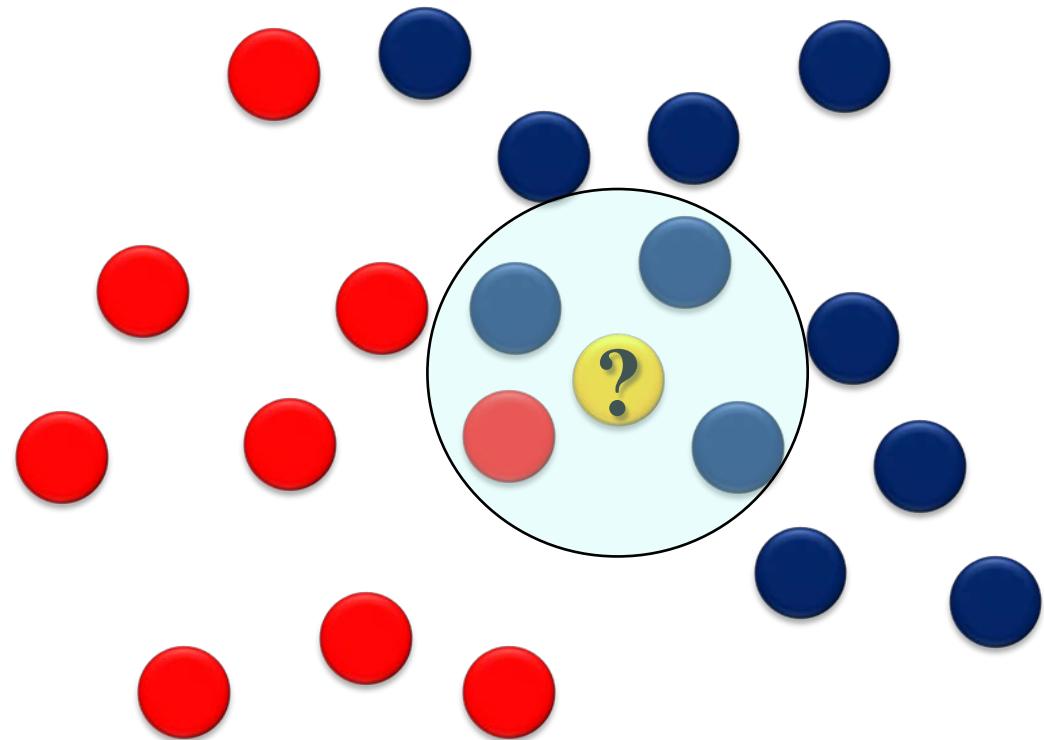
Le modèle Naïve Bayes

$$P(\text{rouge}) = 9 / 20$$

$$P(\text{bleu}) = 11 / 20$$

$$P'(\text{?} | \text{rouge}) = 1 / 9$$

$$P'(\text{?} | \text{bleu}) = 3 / 11$$



$$\begin{aligned} P''(\text{? est rouge}) &= P(\text{rouge}) * P'(\text{?} | \text{rouge}) \\ &= 9 / 20 * 1 / 9 = 1 / 20 = 0.05 \end{aligned}$$

$$\begin{aligned} P''(\text{? est bleu}) &= P(\text{bleu}) * P'(\text{?} | \text{bleu}) \\ &= 11 / 20 * 3 / 11 = 33 / 220 = 0.15 \end{aligned}$$

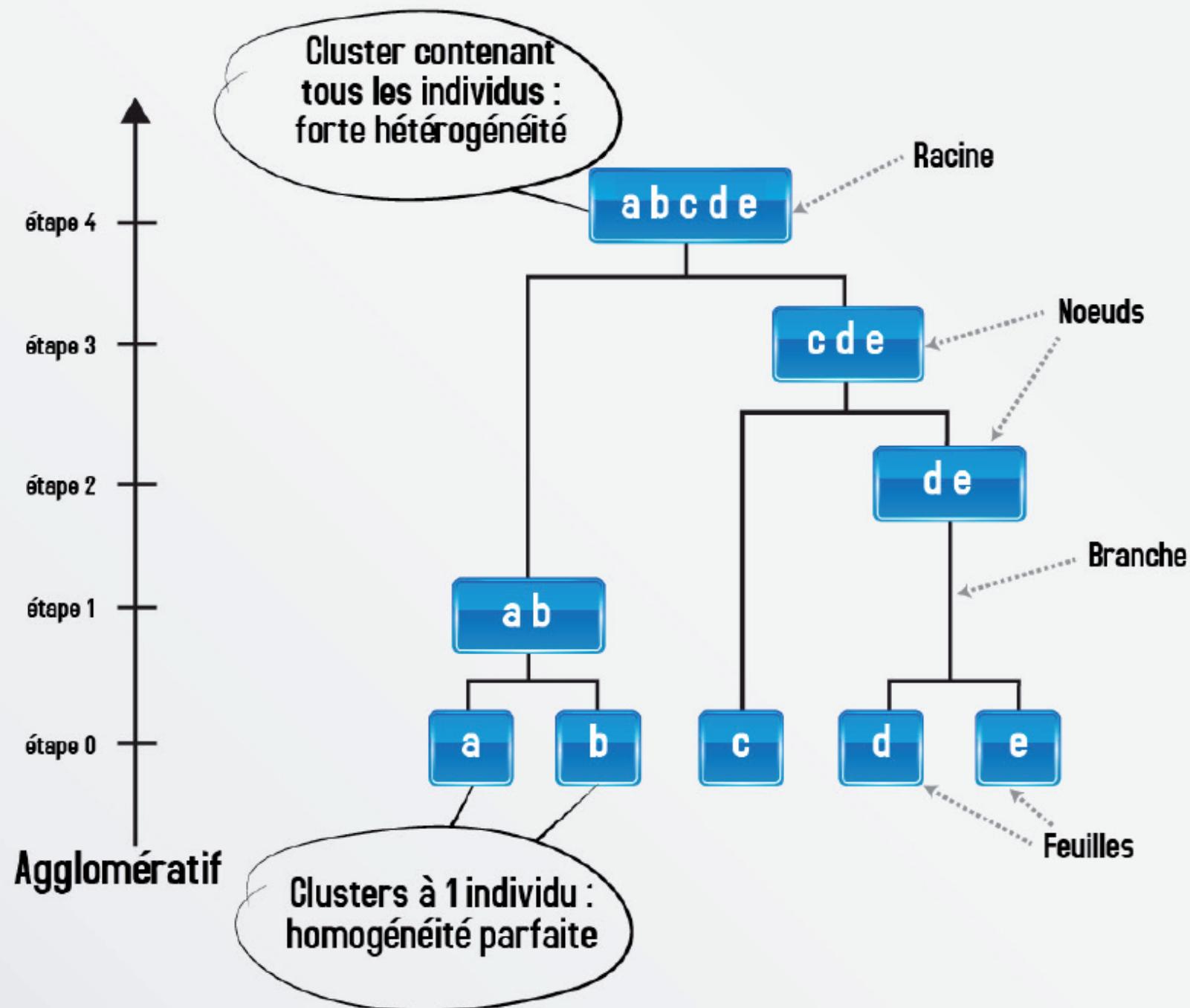
Le Naïve Bayes est l'un des classifieurs les plus simples.

Il repose sur une hypothèse forte : l'indépendance des variables.

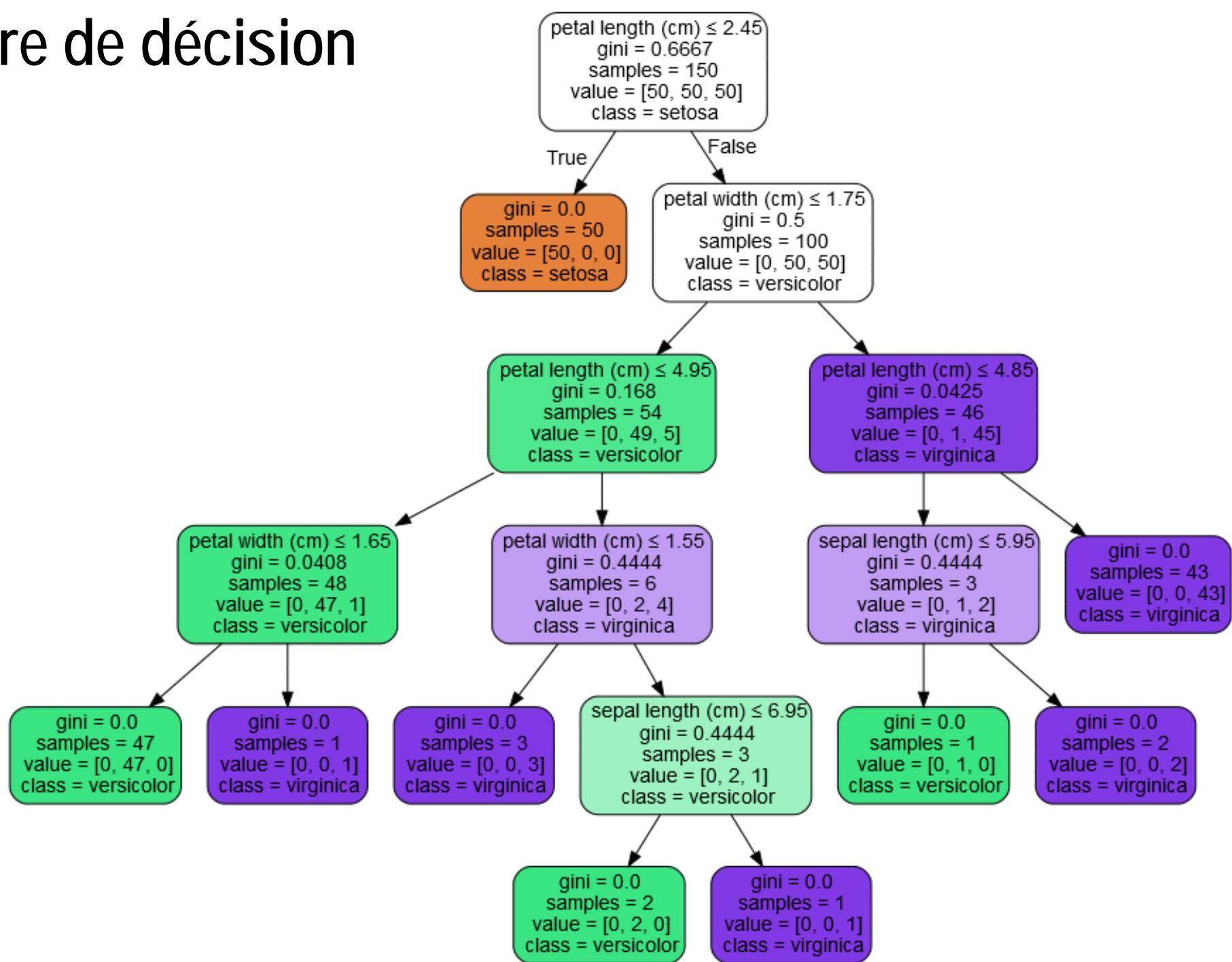
https://www.wikiwand.com/en/Naive_Bayes_classifier

Les arbres de décision

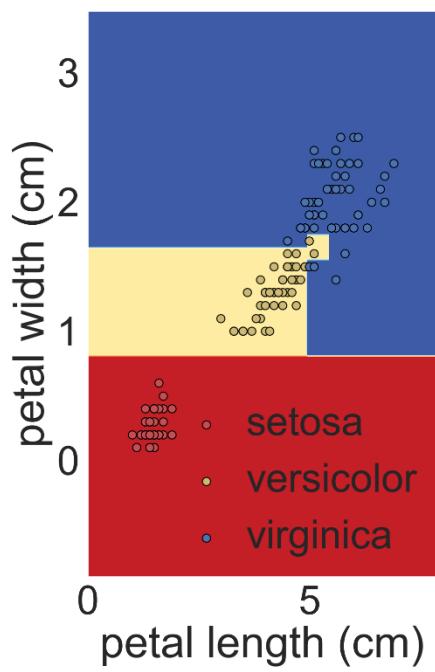
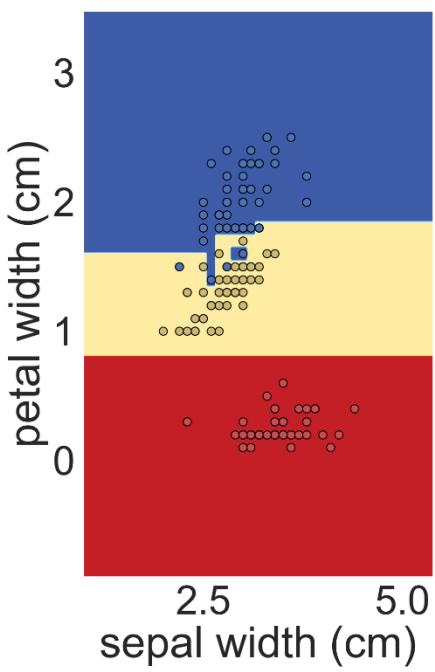
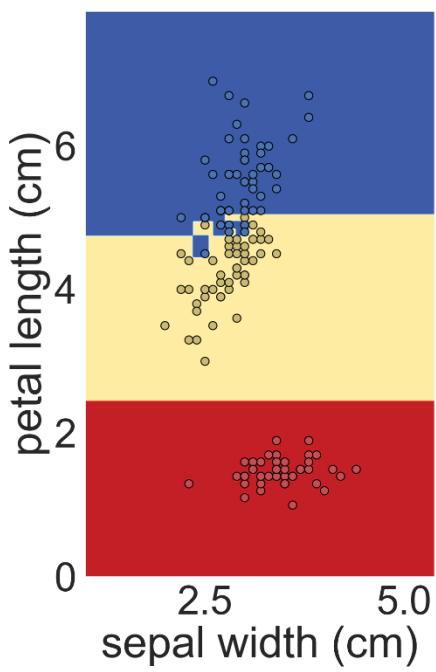
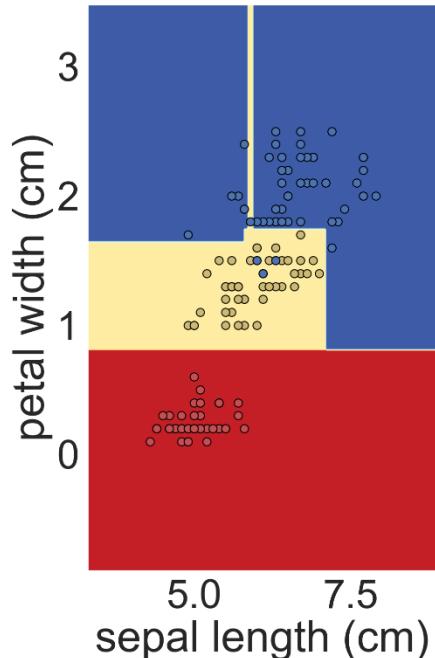
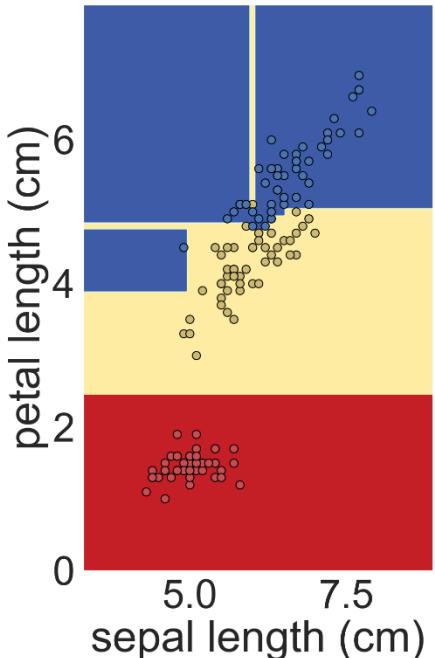
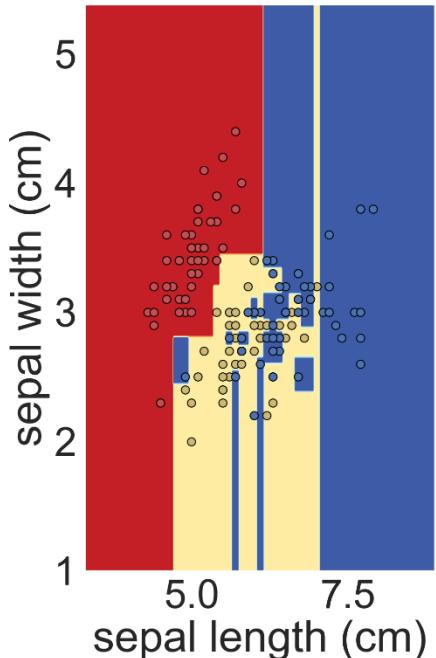
Divisif



Arbre de décision

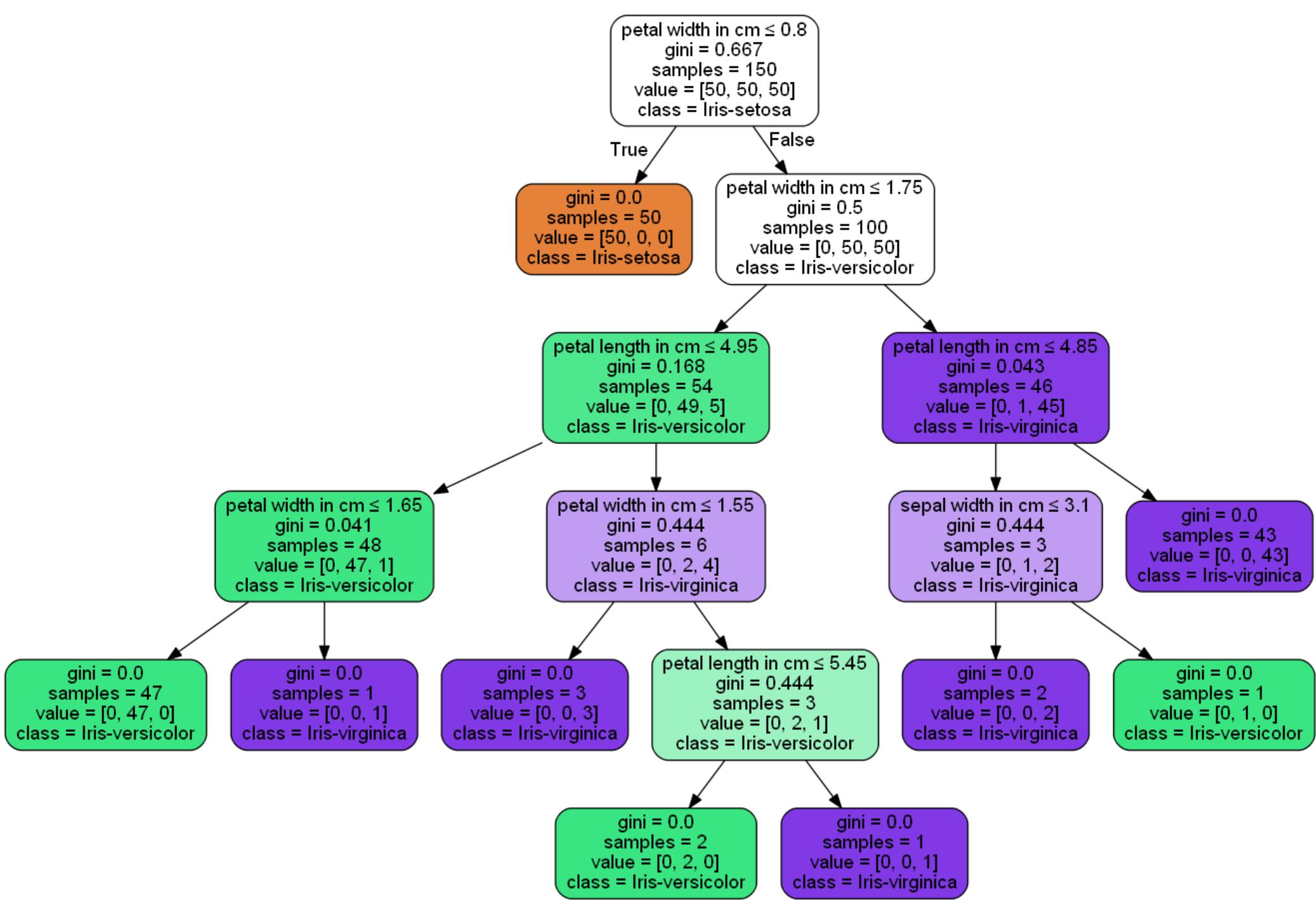


Surface de décision d'un arbre de décision



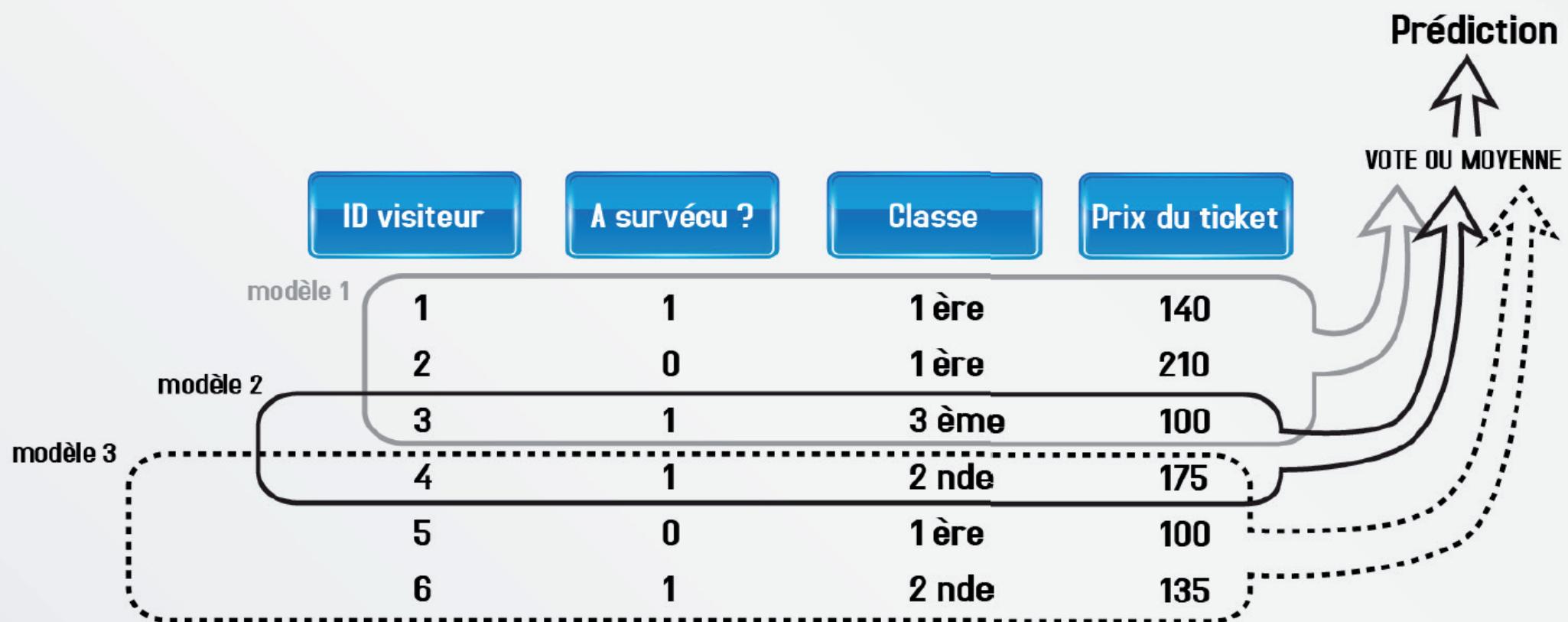
Legend:

- setosa (red circle)
- versicolor (yellow circle)
- virginica (blue circle)

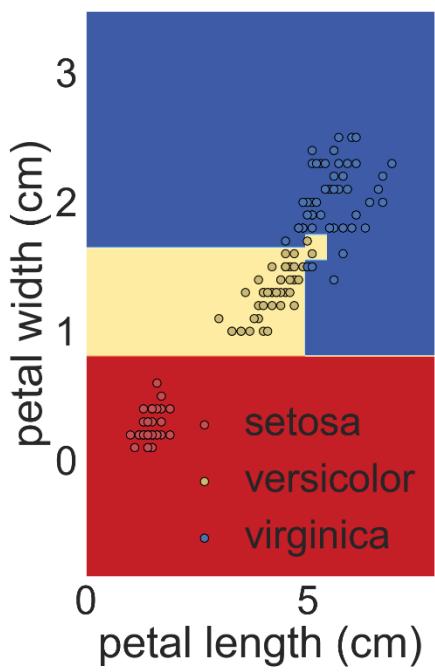
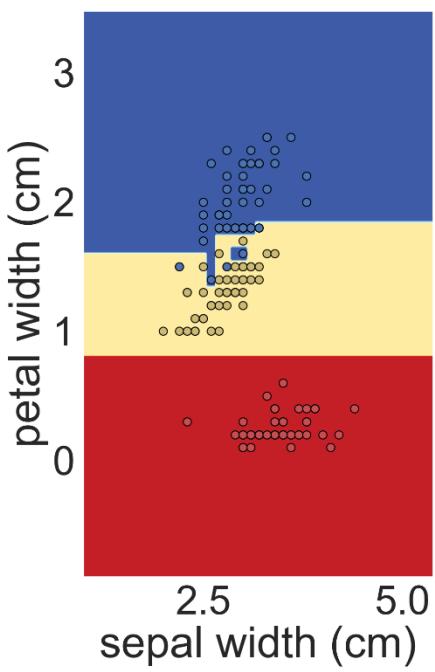
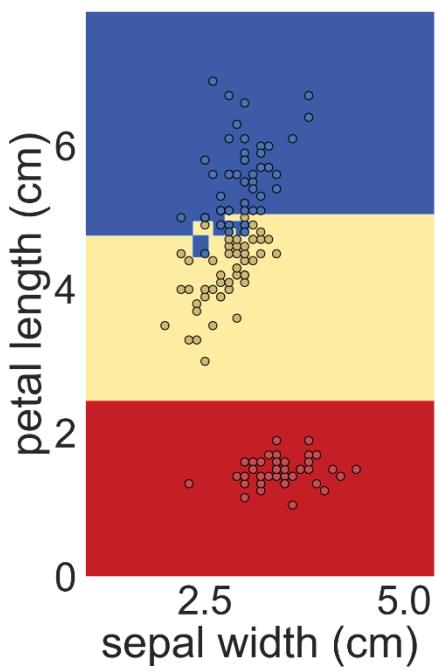
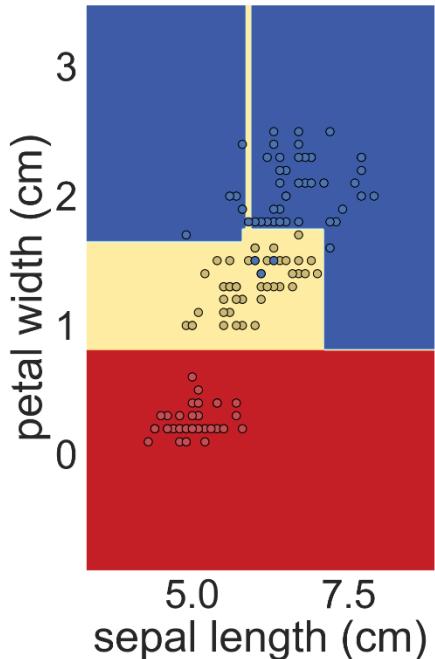
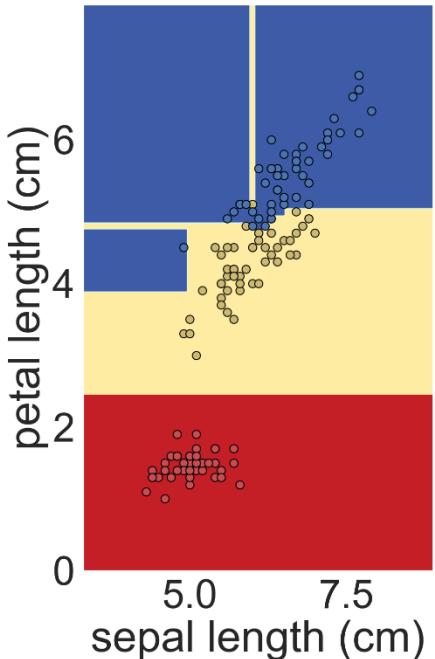
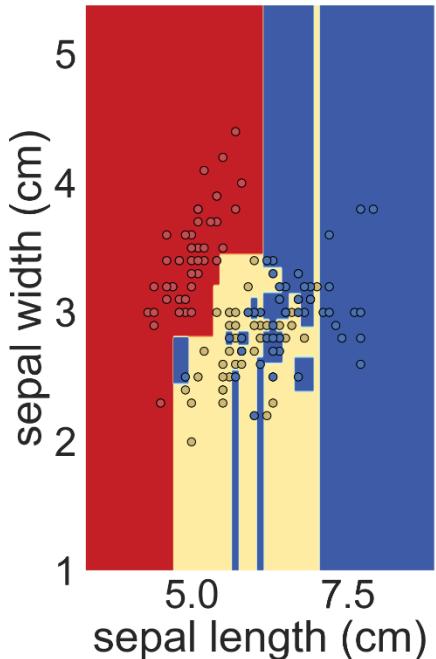


Arbres de décision

Tree Bagging



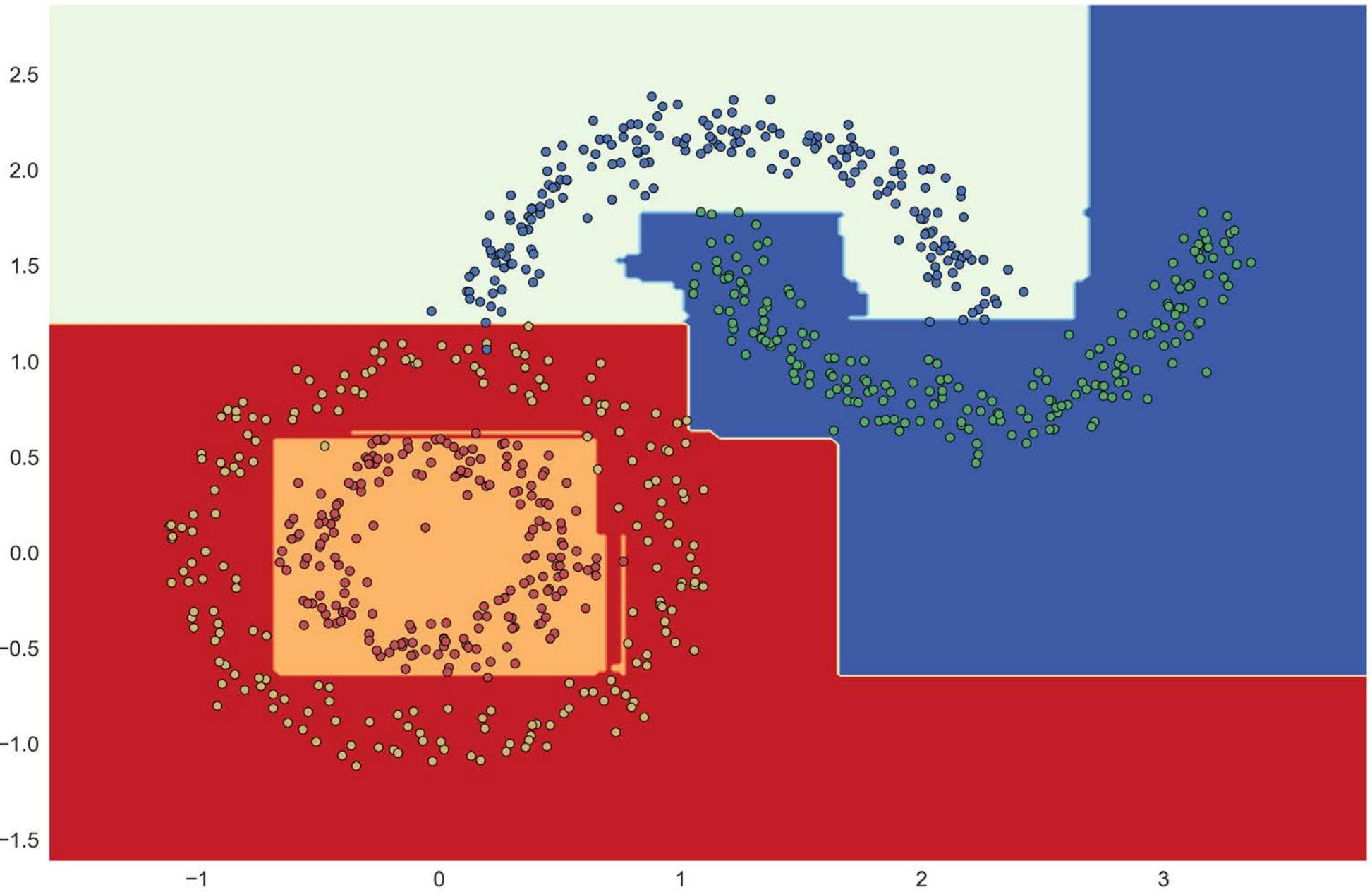
Surface de décision d'un arbre de décision



Legend:

- setosa (red circle)
- versicolor (yellow circle)
- virginica (blue circle)

Tree Bagging



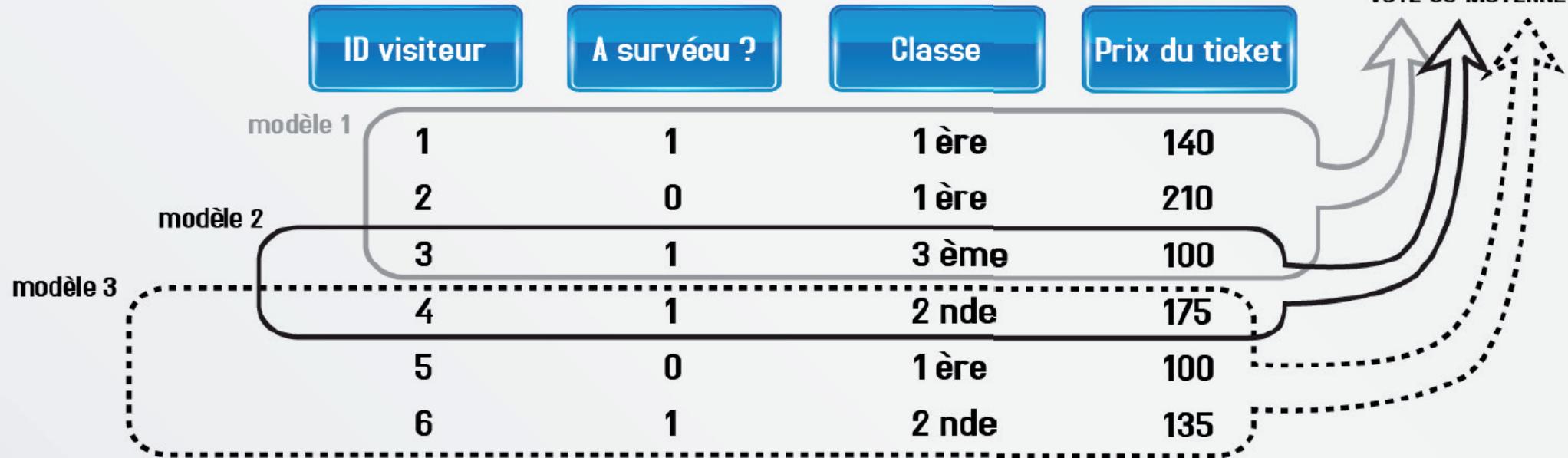
Random forest

Random forest = tree bagging + feature sampling

Prédiction



VOTE OU MOYENNE

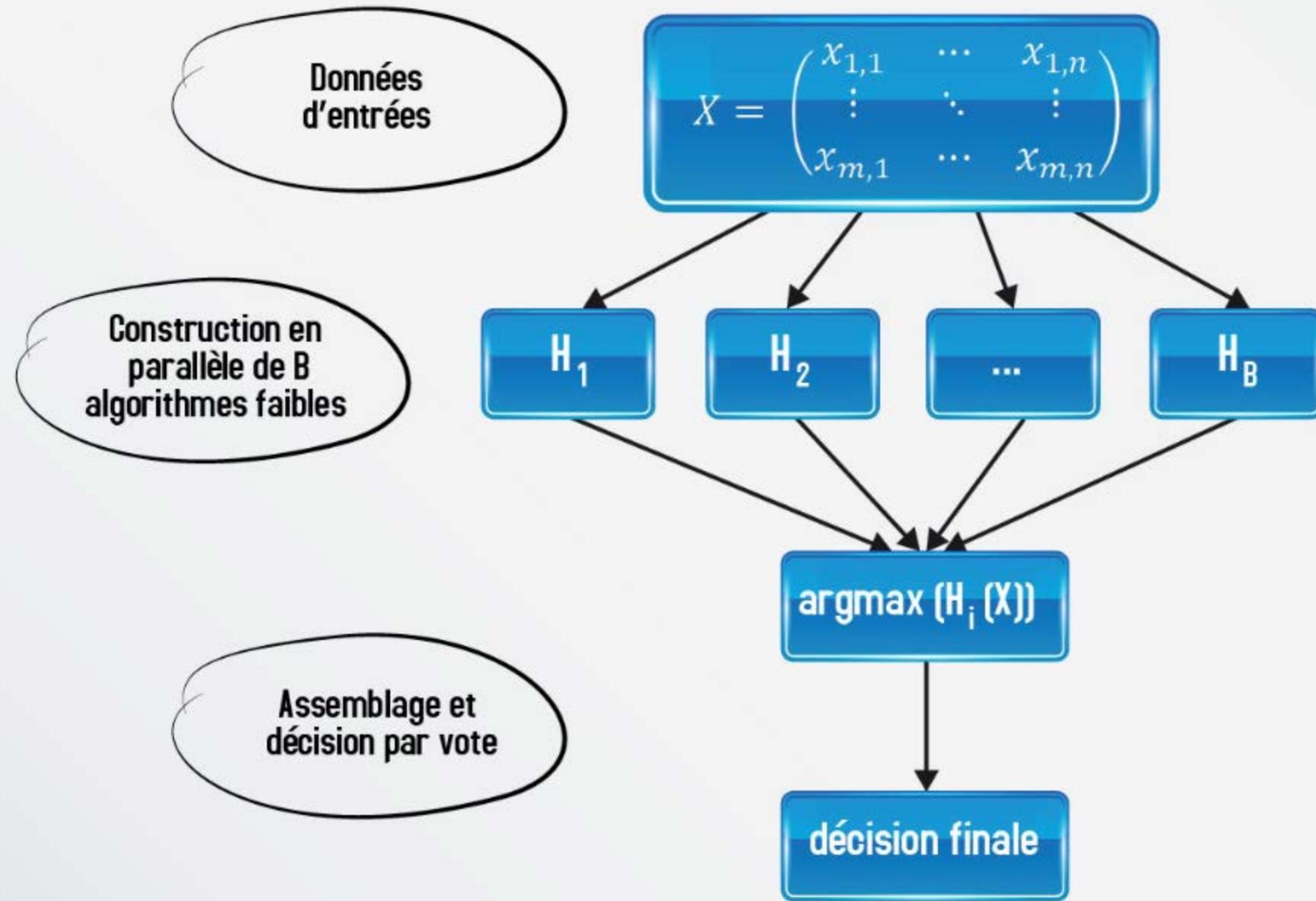


$$V_{forest} = \rho\sigma^2 + \frac{1 - \rho}{B}\sigma^2$$

Feature sampling

Bagging

Random Forest

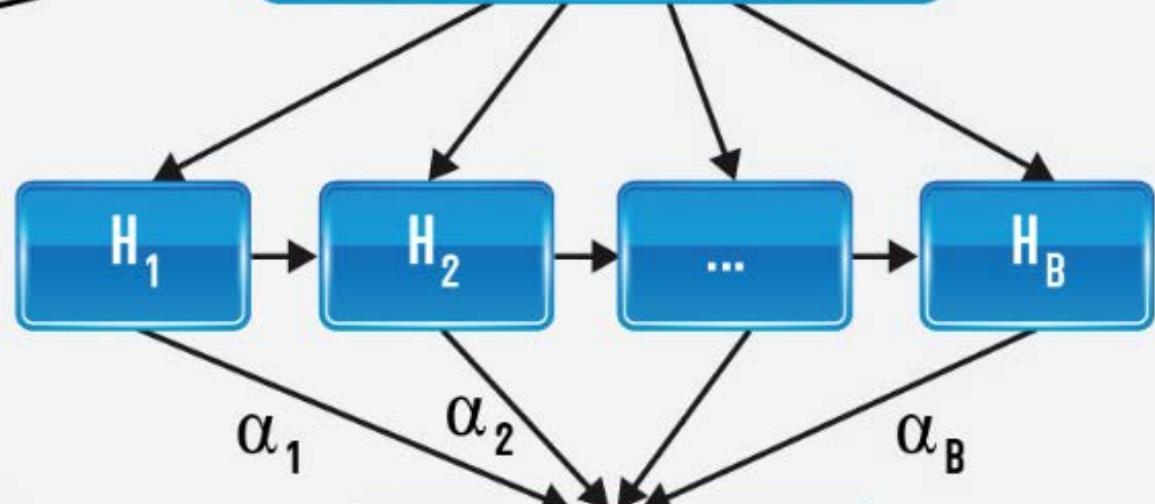


Boosting

Données d'entrées

$$X = \begin{pmatrix} x_{1,1} & \cdots & x_{1,n} \\ \vdots & \ddots & \vdots \\ x_{m,1} & \cdots & x_{m,n} \end{pmatrix}$$

Construction en série de B algorithmes faibles



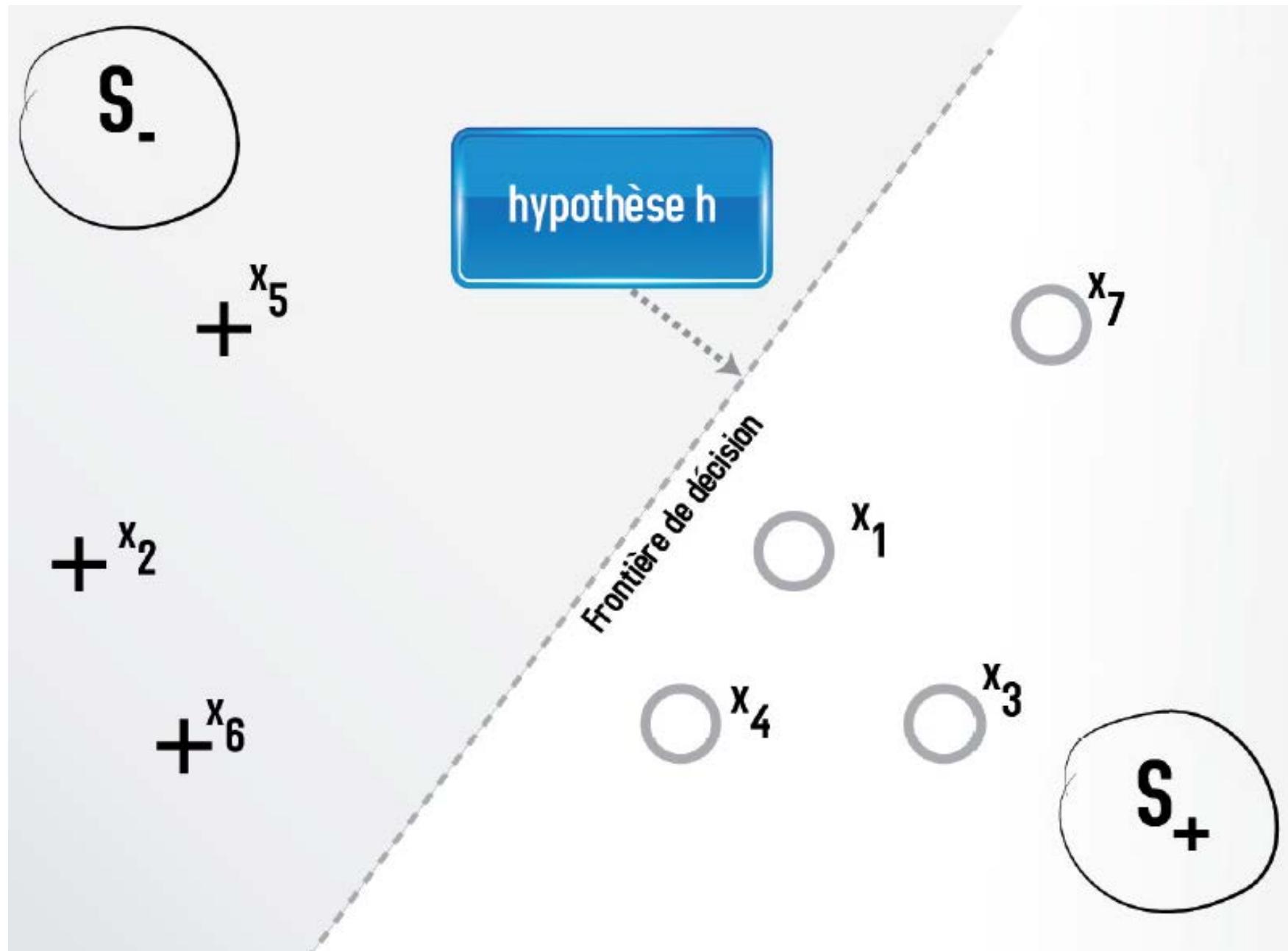
Assemblage et décision par somme pondérée

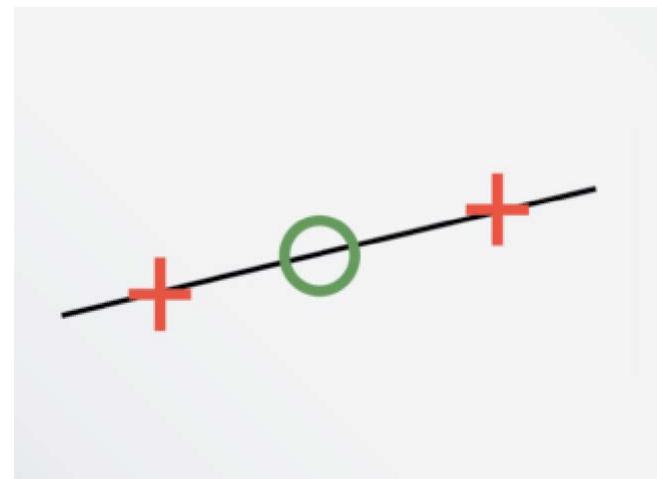
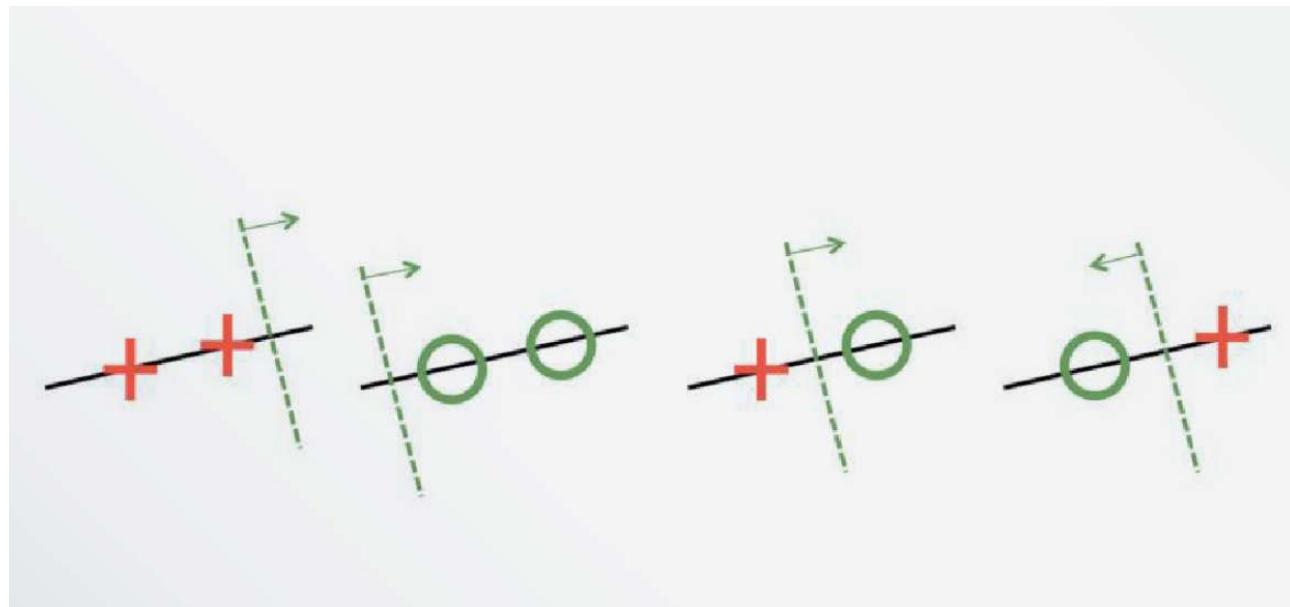
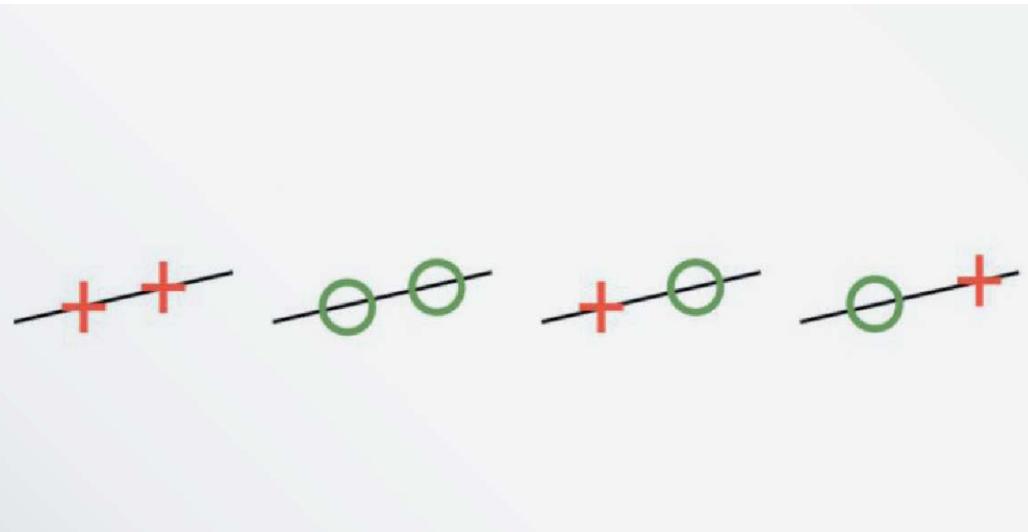
$$F(x) = \sum \alpha_i \cdot h_i(x)$$

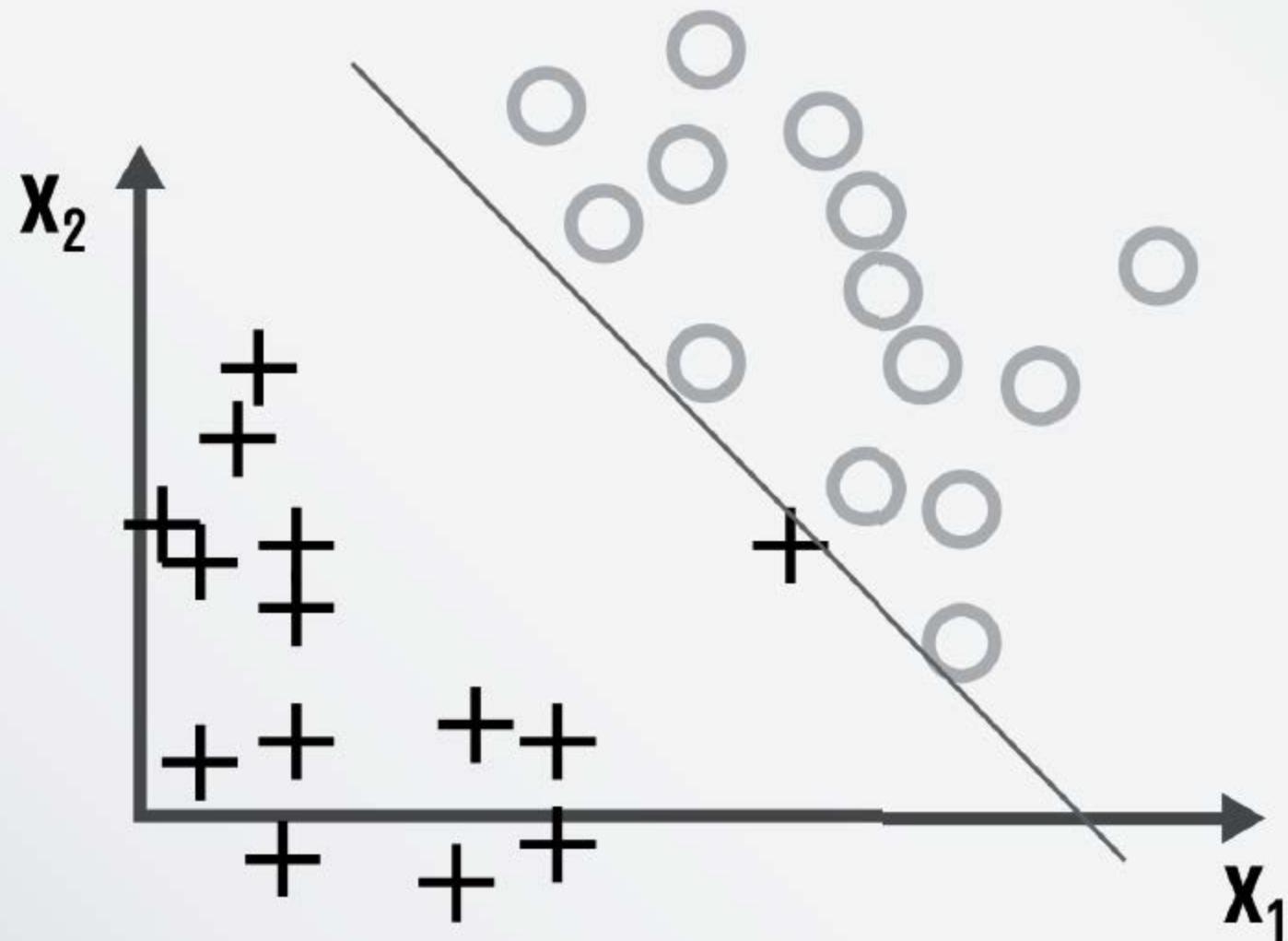
$$H(x) = \text{sign}(F(x))$$

Support Vector Machine

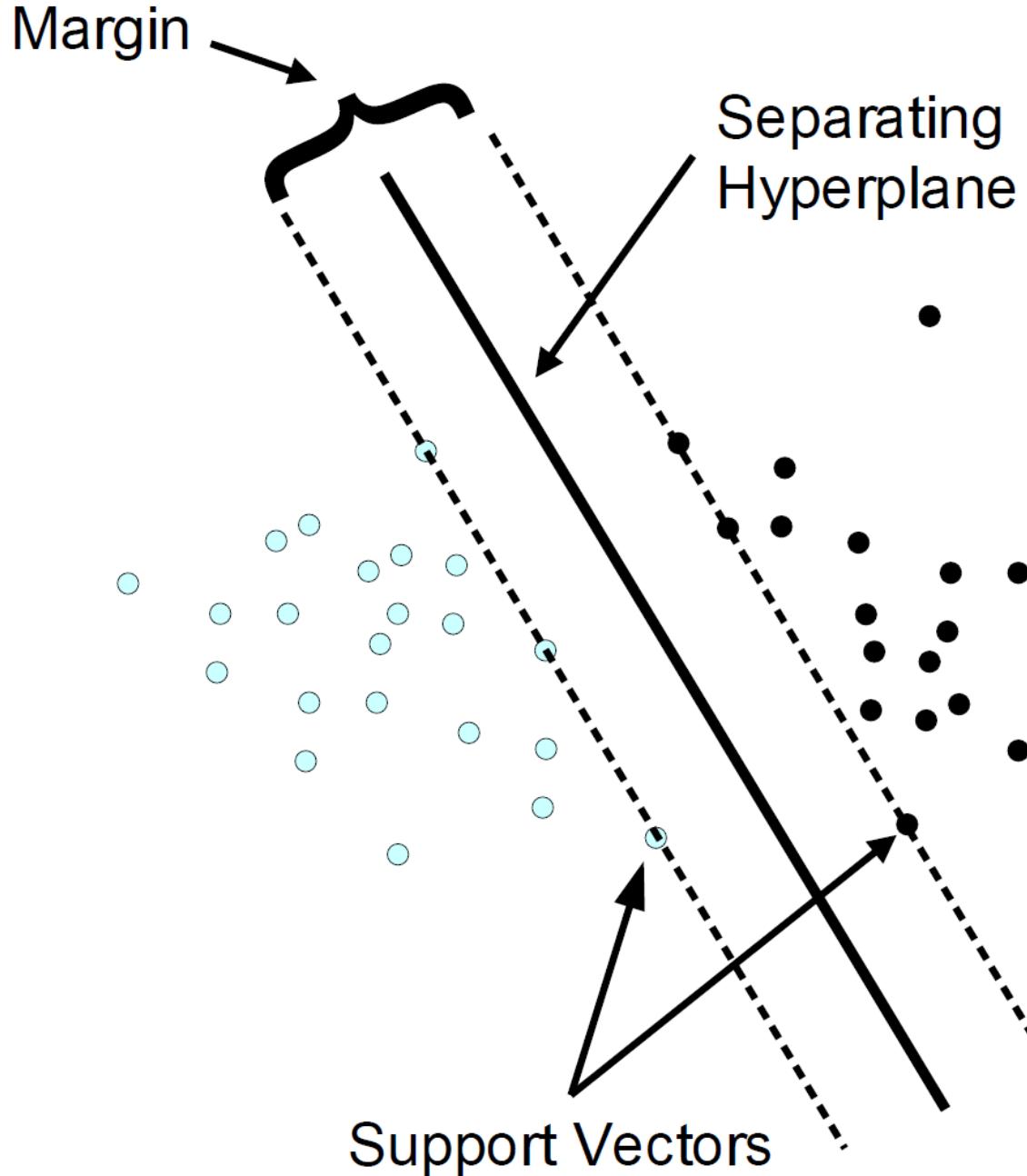
Support Vector Machine



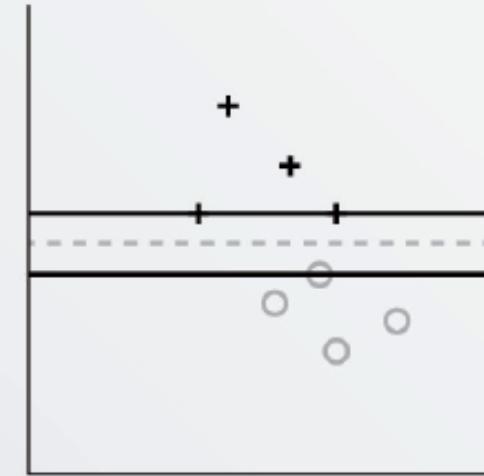




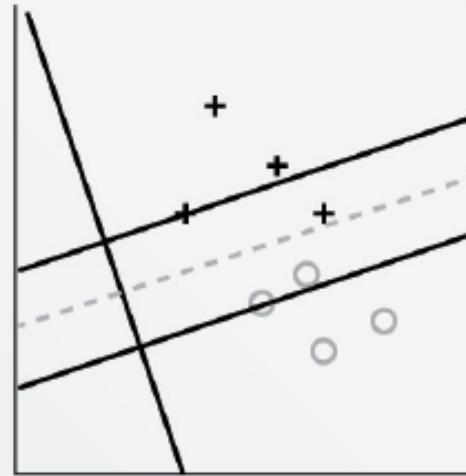
Marge d'un hyperplan séparateur



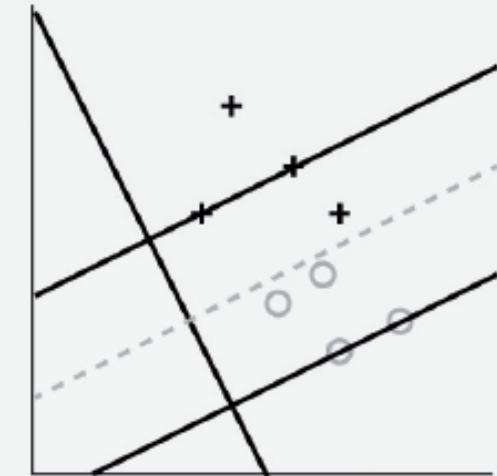
$C=100$



$C=3$

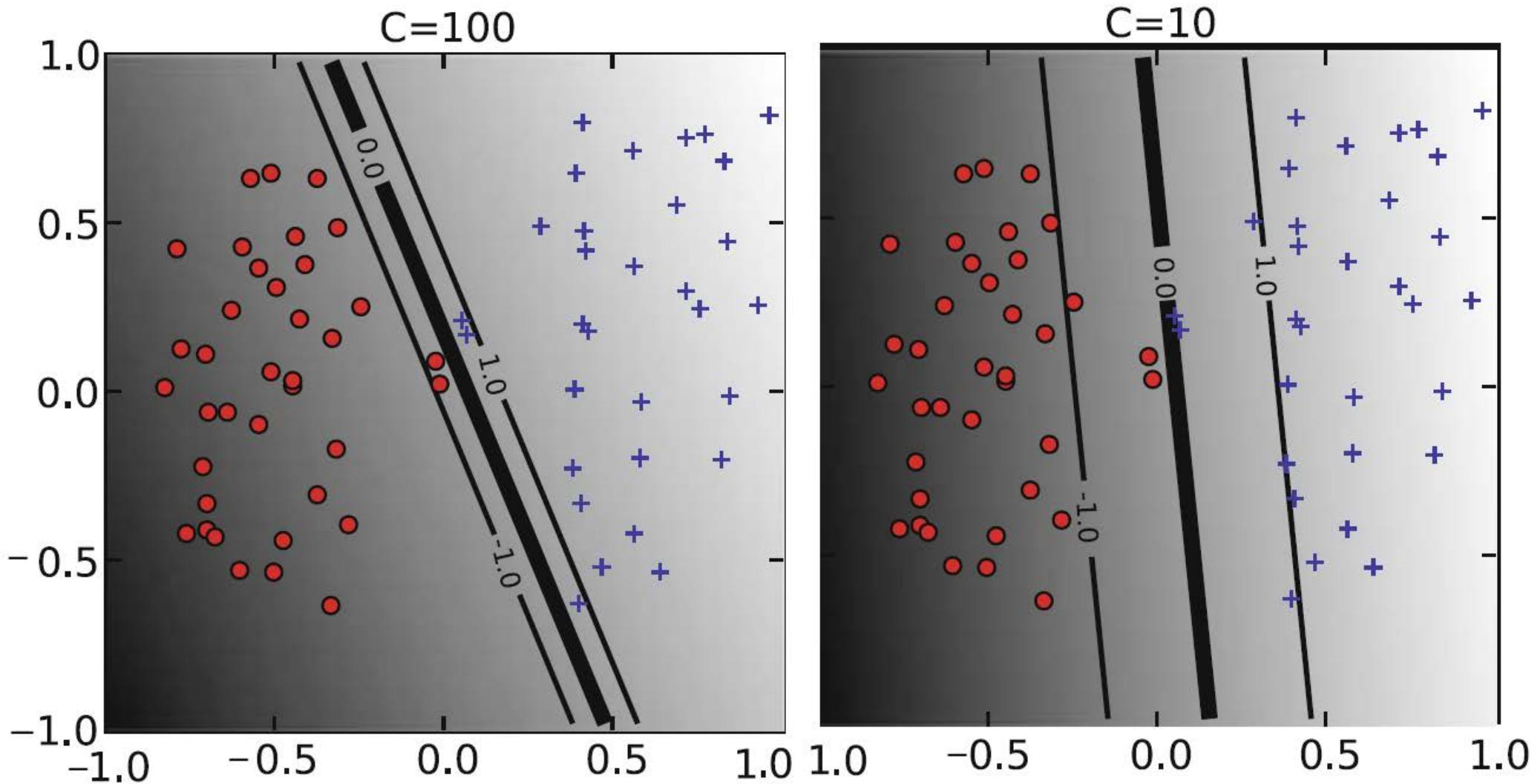


$C=1$



C faible : surface de décision plus régulière, autorisant quelques erreurs de classification, mais une meilleure généralisation

C élevé : surface de décision plus complexe, donc attention à la généralisation



x_2

$$\phi \left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right) = \begin{pmatrix} x_1^2 \\ x_2^2 \end{pmatrix}$$

 x_1 R

$$x_1^2 + x_2^2 - R^2 = 0$$

 $\phi(x)_2$ $\phi(x)_1$ R^2

$$\phi_1 + \phi_2 - R^2 = 0$$

$$K(\mathbf{x}, \mathbf{x}') = \langle \phi(\mathbf{x}), \phi(\mathbf{x}') \rangle_{\mathcal{H}}$$

définie positive

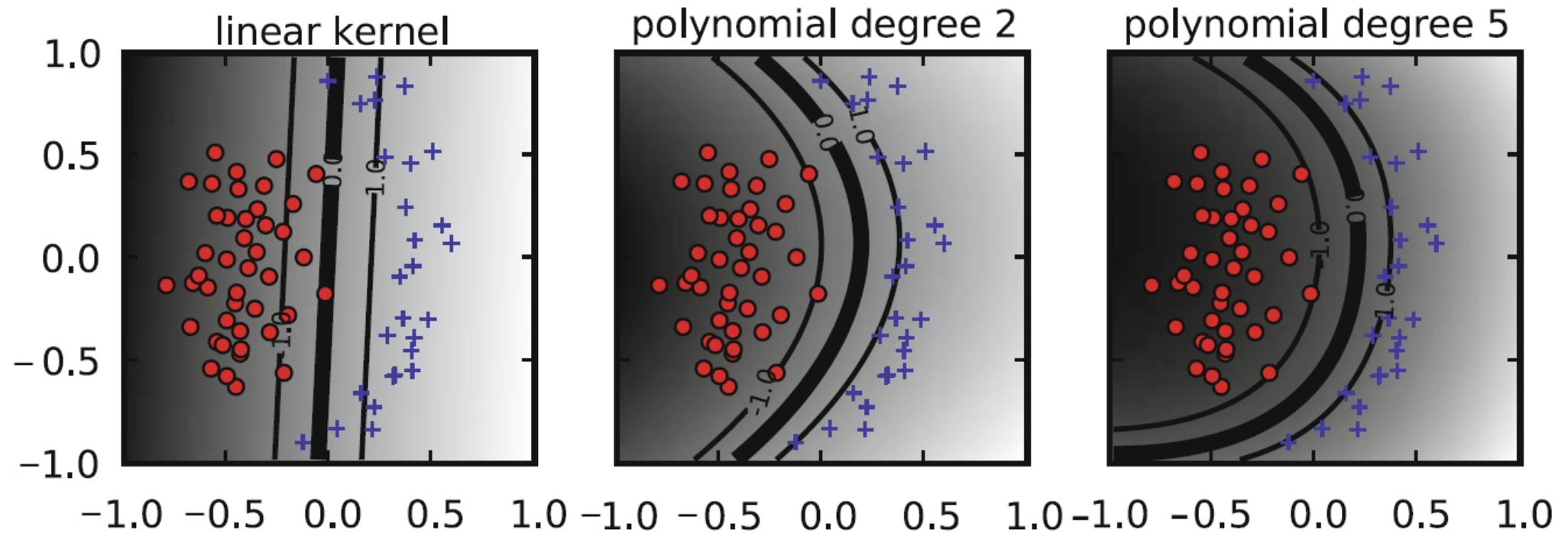
noyau linéaire : $K(\mathbf{x}, \mathbf{x}') = \langle \mathbf{x}, \mathbf{x}' \rangle$

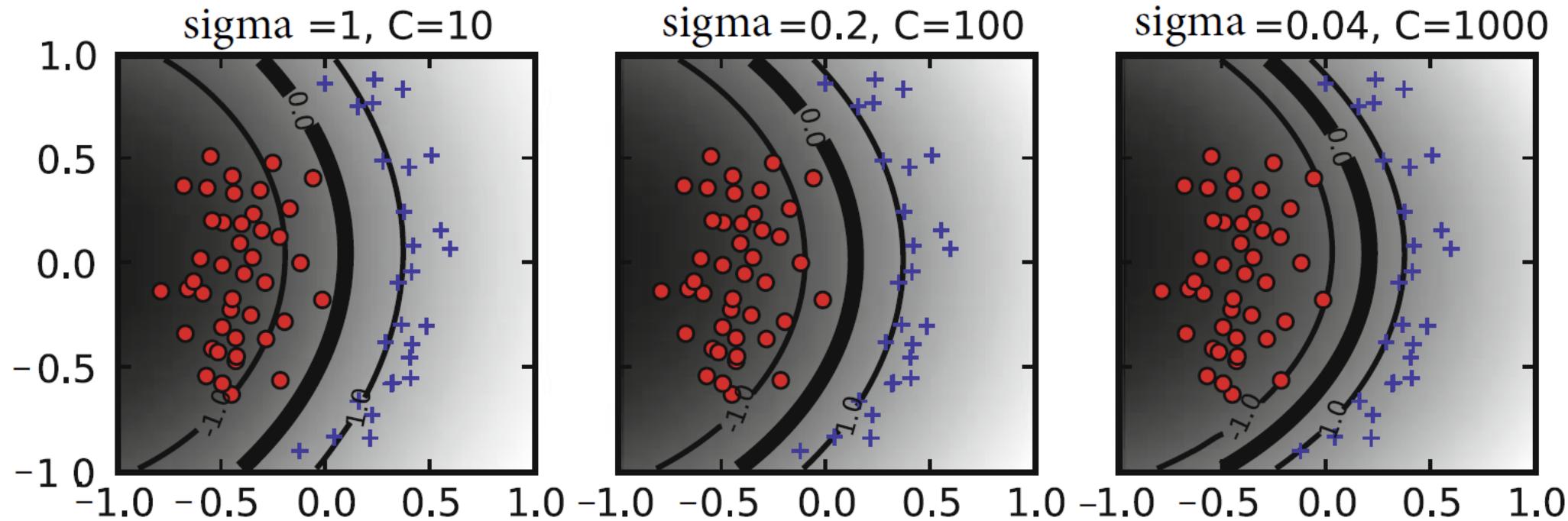
noyau quadratique homogène : $K(\mathbf{x}, \mathbf{x}') = \langle \mathbf{x}, \mathbf{x}' \rangle^2$

noyau polynomial : $K(\mathbf{x}, \mathbf{x}') = (\langle \mathbf{x}, \mathbf{x}' \rangle + c)^d$

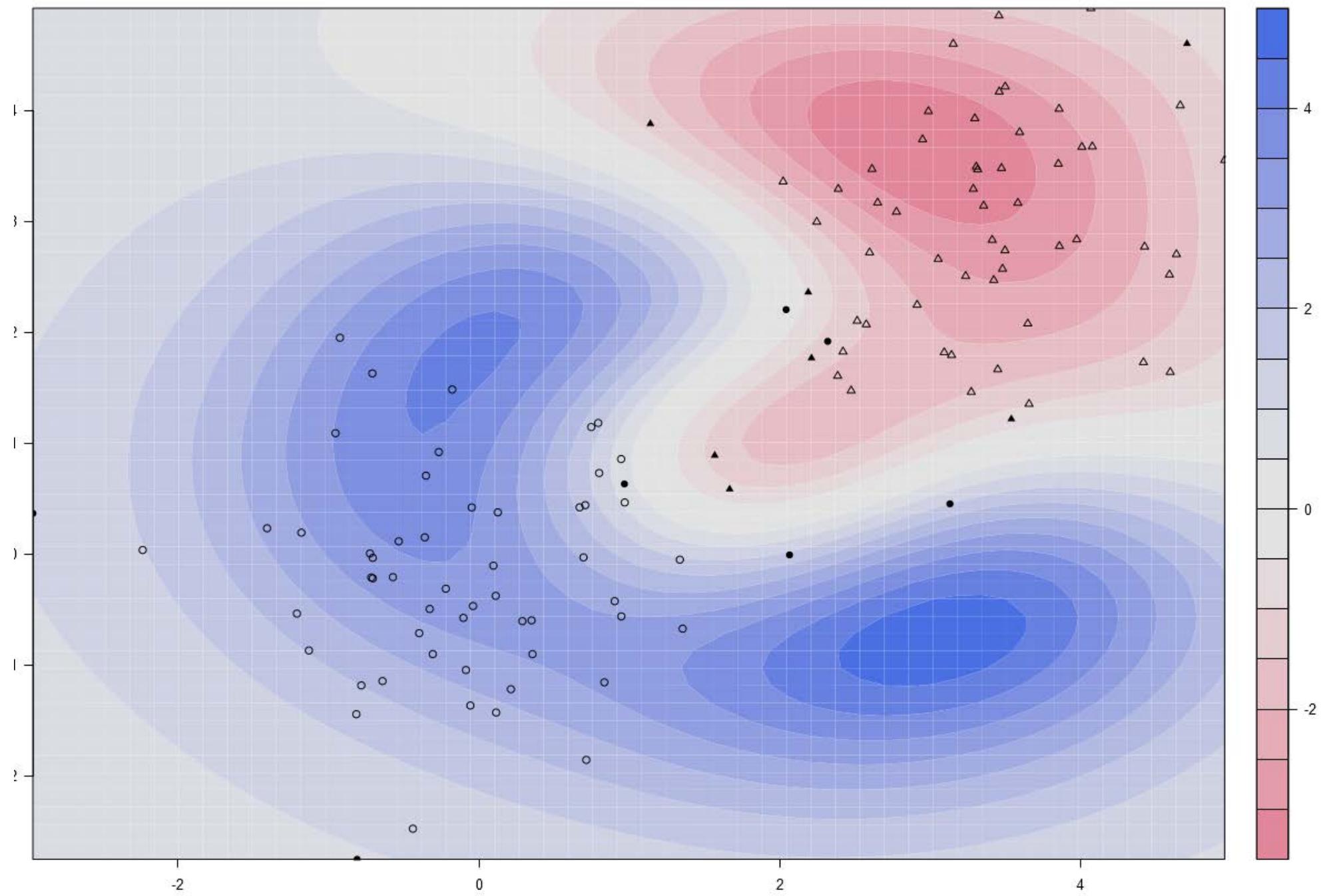
noyau gaussien : $K(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\langle \mathbf{x} - \mathbf{x}', \mathbf{x} - \mathbf{x}' \rangle^2}{2\sigma^2}\right)$

noyau RBF (*Radial Basis Function*) $\mathcal{H} = \mathbb{R}^\infty$

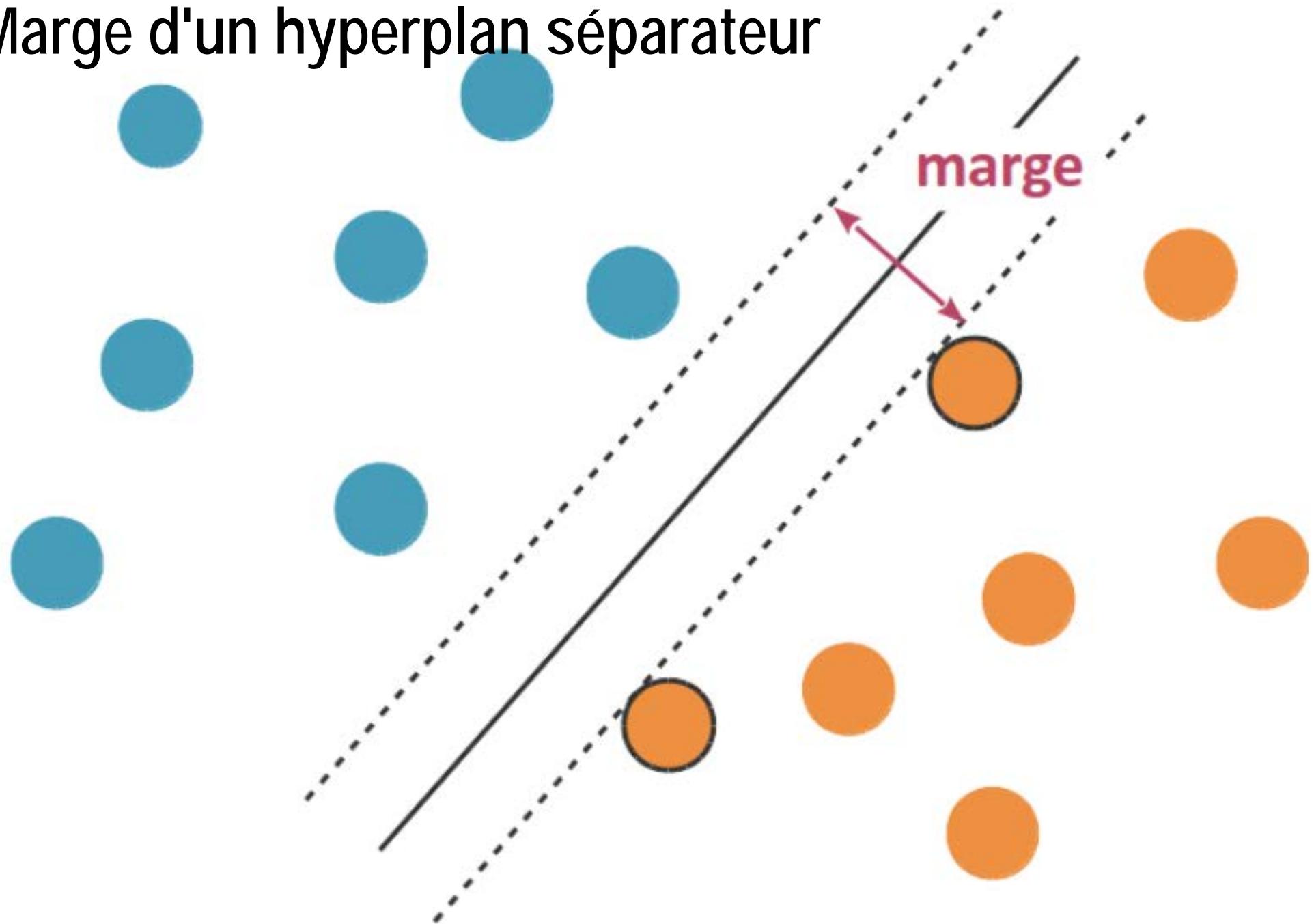




SVM classification plot

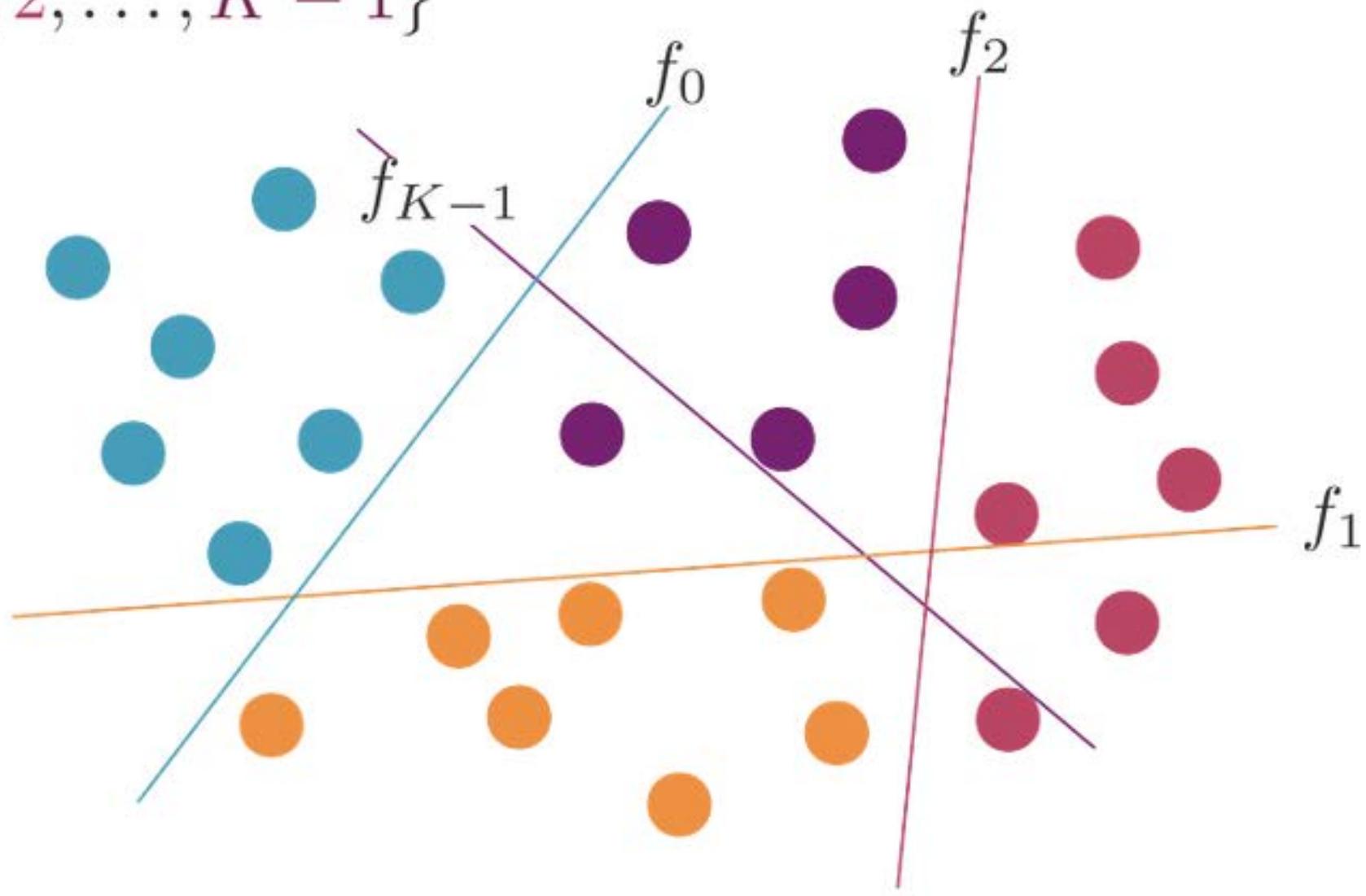


Marge d'un hyperplan séparateur



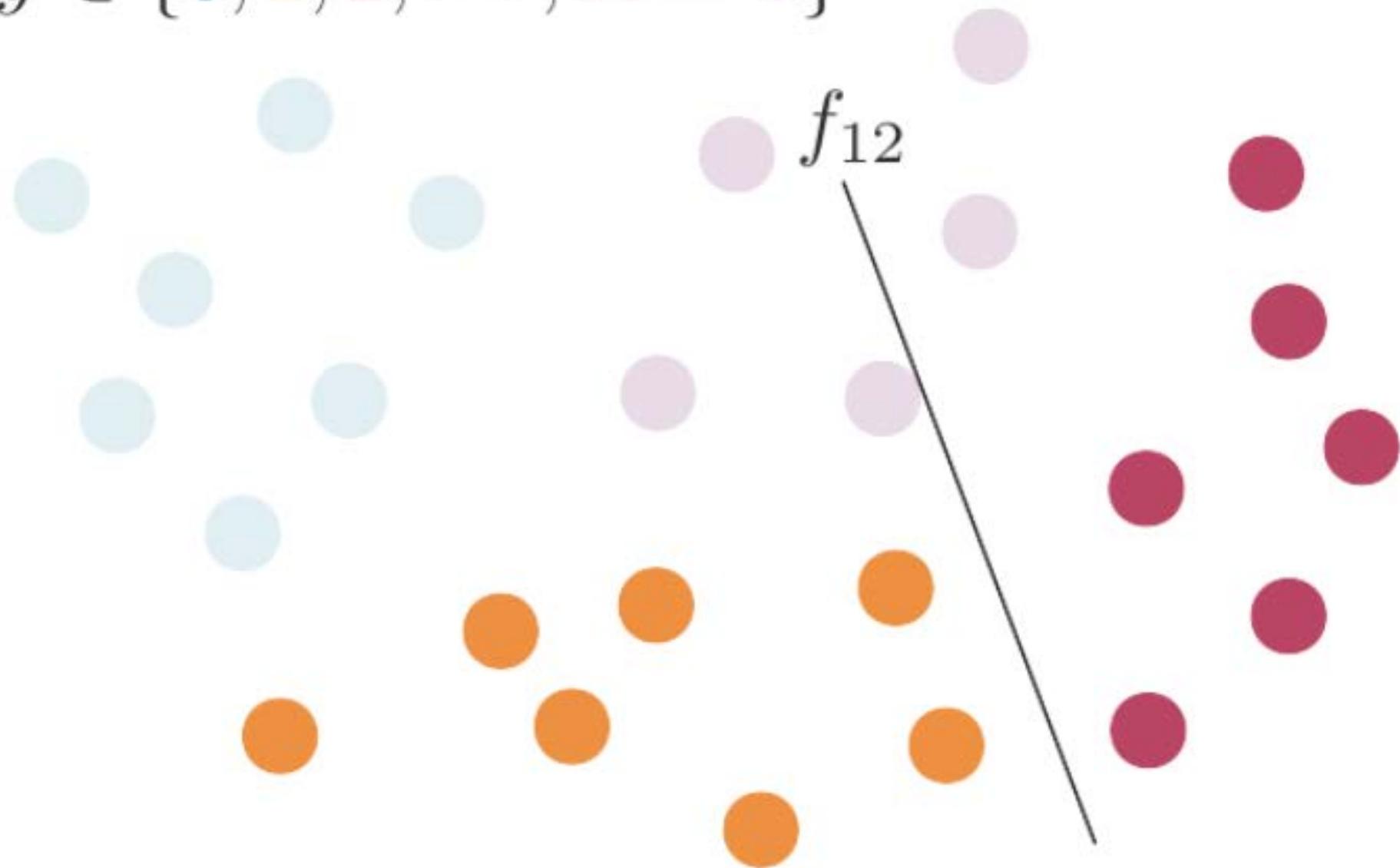
One-versus-rest

$$y \in \{0, 1, 2, \dots, K - 1\}$$



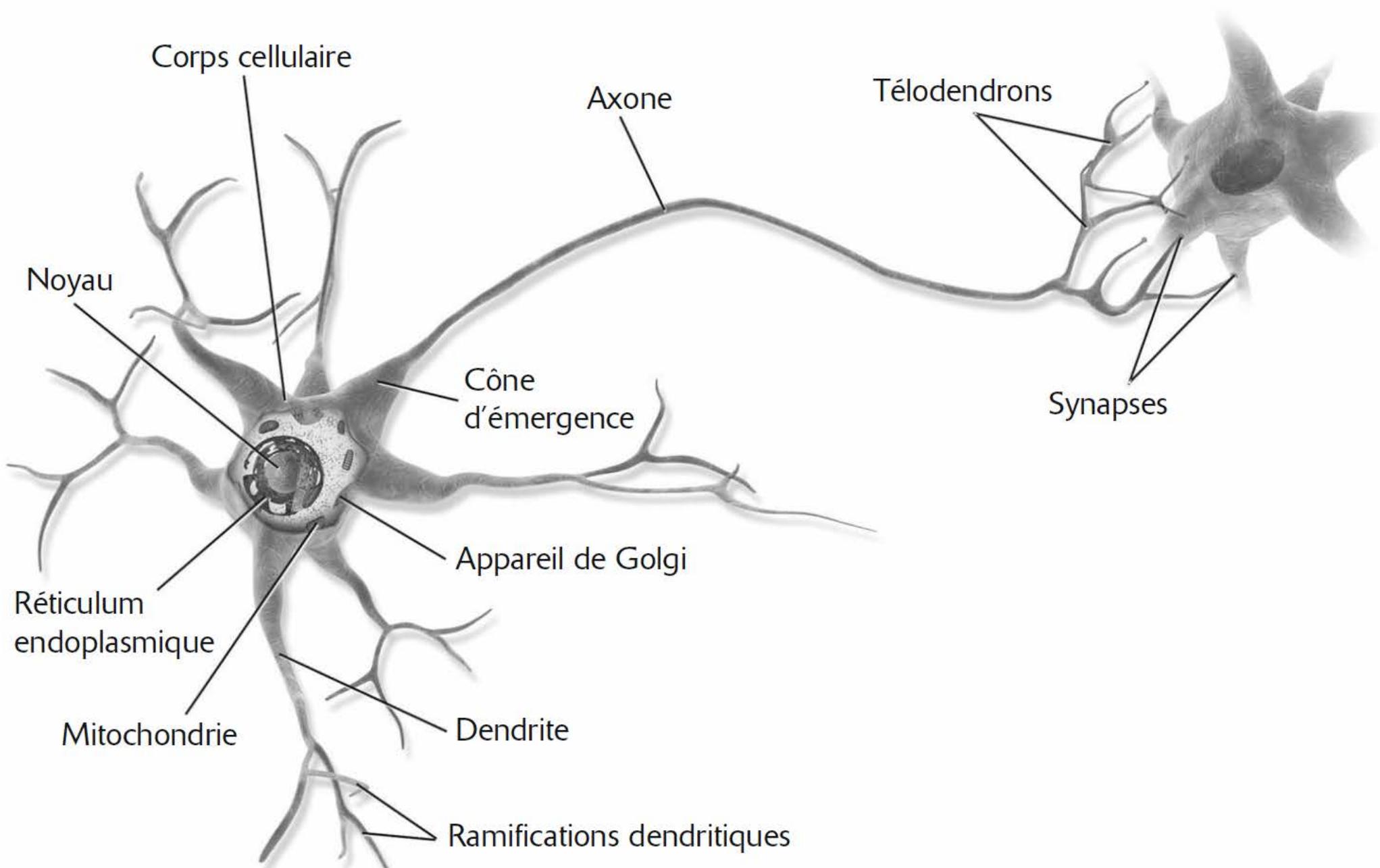
One-versus-one

$$y \in \{0, 1, 2, \dots, K - 1\}$$

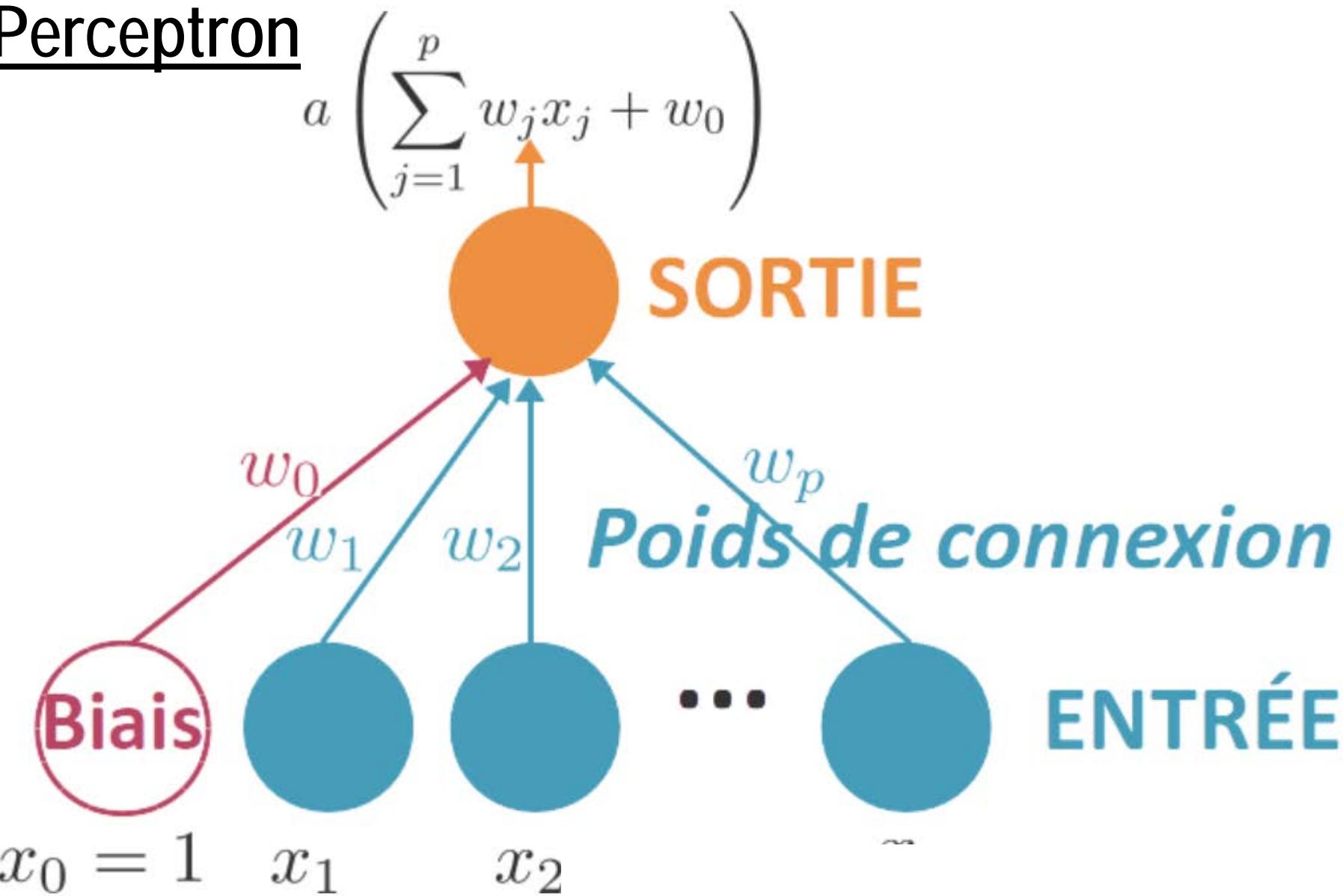


Réseau de neurones

Un neurone biologique



Perceptron

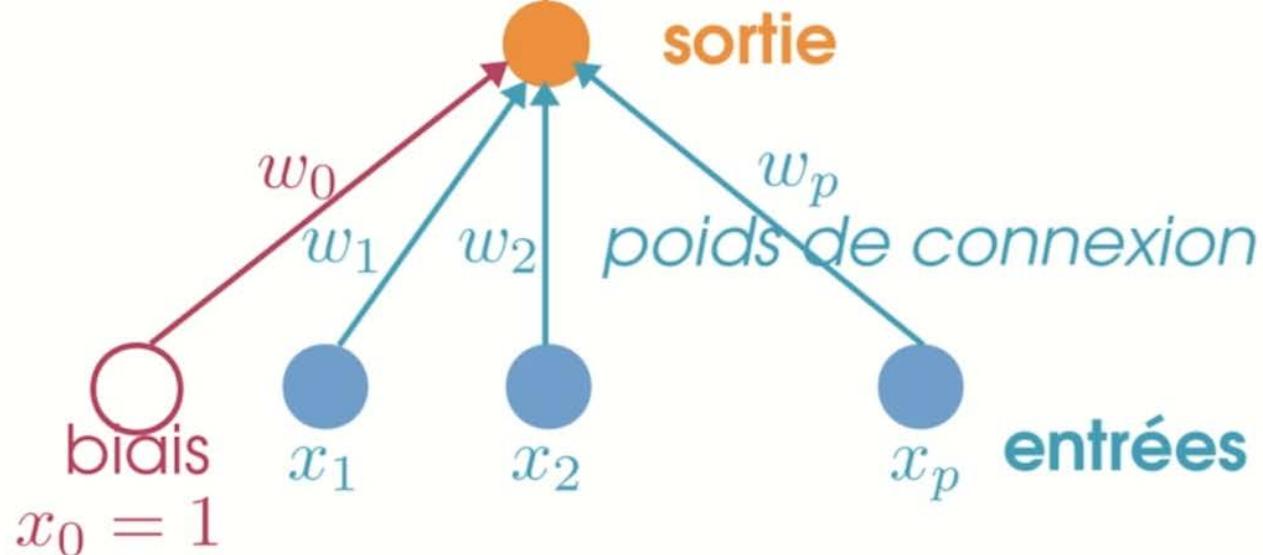


$$s \left(w_0 + \sum_{j=1}^p w_j x_j \right) = \begin{cases} 0 & \text{si } \left(w_0 + \sum_{j=1}^p w_j x_j \right) < 0 \\ 1 & \text{sinon.} \end{cases}$$

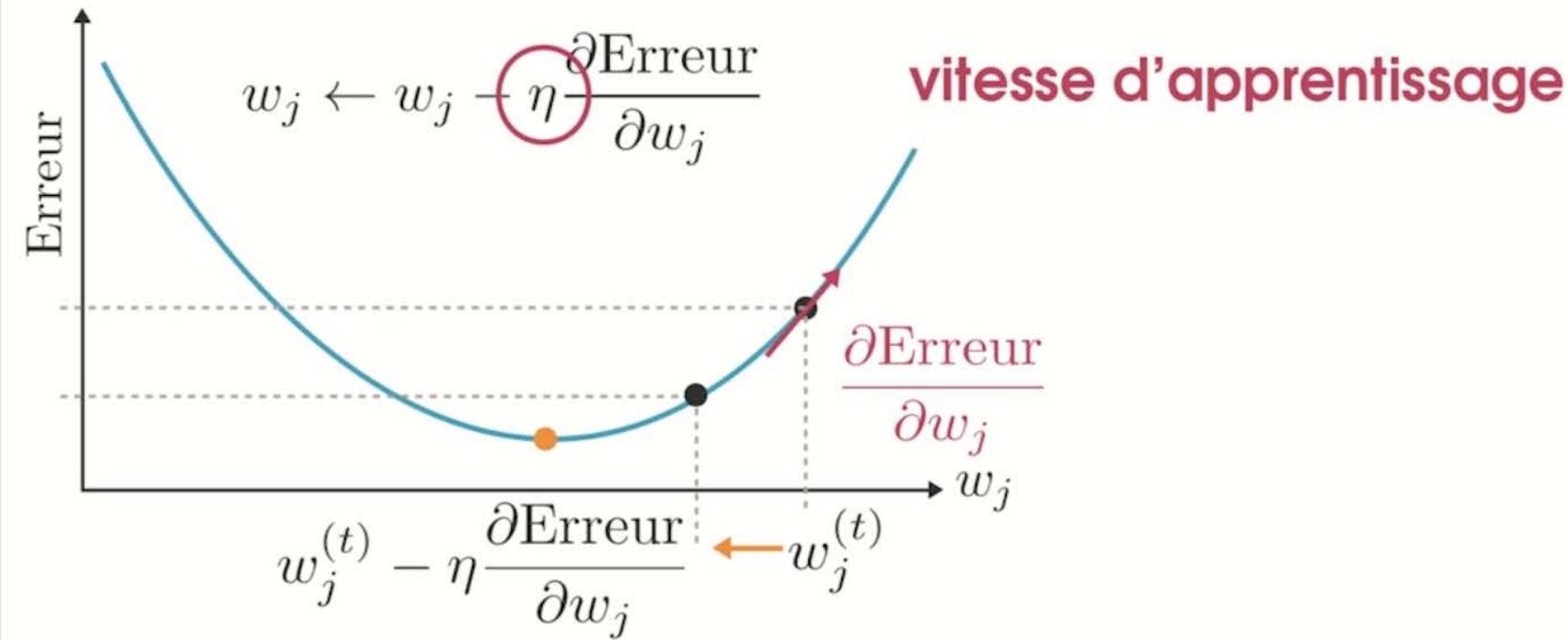
Régression

activation

$$a \left(\sum_{j=1}^p w_j x_j + w_0 \right) = \sum_{j=1}^p w_j x_j + w_0$$



Descente de gradient

 $w_0^{(0)}, w_1^{(0)}, \dots, w_j^{(0)}, \dots, w_p^{(0)}$ $\{(x^{(1)}, y^{(1)}), \dots, (x^{(i)}, y^{(i)}), \dots, (x^{(n)}, y^{(n)})\}$ 

Descente de gradient

Régression

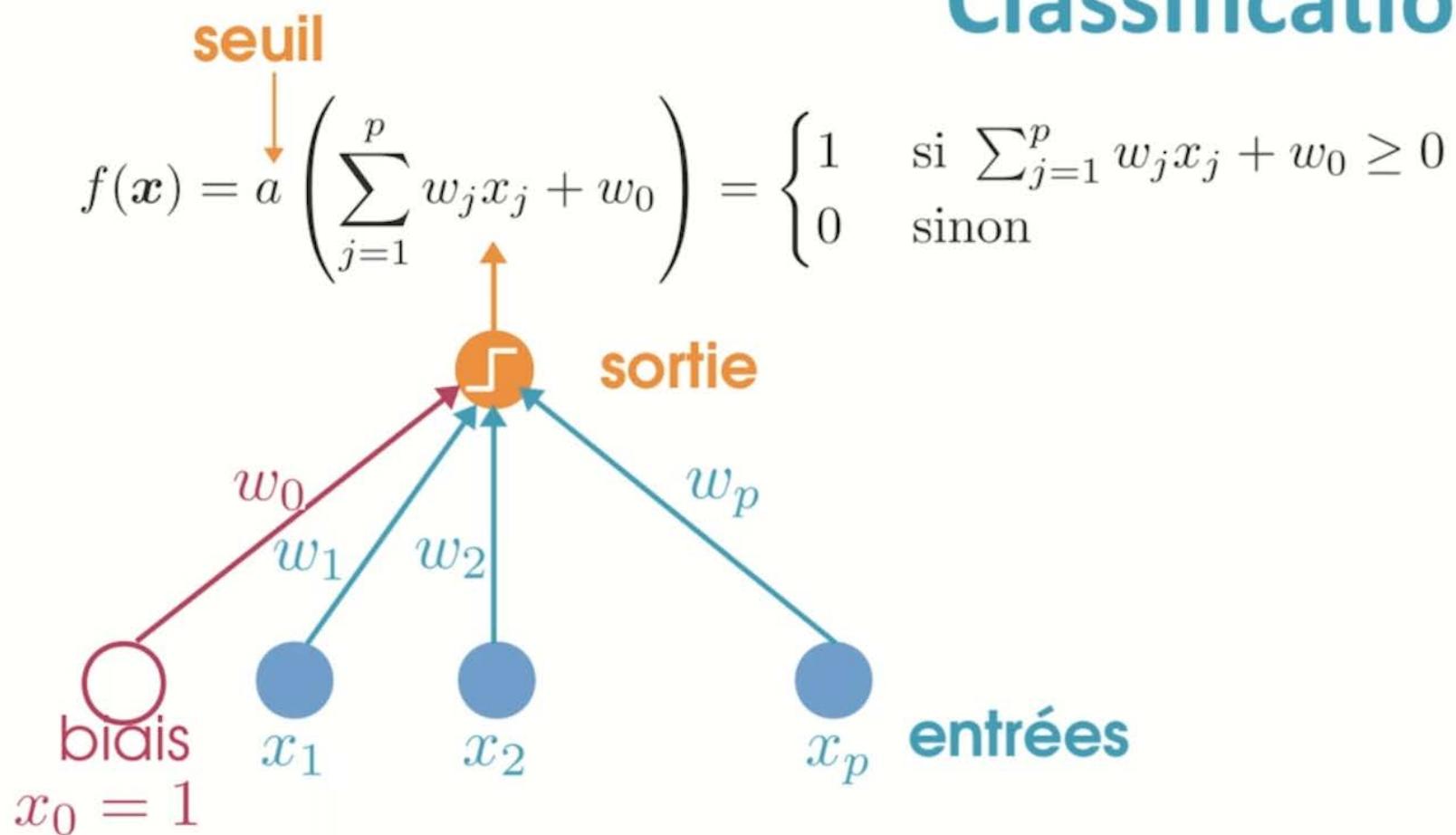
$$w_j \leftarrow w_j - \eta \frac{\partial \text{Erreur}}{\partial w_j}$$

$$\text{Erreur} = \frac{1}{2} \left(f(\boldsymbol{x}^{(i)}) - y^{(i)} \right)^2$$

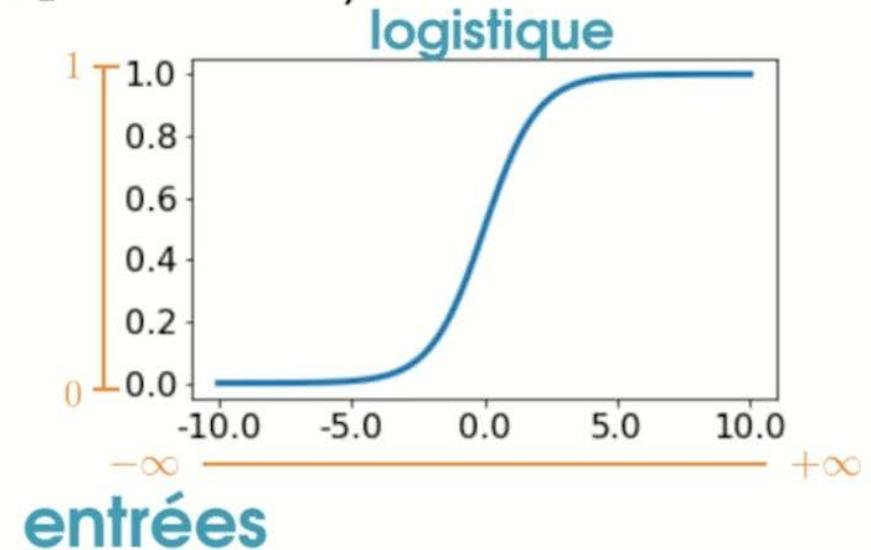
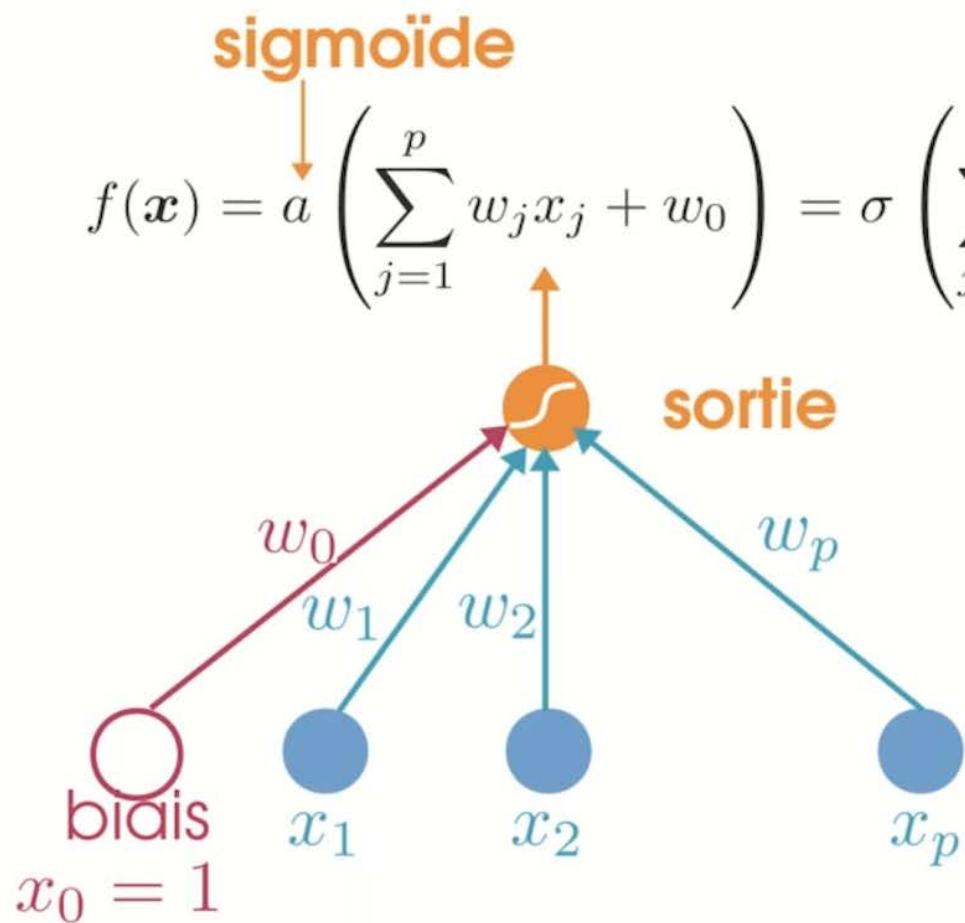
$$f(\boldsymbol{x}) = a \left(\sum_{j=1}^p w_j x_j \right) = \sum_{j=1}^p w_j x_j + w_0$$

$$w_j \leftarrow w_j - \eta \left(y^{(i)} - f(\boldsymbol{x}^{(i)}) \right) x_j^{(i)}$$

Classification



Classification



Descente de gradient Classification

$$\{(\mathbf{x}^{(1)}, y^{(1)}), \dots (\mathbf{x}^{(i)}, y^{(i)}), \dots (\mathbf{x}^{(n)}, y^{(n)})\}$$

$$w_j \leftarrow w_j - \eta \frac{\partial \text{Erreur}}{\partial w_j}$$

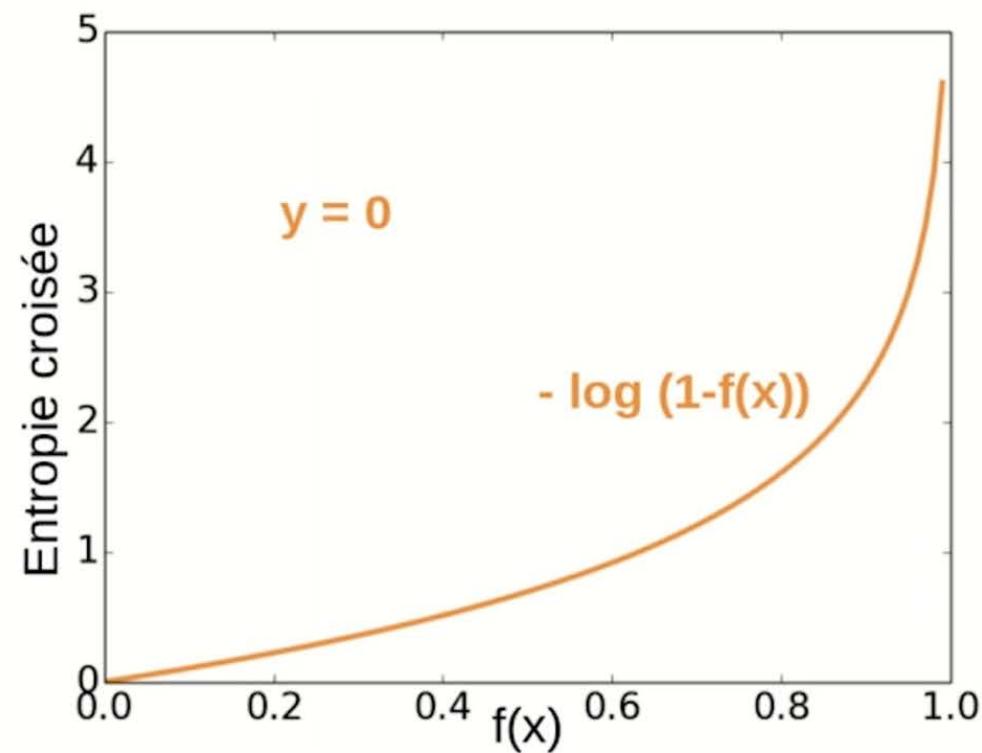
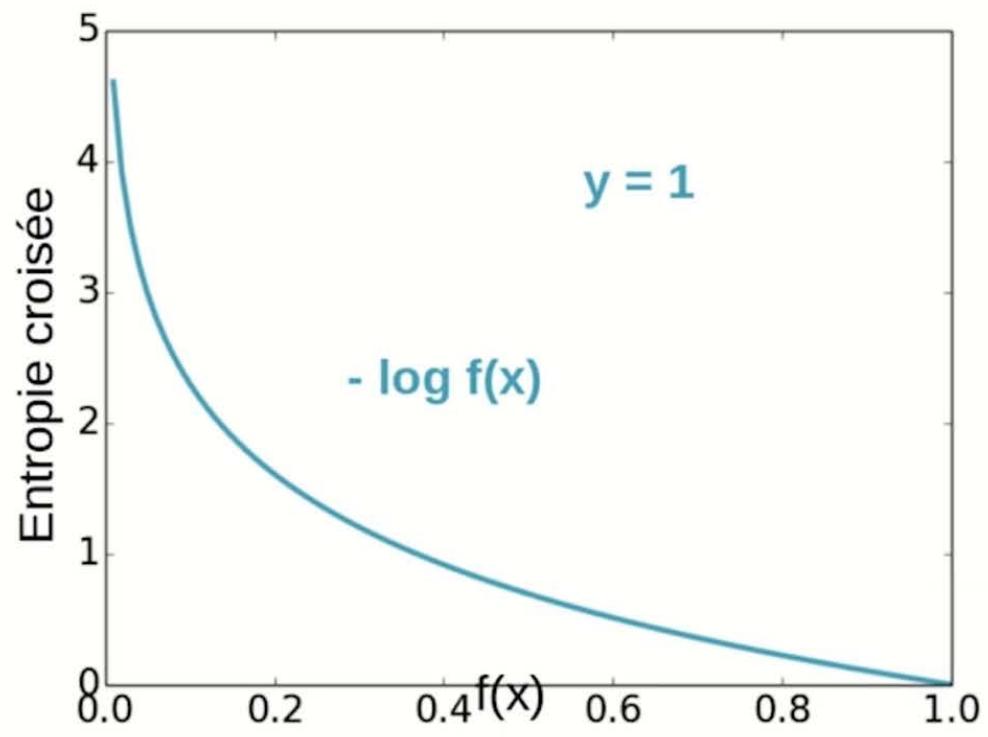
entropie croisée

$$\text{Erreur} = -y^{(i)} \log f(\mathbf{x}^{(i)}) - (1 - y^{(i)}) \log(1 - f(\mathbf{x}^{(i)}))$$

$$f(\mathbf{x}) = a \left(\sum_{j=1}^p w_j x_j + w_0 \right) = \begin{cases} 1 & \text{si } \sum_{j=1}^p w_j x_j + w_0 \geq 0 \\ 0 & \text{sinon} \end{cases}$$

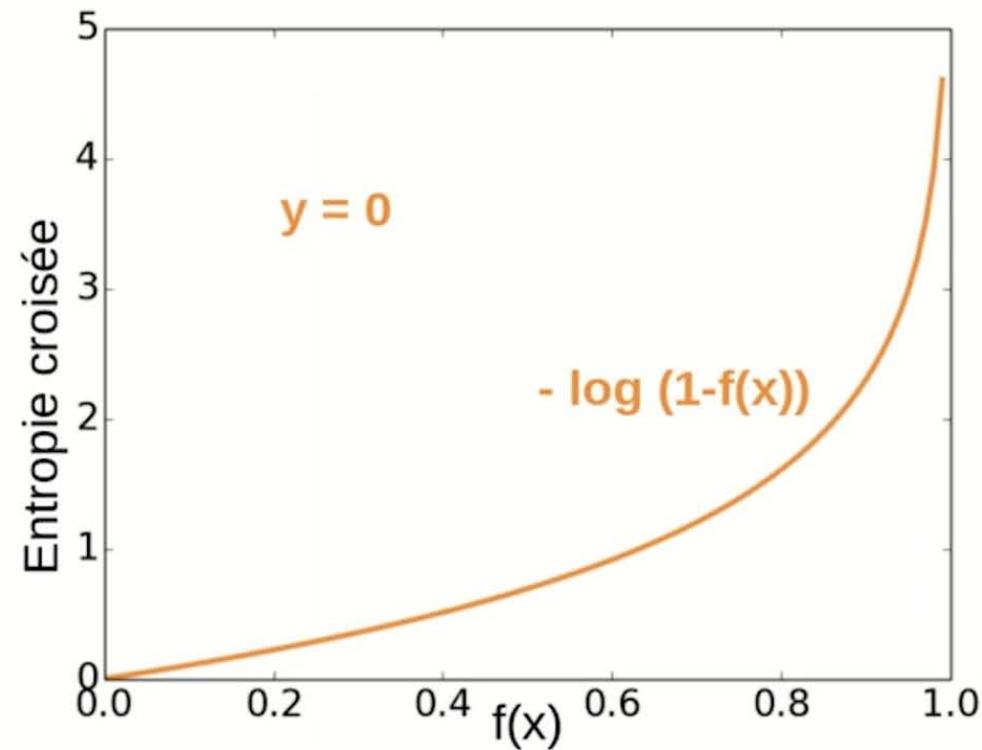
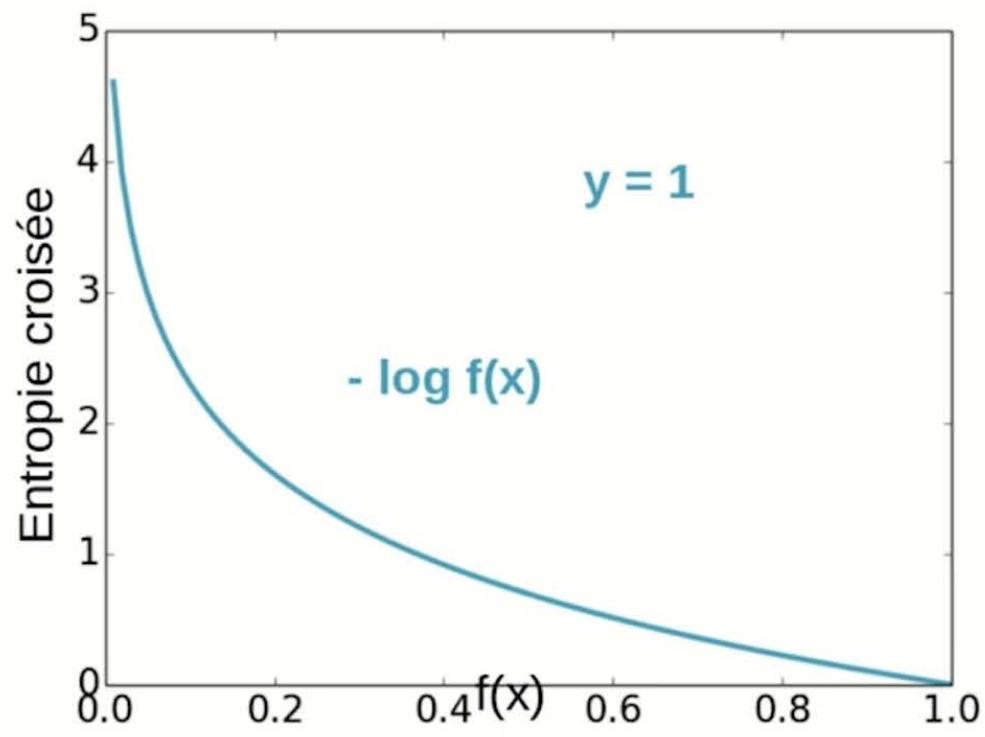
Entropie croisée

$$\text{Erreur} = -y^{(i)} \log f(\mathbf{x}^{(i)}) - (1 - y^{(i)}) \log(1 - f(\mathbf{x}^{(i)}))$$



Entropie croisée

$$\text{Erreur} = -y^{(i)} \log f(x^{(i)}) - (1 - y^{(i)}) \log(1 - f(x^{(i)}))$$



Descente de gradient Classification

$$\{(\mathbf{x}^{(1)}, y^{(1)}), \dots (\mathbf{x}^{(i)}, y^{(i)}), \dots (\mathbf{x}^{(n)}, y^{(n)})\}$$

$$w_j \leftarrow w_j - \eta \frac{\partial \text{Erreur}}{\partial w_j}$$

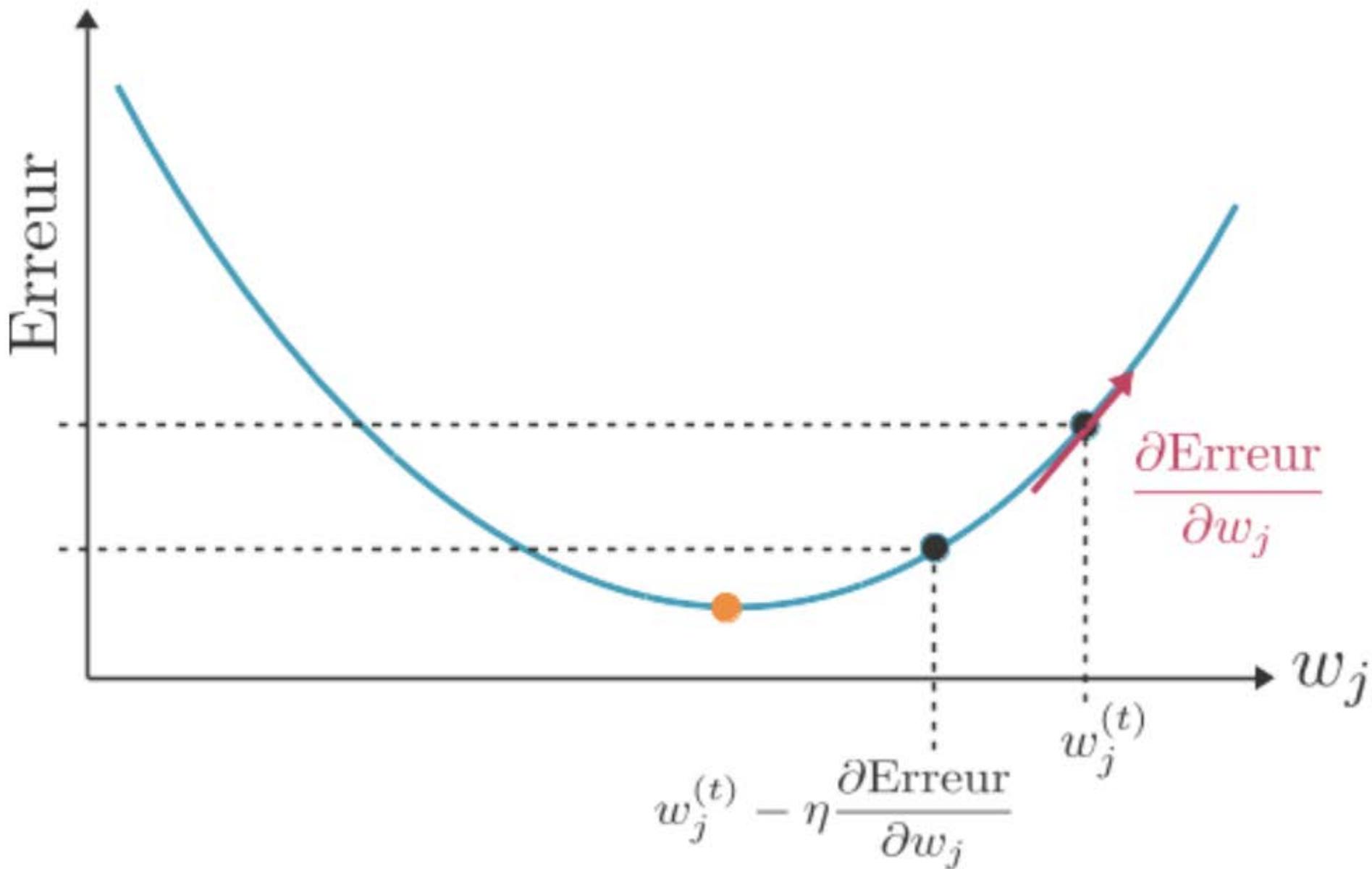
entropie croisée

$$\text{Erreur} = -y^{(i)} \log f(\mathbf{x}^{(i)}) - (1 - y^{(i)}) \log(1 - f(\mathbf{x}^{(i)}))$$

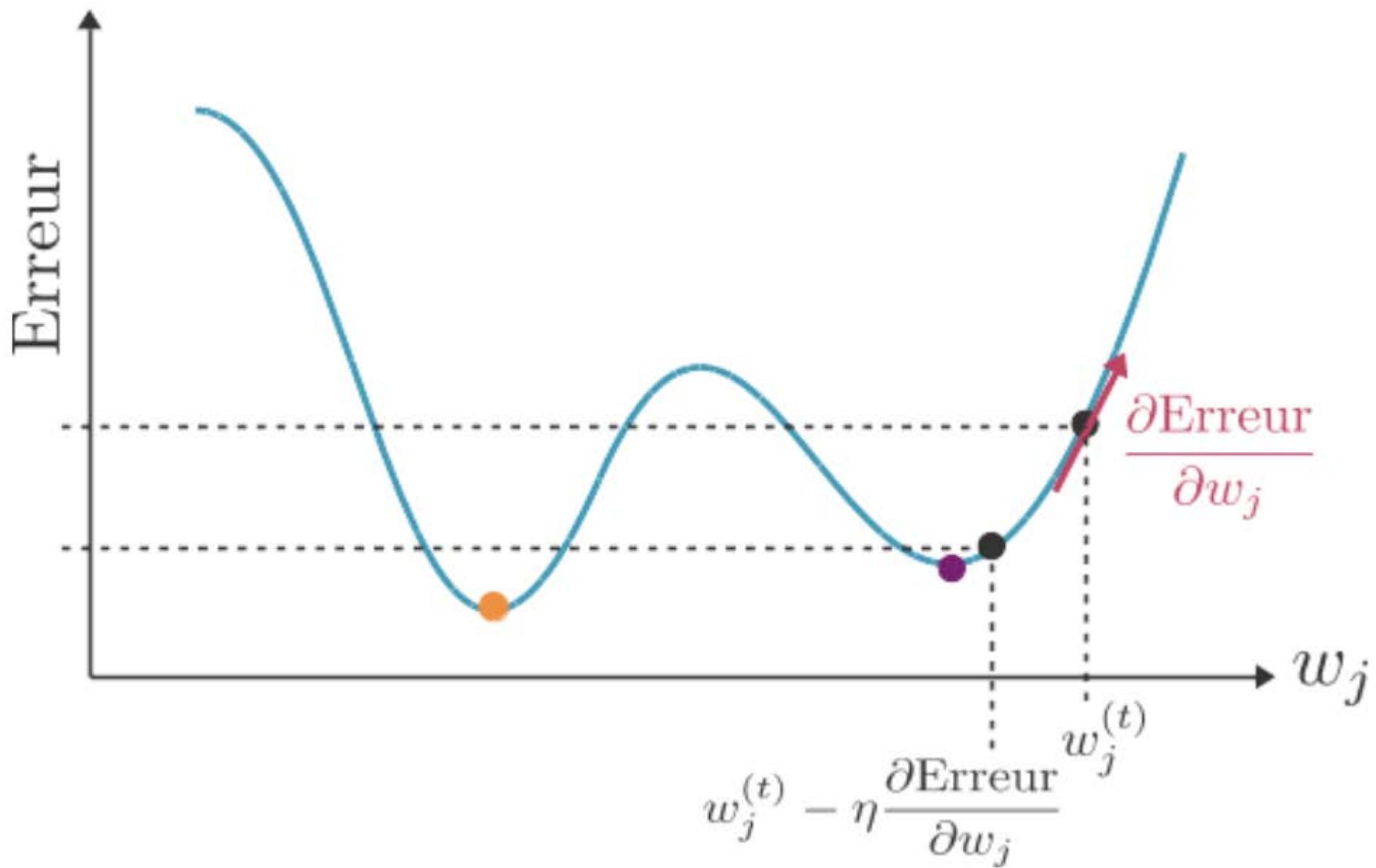
$$f(\mathbf{x}) = a \left(\sum_{j=1}^p w_j x_j + w_0 \right) = \sigma \left(\sum_{j=1}^p w_j x_j + w_0 \right)$$

$$w_j \leftarrow w_j - \eta \left(y^{(i)} - a(\mathbf{x}^{(i)}) \right) x_j^{(i)}$$

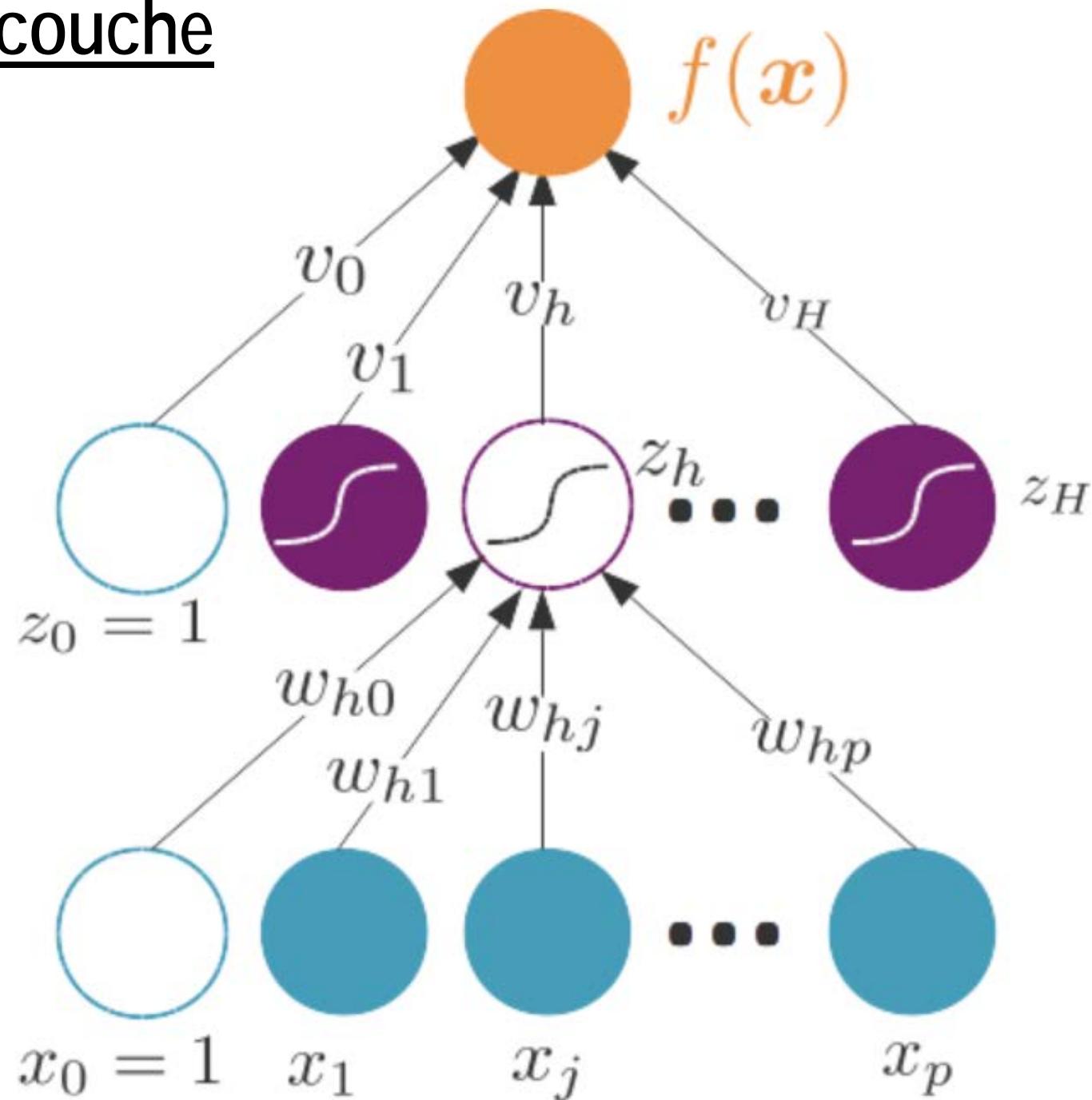
L'entraînement d'un perceptron



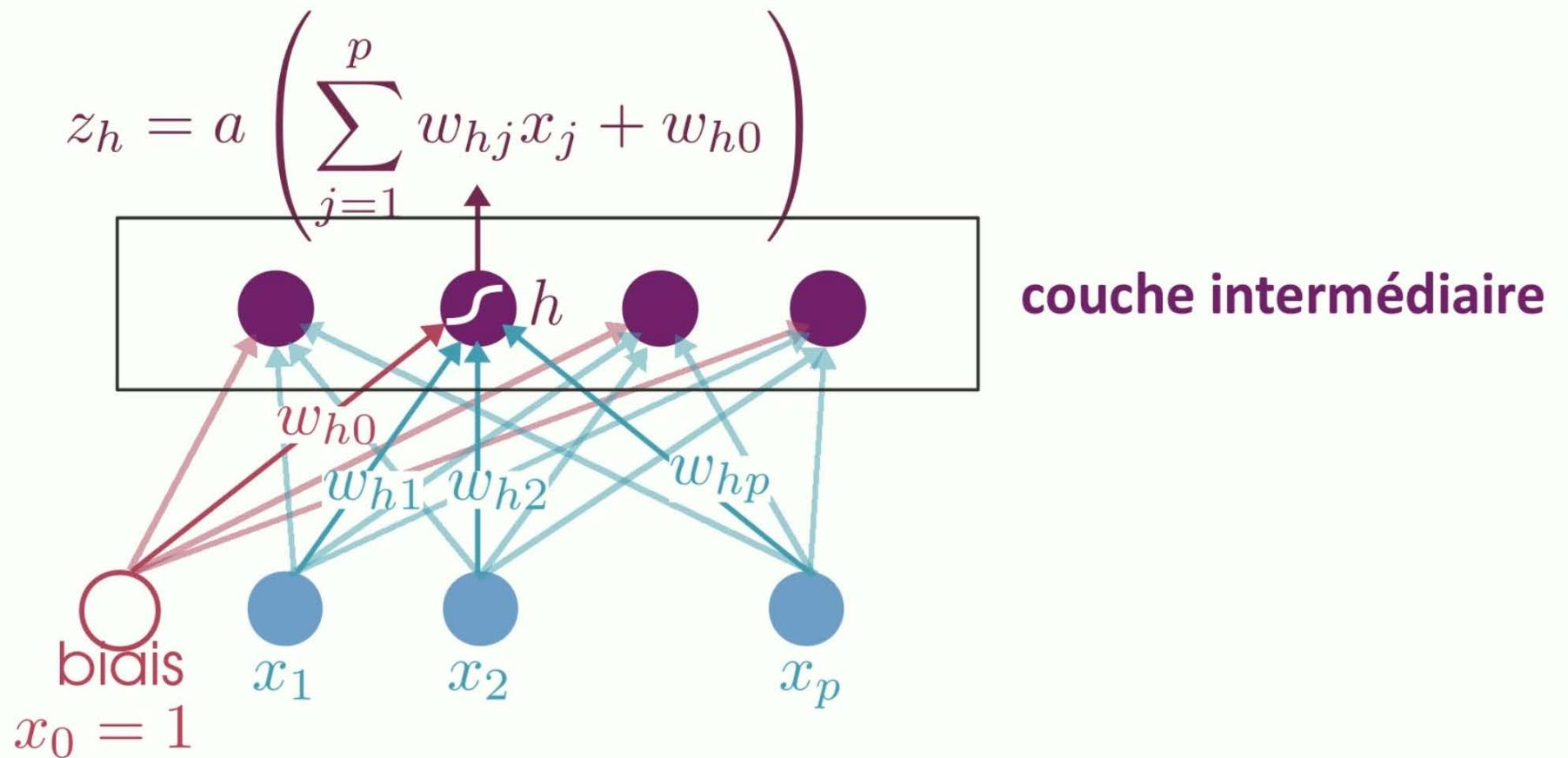
Conditionnement et minimums locaux



Perceptron multi-couche

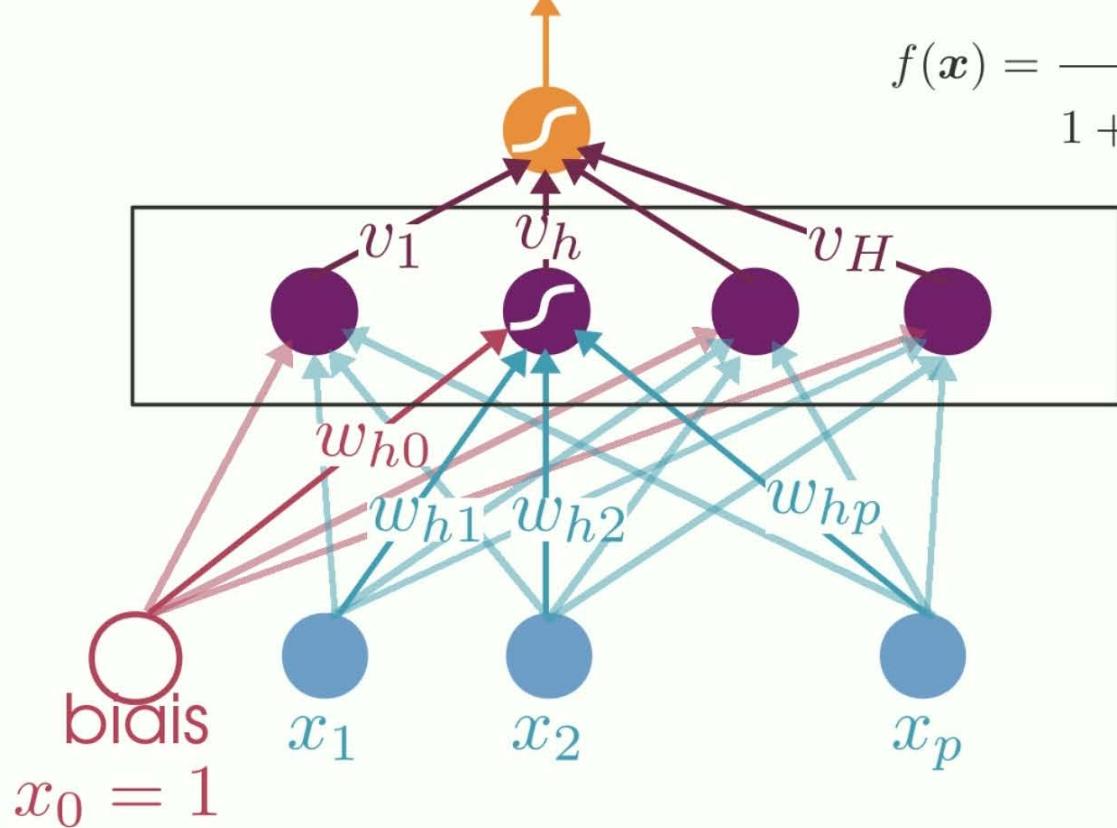


Perceptron multicouche



Perceptron multicouche et rétropropagation

$$f(\mathbf{x}) = a \left(\sum_{h=1}^H v_h z_h \right)$$

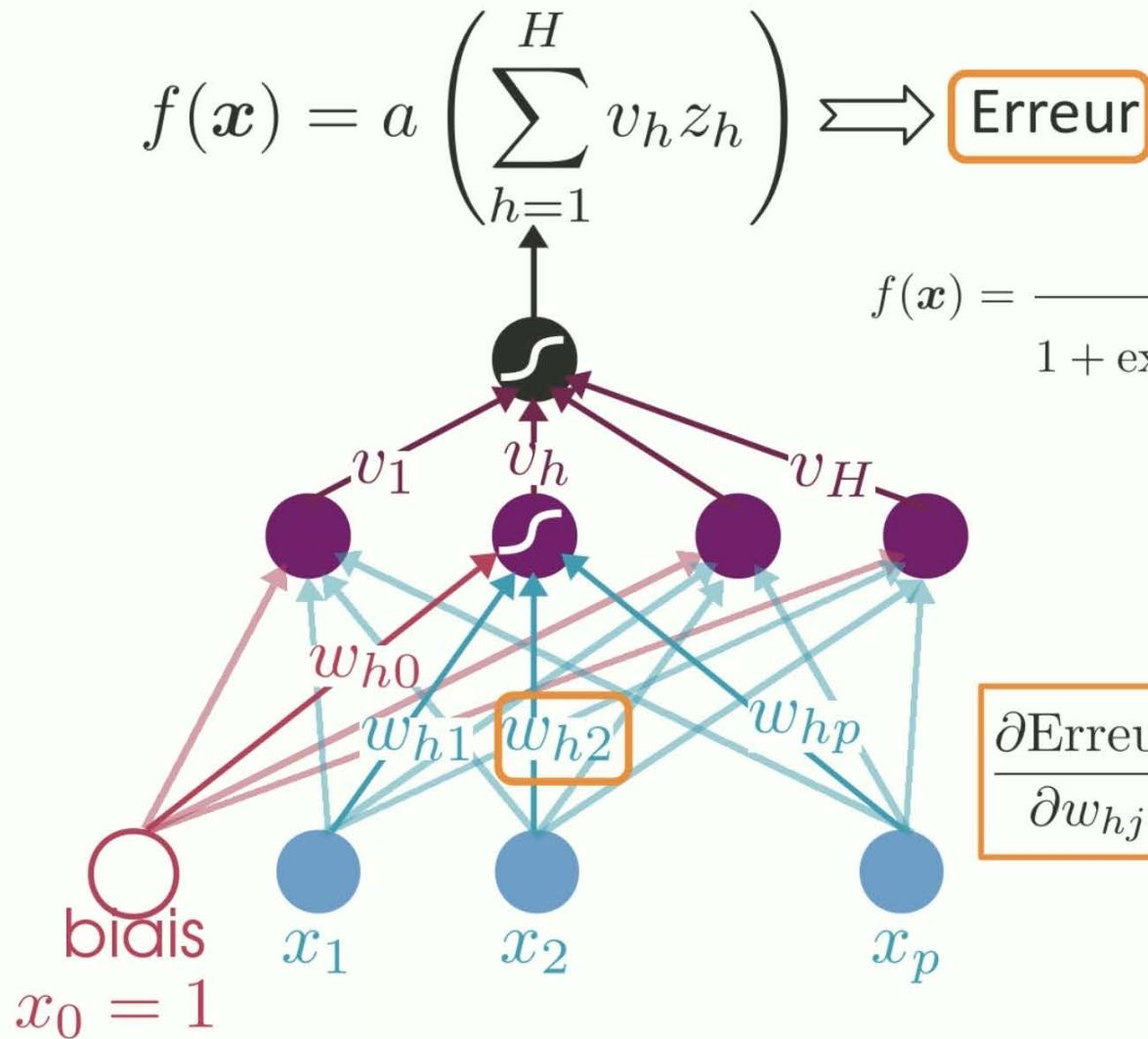


activation logistique

$$a : u \mapsto \frac{1}{1 + e^{-u}}$$

$$f(\mathbf{x}) = \frac{1}{1 + \exp \left(- \sum_{h=1}^H v_h \frac{1}{1 + e^{- \sum_{j=1}^p w_{hj} x_j}} \right)}$$

Perceptron multicouche et rétropropagation



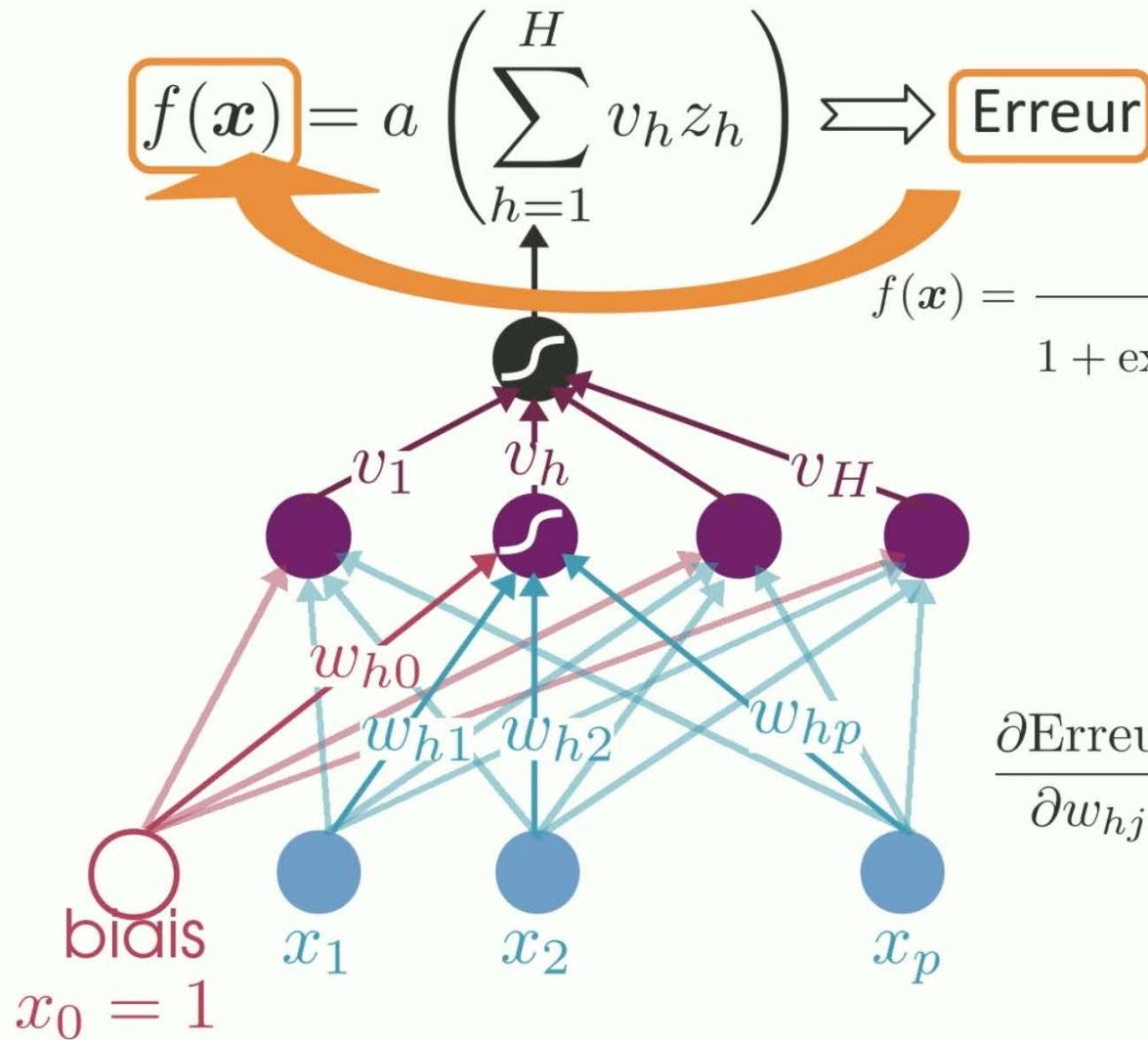
$$f(\mathbf{x}) = a \left(\sum_{h=1}^H v_h z_h \right) \iff \text{Erreur}$$

$$f(\mathbf{x}) = \frac{1}{1 + \exp \left(- \sum_{h=1}^H v_h \frac{1}{1 + e^{- \sum_{j=1}^p w_{hj} x_j}} \right)}$$

$$w \leftarrow w - \eta \frac{\partial \text{Erreur}}{\partial w}$$

$$\frac{\partial \text{Erreur}}{\partial w_{hj}} = \frac{\partial \text{Erreur}}{\partial f} \times \frac{\partial f}{\partial z_h} \times \frac{\partial z_h}{\partial w_{hj}}$$

Perceptron multicouche et rétropropagation

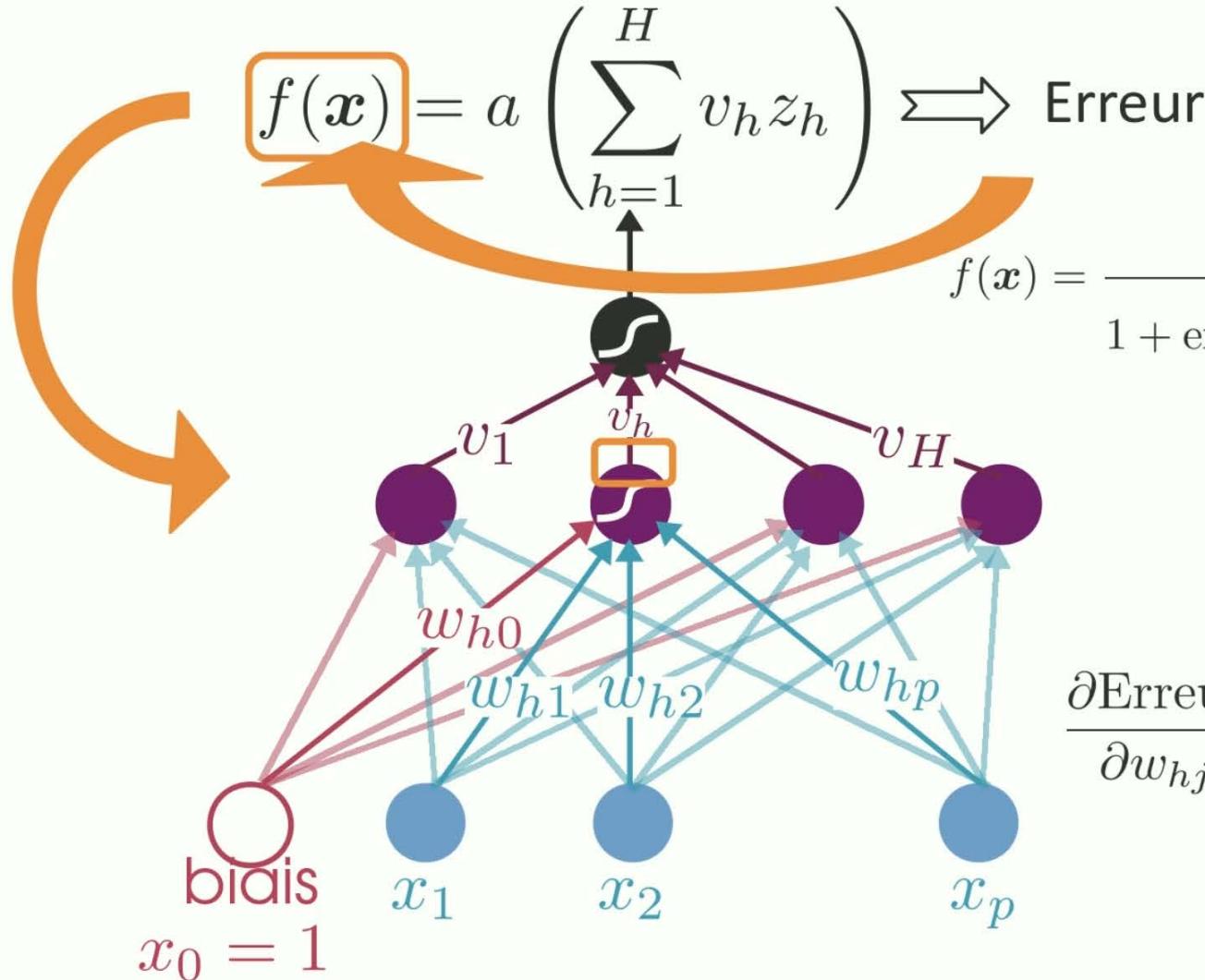


$$f(\mathbf{x}) = \frac{1}{1 + \exp \left(- \sum_{h=1}^H v_h \frac{1}{1 + e^{- \sum_{j=1}^p w_{hj} x_j}} \right)}$$

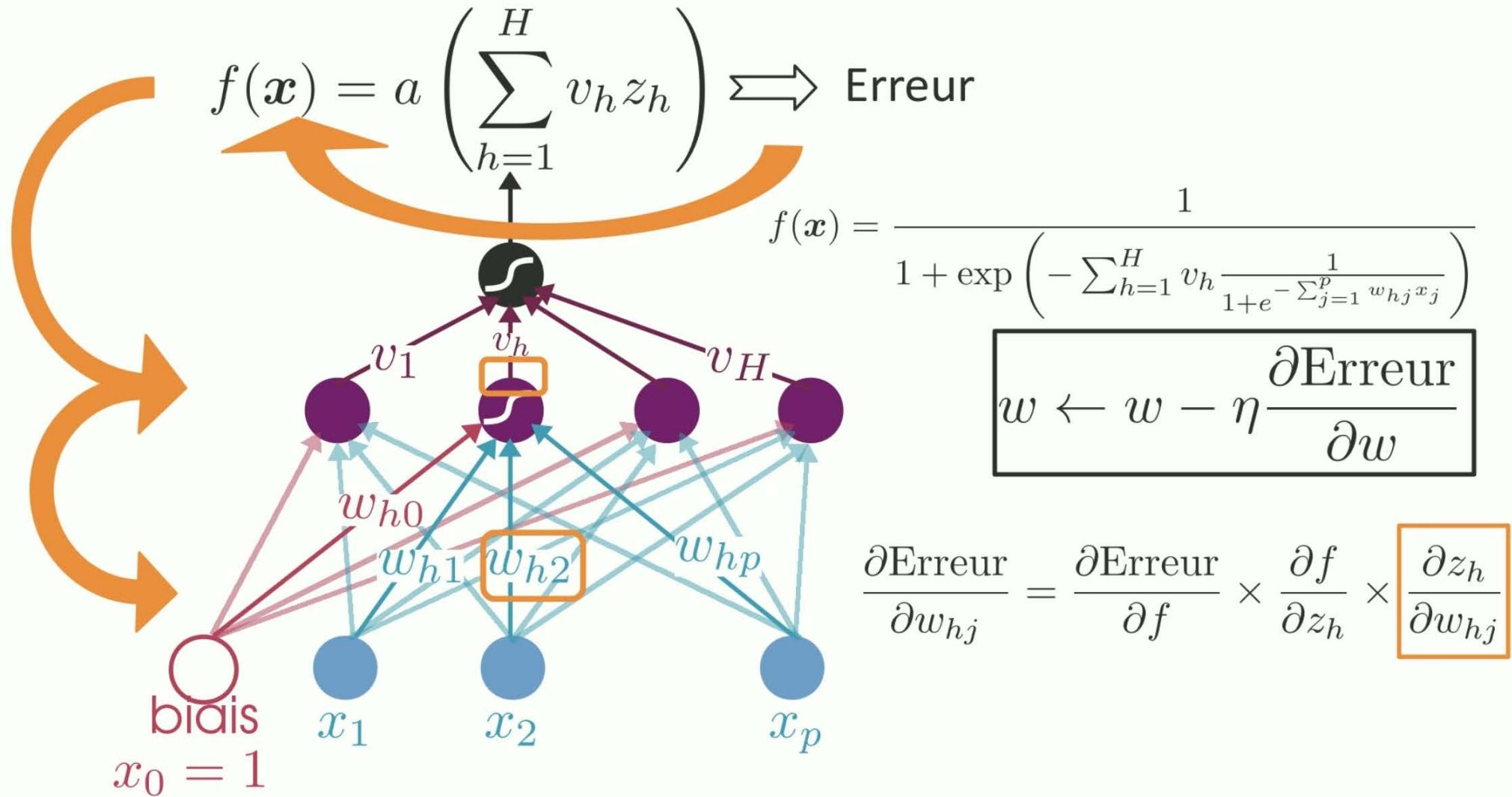
$$w \leftarrow w - \eta \frac{\partial \text{Erreur}}{\partial w}$$

$$\frac{\partial \text{Erreur}}{\partial w_{hj}} = \boxed{\frac{\partial \text{Erreur}}{\partial f}} \times \frac{\partial f}{\partial z_h} \times \frac{\partial z_h}{\partial w_{hj}}$$

Perceptron multicouche et rétropropagation



Perceptron multicouche et rétropropagation



- <http://archive.ics.uci.edu/ml/index.php>
 - <https://registry.opendata.aws/>
 - <https://www.kaggle.com/datasets>
-
- <http://dataportals.org/>
 - <https://opendatamonitor.eu/>