

Modélisation statistique

Préliminaires sur les modèles en général

- Un modèle est une représentation «humaine» de ce que pourrait être le réel.
- Un modèle ne peut décrire la «réalité» des choses mais seulement les relations/rapports entre ces choses.
- Un modèle n'est ni vrai ni faux. Son but est d'être utile.
- On peut même dire qu'un modèle est toujours faux, mais qu'il peut être très efficace pour résoudre/faire avancer des problèmes appliqués ou théoriques.
- Un modèle est constitué d'hypothèses de travail sur des questions appliquées en physique, chimie, biologie, économie, environnement ou autre.

Sur les modèles mathématiques

- Une grande partie des mathématiques appliquées consiste à faire de la modélisation dans diverses disciplines
- De façon schématique, on peut distinguer
 - la modélisation **déterministe** où on ne prend pas en compte des variations aléatoires par le biais d'outils (EDO, EDP, filtrage,...)
 - et la modélisation **stochastique/statistique** (qui prend en compte ces variations aléatoires en essayant de leur associer des lois de probabilité)
- Qualités d'un modèle: (parfois antagonistes)
 - Bon ajustement aux données et économe en paramètres
 - Robustesse (aux erreurs de mesures, à de petits écarts aux hypothèses de travail)
 - Facilité de manipulation (calculs numériques et stabilité des résultats)
 - Caractère prédictif
 - Interprétation des paramètres

Les modèles statistiques n'existent pas!

- Ce sont les modèles stochastiques (probabilistes) qui importent le plus.
- Les statistiques complètent la modélisation stochastique en détaillant par des paramètres (vecteurs, fonctions numériques) les choix précis de lois déjà adoptées.
- Les «modèles statistiques» relèvent surtout de techniques permettant de préciser parmi les modèles en compétition ceux qui seraient les plus «proches», plus «vraisemblables» au vu des données récoltées.
- La modélisation statistique peut commencer déjà «sans modèle» par l'utilisation des «Statistiques Descriptives des données»: calcul de moyenne, de covariance, de distributions empiriques,...

La modélisation stochastique

- **La modélisation stochastique** a pour but essentiel de préciser des classes de lois de probabilité rendant compte des variations aléatoires des phénomènes d'intérêt, variations dues à des causes soit inconnues, soit impossible à mesurer (cachées, trop nombreuses,...)
- Pour cela, elle se donne un cadre formel permettant:
 - de décrire les variations aléatoires mentionnés,
 - Et d'étudier les propriétés générales des phénomènes qu'ils engendrent.
- **La modélisation statistique**, plus appliquée, consiste essentiellement à choisir les outils appropriés pour confronter les données au modèle stochastique.
- Noter que le terme de modélisation statistique est très général et que, à la limite, toute démarche statistique en relève.

La modélisation statistique en constante évolution

- Les méthodes de modélisation statistique sont très nombreuses et en constante évolution/amélioration. Il s'agit plutôt de démarches /méthodes que de modèles figés pour:
 - Traiter des masses de données de plus en plus volumineuses (internet, biologie à haut débit, climat, imagerie, marketing...)
 - Et utiliser les nouveaux moyens de calcul tout aussi considérables
- Mais la question de base reste globalement l'«explication ou la mise en relation stochastique d'une variable privilégiée Y de nature parfois complexe , appelée variable à expliquer ou réponse, avec des variables dites explicatives X , expliquant “au mieux” Y .

Modèles (?) préliminaires

- **Nettoyage des données (data management):**
 - Statut des données manquantes, erronées, variables redondantes
- **L'exploration des données :**
 - Toutes les statistiques descriptives et caractéristiques empiriques : moyennes, covariances, histogrammes,
 - Estimation non paramétrique des densités, des intensités, ...
 - ACP (analyse en Composantes Principales) pour les liaisons des variables, ACM (Analyse des Correspondances Multiples) entre variables qualitative. Analyse discriminante, classification, méthodes CART (découpage de populations en fonction des variables explicatives,...
- Transformer des variables? Regrouper des variables? des modalités de variables?

Méthode statistiques non paramétriques

- On a seulement besoin d'hypothèses de travail assez générales (données iid (indép et identiquement distribuées), ou stationnaires, ou invariantes par permutation,...):
- **Des méthodes d'estimation** par noyaux de densité, d'intensité de processus, de fonctions de covariance d'un champs aléatoire, de la fonction de survie, etc...
- Des tests non paramétriques : égalité des moyennes de 2 échantillons, test des rangs, ...
- Test d'égalité de distributions Kolmogorov-Smirnov, etc...
- Toute la famille des tests par permutations en statistique spatiale sur la détection d'agrégats spatiaux, sur l'homogénéité directionnelle, ...
- **Extension a de la statistique semi-paramétrique**: une partie paramétrique à inférer et une partie non paramétrisée (modèle de Cox,...)

Modèles Paramétriques classiques (échantillon iid ou presque)

- **Le modèle linéaire (gaussien) de base**: le plus simple, le plus ancien
 - Régression linéaire, analyse de *variance/covariance*, les variables explicatives sont déterministes (effets fixes).
 - Cadre gaussien très efficace au niveau du formalisme
- **Le modèle linéaire généralisé**: extension au non gaussien et description des paramètres par des fonctions de liens très générales et plus seulement linéaires: régression logistique, de Poisson, log-linéaire, durée de survie,...
- **Modèles linéaires généralisés**: On explique Y de façon non linéaire à partir de fonctions inconnues des X (on fait alors de la statistique non paramétrique):
 - Exemples: régression non paramétrique, GAM (GeneralizedAdditive Models), Réseaux de neurones
- **Les modèles mixtes**: On étend les modèles précédents au cas où les variables explicatives sont elles aussi aléatoires avec spécification de leurs lois (effets aléatoires): Ceci permet d'expliquer une plus grande variabilité des données

Modèles paramétriques spatio-temporels et autres

- **Les modèles probabilistes plus complexes** se dénomment selon
 - le caractère continu/discret de l'espace et du temps
 - S'il y a une flèche (du temps ou d'ordre) pour les processus temporels et pour certains processus sur graphes (causalité directionnelle)
 - La structure de voisinage des points supports: temporel/ spatial/ spatiotemporel/ sur réseaux
 - Le types d'objets à étudier sur ces espaces
- **Modèle Statistique**: On le décrit par la donnée d'une famille de variables aléatoires $\{X(t), t \in T\}$ (observations) générées selon un ensemble donné de lois conjointes $P_\theta(dx(t), t \in T)$ où θ est un paramètre vectoriel de dimension finie ou infinie inconnu.
- **Le but** inférer sur θ : **l'estimer et estimer la qualité de cette estimation**,
...

Validation d'un modèle

- Plusieurs façons de concevoir la validation d'un modèle:
- Cas classique : Regarder la distribution des résidus (d'une régression par exemple: R^2 statistique, ...)
- Sinon, des considérations plus générales dépendantes du modèle lui même
 - On simule des données selon le modèle estimé et on compare quelques statistiques calculées sur les simulations à celles calculées sur l'échantillon.
 - On utilise une partie des données pour regarder la qualité de prédiction des données non utilisées (on peut répéter cela plusieurs fois : bootstrap)

Il existe deux types de modèles:

- Modèles déterministes
- Modèles statistiques

Il existe deux types de modèles:

- Modèles **déterministes**
- Modèles statistiques

Modèles déterministes

Ils consistent à mettre sous forme d'équations mathématiques les processus identifiés dans le modèle conceptuel

Dans la grande majorité des cas, il n'y a pas de solution analytique au système d'équations mathématiques et il faut développer une solution numérique (approximative)

Il existe deux types de modèles:

- Modèles déterministes
- Modèles **statistiques**

Modèles statistiques

Ils consistent à relier différentes variables (par exemple émissions de polluants et concentrations de polluants dans l'atmosphère) par des équations qui représentent des relations statistiques

Les processus qui relient ces variables sont donc représentés de façon implicite dans ces équations statistiques

Avantages

- Ils sont fondés sur des mesures et sont donc ancrés dans la réalité (avec cependant l'incertitude liée aux mesures)

Inconvénients

- Les processus physico-chimiques ne sont pas traités explicitement

- **La statistique fait partie de la culture**

ANGLO-SAXONNE

- **Culture probabiliste indispensable**
- **Enorme retard en France**

- **Causes :**

- **Statistique: parent pauvre des probabilités et des mathématiques**
- **Forte imprégnation cartésienne**
- **L'«incertain» est anti pédagogique (déstabilisant ...)**
- **Nette préférence pour l'enseignement de certitudes**
- **Association obligatoire avec une autre discipline (bio., gestion,...)
(l'approche pluridisciplinaire est encore mal acceptée)**

- **Conséquences :**

- * **Bon sens statistique insuffisamment développé**

- * **La présentation de données statistiques :
un exercice « difficile »**

- * **« *Un chiffre ne peut être qu'exact ,
exempt de toute indétermination* »**

→ **Risque de manipuler l'opinion**

Exemples :

- **moyenne, médiane**
- **cote de popularité des hommes politiques**
- **classement des hôpitaux, des lycées...**
- **rapports d'«experts» ...**

Connaître, c'est mesurer (Brunschvicg)

- **La statistique nous dit:**
 - * **Comment effectuer les mesures**
 - * **Comment extraire l'**information** des mesures**

INFORMATION

Principale matière première du XXI^e siècle

Sa production et son exploitation



statisticiens

LA VARIABILITE

- ❖ En général, dans une population, la **variabilité** est un critère de **qualité** , voire une nécessité (pour la survie)

Exception : l'industrie

Qualité \equiv  de la variabilité

- ❖ Le statisticien doit savoir :
 - appréhender
 - analyser
 - « gérer » la variabilité
- ❖ La variabilité est une entrave à l'inférence statistique

LA MODELISATION STATISTIQUE

- En général, une observation dépend
d'un grand nombre de facteurs
- cause **multifactorielle**
- les effets des facteurs ne sont pas simplement
additifs
(présence d'**interactions**)

LA MODELISATION (suite)

→ Le statisticien va tenter :

- d'identifier les facteurs prépondérants
- évaluer leur importance relative
- expliciter le lien de ces facteurs « causaux »
avec le caractère étudié, à l'aide d'un

modèle mathématique

Exemple :

variable expliquée : prix d'une voiture

variables explicatives : cylindrée, puissance, vitesse,
largeur, longueur, poids ...

$$\text{prix} = 1775 + 103,8 \text{ puissance} + 9,8 \text{ poids}$$

(CV DIN) (kg)

Statisticiens → Interlocuteurs privilégiés des décideurs

- dans **tous** les secteurs d'activité :

(politique, économique, scientifique, industriel ...)

et

- à **tous** les niveaux :

(collecte de données, conception des systèmes d'information, contrôle de la production, analyse et restitution des données, etc.)



Diversification des métiers de la statistique

Discipline transversale par excellence

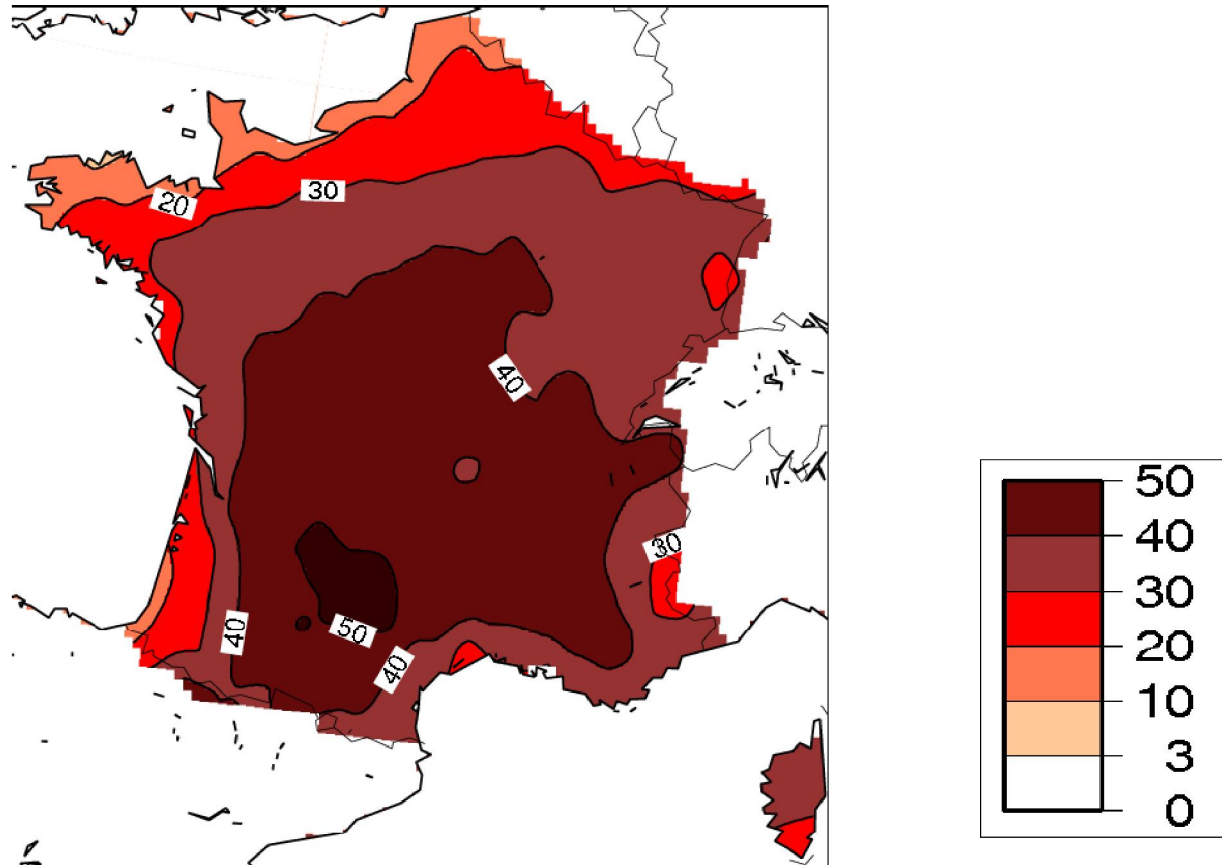
DOMAINES D'UTILISATION DE LA STATISTIQUE

- **STATISTIQUES OFFICIELLES (I.N.S.E.E.)**
- **PRESSE – MEDIAS**
- **BANQUES – ASSURANCES**
- **SCIENCES DE LA VIE**
- **ENVIRONNEMENT (Foresterie, pêche ...)**
- **SANTE**
- **SCIENCES HUMAINES**
- **ENTREPRISES – INDUSTRIE (R&D, contrôle de qualité, études de marché, management ...)**
- **FINANCE**
- **RECHERCHE FONDAMENTALE ET APPLIQUEE**
etc ...

SCIENCES DE L'ENVIRONNEMENT

- **Surveillance d'un milieu (qualité d'une eau,...)**
- **Estimation de l'effectif d'une population**
- **Toxicologie de l'environnement, analyse du risque**
- **Epidémiologie environnementale**
- **Prévision des effets du réchauffement climatique**
- **Phénomènes extrêmes (vagues, cyclones, ...)**
- **Prévision des épisodes de pollution (ozone)**
- **etc.**

Estimation du nombre de jours/an avec des températures $> 35^{\circ}\text{C}$ Période 2090-2099 (Météo-France, 2007)



La démarche de la modélisation statistique

- **La connaissance est dans les données ...**
 - C'est la démarche de la science expérimentale : modélisation empirique
 - Cette démarche suppose la stabilité des phénomènes : existence de lois
- Les problèmes portent le plus souvent sur des données issues d'usages humains : **les comportements sont instables**
- Le besoin existe surtout dans les phases d'instabilité !!

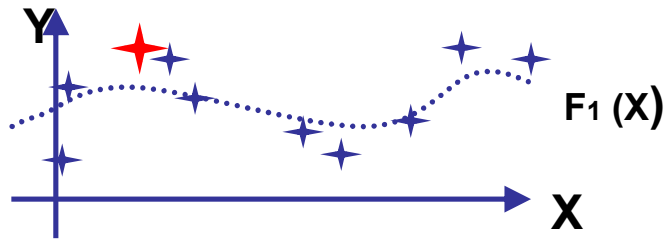
Techniques utilisées

- Analyse de données
- Probabilités / statistiques
- Théorie de l'information
- Intelligence artificielle
- Base de données
- Visualisation

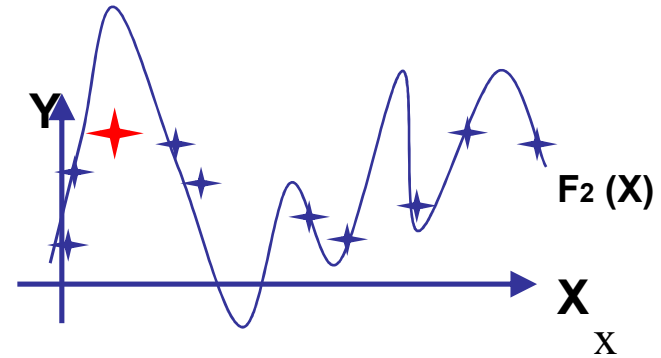
Techniques utilisées

- AD
 - Analyse en composantes principales
 - Analyse discriminante
 - Classification ascendante hiérarchique
- Intelligence artificielle
 - Réseaux neuronaux, réseaux bayésiens
- Probabilités, théorie de l'information
 - Arbres de décision, réseaux bayésiens
- Autres
 - Règles d'association
 - Filtrage

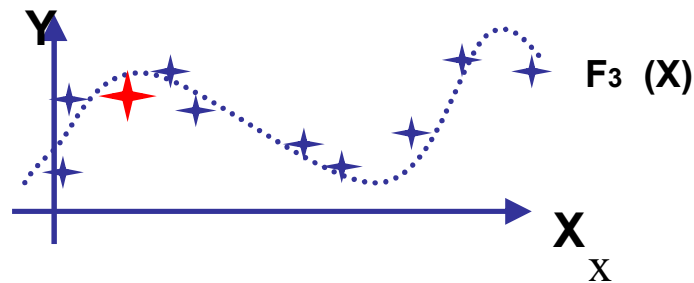
Challenge de la modélisation prédictive



Modèle très robuste



Modèle très précis



Tradeoff fit-robustesse

Challenge : construire à partir d'un échantillon un modèle qui soit aussi bon sur l'échantillon que sur de nouvelles données.

Evaluation des modèles

Comparaison Mesures / Modèles

Observation_i....Modélisation_i

O_i / M_i

Evaluation des modèles

Comparaison Mesures / Modèles

$$M_i = O_i + e$$

$$M_i \pm \Delta M_i = O_i \pm \Delta O_i + e \rightarrow$$

écart modèle/mesure

Erreur interne au modèle

Erreur analytique sur les mesures

Evaluation des modèles

Comparaison Mesures / Modèles

$$M_i = O_i + e$$

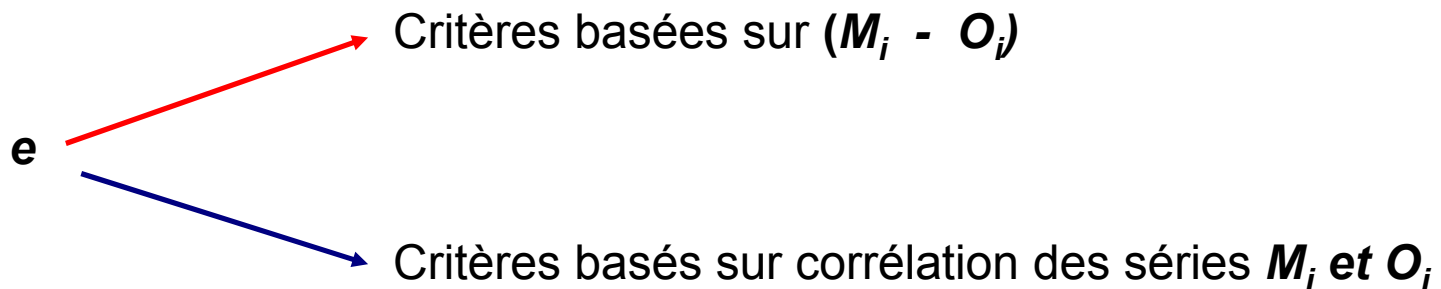
$$e = M_i - O_i$$

Evaluation des modèles

Comparaison Mesures / Modèles

$$M_i = O_i + e$$

$$e = M_i - O_i$$



1) Critères sur la différence:

- $e \rightarrow 0$

$$M_i \rightarrow O_i$$

comparer \bar{e} à 0

Test t Student

e Normalement distribuées

→ Test de Shapiro-Wilk,
Test X^2
Test de Kolmogorov-Smirnov

1) Critères sur la différence:

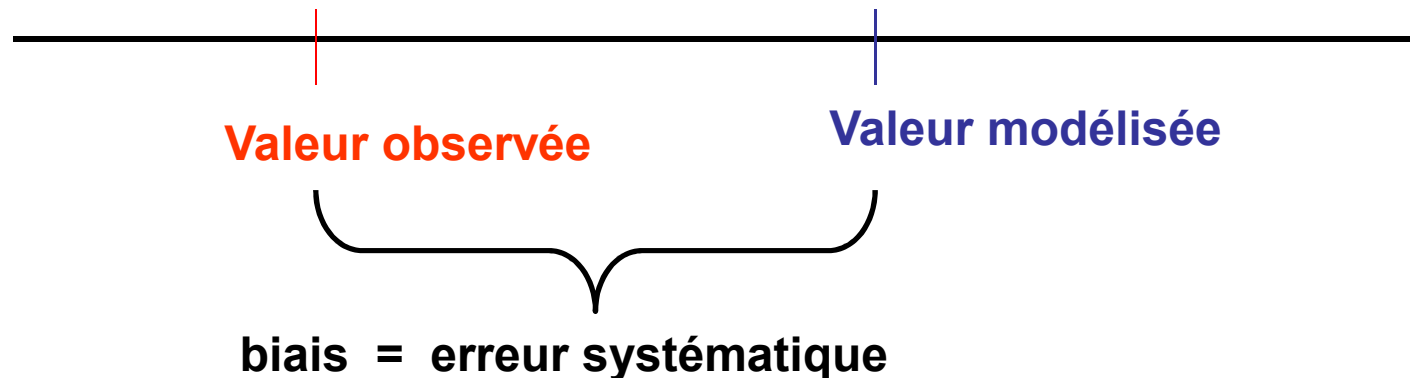
- Erreur Biais Moyen(*EBM*)

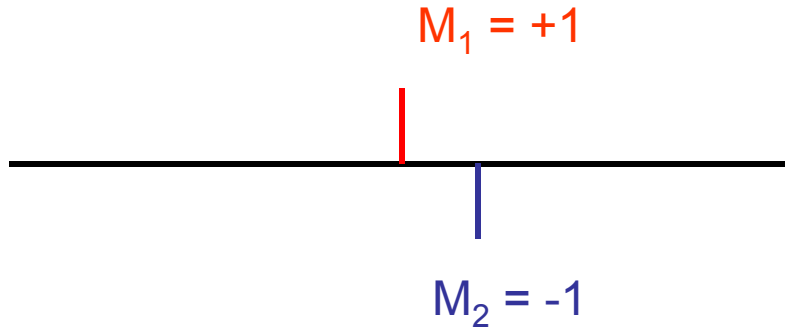
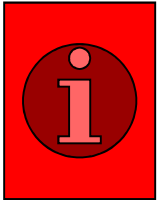
=

Mean Bias Error(*MBE*)

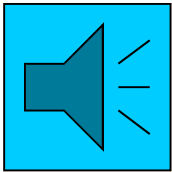
=

$$\bar{M} - \bar{O} = \frac{1}{n} \sum (M_i - O_i)$$





$$MBE = +1 - 1 = 0 \quad \longrightarrow \quad \text{compensation des erreurs}$$



MBE en unités de mesure

$$MNB = \frac{1}{n} \sum \left(\frac{M_i - O_i}{O_i} \right)$$

Mean Normalized Bias

Biais Moyen Normalisé

en %

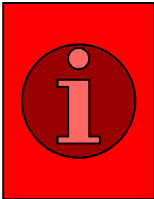
Erreur Moyenne Absolue (*EMA*)

=

Mean Absolute Error (*MAE*)

=

$$MAE = \frac{1}{n} \sum |M_i - O_i|$$

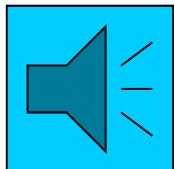


compensation des erreurs disparaît

critère plus exigeant



$$MAE \geq MBE$$



en % : $MANE = \frac{1}{n} \sum \left(\frac{|M_i - O_i|}{|O_i|} \right)$ **Erreur absolue normalisée moyenne**
Mean absolute normalized error

-Erreur quadratique

$$(M - O)^2$$

$$\frac{1}{n} \sum (M_i - O_i)^2 = \text{Erreur Quadratique Moyenne (EQM)}$$

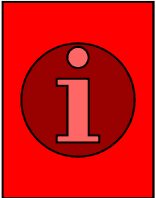
Mean Square Error (MSE)



Pénalise les fortes erreurs

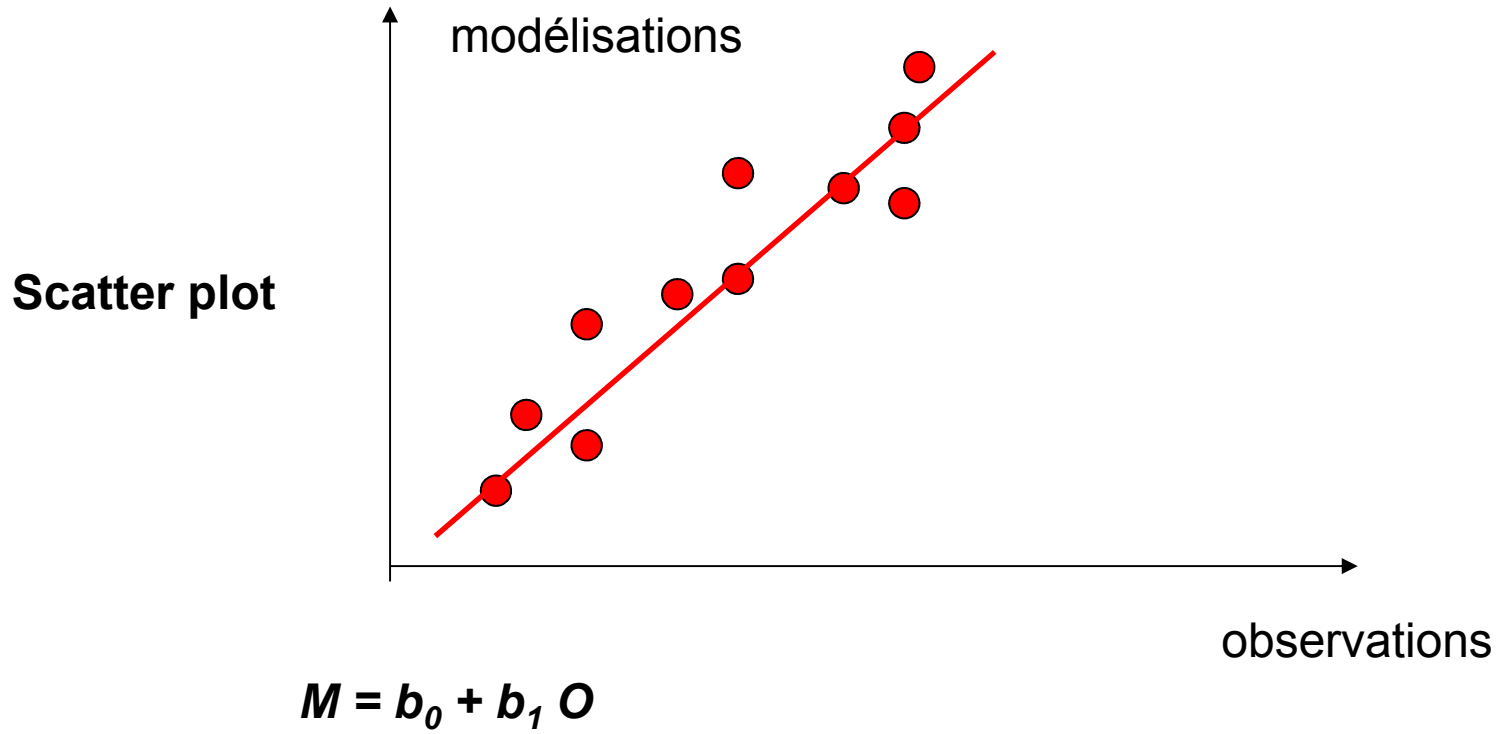
$$\sqrt{\frac{1}{n} \sum (M_i - O_i)^2} = \text{Racine Erreur Quadratique Moyenne (REQM)}$$

Root Mean Square Error (RMSE)

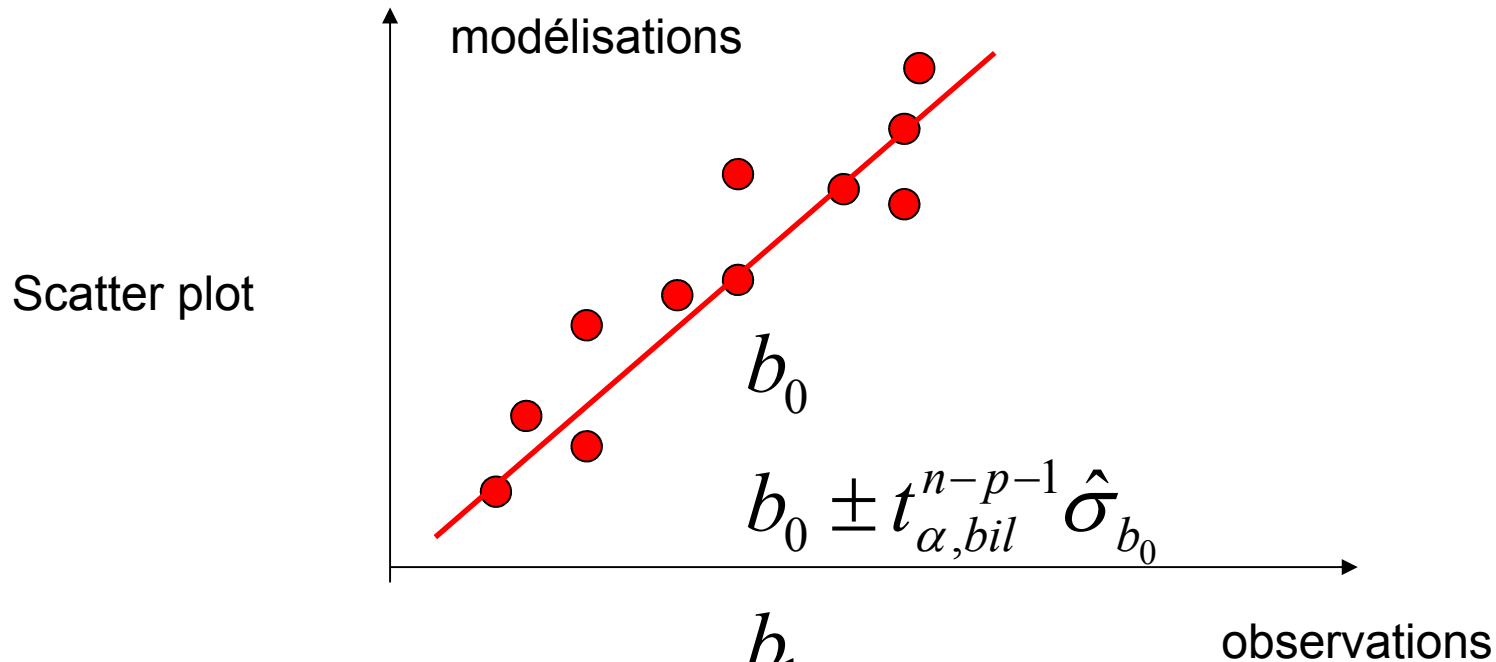


en unité de mesure

2) Critères sur les corrélations entre séries des observations et modélisations:



2) Critères sur les corrélations entre séries des observations et modélisations:



$$M = b_0 + b_1 O$$

$$b_1 \pm t_{\alpha, bil}^{n-p-1} \hat{\sigma}_{b_1}$$

$$\hat{\sigma}_{estimation} = \hat{\sigma}_{\hat{y}}$$

$$R^2$$

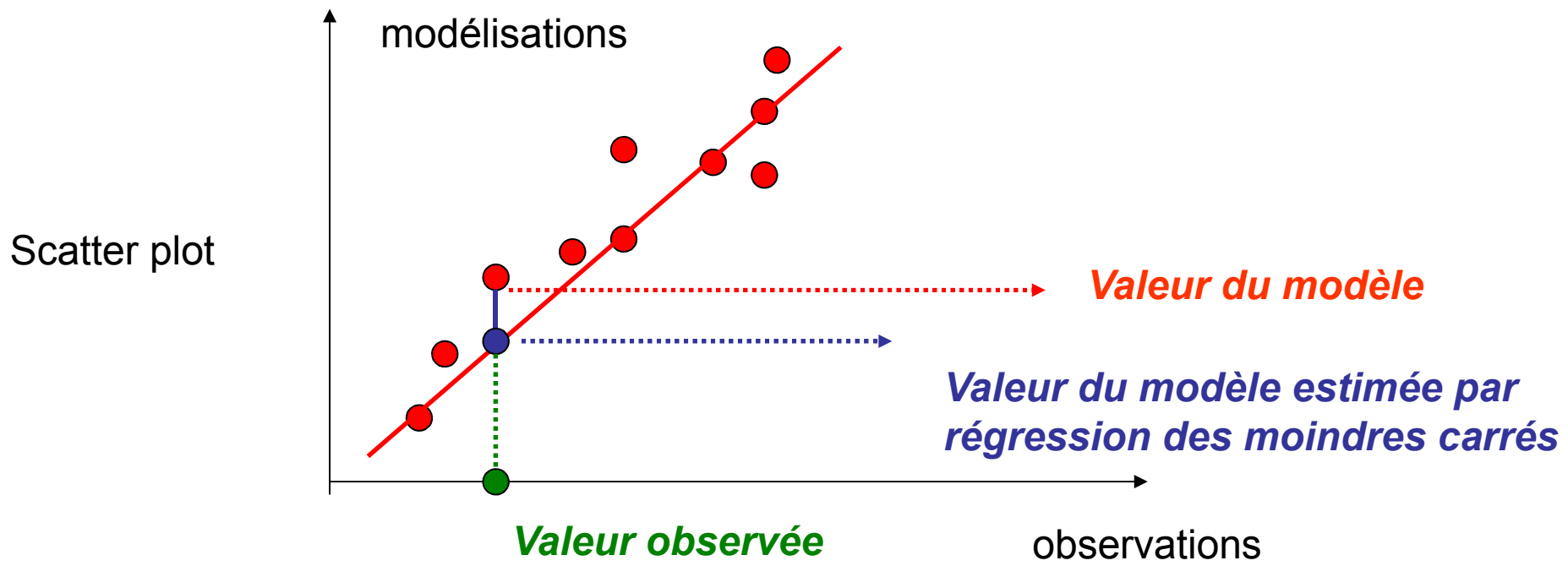
$$F$$

$$b_1 = 1$$

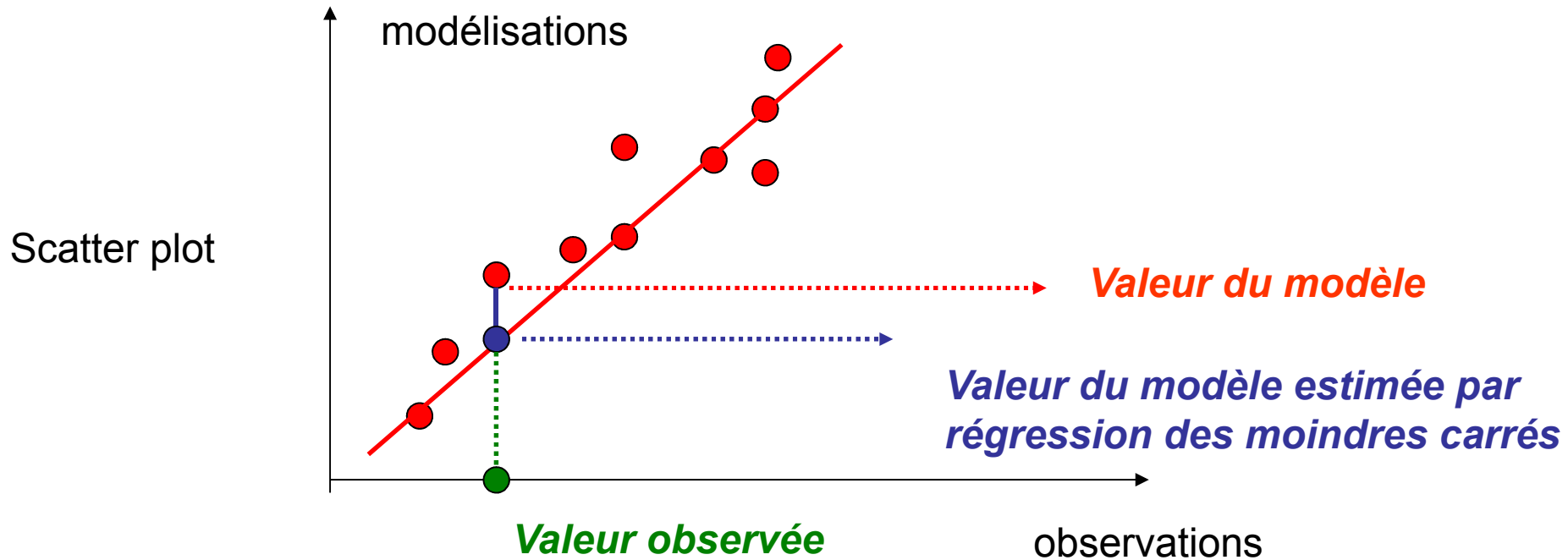
$$b_1 \in \left[1 - t\hat{\sigma}_{b_1}; 1 + t\hat{\sigma}_{b_1} \right]$$

$$b_0 = 0$$

$$b_0 \in \left[-t\hat{\sigma}_{b_0}; t\hat{\sigma}_{b_0} \right]$$



$$M = b_0 + b_1 O$$



O valeur observée

M valeur modèle

\hat{M} valeur rég MCO

$$\sqrt{\frac{1}{n} \sum (M_i - O_i)^2} = \text{RMSE} = \text{RMSE}_s + \text{RMSE}_\varepsilon$$

erreurs aléatoires (bruit)

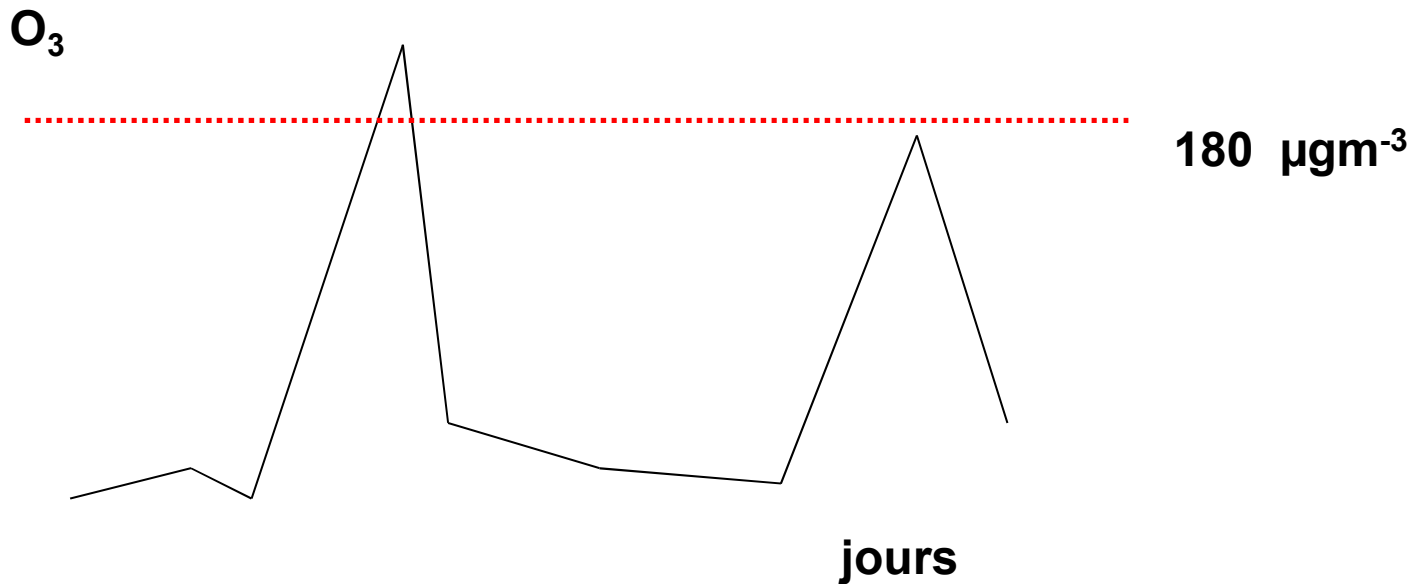
RMSE systématique = biais du modèle

$$\text{RMSE}_s = \sqrt{\frac{1}{n} \sum (\hat{M}_i - O_i)^2} \rightarrow \text{RMSE}$$

biais fonctionnel
biais linéaire

$$\text{RMSE}_\varepsilon = \sqrt{\frac{1}{n} \sum (M_i - \hat{M}_i)^2} \rightarrow 0$$

Critères d'évaluation de dépassement de seuil



dépassements correctement modélisé = A

dépassements réels totaux = M

dépassements prédits totaux = F

nombre total de prévisions = n

Tableau de contingence

Réel → Prédiction (modèle) ↓	dépassement	non dépas.
dépassement	A dépas. bien modélisés	Y fausses alertes
non dépas.	Z dépas. non prévus	U non dépas. bien prévus

F

A+Y=F

A+Z=M

M

Total=n

$$\text{taux de bonne prédiction} = TPR = \frac{A}{M}$$

$$\text{taux fausse alerte} = FAR = \frac{F - A}{n - M}$$

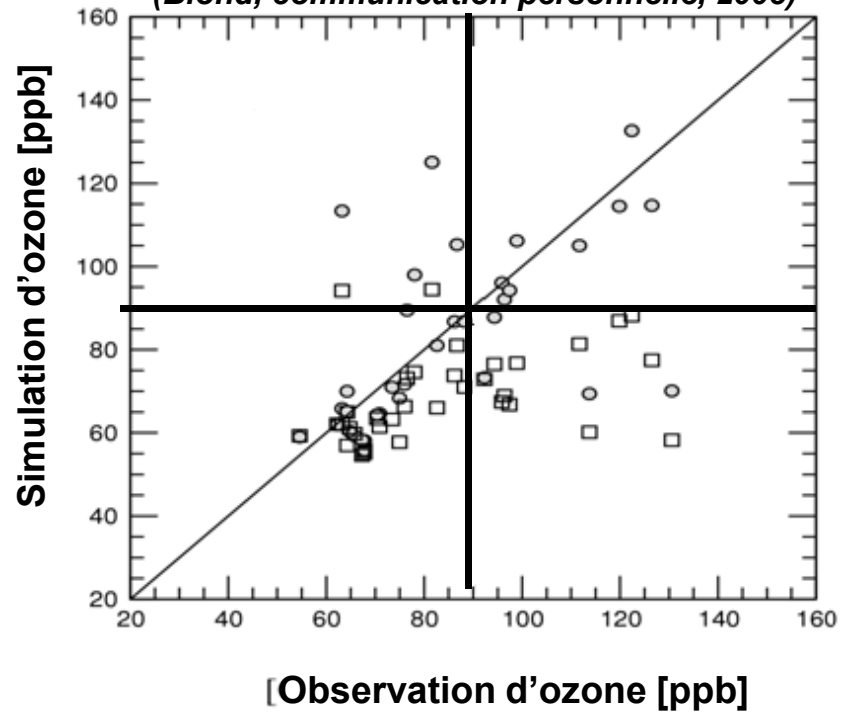
$$\text{Index de Succès} = SI = TPR - FAR$$

- Indice d'agrément:

$$\text{index of agreement} = 1 - \frac{\sum (M_i - O_i)^2}{\sum (|M_i - \bar{O}| + |O_i - \bar{O}|)^2} \rightarrow 1$$

90 ppb

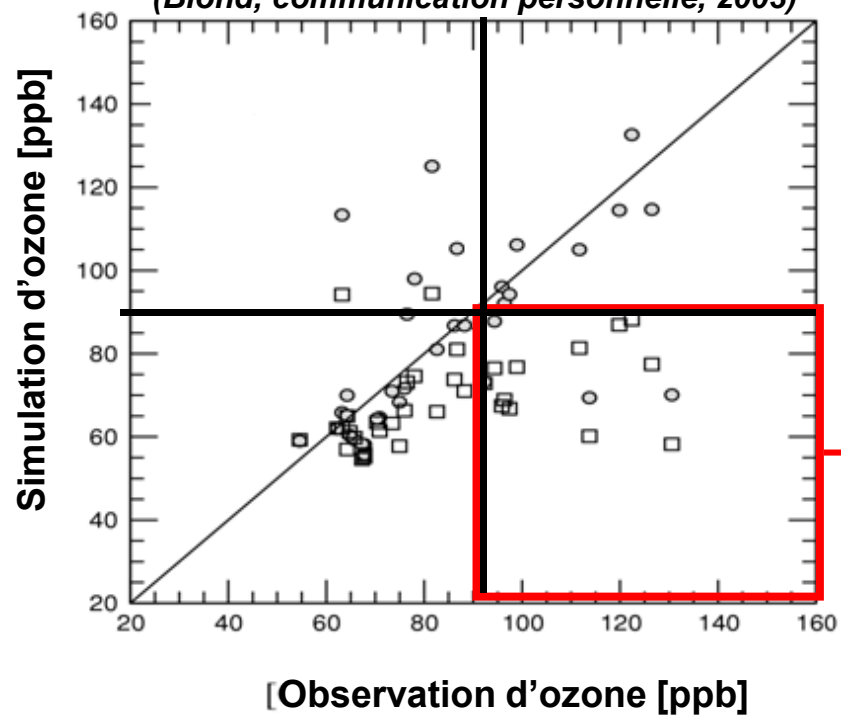
(Blond, communication personnelle, 2003)



90 ppb

90 ppb

(Blond, communication personnelle, 2003)



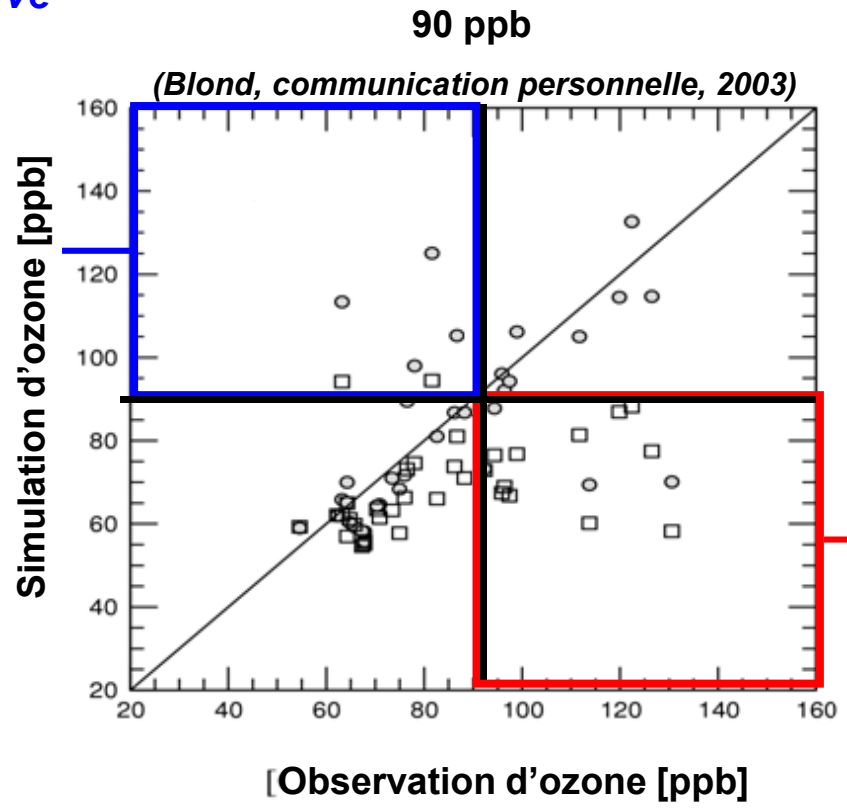
90 ppb

Observé mais mal prévu

Risques sanitaires

Surcoût

Prévu mais pas observé



90 ppb

Observé mais mal prévu

Risques sanitaires

$$Y = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p$$

$$Y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_p x_p$$

Variables dont dépend Y

=

Régresseurs de la fonction

$$Y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_p x_p$$

Coefficients de régression

$$Y = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p$$

**Fonction de régression multiple
de Y par p variables:**

x_1, x_2, \dots, x_p

$$Y = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p$$

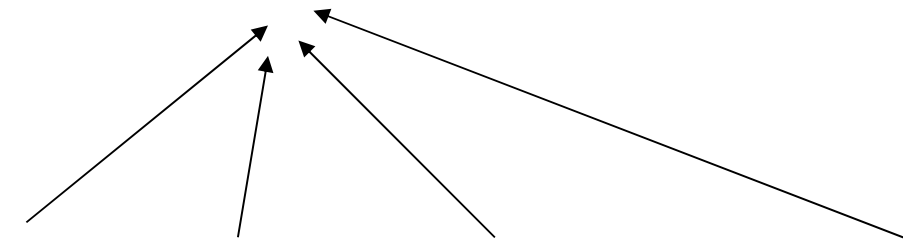
**Fonction de régression multiple
de Y par p variables:**

x_1, x_2, \dots, x_p

=

**Modèle statistique
de prévision de Y
par x_1, x_2, \dots, x_p**

$P+1$ coefficients à déterminer



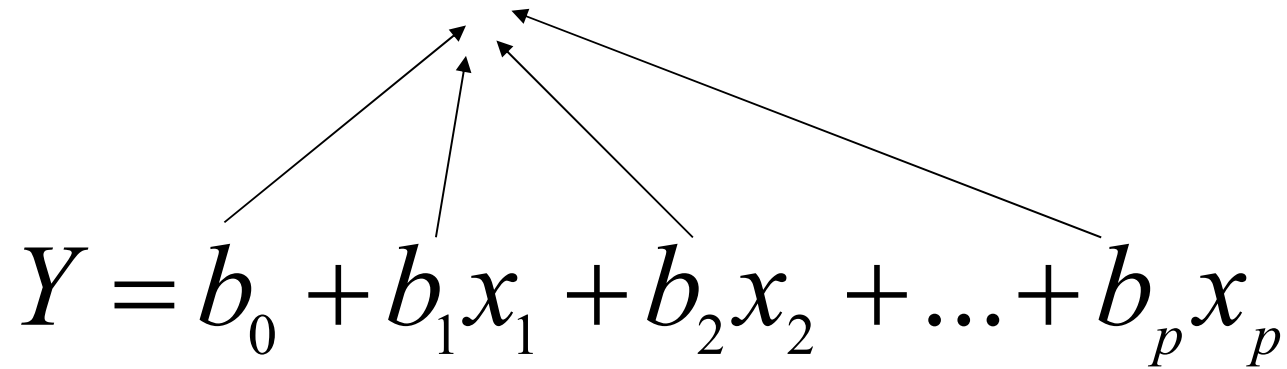
$$Y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_p x_p$$

Fonction de régression linéaire multiple

=

Mesure les relations linéaires

$P+1$ coefficients à déterminer



The diagram shows the equation $Y = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p$. Four arrows originate from the coefficients b_0 , b_1 , b_2 , and b_p and point towards the text " $P+1$ coefficients à déterminer" located above the equation.

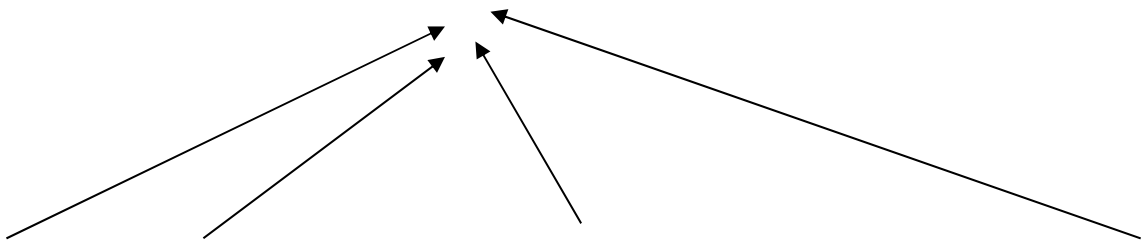
$$Y = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p$$

Fonction de régression multiple

Linéaire en ses coefficients

Et linéaire en x_1, x_2, \dots, x_p

$P+1$ coefficients à déterminer


$$Y = b_0 + b_1 \sin x_1 + b_2 \operatorname{Ln} x_2 + \dots + b_p \exp^{x_p}$$

Fonction de régression multiple

Linéaire en ses coefficients

Et non linéaire en x_1, x_2, \dots, x_p

$$Y = b_0 + b_1 \text{Ln}(b_2 x_1) + b_3 \sin(b_4 x_2) + \dots$$

Fonction de régression multiple

Non linéaire en ses coefficients

Et non linéaire en x_1, x_2, \dots, x_p

**Fonction régression
linéaire en ses coefficients**

**fonction régression
non linéaire en ses coefficients**

Détermination des valeurs b_1, b_2, \dots, b_p

Par méthode des moindres carrés ordinaires

Méthode d'estimation itérative

Matrice connue de Y en fonction de x_1, x_2, \dots, x_p

	Y	x_1	x_2	...	x_p
Mes.1					
Mes. 2					
...					
Mes. n					

Détermination des valeurs b_1, b_2, \dots, b_p

Matrice connue de Y en fonction de x_1, x_2, \dots, x_p

	Y	x_1	x_2	...	x_p
Mes.1					
Mes. 2	Valeurs réelles observées				
...					
Mes. n					

Détermination des valeurs b_1, b_2, \dots, b_p

Estimation des coefficients de régression

si:

1. $n < \text{nombre de variables } p \Rightarrow$ pas de solution

Estimation des coefficients de régression

quelque soit la régression si:

1. $n < \text{nombre de variables } p \Rightarrow$ pas de solution
2. $n = p \Rightarrow$ 1 solution unique, *illusoire*

Estimation des coefficients de régression

quelque soit la régression si:

1. $n < \text{nombre de variables } p \Rightarrow$ pas de solution
2. $n = p \Rightarrow$ 1 solution unique, *illusoire*
3. $n \gg p \Rightarrow$ plusieurs solutions

Estimation des coefficients de régression

quelque soit la régression si:

1. $n < \text{nombre de variables } p \Rightarrow$ pas de solution
2. $n = p \Rightarrow$ 1 solution unique, *illusoire*
3. $n > p \Rightarrow$ plusieurs solutions
 \Rightarrow la meilleure trouvée par minimisation
fonction d'erreur:

(valeur observée – valeur du modèle)

Estimation des coefficients de régression

quelque soit la régression si:

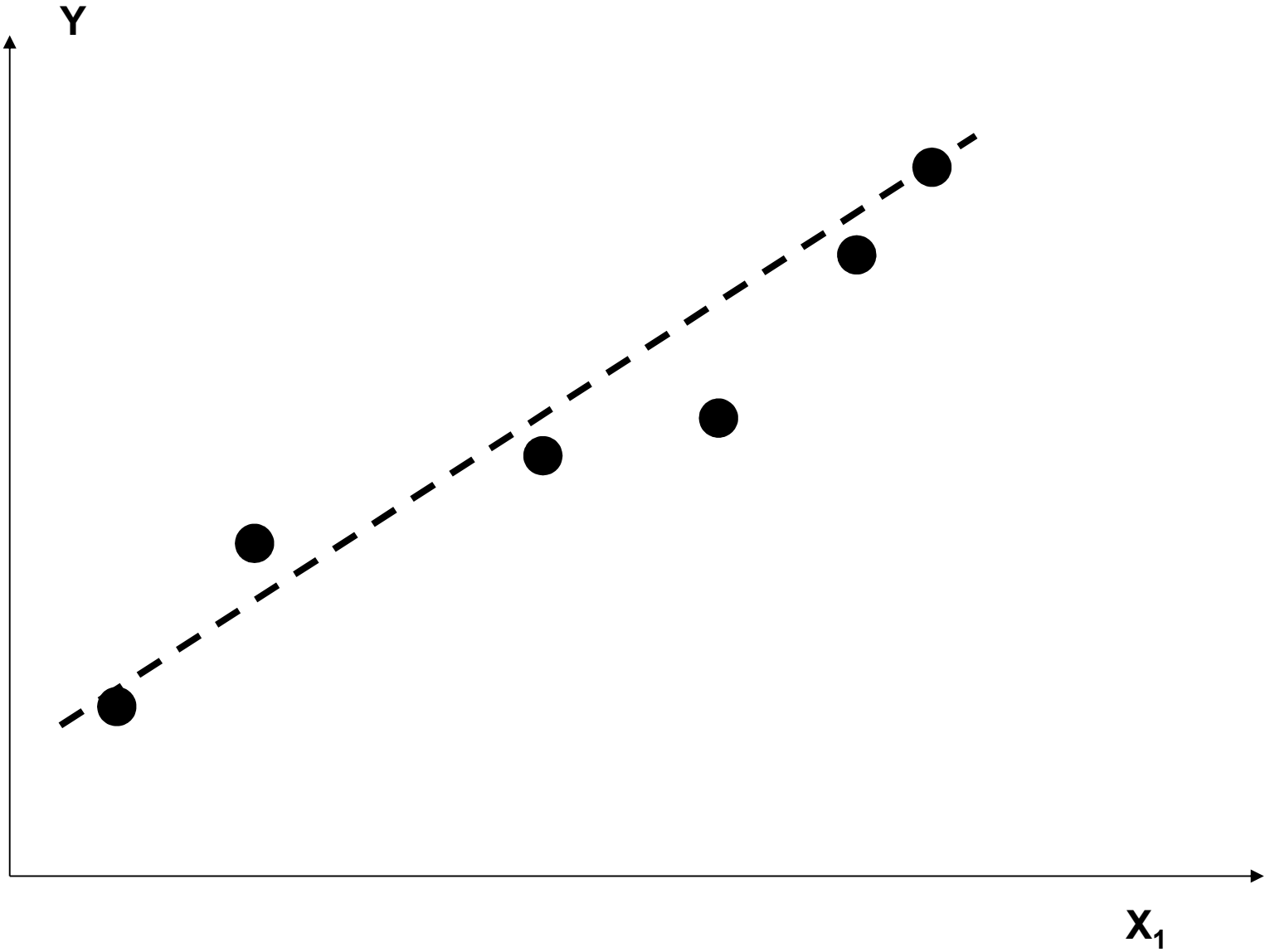
1. $n < \text{nombre de variables } p \Rightarrow$ pas de solution
2. $n = p \Rightarrow$ 1 solution unique, *illusoire*
3. $n > p \Rightarrow$ plusieurs solutions
 \Rightarrow la meilleure trouvée par minimisation
fonction d'erreur Δ :

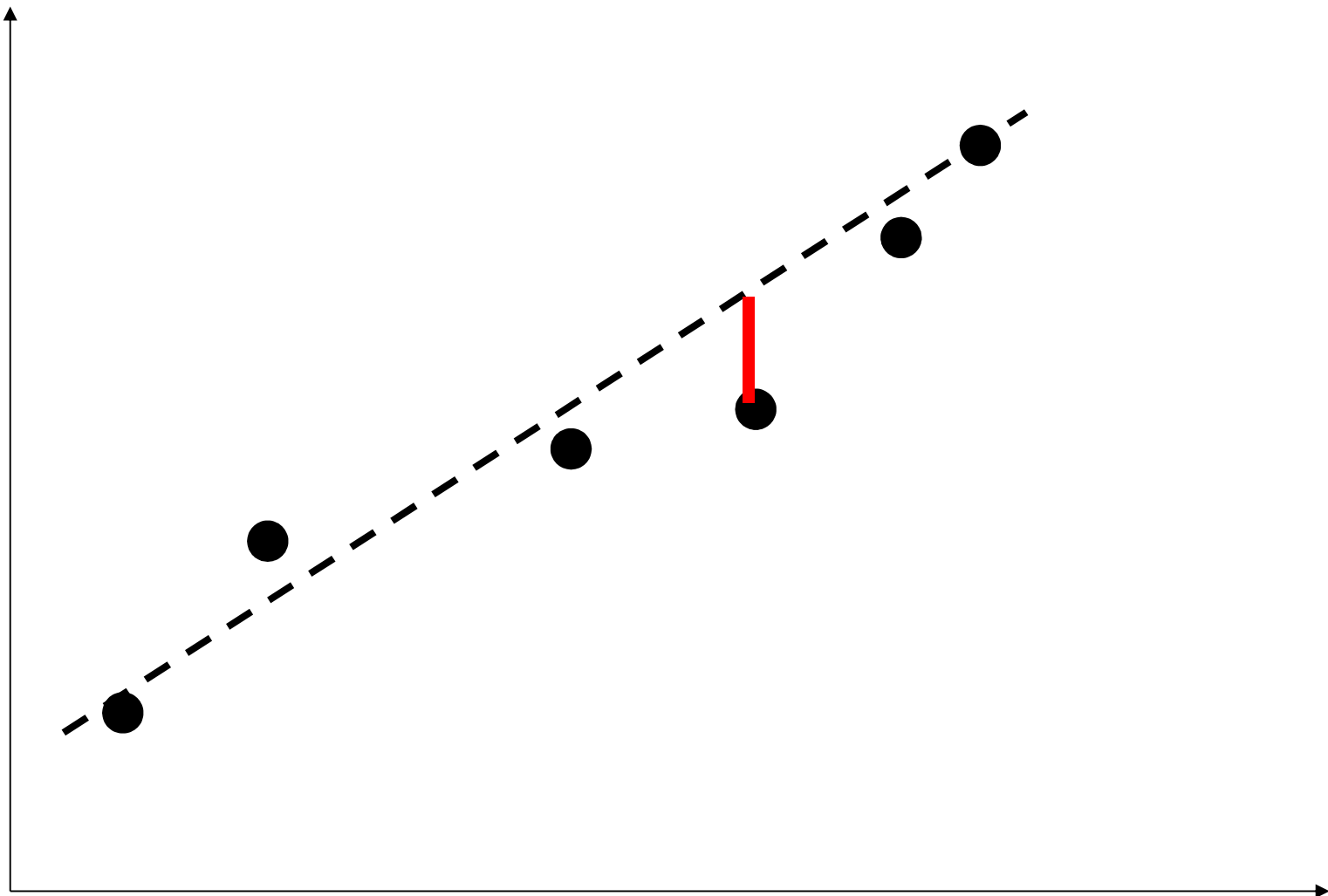
$$\Delta = (\text{valeur observée} - \text{valeur du modèle})$$

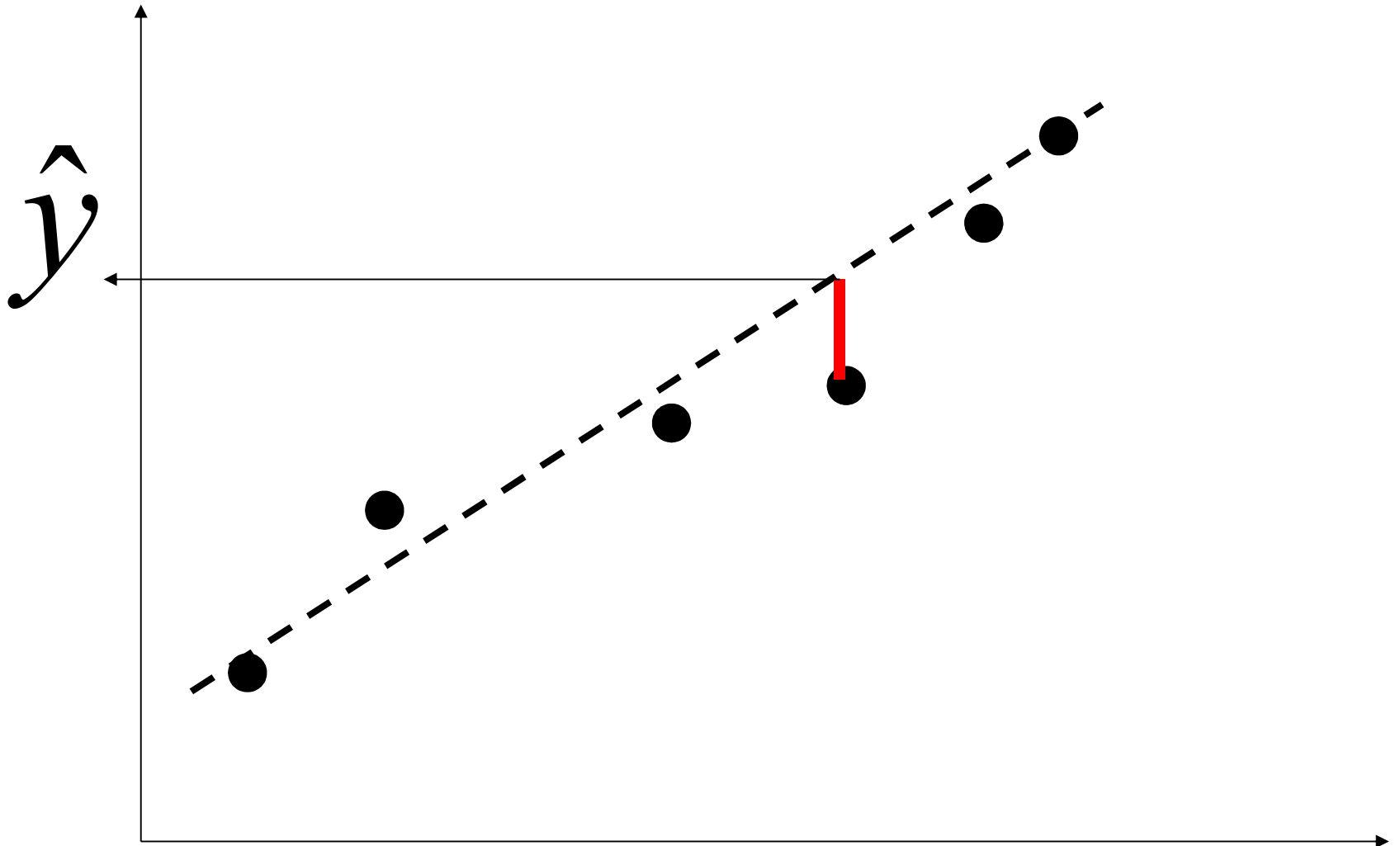
$$y = f(x_1, x_2 \dots x_p)$$

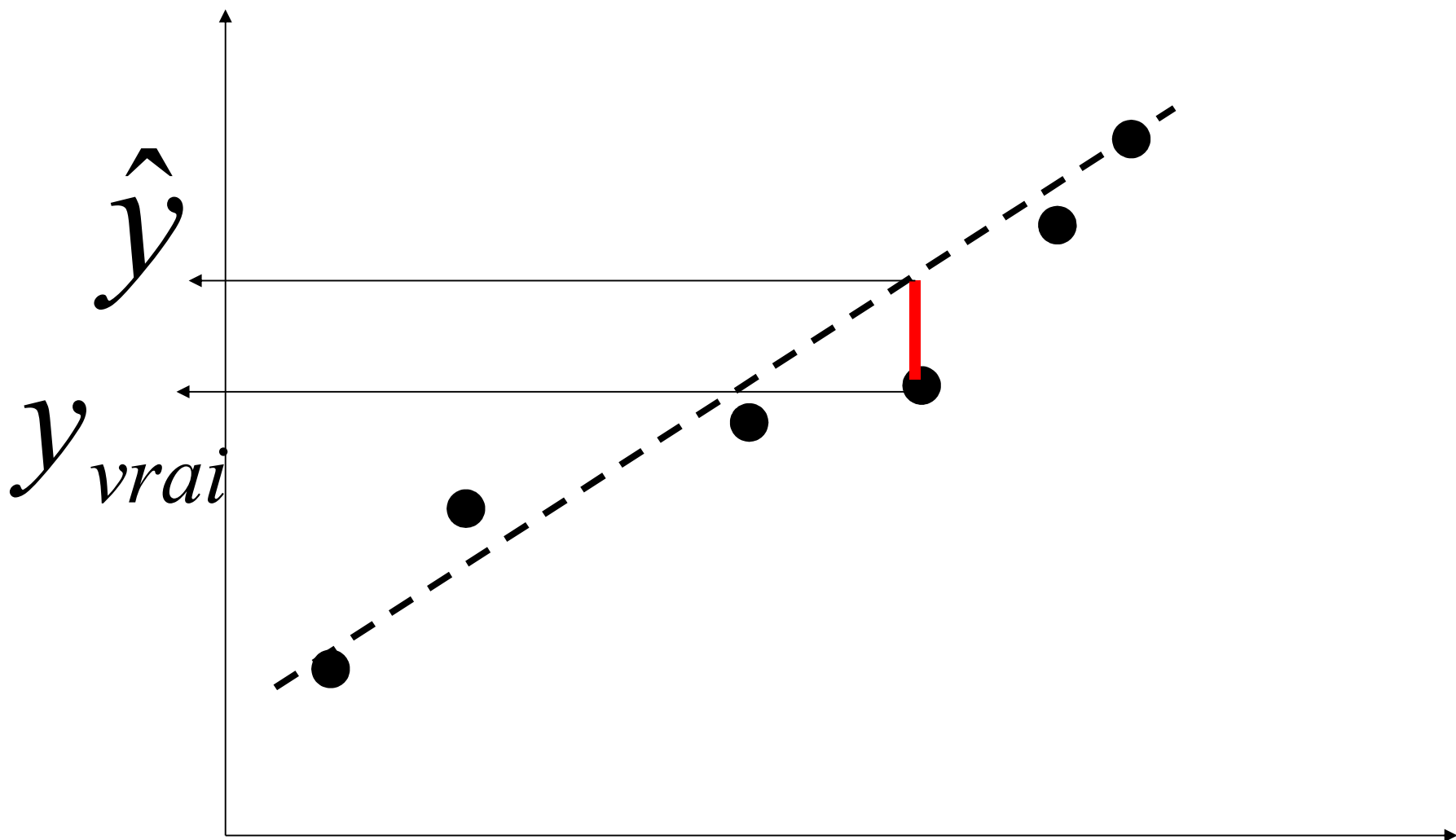
$$y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_p x_p$$

**Méthode d'estimation
par les moindres carrés ordinaires**







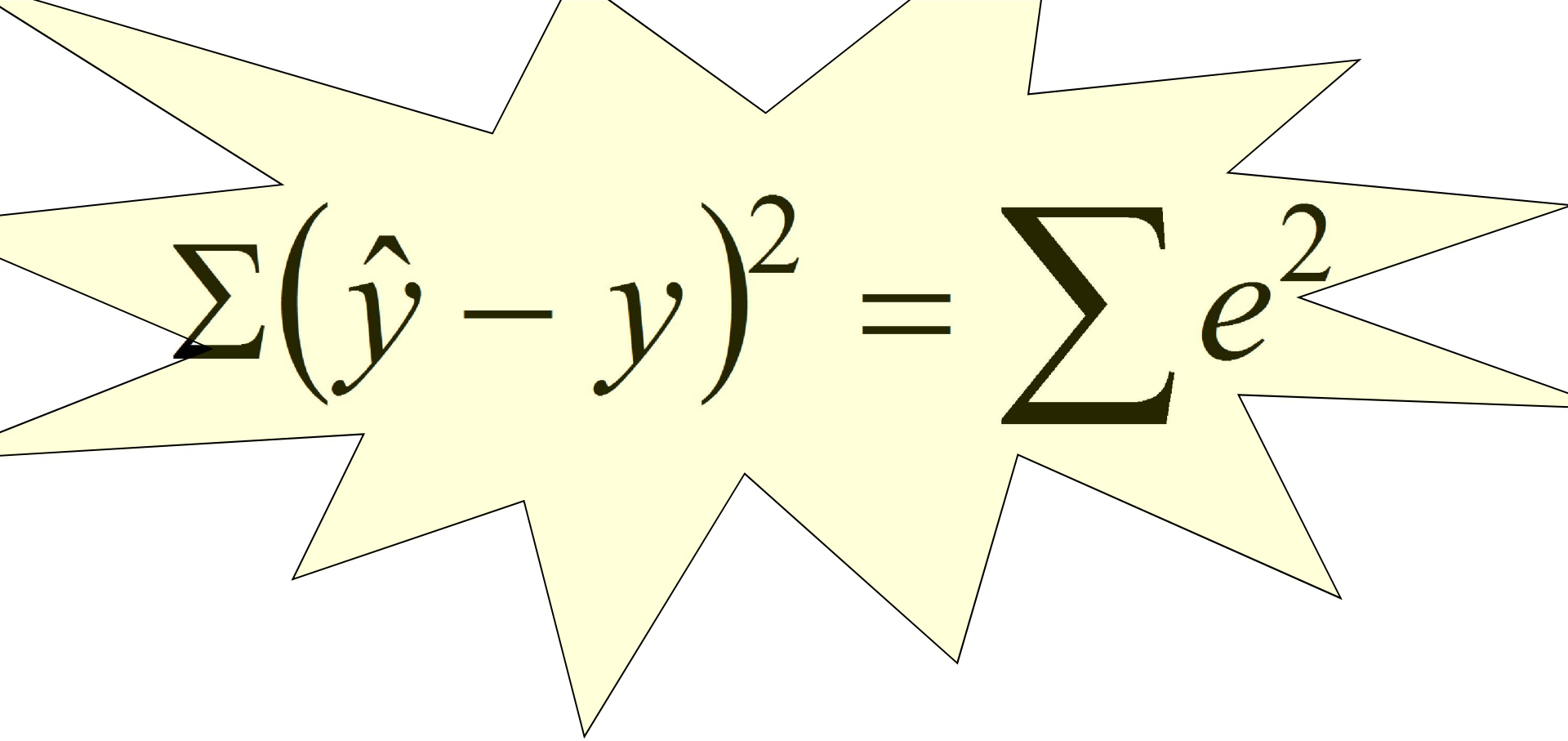


$$(\hat{y} - y) = e$$

$$(\hat{y} - y) = e$$

e = résidus

$$(\hat{y} - y)^2 = e^2$$


$$\sum (\hat{y} - y)^2 = \sum e^2$$

Minimisation des moindres carrés ordinaires

$$\Sigma \left(b_0 + b_1 x_1 + b_2 x_2 + \dots + b_p x_p - y_{vrai} \right)^2$$

Minimisation des moindres carrés ordinaires

Pour $p=1$

Ceci s'obtient en résolvant les dérivées de la somme des carrés des résidus par rapport aux coefficients cherchés:

$$\frac{\partial \sum (e)^2}{\partial b_0} = -2 \sum (y - b_0 - b_1 x) = 0$$

$$\frac{\partial \sum (e)^2}{\partial b_1} = -2 \sum [x(y - b_0 - b_1 x)] = 0$$

En arrangeant, on appelle
équations normales des moindres carrés ordinaires
les expressions suivantes:

$$\sum_{i=1}^n y_i = nb_0 + b_1 \sum_{i=1}^n x_i$$

$$\sum_{i=1}^n x_i y_i = b_0 \sum_{i=1}^n x_i + b_1 \sum_{i=1}^n x_i^2$$

estimation des coefficients b_0 et b_1 :

$$b_1 = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

Ou:

**La résolution de ces équations normales
fournit l'estimation des coefficients b_0 et b_1 :**

$$b_1 = R \frac{\hat{\sigma}_Y}{\hat{\sigma}_X}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

Estimation erreur-type associée

$$\hat{\sigma}_{b_0}$$

$$\hat{\sigma}_{b_1}$$

Erreur-type (ou incertitude-type) des coefficients b_0 et b_1

$$\hat{\sigma}_{b_0} = \frac{\hat{\sigma}_y \sqrt{(1 - R^2) \sum x_i^2}}{\hat{\sigma}_x \sqrt{n(n - 2)}}$$

Erreur-type (ou incertitude-type) des coefficients b_0 et b_1

$$\hat{\sigma}_{b_0} = \frac{\hat{\sigma}_y \sqrt{(1 - R^2) \sum x_i^2}}{\hat{\sigma}_x \sqrt{n(n - 2)}}$$

$$\hat{\sigma}_{b_1} = \frac{\hat{\sigma}_y \sqrt{(1 - R^2)}}{\hat{\sigma}_x \sqrt{(n - 2)}} = \frac{\hat{\sigma}_{\hat{Y}}}{\sqrt{\sum (X_i - \bar{X})^2}}$$

Intervalles de confiance à $\alpha\%$ près :

Intervalles de confiance à $\alpha\%$ près :

$$b_0 \pm t_{bil,\alpha}^{n-2} \hat{\sigma}_{b_0}$$

$$b_1 \pm t_{bil,\alpha}^{n-2} \hat{\sigma}_{b_1}$$

$H_0 : b_1 = a$

$H_1 : b_1 \neq a$

Si:

$$a - t_{\alpha, bil}^{n-2} \hat{\sigma}_{b_1} \leq b_1 \leq a + t_{\alpha, bil}^{n-2} \hat{\sigma}_{b_1}$$

H_0 acceptée

$H_0 : b_0 = a$

$H_1 : b_0 \neq a$

Si:

$$a - t_{\alpha, bil}^{n-2} \hat{\sigma}_{b_0} \leq b_0 \leq a + t_{\alpha, bil}^{n-2} \hat{\sigma}_{b_0}$$

H_0 acceptée

Erreur-type (incertitude-type) d'estimation du modèle

=

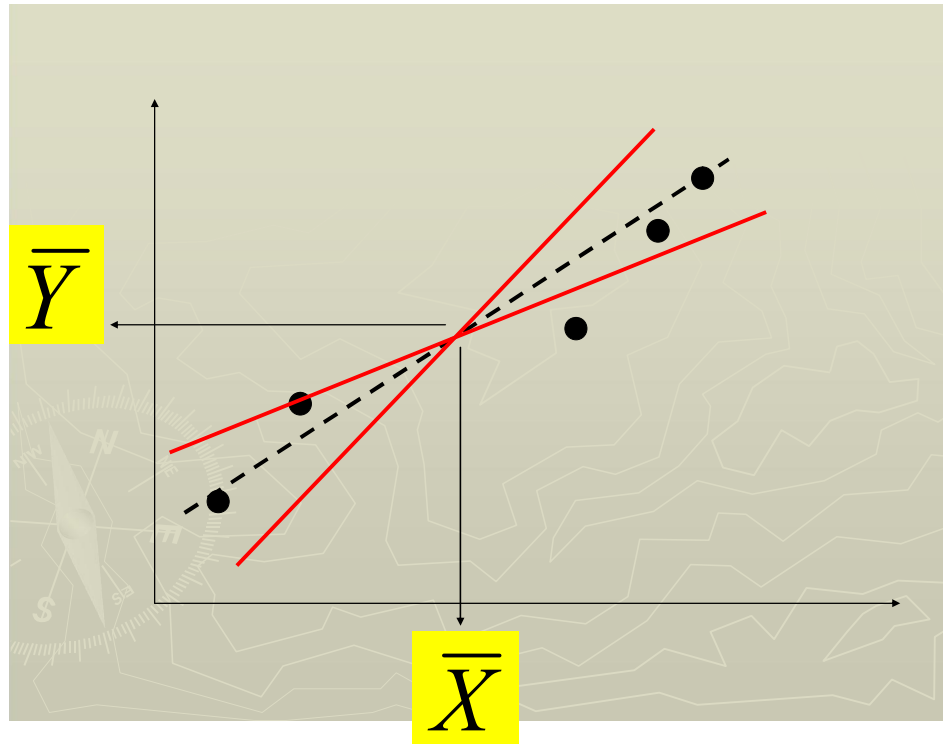
$$\hat{\sigma}_{estimation}$$

=

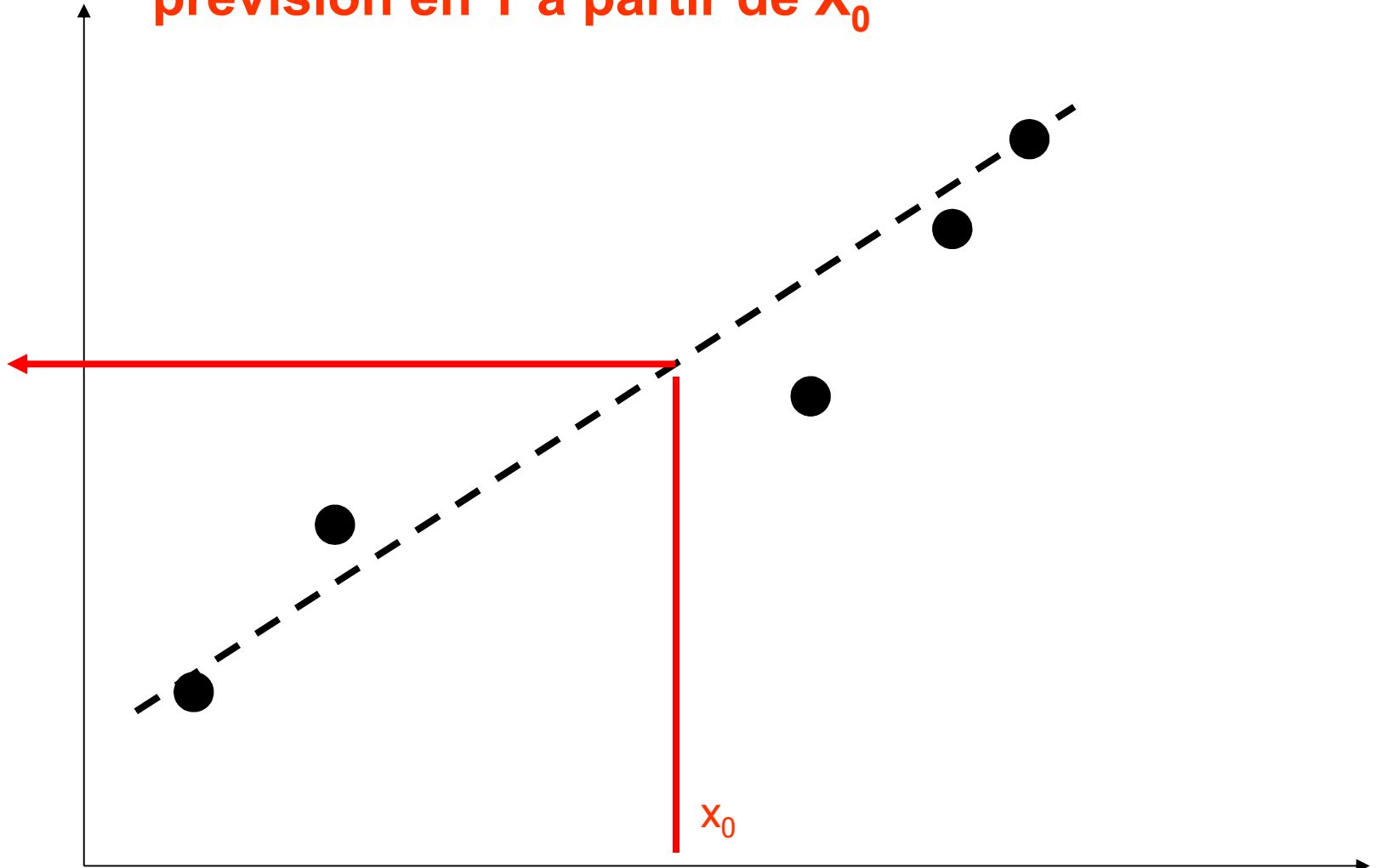
$$\hat{\sigma}_{\hat{y}} = \sqrt{\frac{\sum (y - \hat{y})^2}{n - 2}}$$

$$R = b_1 \frac{\sigma_x}{\sigma_y}$$

Coefficient de corrélation de Pearson

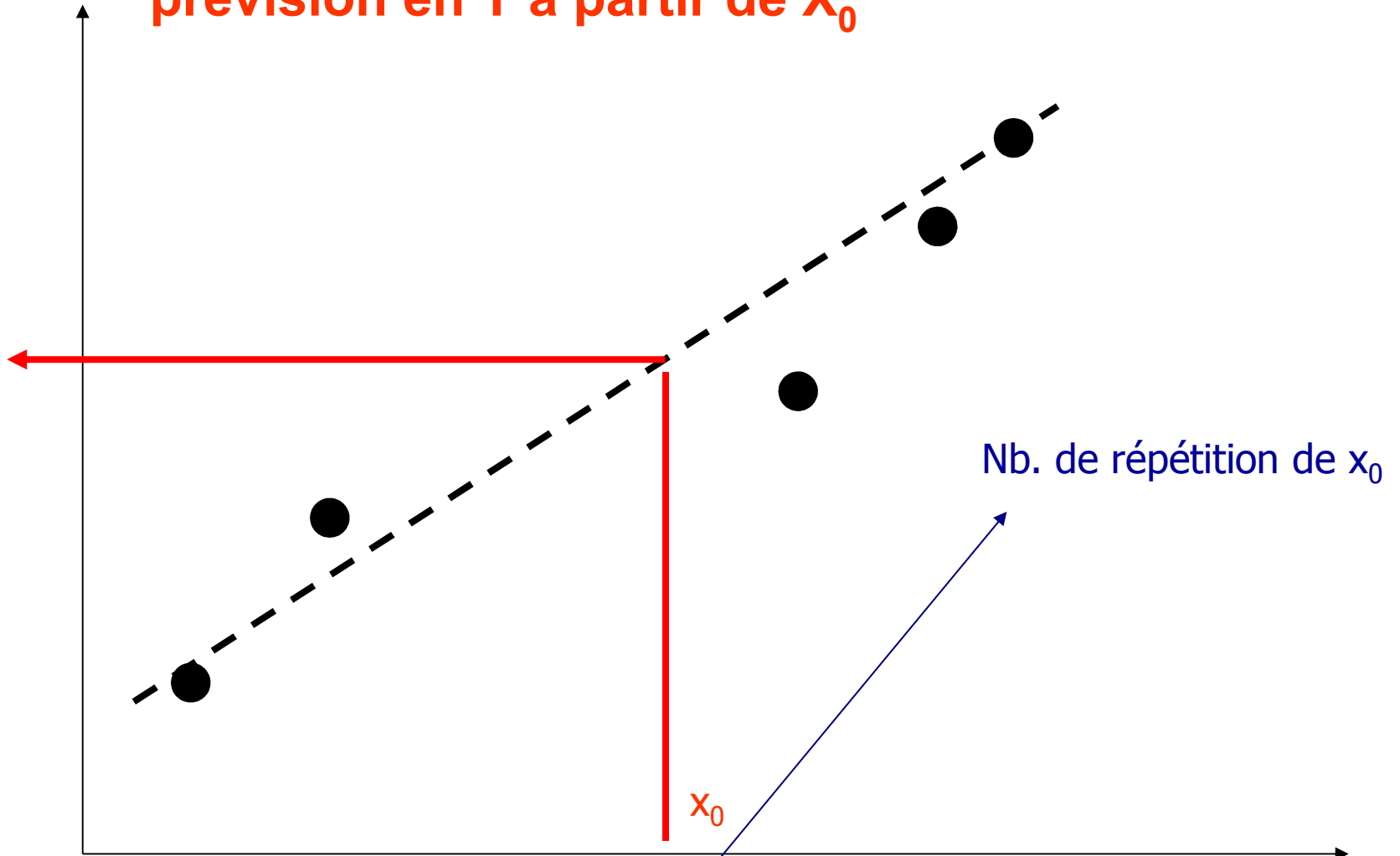


prévision en Y à partir de X_0



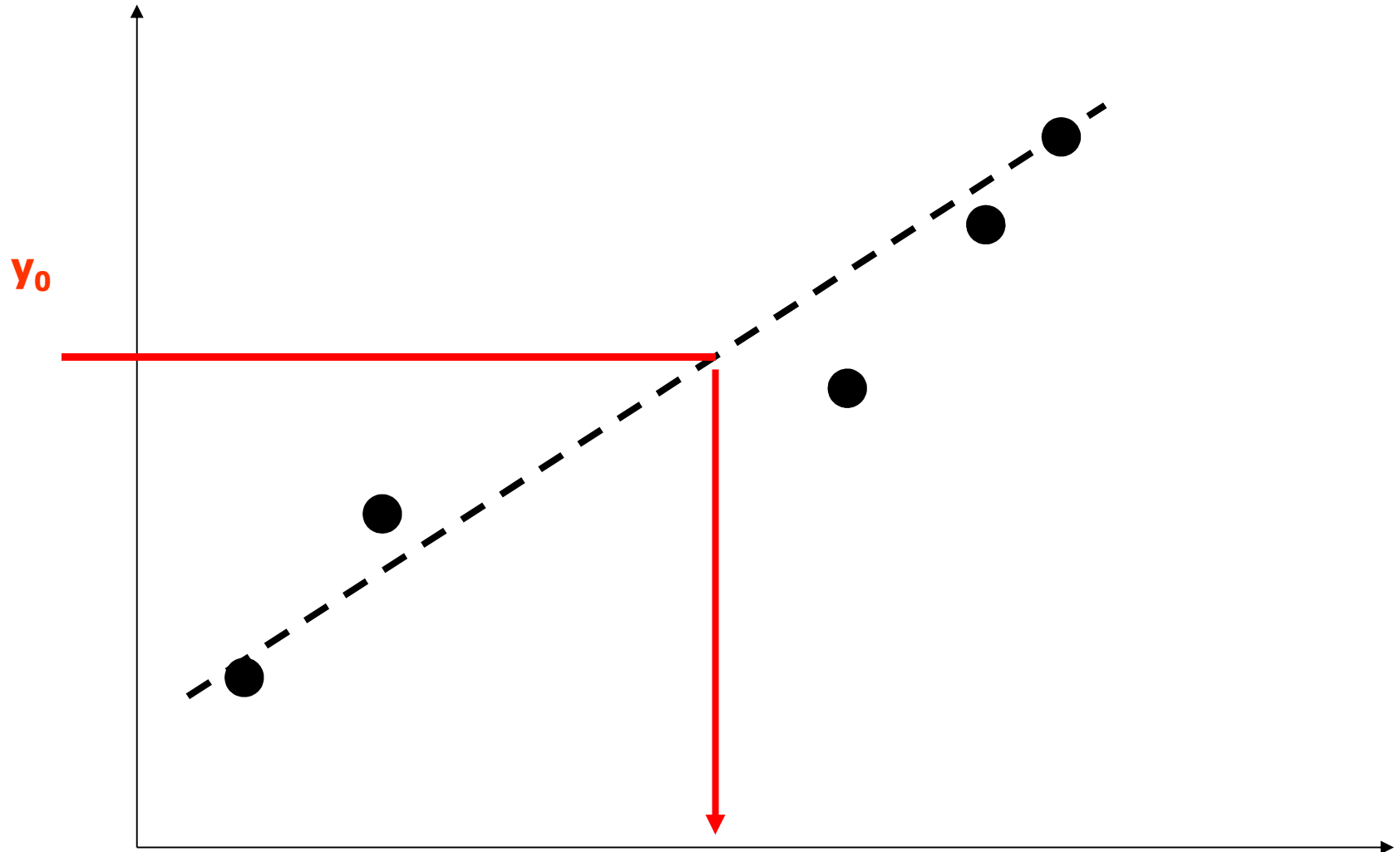
$$\hat{Y} = b_0 + b_1 X \pm t^{n-p-1} \hat{\sigma}_{estimation} \sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum (x_i - \bar{x})^2}}$$

prévision en Y à partir de X_0

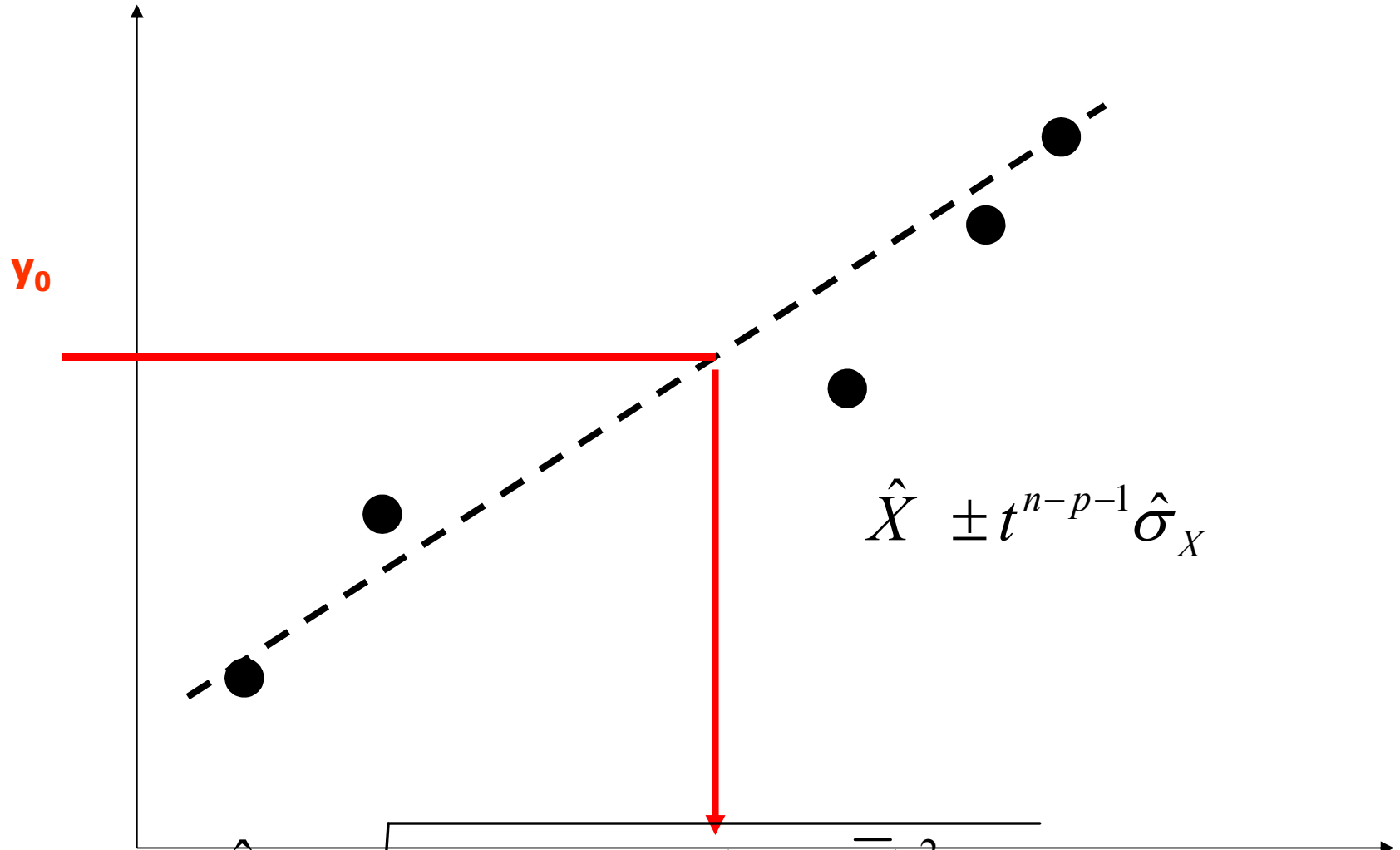


$$\hat{Y} = b_0 + b_1 X \pm t^{n-p-1} \hat{\sigma}_{estimation} \sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum (x_i - \bar{x})^2}}$$

prévision en X à partir de y_0



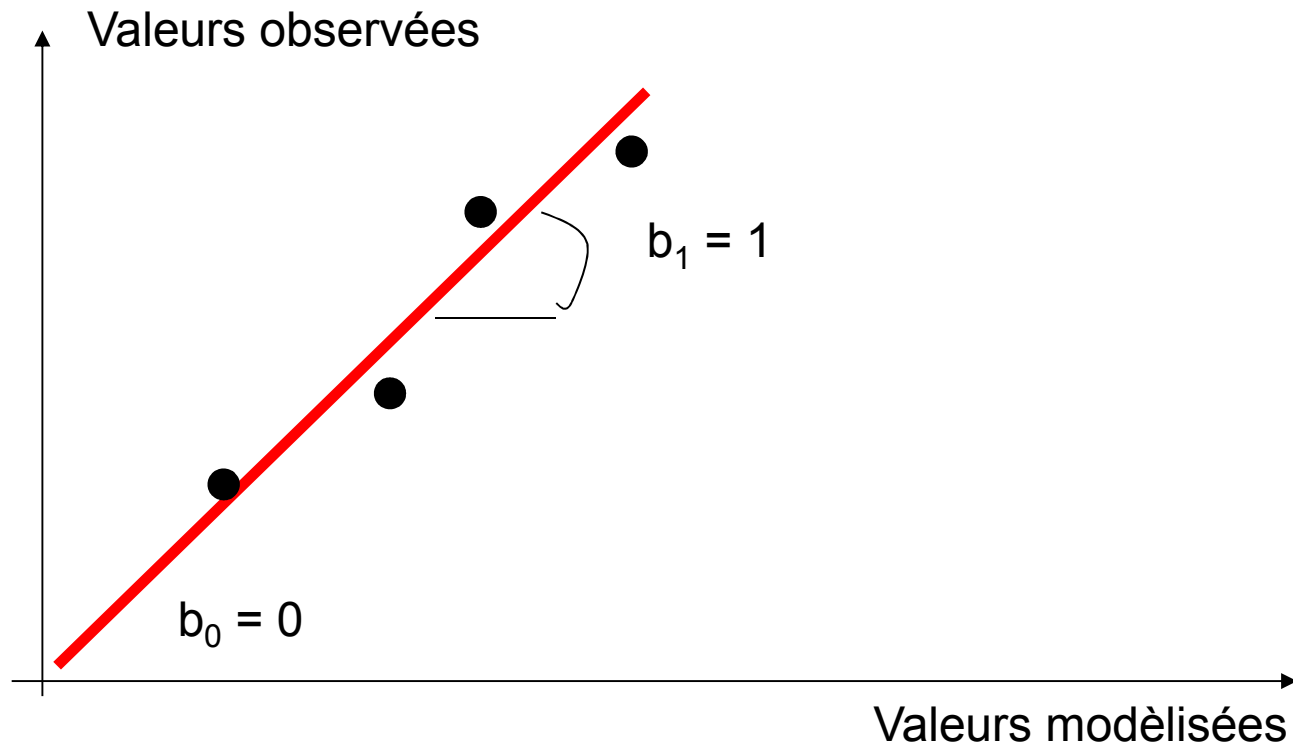
prévision en X à partir de y_0



$$\hat{\sigma}_X = \frac{\hat{\sigma}_{\hat{y}}}{b_1} \times \sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(Y_0 - \bar{Y})^2}{b_1^2 \sum_i (X_i - \bar{X})^2}}$$

Qualité statistique de la régression:

$R^2 \rightarrow 1$ coefficient de détermination



Qualité statistique de la régression:

$R^2 \rightarrow 1$ coefficient de détermination

$$R^2 = \frac{\text{variance expliquée par le modèle}}{\text{variance totale}}$$

$R^2 \rightarrow 1$ coefficient de détermination

Test R^2 :

$H_0 : R^2=0$ reg. Non significativ

$H_1 : R^2=1$ rég. Significative

Si:

$$-t_{\alpha, bil}^{n-p-1} \leq \frac{|R| \sqrt{n-p-1}}{\sqrt{1-R^2}} \leq +t$$

Nb. de points de la régression

$R^2 \rightarrow 1$ coefficient de détermination

Test R^2 :

$H_0 : R^2=0$ reg. Non significativ

$H_1 : R^2=1$ rég. Significative

Si:

$$-t_{\alpha, bil}^{n-p-1} \leq \frac{|R| \sqrt{n-p-1}}{\sqrt{1-R^2}} \leq +t$$

il n' y a pas de raison de rejeter H_0

Intervalles de confiance à $\alpha\%$ près, des coefficients b_i :

$$b_0 \pm t \hat{\sigma}_{b_0}$$

$$b_i \pm t \hat{\sigma}_{b_i}$$

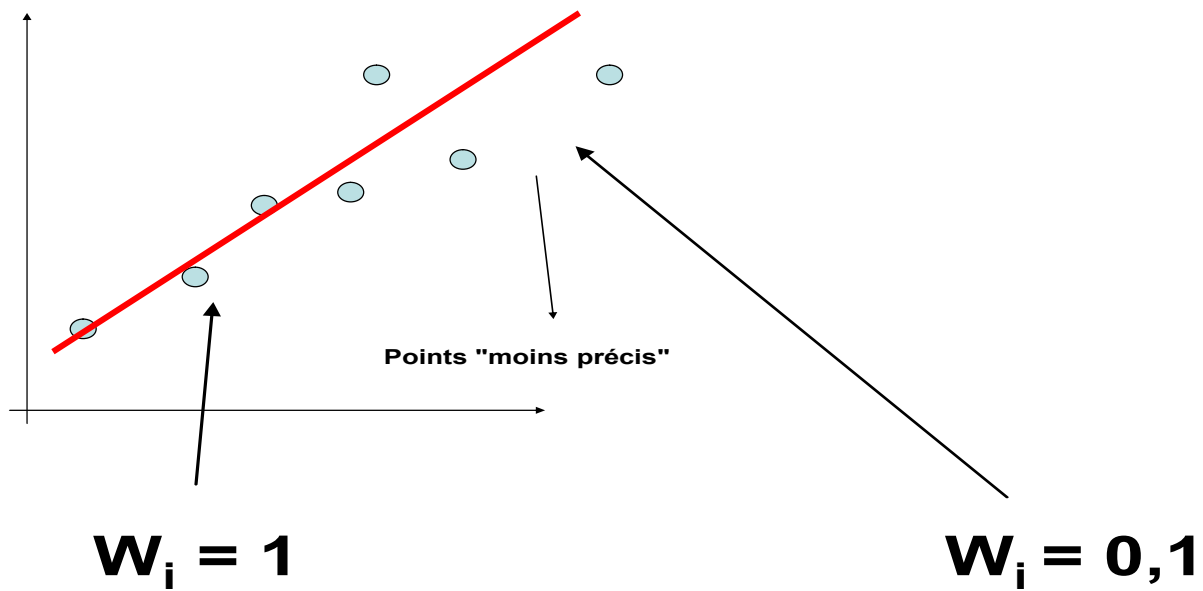
Étude des résidus réduits:

$$e = y - \hat{y}$$

$$er = \frac{y - \hat{y}}{\hat{\sigma}_e}$$

Minimisation des moindres carrés pondérés

par exemple:



Minimisation des moindres carrés pondérés

$$\sum w_i [y_i - (b_0 + b_1 x + \dots)]^2$$

Quantité à minimiser

La résolution de ces équations normales pondérées fournit l'estimation des coefficients b_0 et b_1 :

$$b_1 = \frac{\sum w_i (y_i - \bar{y})(x_i - \bar{x})}{\sum w_i (x_i - \bar{x})^2}$$

La résolution de ces équations normales pondérées fournit l'estimation des coefficients b_0 et b_1 :

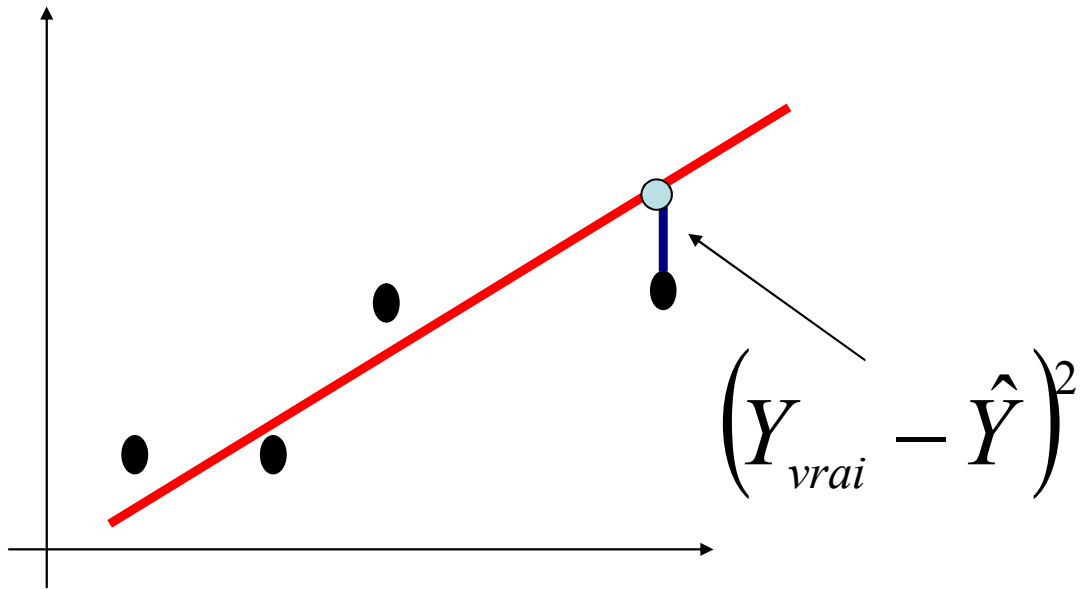
$$b_1 = \frac{\sum w_i (y_i - \bar{y})(x_i - \bar{x})}{\sum w_i (x_i - \bar{x})^2}$$

La résolution de ces équations normales pondérées fournit l'estimation des coefficients b_0 et b_1 :

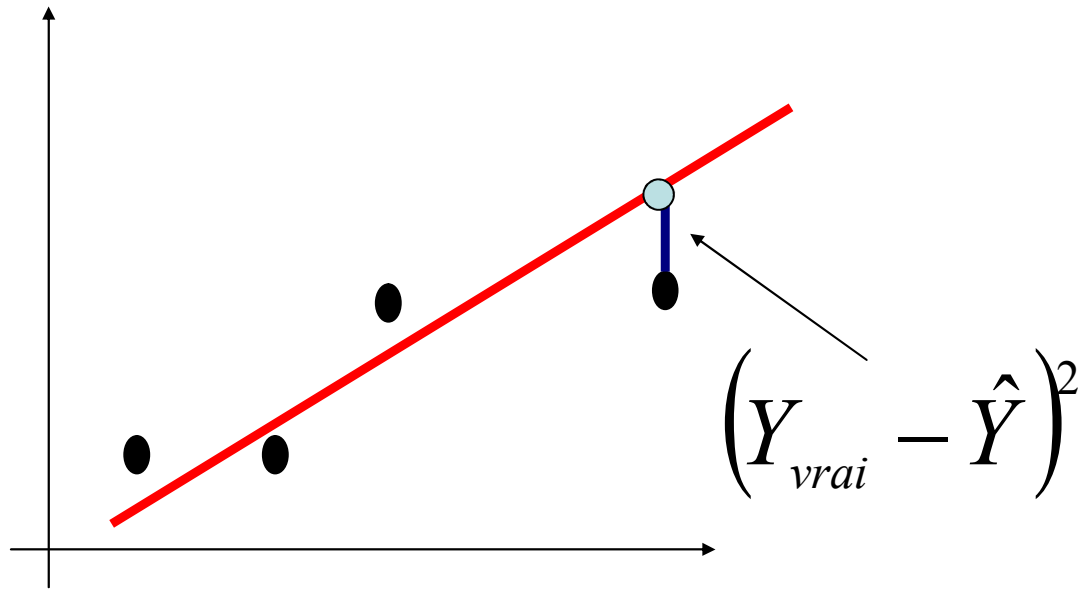
$$b_1 = \frac{\sum w_i (y_i - \bar{y})(x_i - \bar{x})}{\sum w_i (x_i - \bar{x})^2}$$

$$b_0 = \bar{y} + b_1 \bar{x}$$

Souvent:



Souvent:



$$W_i = \frac{1}{(Y_{vrai} - \hat{Y})^2}$$

Régression et colinéarité des variables prédictives.

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p$$

Régression et colinéarité des variables prédictives.

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p$$

$$[O_3] = b_0 + b_1T_{air} + b_2VitVent$$

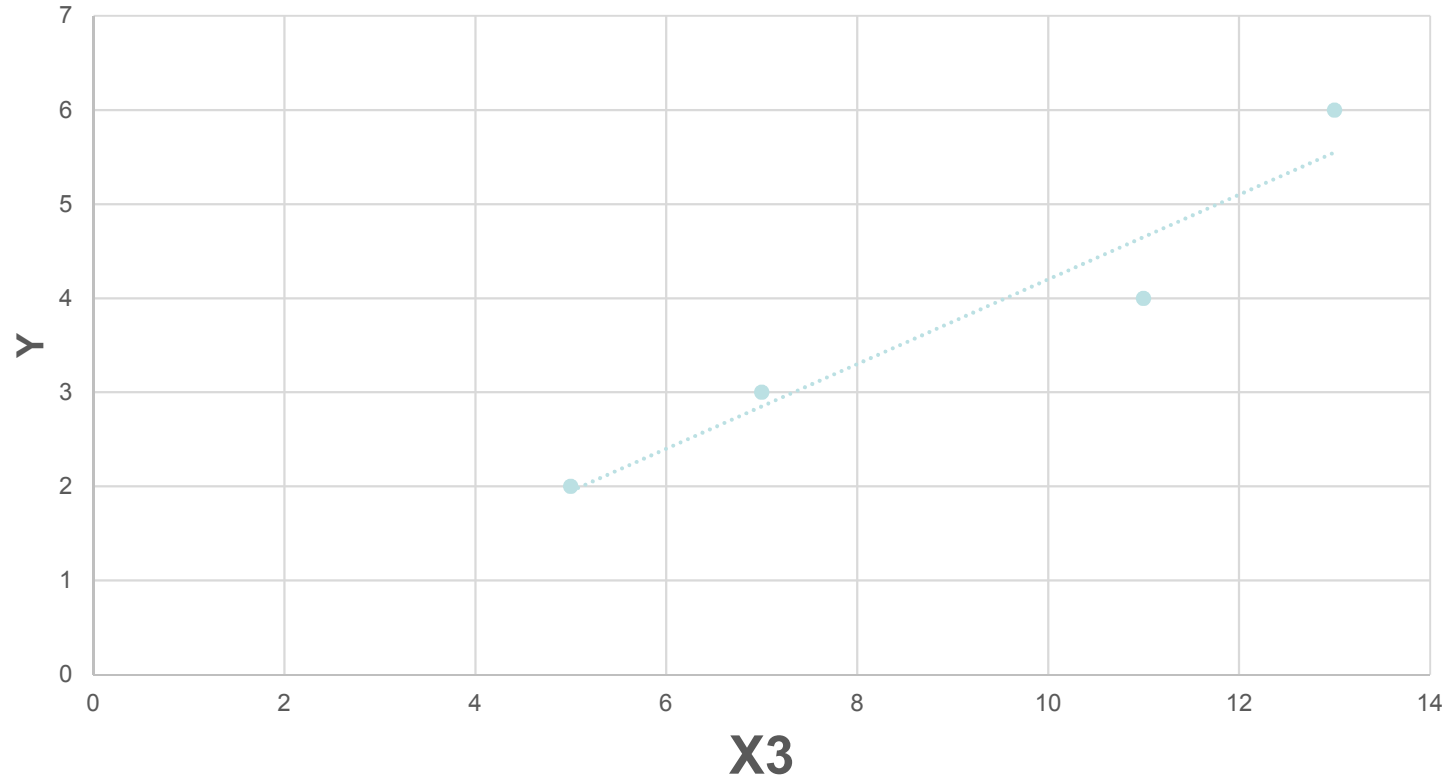
$$+ b_3Radiation + b_4[HC] + b_5[NOx]$$

Y	x₁	x₂	x₃
2	4	1	5
3	6	8	7
4	9	1	11
6	11	5	13

modèle par M.C.O.

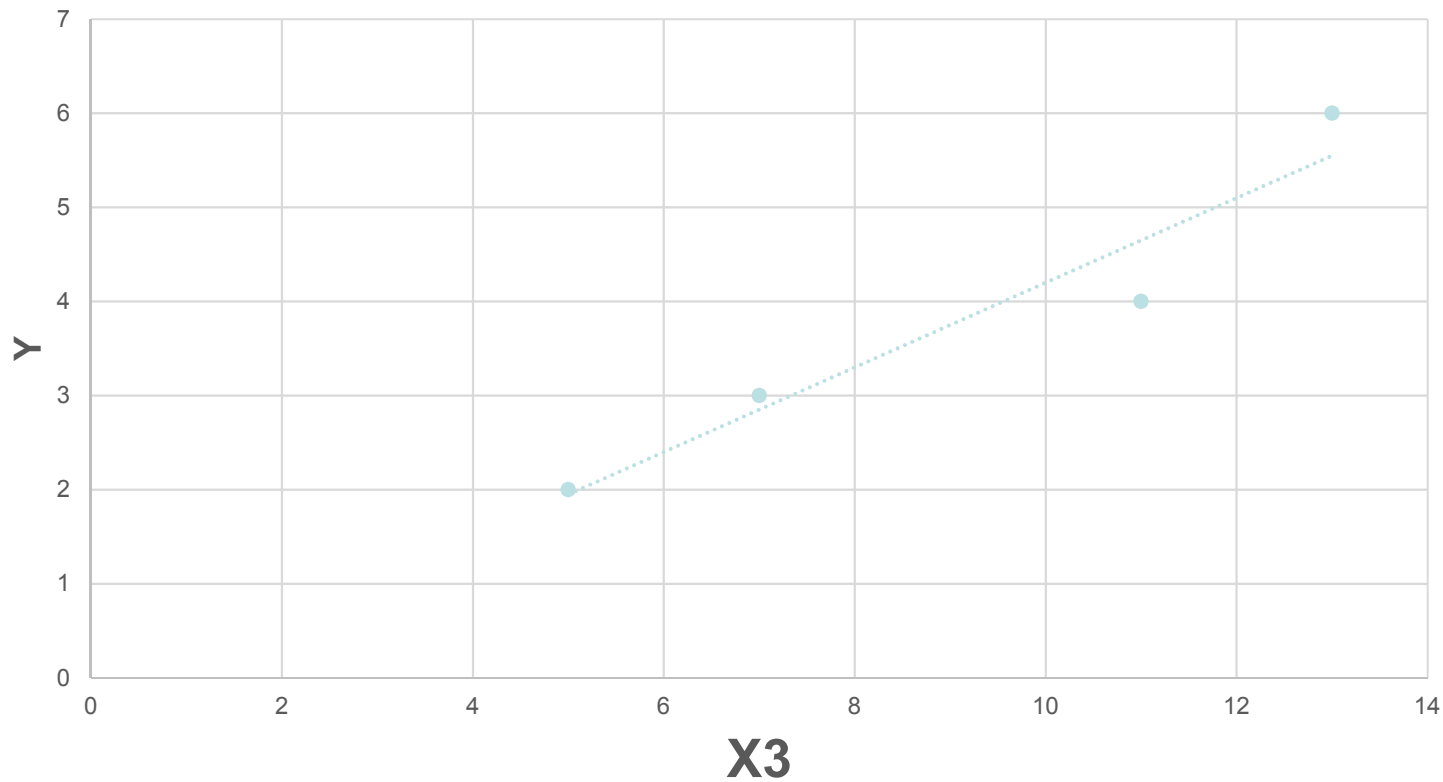
$$**Y = 2 + 8 x_1 - 0,33 x_2 - 0,3 x_3**$$

$$Y=f(X_3)$$



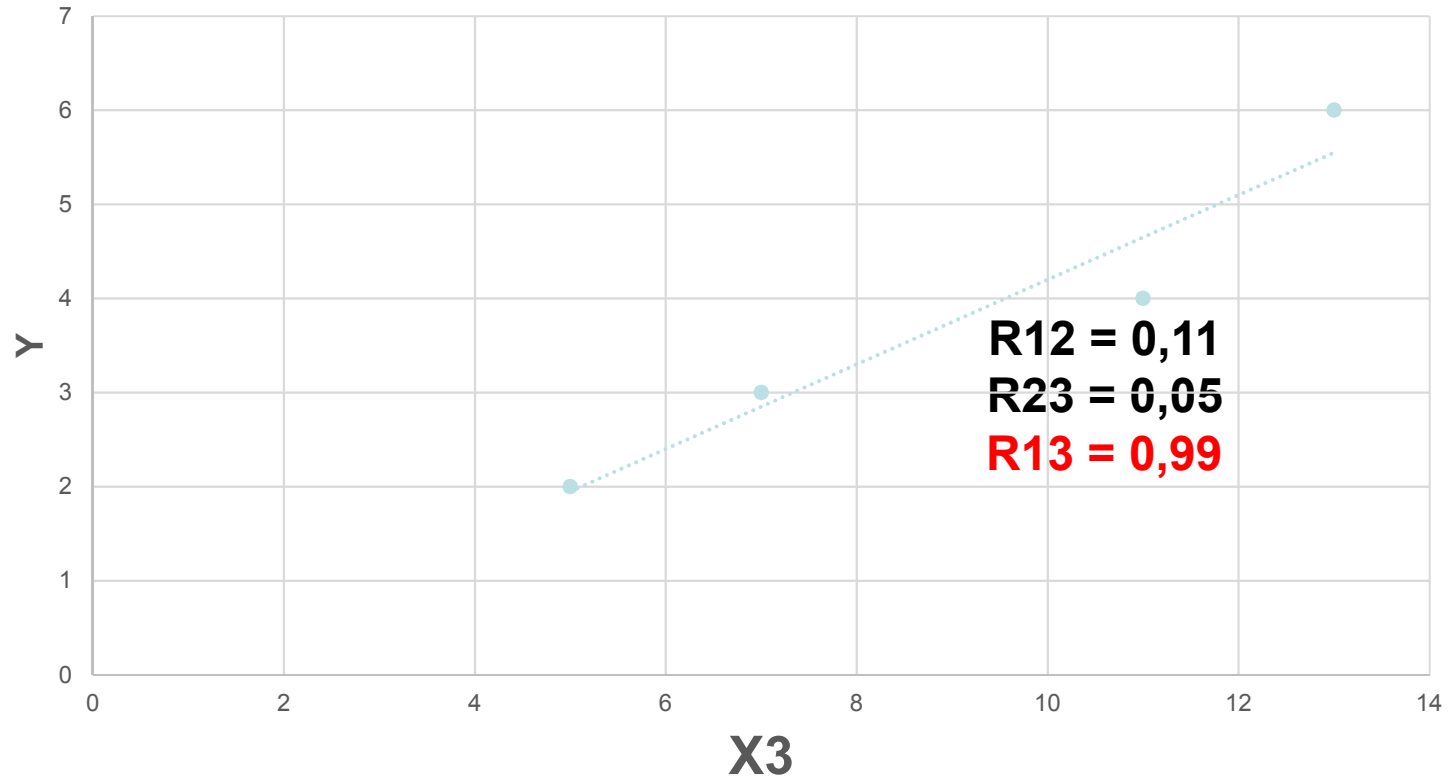
Pente = b3 = -0,3 !!!!

$Y=f(X3)$



Pente = b3 = -0,3 !!!!

$Y=f(X3)$



**La colinéarité des variables X_i
peut être détectée par exemple :
par la matrice des corrélations des variables X_i**

**La colinéarité des variables X_i
peut être détectée par exemple :
par la matrice des corrélations des variables X_i**

$$R_{X_i, X_j}$$

**Si : $R \rightarrow +1$ ou -1
Il y a colinéarité entre X_i et X_j**

- **le signe des coefficients devient instable**
- **les tests t deviennent non significatifs**

- **l'élimination des variables corrélées (moins une bien entendu).**
- **ou mieux, l'emploi de la technique de la « Ridge regression ».**

Les variables X_i doivent être centrées-normées :

$$x_{\text{transformé}} = \frac{x_i - \bar{x}}{\hat{\sigma}_{x_i}}$$

la réponse Y , centrée :

$$y_{transformé} = y_i - \bar{y}$$

On estime de façon itérative
une pondération λ de la régression des moindres carrés
en minimisant :

$$\sum \lambda [y_{i,trans} - (b_0 + b_i x_{i,trans})]^2$$

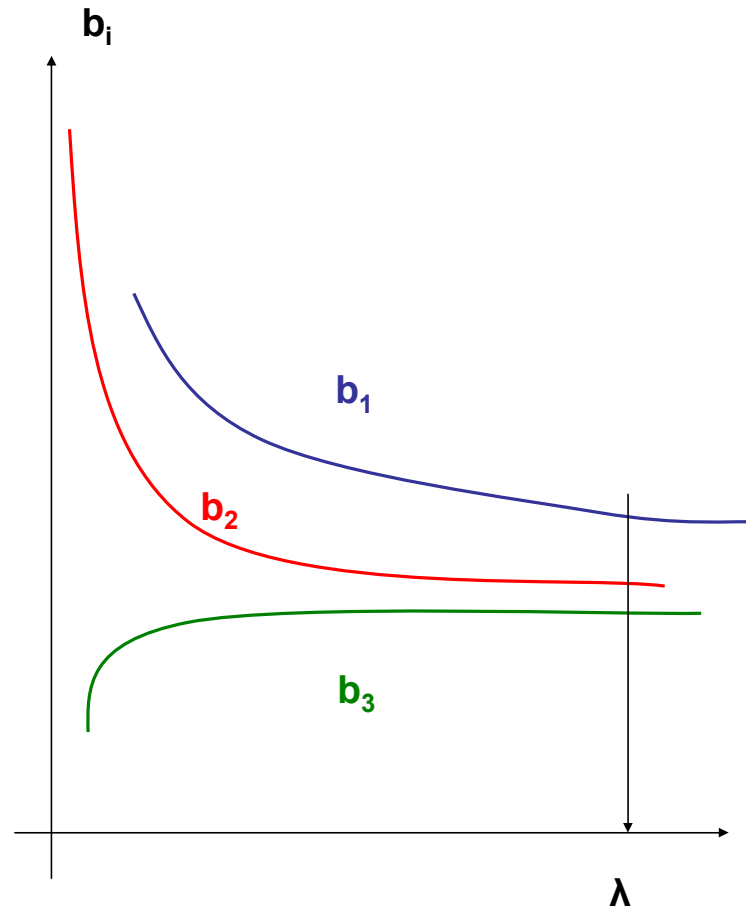
λ varie généralement entre 0 et 1

La résolution de ces équations normales pondérées fournit l'estimation des coefficients b_0 et b_1 :

$$b_1 = \frac{\sum \lambda (y_{i,trans} - \bar{y}_{trans})(x_{i,trans} - \bar{x}_{trans})}{\sum w_i (x_{i,trans} - \bar{x}_{trans})^2}$$

$$b_0 = \bar{y}_{trans} + b_1 \bar{x}_{trans}$$

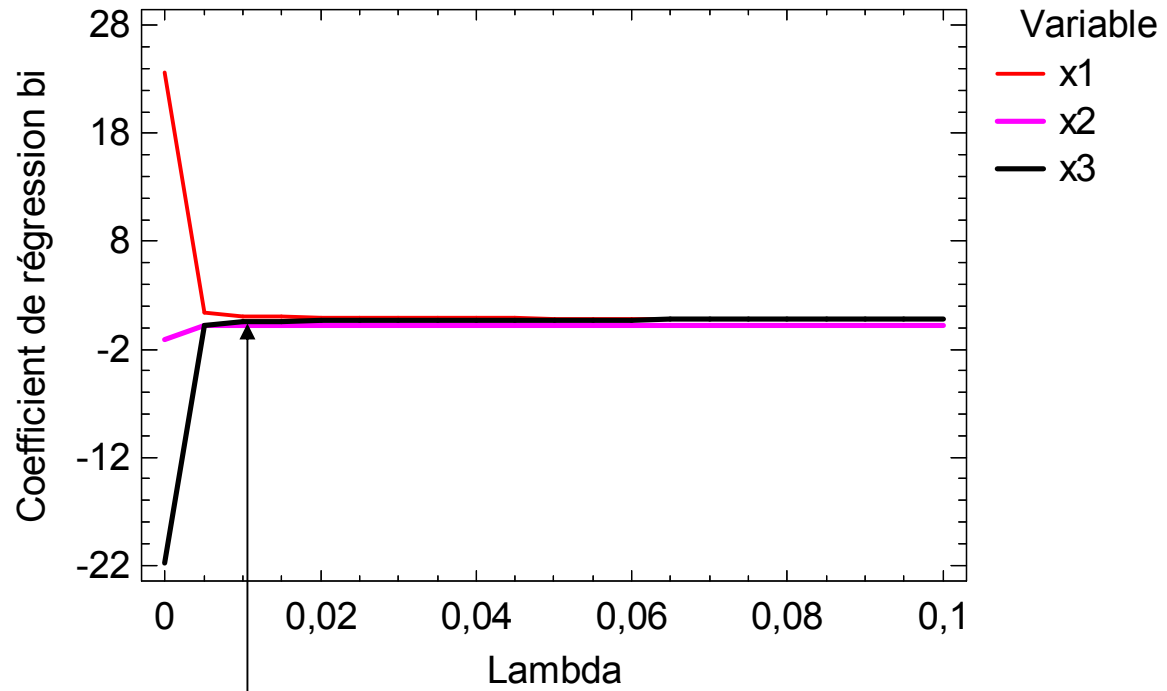
Pour déterminer le choix de λ
on trace les b_i en fonction de λ
et on retient la valeur de λ
pour laquelle il y a stabilisation de l'estimation des b_i .



**Matrice données transformées:
X centrées normées
et Y centrée**

Y	X1	X2	X3
-1,75	-1,13	-0,81	-1,10
-0,75	-0,48	1,25	-0,55
0,25	0,48	-0,81	0,55
2,25	1,13	0,37	1,10

ridge regression



stabilisation des b_i pour
 $\lambda = 0.01$

$$Y = 0 + 1 x_1 + 0,2 x_2 + 0,5 x_3$$

Précédemment:



$$Y = 2 + 8 x_1 - 0,33 x_2 - 0,3 x_3$$

Hypothèses nécessaires pour utiliser les MCO :

- **Exogénéité** $e \neq f(|X_i|)$

l'absence de corrélation entre les variables explicatives et le terme d'erreurs de ce modèle

Hypothèses nécessaires pour utiliser les MCO :

- Exogénéité $e \neq f(|X_i|)$
- $e \sim (N, 0)$
- non colinéarité Σx_i
- non corrélation Σe_i
- homoscedasticité : égalité des variances
- $\Delta Y \gg \Delta X$

Problématique

- Si n individus et seulement 2 variables X et Y , il est facile de représenter l'ensemble des données sur un graphique plan : chaque individu i est un point de coordonnées X_i et $Y_i \rightarrow$ nuage
- L'allure du nuage renseigne sur l'intensité et la nature de la relation entre X et Y .
- **Si plus de 3 variables**, il faut **trouver de « bonnes » approximations du nuage** pour l'appréhender dans sa globalité.

Analyses exploratoires de données

Définition : statistiques descriptives
multidimensionnelles
(beaucoup de dimensions)

Objectif : extraire l'information principale
d'un tableau à double entrée,
y compris quand il est très grand

Méthode : consentir une perte ...
d'information
pour gagner ... en efficacité

Deux grands types de méthodes

*Tableau de données
à double entrée
(n individus *
 p variables)*

**Analyses
factorielles**
*(nuages et
axes factoriels)*

Classifications
*(agrégations
et classes)*

Différentes méthodes factorielles (fc du type de variable)

Analyse en Composantes Principales ACP

établie sur variables **quantitatives**

Si variables qualitatives

=

Analyse des correspondances

Plusieurs variables quantitatives et/ou qualitatives

=

Analyse de correspondance multiples

Analyse factorielle

- Etude de la position d'un **nuage de points** dans l'espace et **description de sa forme**
- Pour mieux voir :
 - **se placer au milieu du nuage**, c'est-à-dire déplacer l'origine au centre de gravité (= individu fictif « moyen »)
 - **regarder dans les directions d'allongement principal**, c'est-à-dire changer d'axes
- Techniquement, **changer de repère** (→ diagonaliser une matrice)

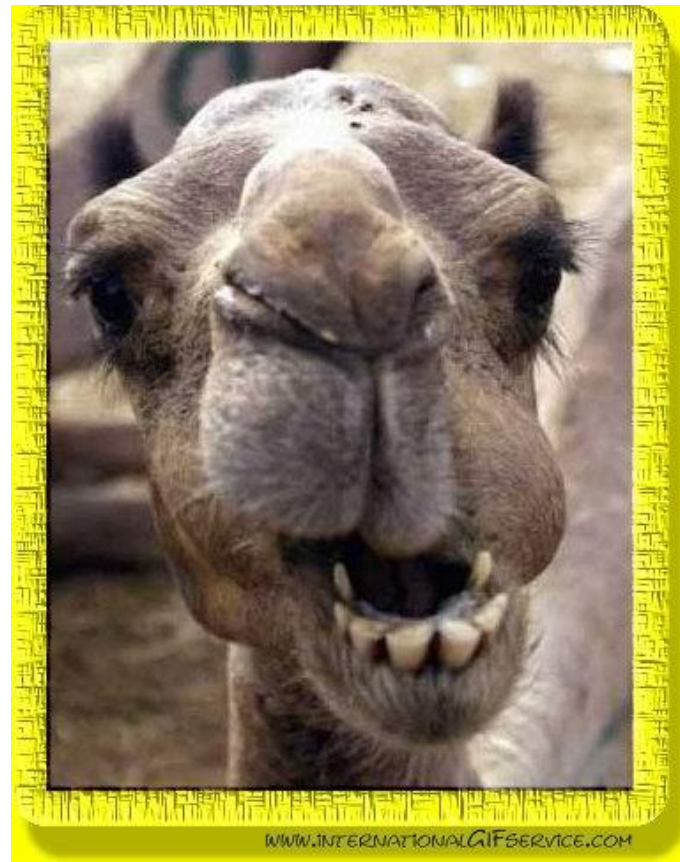
Analyses factorielles

- Un tronc commun :
 - Analyse des proximités
au sein d'un nuage de points « pesants »
selon une distance à déterminer
- Plusieurs analyses différentes
selon la distance choisie :
 - Composantes principales (ACP)
 - Correspondances simples (AFC)
 - Correspondances multiples (ACM)
 - ...

Analyse des Données (ADD)

ensemble de méthodes descriptives ayant pour objectif de ***résumer*** et ***visualiser*** l'information ***pertinente*** contenue dans un *grand tableau de données*

C'est le même principe que lorsque l'on prend une photographie, on passe d'un espace à 3 dimensions à un espace à 2 dimensions.



De fait, l'angle de la prise de vue a une importance capitale car c'est lui qui détermine « l'information exploitable ».



De fait, l'angle de la prise de vue a une importance capitale car c'est lui qui détermine « l'information exploitable ».



L'ACP est basée sur le même principe. La première droite proposée a pour but de voir au mieux les individus, c'est-à-dire celle pour laquelle la variance des individus est maximale. Cette droite est appelée « 1ère composante principale » ou « 1er axe principal ». Elle permet d'observer une certaine proportion de la variation totale des individus.

Ensuite, une 2ème composante principale (axe 2) est recherchée en répondant à 2 conditions :

- avoir une corrélation linéaire nulle avec la première, c'est-à-dire être totalement indépendante de l'axe 1,

- avoir, à son tour, la plus grande variance. Ainsi, la plus grande information complémentaire à l'axe 1 est apportée par l'axe 2.

- Le processus est répété autant de fois que de variables. Donc, pour p variables, on obtient p composantes principales.

Techniquement, l'ACP transforme les p variables quantitatives initiales, plus ou moins corrélées entre elles, en p nouvelles variables quantitatives, non corrélées : les composantes principales.

Ensuite, une 2ème composante principale (axe 2) est recherchée en répondant à 2 conditions :

- avoir une corrélation linéaire nulle avec la première, c'est-à-dire être totalement indépendante de l'axe 1,

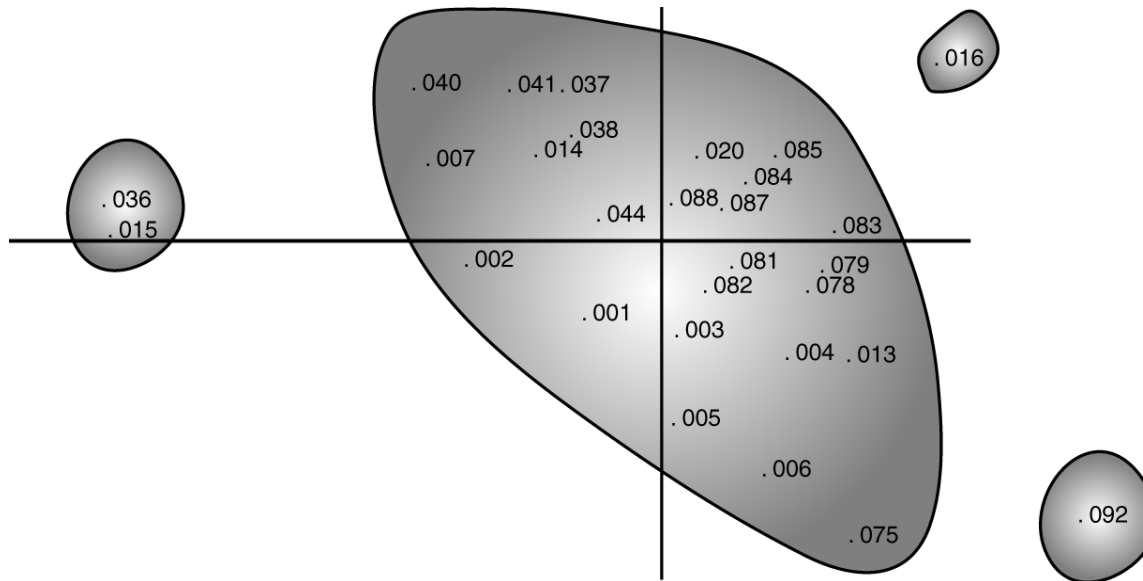
- avoir, à son tour, la plus grande variance. Ainsi, la plus grande information complémentaire à l'axe 1 est apportée par l'axe 2.

- Le processus est répété autant de fois que de variables. Donc, pour p variables, on obtient p composantes principales.

Techniquement, l'ACP transforme les p variables quantitatives initiales, plus ou moins corrélées entre elles, en p nouvelles variables quantitatives, non corrélées : les composantes principales.

L'examen préalable de la représentation des individus sur le plan principal (axes 1 et 2) permet de vérifier l'homogénéité de la population observée. Généralement, trois situations se présentent :

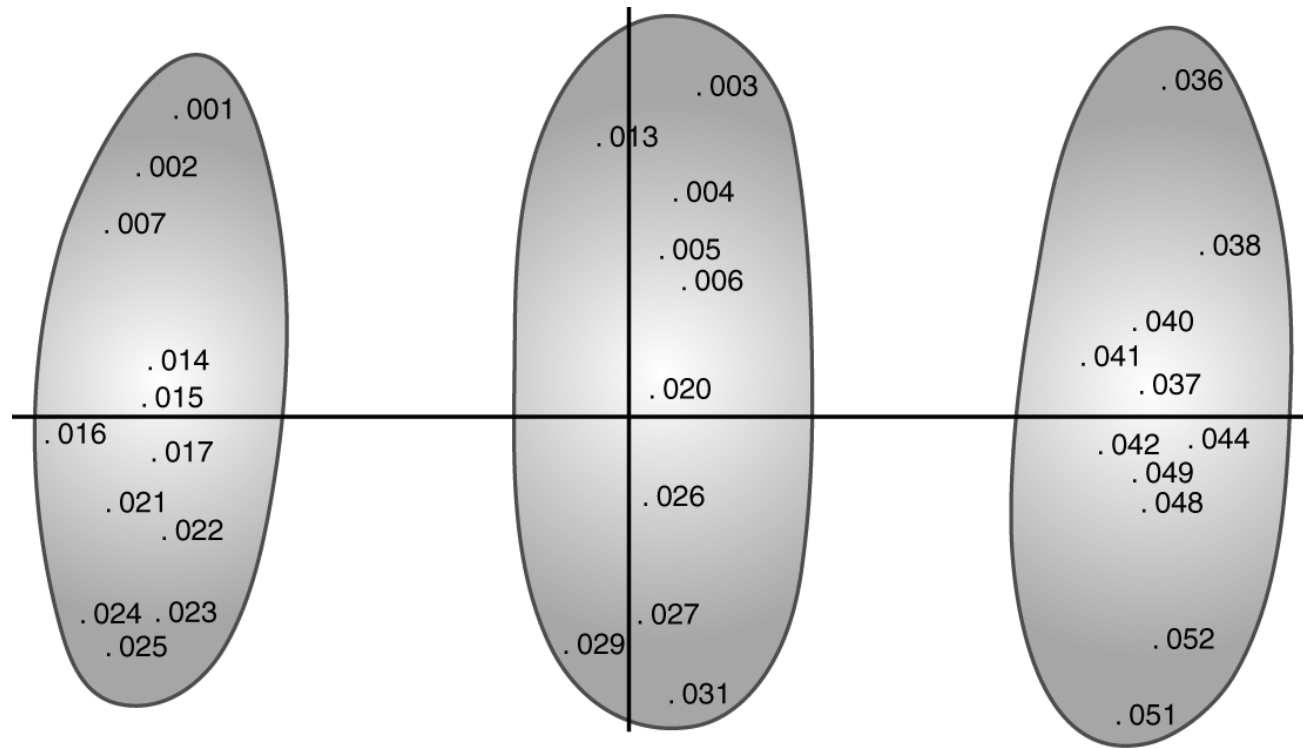
Figure 1 : Vous observez des individus isolés du reste de la population.



Cela peut signifier 2 choses :

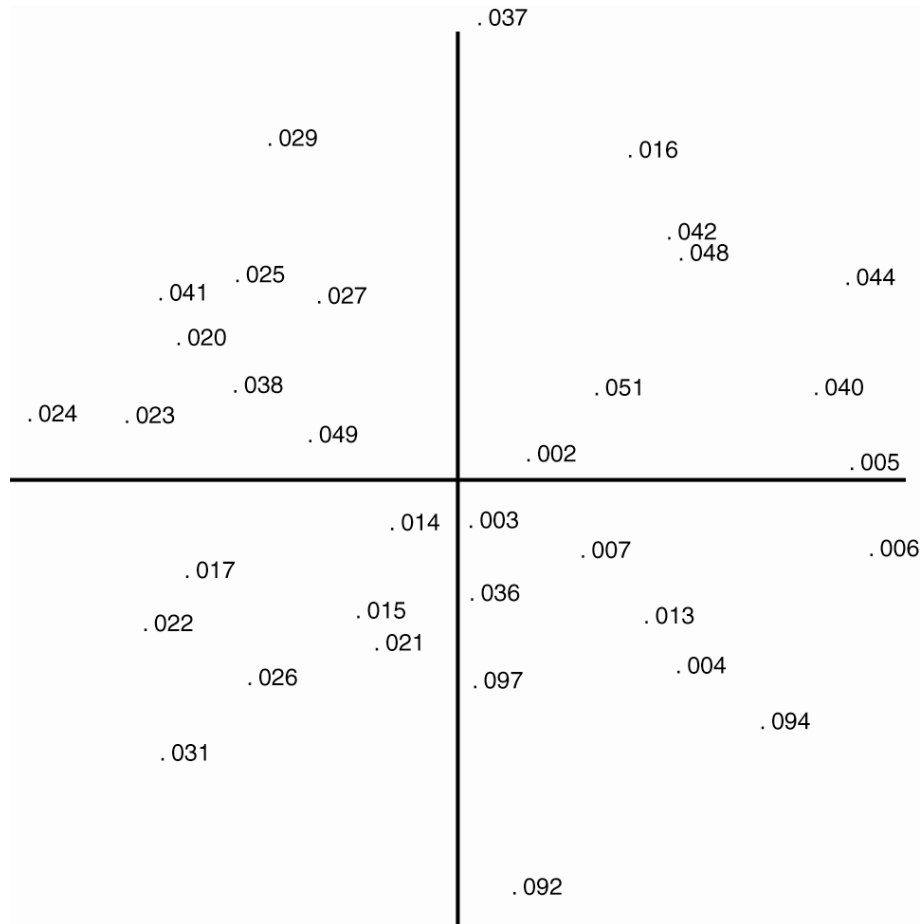
- soit certaines valeurs sont erronées > il faut corriger et refaire l'ACP ;
- soit certains individus sont très différents, il vaut mieux les étudier à part > il faut les retirer de l'analyse.

Figure 2 : Plusieurs groupes distincts sont constitués : il s'agit de sous-populations.



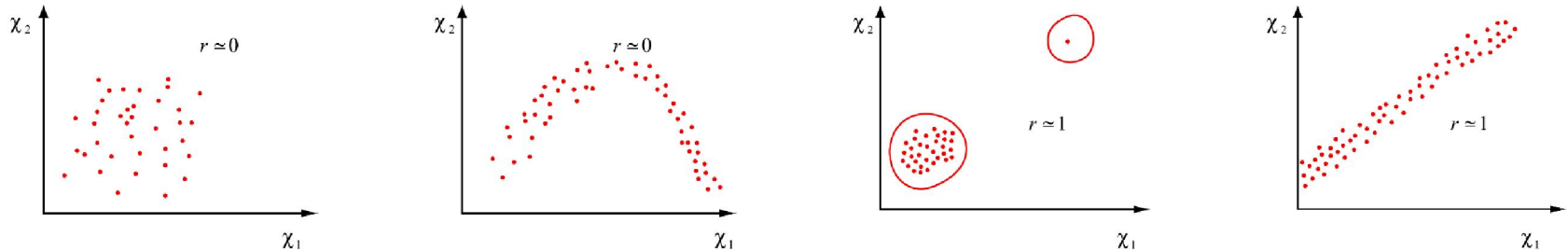
- Si l'objectif est d'étudier les sous-populations, il faut alors réaliser une ACP pour chacune d'elles.
- Si l'objectif est de décrire les individus et de mettre en évidence les caractéristiques de ces groupements > interprétation de l'ACP.

Figure 3 : Les individus se répartissent sur l'ensemble du plan



> interprétation de l'ACP.

2 – Corrélations linéaires



2. Matrice d'inertie (* 1000)
[Coefficients de corrélation linéaire de Pearson]

	V01	V02	V03	V04	V05
V01 TX_AIND	1000	-939	599	-425	-588
V02 TX_ASERV	-939	1000	-567	424	449
V03 TX_EXPNONM	599	-567	1000	-53	-105
V04 TX_NUITEES	-425	424	-53	1000	359
V05 EXPL_BIO	-588	449	-105	359	1000

Vérifier la corrélation des variables deux à deux, grâce à la table de Bravais-Pearson, puis, dans le cas d'une forte corrélation, enlever de l'analyse l'une des deux variables.

Ce travail préalable réalisé, on peut alors relancer l'ACP avec des variables appropriées et validées.

L'ACP propose donc des projections du nuage de points dans des plans de représentation « judicieusement choisis ».

La procédure dite de « diagonalisation » permet de choisir le nombre de plans à retenir.

Afin d'explicitier à quoi correspondent ces plans, il faut donner un sens à leurs axes en procédant à la caractérisation de ces derniers à l'aide des variables initiales. Ceci se fait par l'étude du cercle des corrélations.

Par la suite, on peut discuter de la position et de la proximité des individus dans les plans selon le sens donné aux axes.

Conclusion

L'ACP est une méthode qui a pour but de rendre plus compréhensible l'ensemble des variables contenues dans un tableau de données.

Ne lancer pas d'analyse sans avoir pris soin de sélectionner les variables, supprimer toutes les variables redondantes ou très fortement corrélées.

L'ACP est une méthode parmi d'autres, elle révèle simplement ce qui existe dans vos données.

Entraînez-vous dans un premier temps avec des données et sur des problématiques que vous connaissez bien afin de bien saisir les possibilités offertes par une telle analyse.

Rappels sur les distances

$$\times i \text{ <----- } D(i,j) \text{ -----> } \times j$$

- En géométrie :

Distance euclidienne classique

$$D^2(i,j) = (X_i - X_j)^2 + (Y_i - Y_j)^2$$

- En statistique :

- p variables quantitatives
- n individus, points d'un espace de dimension p
- mesure des distances entre couples d'individus
- la distance euclidienne classique ne convient pas
→ on pondère

Forme générale d'une distance euclidienne

$$D^2(i,j) = \sum \sum M_{ab} (X_{ia} - X_{ja}) (X_{ib} - X_{jb})$$

avec X_{ia} = valeur de la variable a pour l'individu i

et M_{ab} = coefficient de pondération
de l'interaction des variables a et b

On peut lui associer une **métrique**, c'à d une matrice carrée à p lignes et p colonnes contenant les coefficients M_{ab} .

Distances non euclidiennes

- Exemples :
 - Écart moyen $D(i,j) = (\sum | X_{ia} - X_{ja} |) / p$
 - City block $D(i,j) = \sum | X_{ia} - X_{ja} |$
 - Saut maximum $D(i,j) = \max | X_{ia} - X_{ja} |$
 - Saut minimum $D(i,j) = \min | X_{ia} - X_{ja} |$
- On ne peut pas leur associer de métriques (matrices carrées)

Principes d'une analyse factorielle

- n individus sont décrits par p variables
→ tableau de données à double entrée
- On a choisi une distance pour mesurer les distances entre les points du nuage.
- On cherche la meilleure « image approchée » du nuage en projection sur une droite Δ .
- C'est celle qui respecte au mieux les distances entre tous les couples de points

Composantes principales (ACP)

- n individus décrits par p variables quantitatives
- Tableau $X_{n \times p} = (x_{ij})$ des données brutes
- x_{ij} = valeur de la variable j pour l'individu i
- Distance euclidienne canonique
- On cherche la meilleure « image approchée » du nuage en projection sur une droite Δ ; c'est celle qui respecte au mieux les distances entre tous les couples de points : le 1er axe factoriel F_1 .
- Puis on cherche orthogonalement la 2èmePage 162

ACP : objectifs

- Faire le bilan des **ressemblances entre individus** et des **liaisons entre variables**
- Rechercher un **nombre limité de « variables » fictives** appelées « **composantes principales** », non corrélées entre elles et résumant le mieux possible l'information contenu dans le tableau des données brutes

ACP : cercle des corrélations

- Sur un plan factoriel, c'est le **grand cercle de rayon 1** centré sur l'origine.
- Les points-variables tombent tous à l'intérieur.
- Les points-variables situés près du cercle des corrélations sont bien expliqués par le plan factoriel correspondant.
- Deux variables indépendantes forment un angle droit avec l'origine.

ACP : perte et taux d'inertie

- Les axes factoriels sont ordonnés : du plus informatif au moins informatif.
- Chacun représente une part λ_k de l'inertie totale.
- Si l'on ne retient que les premiers axes, on perd de l'inertie : celle des derniers axes.
- On repère dans la décroissance des taux d'inertie (sur l'histogramme des valeurs propres), la plus grande rupture et on ne retient que les axes situés avant elle.

ACP : variables expliquant un axe

- Les corrélations entre un axe factoriel et les variables initiales renseignent sur la signification de l'axe.
- Pour chaque axe, on retient les variables actives présentant les plus fortes corrélations en valeur absolue avec lui. Ce sont elles qui expliquent cet axe.

ACP : nuage des individus

- Il est centré sur le centre de gravité.
- Pour chaque axe, on repère les individus ayant les contributions à l'inertie les plus fortes. Leurs coordonnées (positives ou négatives) sur cet axe permettent de les situer.
- Même si sa contribution à l'inertie est faible, un individu dont le cosinus carré avec un axe est proche de 1, est bien représenté sur cet axe.

ACP : qualité de la représentation

- Globalement, elle dépend du taux d'inertie cumulé sur les premiers axes factoriels retenus.
- Sur un axe donné, la corrélation de chaque variable indique si elle est bien liée à cet axe.
- Sur un axe donné, le cosinus carré de chaque individu indique s'il est bien représenté sur cet axe.

ACP normée

- Pour que toutes les variables jouent le même rôle dans le calcul des distances entre individus
- Pour que les distances entre individus soient indépendantes des unités de mesure des variables
- On centre et on réduit des données :
$$X_r = (z_{ij}) \quad \text{avec} \quad z_{ij} = (x_{ij} - \bar{x}_j) / s_j$$
- On mesure un écart à la moyenne (\bar{x}_j) de la variable j en nombre d'écart-type de cette variable (s_j).
- Tous les variables centrées sont comparables ; elles ont la même dispersion (égale à 1).

Exemple: ACP

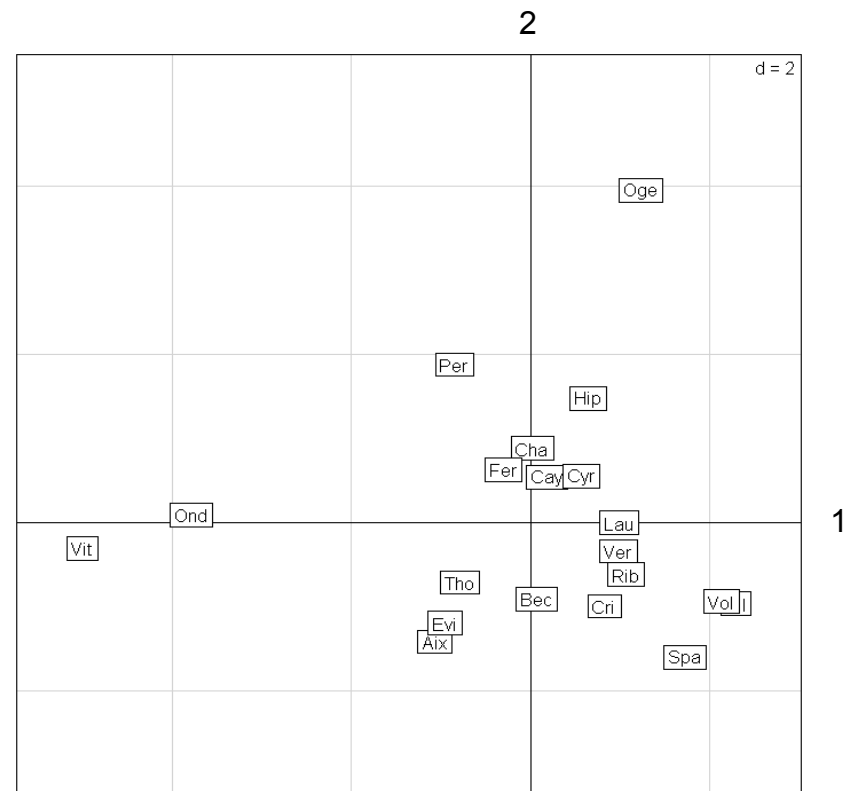
Tableau X

20 eaux minérales (individus)

6 variables composition chimique

Nom	HCO3-	SO4-	Cl-	Ca+	Mg+	Na+
Aix	341.0	27.0	3.0	84.0	23.0	2.0
Bec	263.0	23.0	9.0	91.0	5.0	3.0
Cay	287.0	3.0	5.0	44.0	24.0	23.0
Cha	298.0	9.0	23.0	96.0	6.0	11.0
Cri	200.0	15.0	8.0	70.0	2.0	4.0
Cyr	250.0	5.0	20.0	71.0	6.0	11.0
Evi	357.0	10.0	2.0	78.0	24.0	5.0
Fer	311.0	14.0	18.0	73.0	18.0	13.0
Hip	256.0	6.0	23.0	86.0	3.0	18.0
Lau	186.0	10.0	16.0	64.0	4.0	9.0
Oge	183.0	16.0	44.0	48.0	11.0	31.0
Ond	398.0	218.0	15.0	157.0	35.0	8.0
Per	348.0	51.0	31.0	140.0	4.0	14.0
Rib	168.0	24.0	8.0	55.0	5.0	9.0
Spa	110.0	65.0	5.0	4.0	1.0	3.0
Tho	332.0	14.0	8.0	103.0	16.0	5.0
Ver	196.0	18.0	6.0	58.0	6.0	13.0
Vil	59.0	7.0	6.0	16.0	2.0	9.0
Vit	402.0	306.0	15.0	202.0	36.0	3.0
Vol	64.0	7.0	8.0	10.0	6.0	8.0

Représentation des individus dans un plan engendré par les 2 premières composantes principales



Nom	HCO3-	SO4-	Cl-	Ca+	Mg+	Na+
Aix	341.0	27.0	3.0	84.0	23.0	2.0
Bec	263.0	23.0	9.0	91.0	5.0	3.0
Cay	287.0	3.0	5.0	44.0	24.0	23.0
Cha	298.0	9.0	23.0	96.0	6.0	11.0
Cri	200.0	15.0	8.0	70.0	2.0	4.0
Cyr	250.0	5.0	20.0	71.0	6.0	11.0
Evi	357.0	10.0	2.0	78.0	24.0	5.0
Fer	311.0	14.0	18.0	73.0	18.0	13.0
Hip	256.0	6.0	23.0	86.0	3.0	18.0
Lau	186.0	10.0	16.0	64.0	4.0	9.0
Oge	183.0	16.0	44.0	48.0	11.0	31.0
Ond	398.0	218.0	15.0	157.0	35.0	8.0
Per	348.0	51.0	31.0	140.0	4.0	14.0
Rib	168.0	24.0	8.0	55.0	5.0	9.0
Spa	110.0	65.0	5.0	4.0	1.0	3.0
Tho	332.0	14.0	8.0	103.0	16.0	5.0
Ver	196.0	18.0	6.0	58.0	6.0	13.0
Vil	59.0	7.0	6.0	16.0	2.0	9.0
Vit	402.0	306.0	15.0	202.0	36.0	3.0
Vol	64.0	7.0	8.0	10.0	6.0	8.0


Tableau X







20 eaux minérales (individus)

6 variables composition
chimique

ACP : Les axes

Interprétation

$$\frac{\lambda_\alpha}{\sum \lambda_i}$$


n	Valeur	Pourcent	Cumul	0	3.0941
1	3.0941	51.57	51.57		
2	1.6876	28.13	79.69		
3	0.5965	9.94	89.64		
4	0.5028	8.38	98.02		
5	0.0932	1.55	99.57		
6	0.0258	0.43	100.00		

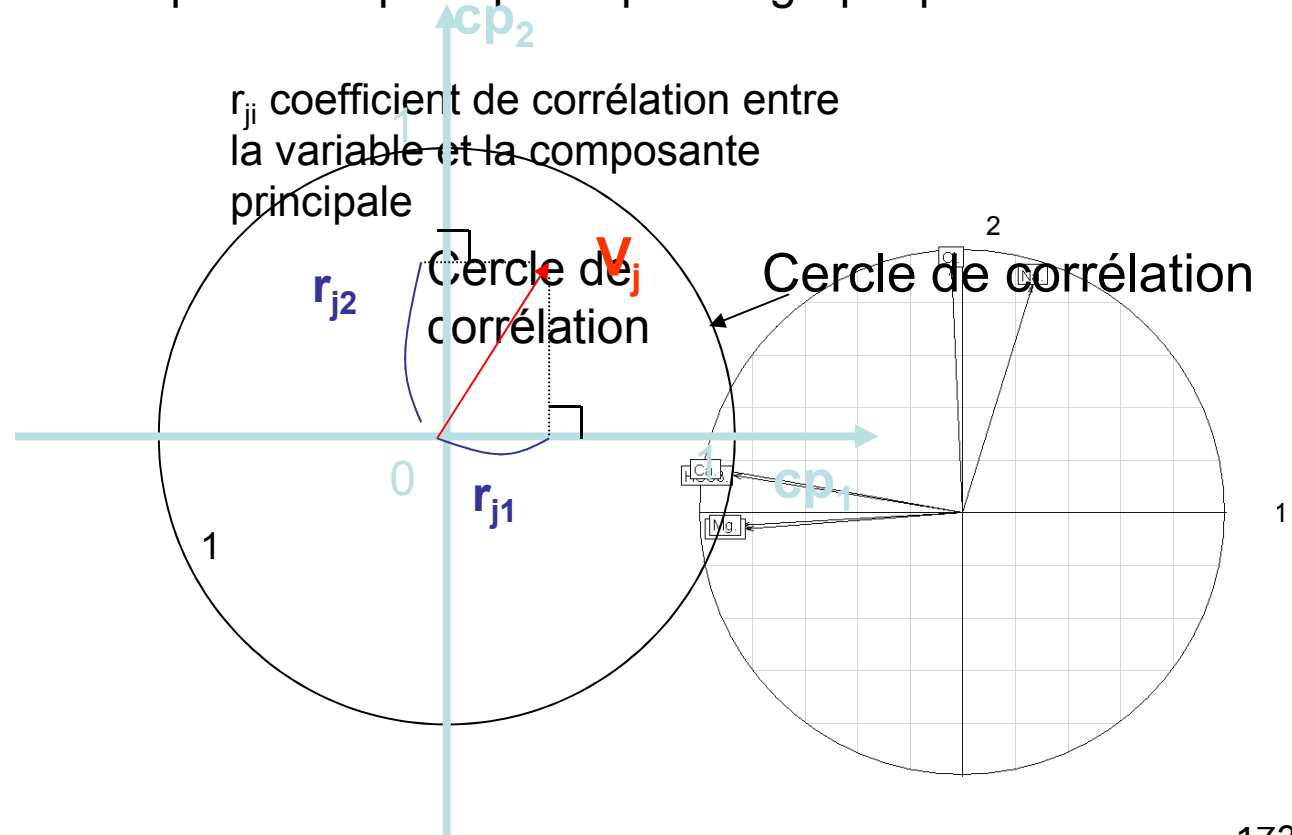
Variance totale = 6.0

La quantité maximisée (valeur propre) par la première composante principale est $\lambda_1 = 3.0941$

On dit que la première composante explique 51.57% de la variabilité des données (la variance totale = nombre de variables en **ACP normée**)

La deuxième explique 28.13% de la variance totale. Les deux premières expliquent 79.69%.

- Pour apprécier dans quelle mesure une composante principale représente bien les variables on calcule, le coefficient de corrélation entre cette composante et chacune des variables initiales.
- En pratique on visualise l'ensemble des corrélations entre chaque variable et deux composantes principales par un graphique: le **cercle des corrélations**



ACP

composante principale = combinaison des variables

U matrice des vecteurs propres

u1	u2	u3	u4	u5	u6
-0.498	-0.11	0.243	0.578	0.004	-0.589
-0.472	0.047	-0.348	-0.675	0.165	-0.415
-0.025	-0.721	-0.39	0.021	-0.572	0.012
-0.524	-0.128	-0.323	0.263	0.427	0.595
-0.481	0.036	0.601	-0.296	-0.438	0.355
0.151	-0.67	0.449	-0.231	0.521	-0.04

coefficients = loadings



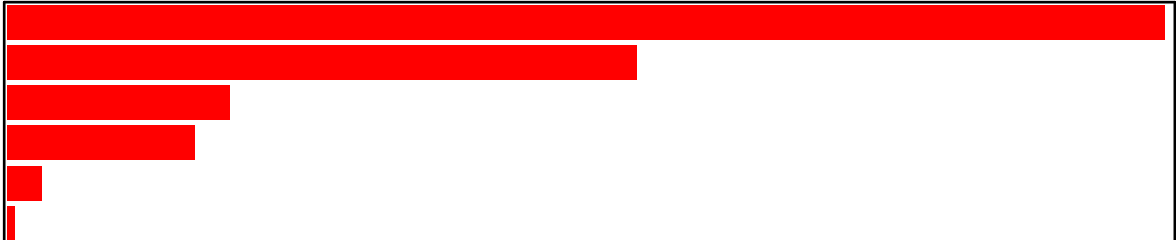
U1 contient les coefficients de la première composante principale
 $Cp1 = -0.498 V1 - 0.472 V2 - 0.025 V3 - 0.524 V4 - 0.481 V5 + 0.151 V6$
 U22ème

V1 n'est pas exactement la variable HCO3-

Dans l'ACP le tableau est centré ou bien centré réduit (normalisé)

valeurs propres λ

n	Valeur	Pourcent	Cumul	0	3.0941
1	3.0941	51.57	51.57		
2	1.6876	28.13	79.69		
3	0.5965	9.94	89.64		
4	0.5028	8.38	98.02		
5	0.0932	1.55	99.57		
6	0.0258	0.43	100.00		



Variance totale = 6.0

U matrice des vecteurs propres					
u1	u2	u3	u4	u5	u6
-0.498	-0.11	0.243	0.578	0.004	-0.589
-0.472	0.047	-0.348	-0.675	0.165	-0.415
-0.025	-0.721	-0.39	0.021	-0.572	0.012
-0.524	-0.128	-0.323	0.263	0.427	0.595
-0.481	0.036	0.601	-0.296	-0.438	0.355
0.151	-0.67	0.449	-0.231	0.521	-0.04

Matrice de corrélation						
	HCO3-	SO4-	CL-	CA+	MG+	NA+
HCO3-	1	0.478	0.122	0.85	0.731	-0.109
SO4-	0.478	1	0.045	0.73	0.671	-0.279
CL-	0.122	0.045	1	0.25	-0.13	0.668
CA+	0.852	0.733	0.252	1	0.606	-0.196
MG+	0.731	0.671	-0.13	0.61	1	-0.091
NA+	-0.109	-0.279	0.668	-0.2	-0.09	1

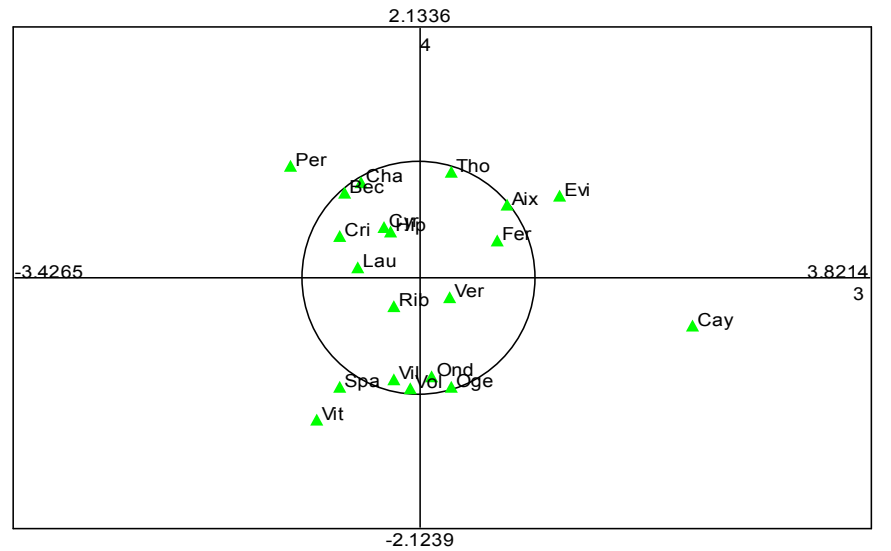
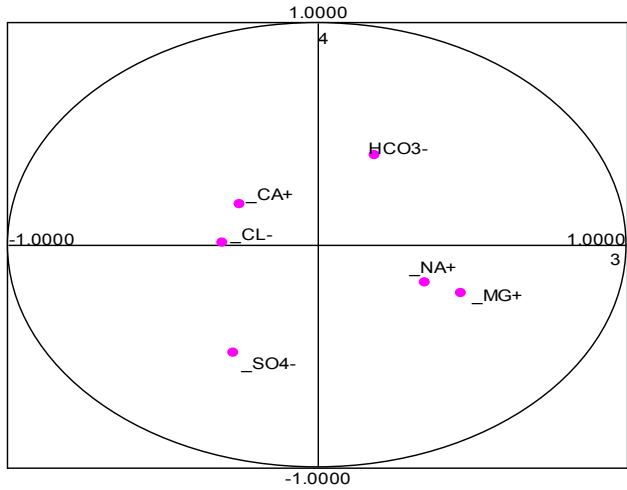
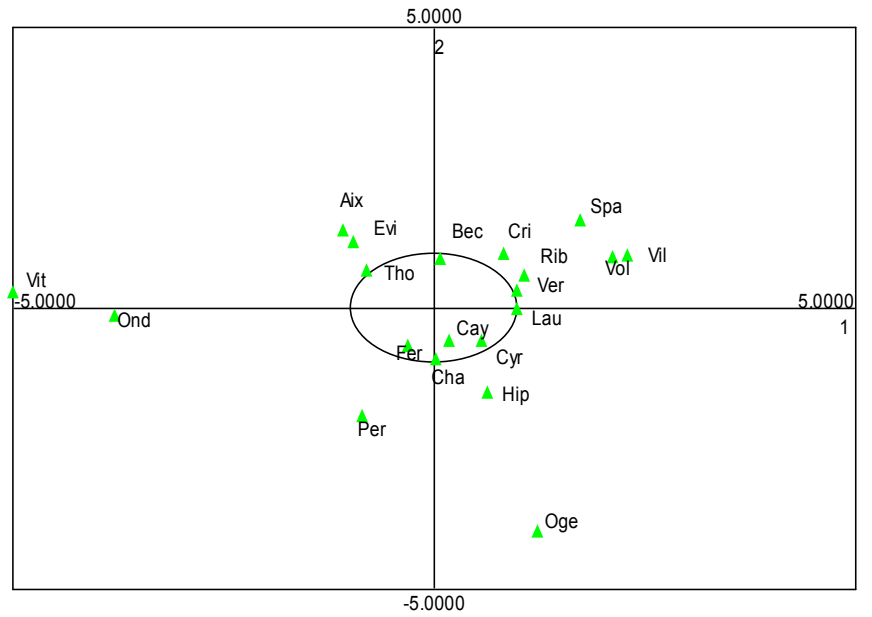
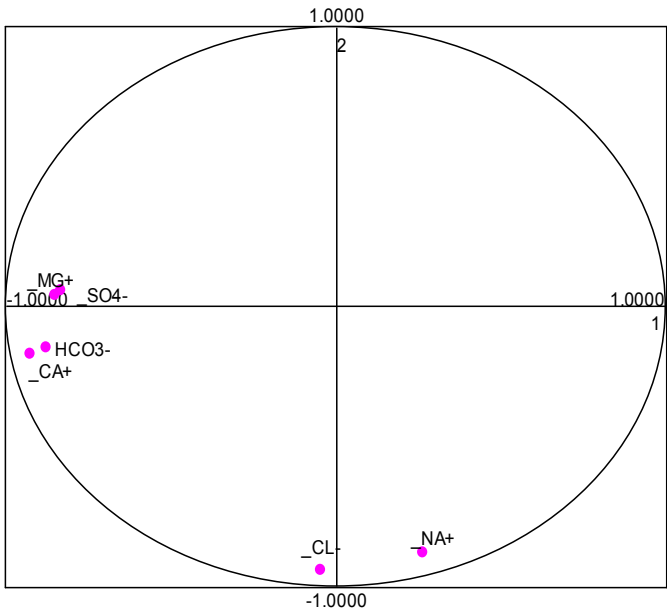
$$\text{Cor}(\text{CA+}, \text{HCO3-}) = 0.852$$

.....

ACP normée diagonalisation matrice de corrélation
 ⇒ valeurs propres et vecteurs propres

acp : facteurs pour les variables principales									
		Axe : 1	Axe : 1	Axe : 2	Axe : 2	Axe : 3	Axe : 3	Axe : 4	Axe : 4
	Qlt	Coord	Cor	Coord	Cor	Coord	Cor	Coord	Cor
HCO3-	0.991	-0.876	0.767	-0.143	0.02	0.188	0.035	0.41	0.168
_SO4-	0.993	-0.83	0.688	0.061	0.004	-0.268	0.072	-0.478	0.229
_CL-	0.969	-0.044	0.002	-0.936	0.876	-0.301	0.091	0.015	0
_CA+	0.974	-0.922	0.849	-0.166	0.028	-0.249	0.062	0.186	0.035
_MG+	0.979	-0.847	0.717	0.046	0.002	0.464	0.216	-0.21	0.044
_NA+	0.975	0.265	0.07	-0.87	0.757	0.347	0.12	-0.164	0.027

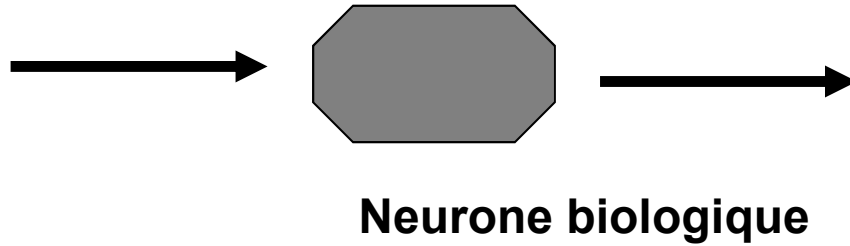
acp : facteurs pour les individus principaux													
		Axe : 1	Axe : 1	Axe : 1	Axe : 2	Axe : 2	Axe : 2	Axe : 3	Axe : 3	Axe : 3	Axe : 4	Axe : 4	Axe : 4
	Qlt	Coord	Cor	Ctr	Coord	Cor	Ctr	Coord	Cor	Ctr	Coord	Cor	Ctr
Aix	0.958	-1.069	0.266	0.018	1.414	0.466	0.059	0.758	0.134	0.048	0.631	0.093	0.04
Bec	0.995	0.074	0.003	0	0.906	0.467	0.024	-0.626	0.223	0.033	0.728	0.302	0.053
Cay	0.95	0.19	0.006	0.001	-0.54	0.047	0.009	2.315	0.872	0.449	-0.395	0.025	0.016
Cha	0.99	0.02	0	0	-0.88	0.454	0.023	-0.481	0.136	0.019	0.826	0.4	0.068
Cri	0.99	0.829	0.302	0.011	0.994	0.435	0.029	-0.664	0.194	0.037	0.365	0.059	0.013
Cyr	0.957	0.571	0.344	0.005	-0.552	0.321	0.009	-0.294	0.091	0.007	0.437	0.202	0.019
Evi	0.982	-0.955	0.209	0.015	1.194	0.327	0.042	1.199	0.329	0.12	0.714	0.117	0.051
Fer	0.875	-0.298	0.075	0.001	-0.627	0.333	0.012	0.668	0.378	0.037	0.324	0.089	0.01
Hip	0.941	0.641	0.138	0.007	-1.476	0.73	0.065	-0.226	0.017	0.004	0.409	0.056	0.017
Lau	0.979	0.993	0.769	0.016	0.001	0	0	-0.51	0.203	0.022	0.095	0.007	0.001
Oge	0.989	1.234	0.084	0.025	-3.946	0.855	0.461	0.282	0.004	0.007	-0.911	0.046	0.082
Ond	1	-3.791	0.953	0.232	-0.092	0.001	0	0.112	0.001	0.001	-0.824	0.045	0.068
Per	0.99	-0.849	0.113	0.012	-1.876	0.551	0.104	-1.079	0.182	0.098	0.963	0.145	0.092
Rib	0.956	1.072	0.678	0.019	0.616	0.224	0.011	-0.198	0.023	0.003	-0.228	0.031	0.005
Spa	0.95	1.735	0.417	0.049	1.604	0.356	0.076	-0.663	0.061	0.037	-0.915	0.116	0.083
Tho	0.987	-0.793	0.306	0.01	0.71	0.245	0.015	0.276	0.037	0.006	0.906	0.399	0.082
Ver	0.741	0.981	0.609	0.016	0.341	0.073	0.003	0.26	0.043	0.006	-0.157	0.016	0.002
Vil	0.992	2.294	0.751	0.085	0.961	0.132	0.027	-0.208	0.006	0.004	-0.848	0.103	0.071
Vit	0.999	-5.011	0.917	0.406	0.307	0.003	0.003	-0.85	0.026	0.061	-1.195	0.052	0.142
Vol	0.976	2.133	0.706	0.073	0.94	0.137	0.026	-0.071	0.001	0	-0.925	0.133	0.085



La réduction d'une matrice de données à quelques grands groupes, voilà ce qu'est une analyse factorielle (pure ou en composantes principales). Mais qu'est-ce qu'on fait? On essaie, par une méthode quantitative (analyse multivariée), d'expliquer la portion la plus grande de la variance totale (cas d'une analyse en composantes principales) ou de la covariance (cas de l'analyse factorielle) par de nouvelles variables (créées de toute pièce par l'analyse - variables construites) que l'on appelle "composantes", "facteurs" ou "axes factoriels"; ces variables dites "latentes" ne sont donc pas mesurées mais plutôt générées. Ces composantes ou facteurs regroupent, dans une certaine mesure, des éléments (variables) ou plutôt des variables corrélées dans le but d'expliquer un phénomène par une quantité plus restreinte d'éléments.

Réseau Neurons Artificiels

Réseau **N**eurones **A**rtificiels



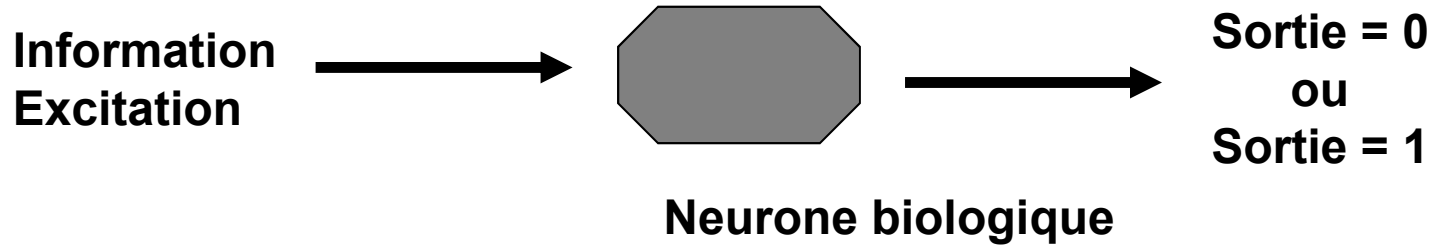
Réseau Neurons Artificiels

Information
Excitation

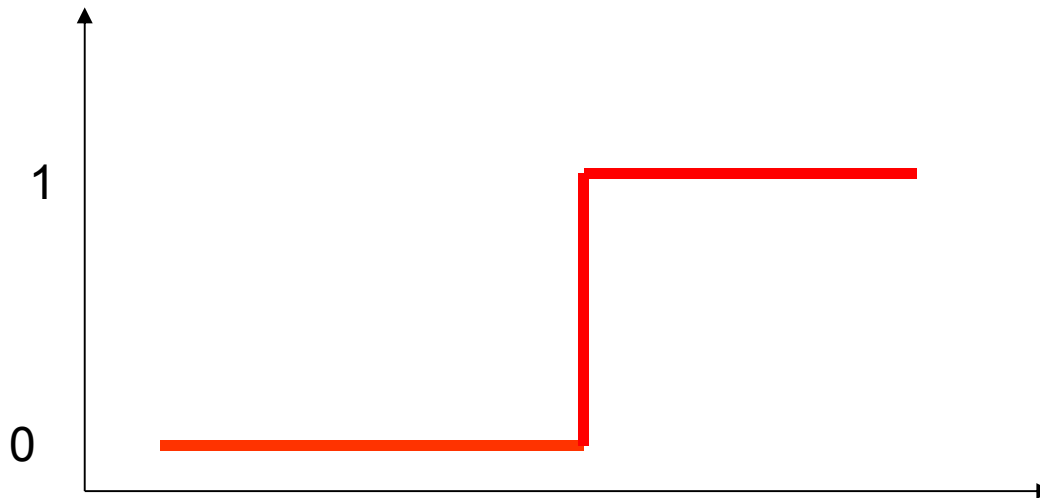
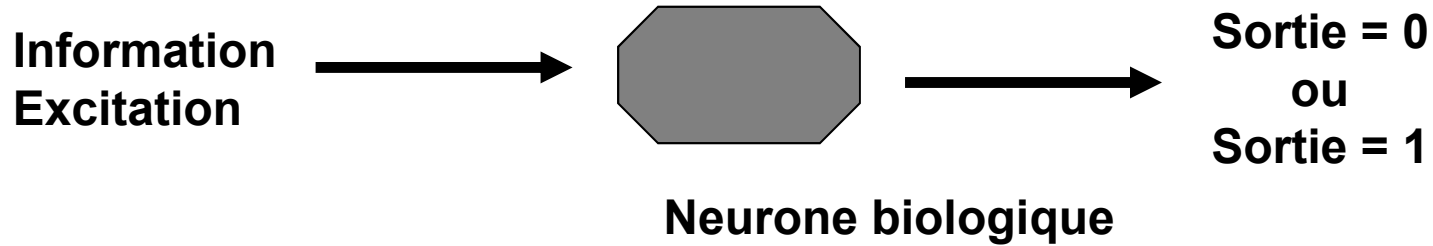


Neurone biologique

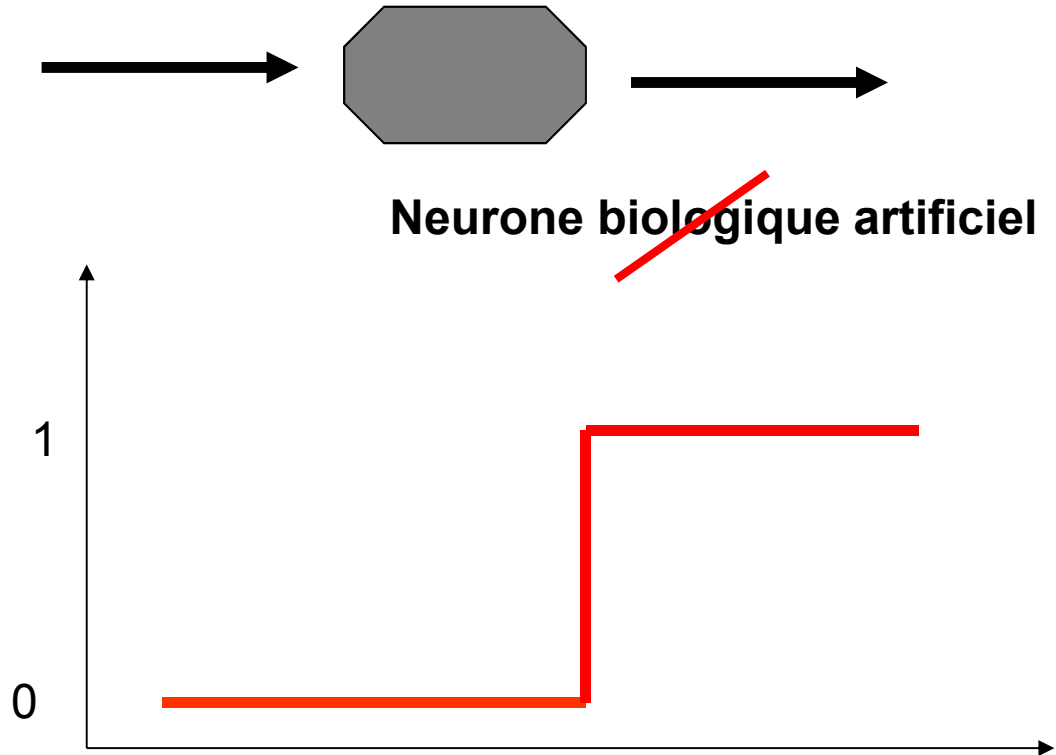
Réseau Neurons Artificiels



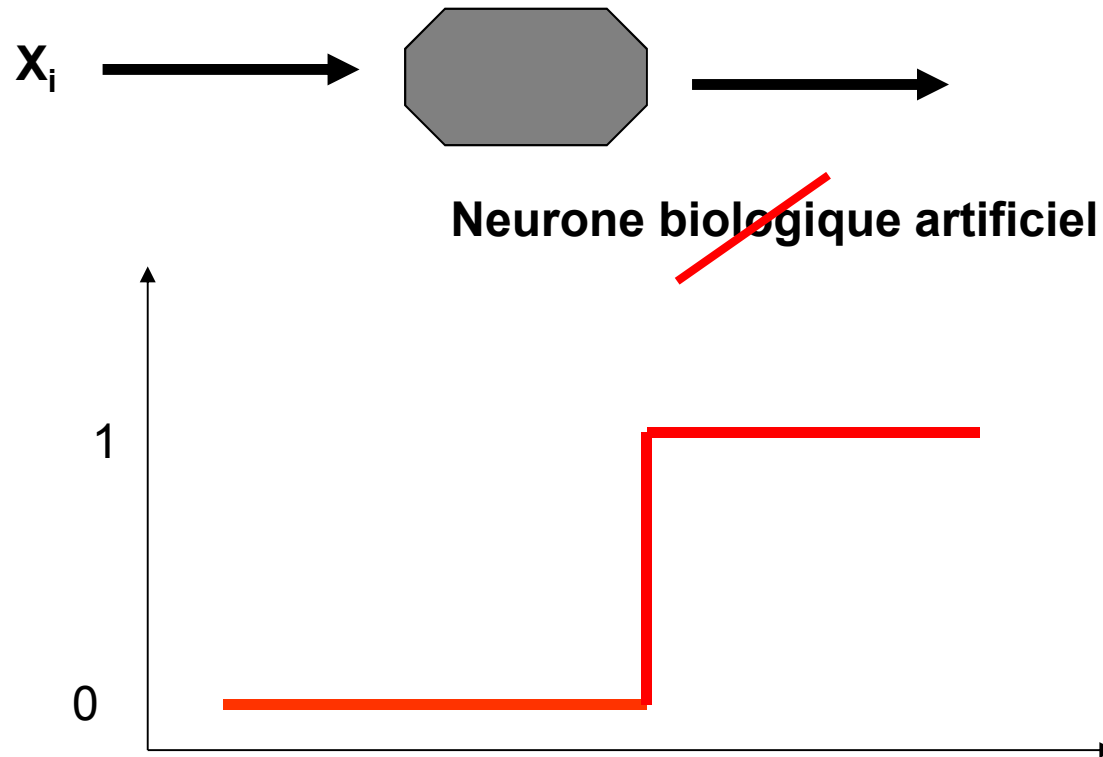
Réseau Neurons Artificiels



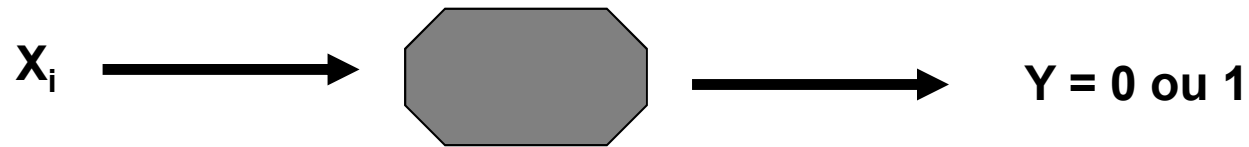
Réseau **N**eurones **A**rtificiels



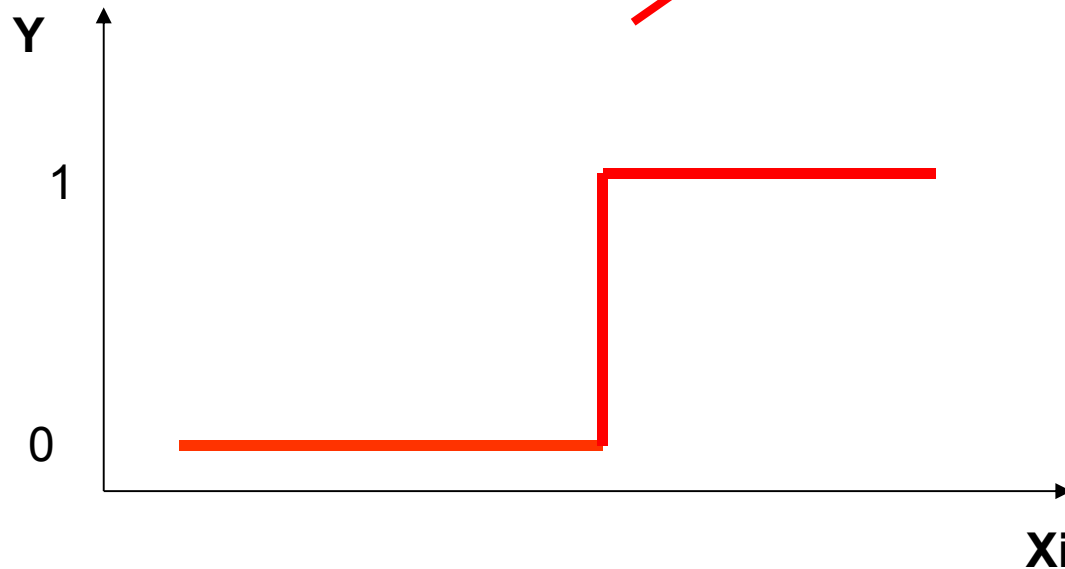
Réseau **N**eurones **A**rtificiels



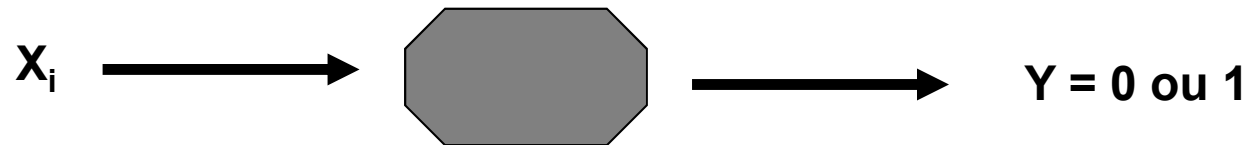
Réseau Neurons Artificiels



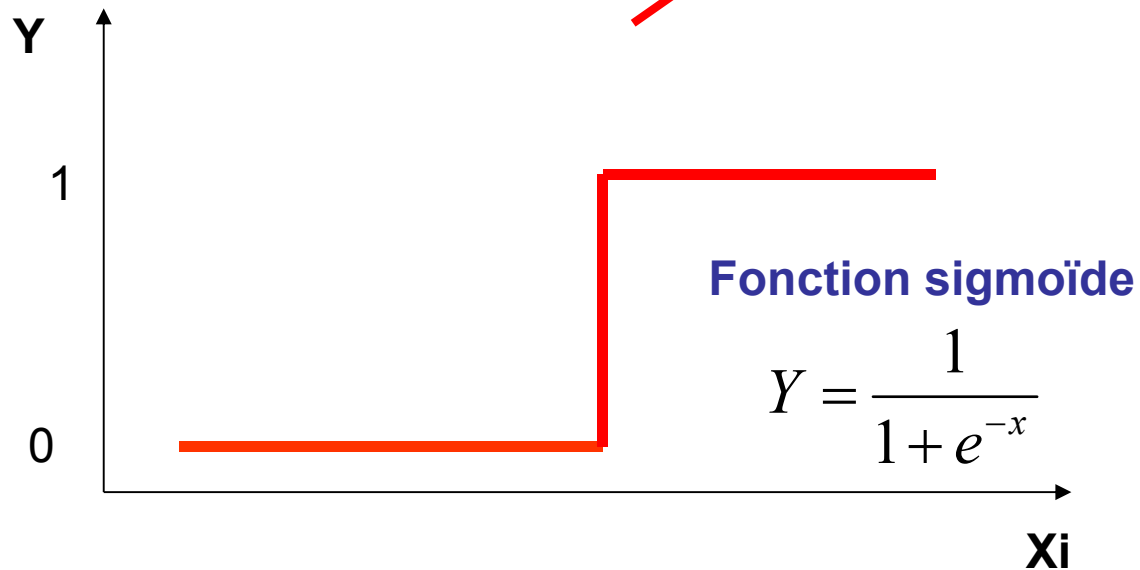
~~Neurone biologique artificiel~~



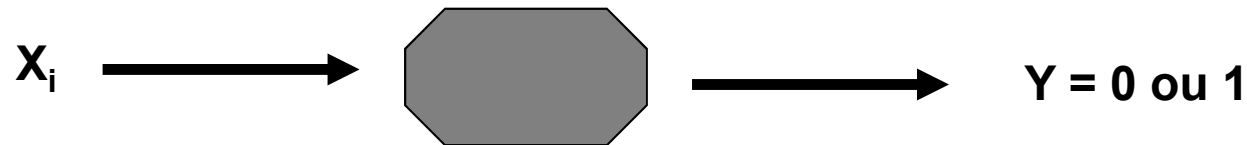
Réseau Neurons Artificiels



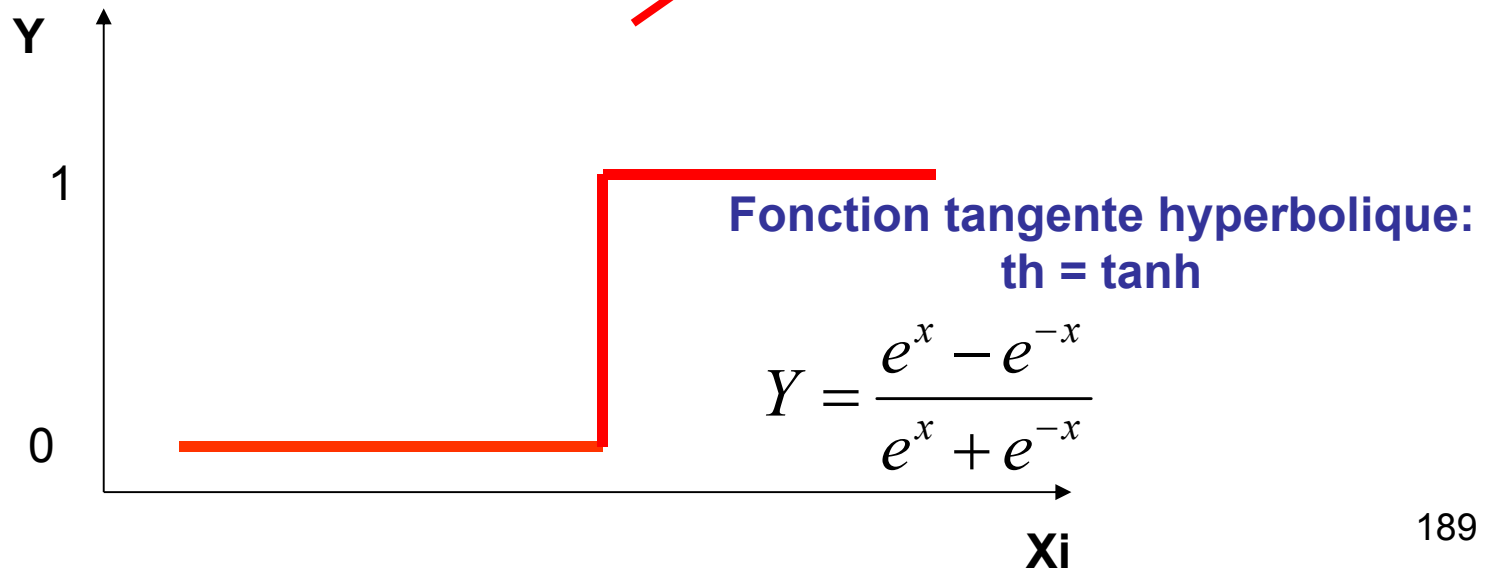
~~Neurone biologique artificiel~~



Réseau Neurons Artificiels



~~Neurone biologique artificiel~~



Réseau Neurons Artificiels

Perceptron Multicouches

Modélisation non linéaire par Apprentissage

$$y = f\left(\sum x_i\right)$$

Réseau Neurons Artificiels

Perceptron Multicouches

Modélisation non linéaire par Apprentissage

$$y = f\left(\sum x_i\right)$$

$$y = th\left(b_0 + \sum (b_i x_i)\right)$$

$$y = \frac{\exp\left(b_0 + \sum (b_i x_i)\right) - \exp\left(-b_0 - \sum (b_i x_i)\right)}{\exp\left(b_0 + \sum (b_i x_i)\right) + \exp\left(-b_0 - \sum (b_i x_i)\right)}$$

Réseau Neurons Artificiels

Perceptron Multicouches

Modélisation non linéaire par Apprentissage

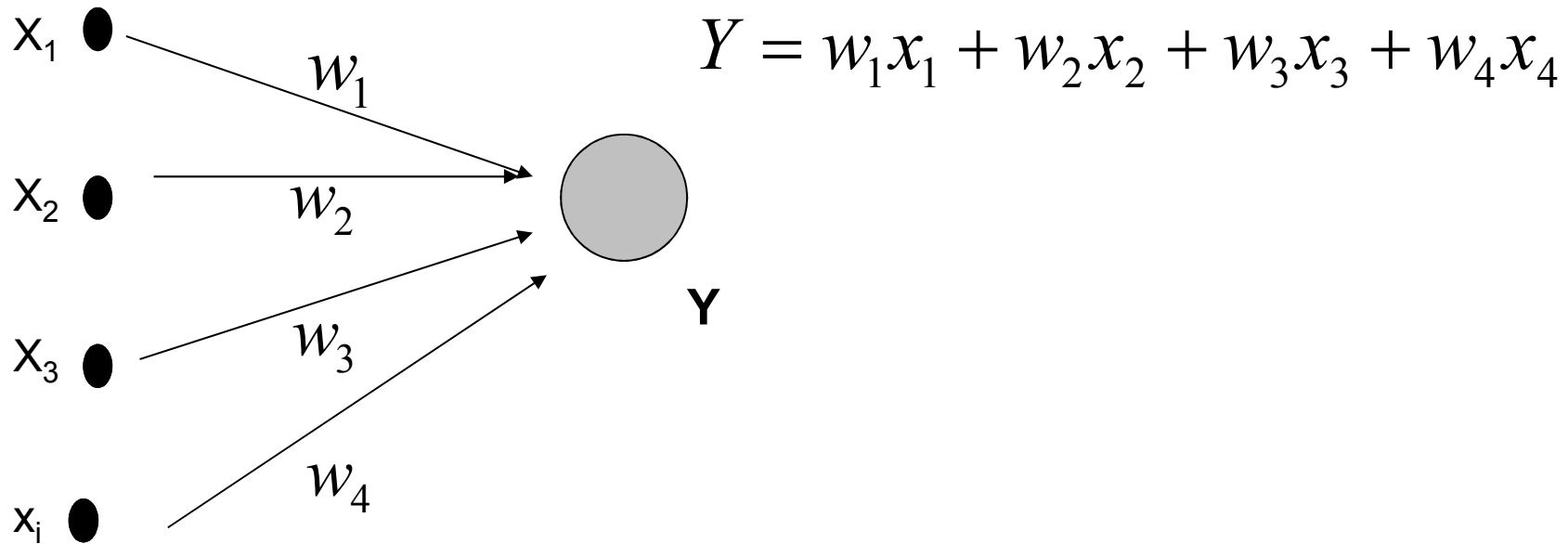
$$y = f \left(\sum x_i \right)$$

$$y = th \left(b_0 + \sum (b_i x_i) \right)$$

$$y = \frac{\exp \left(b_0 + \sum (b_i x_i) \right) - \exp \left(- b_0 - \sum (b_i x_i) \right)}{\exp \left(b_0 + \sum (b_i x_i) \right) + \exp \left(- b_0 - \sum (b_i x_i) \right)}$$

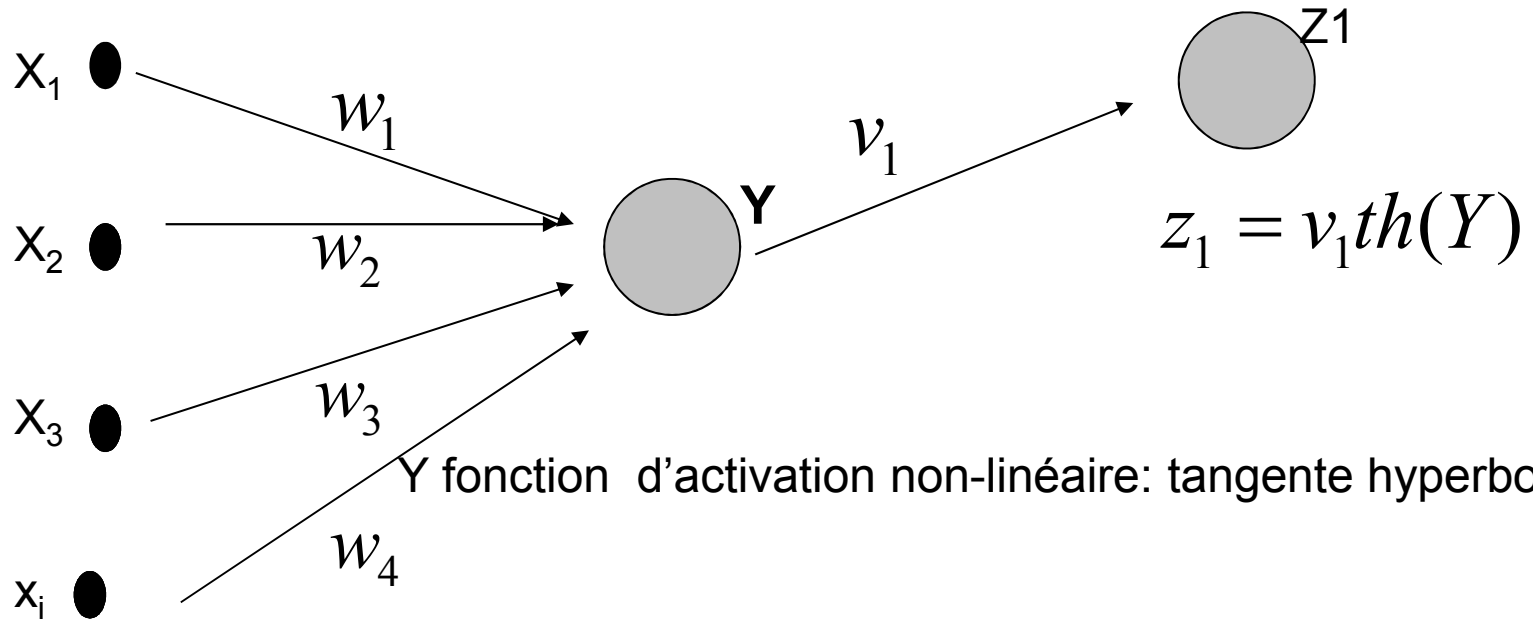
$$y = \frac{1 - \exp \left(- 2 \left(b_0 + \sum (b_i x_i) \right) \right)}{1 + \exp \left(- 2 \left(b_0 + \sum (b_i x_i) \right) \right)}$$

Perceptron Multicouches

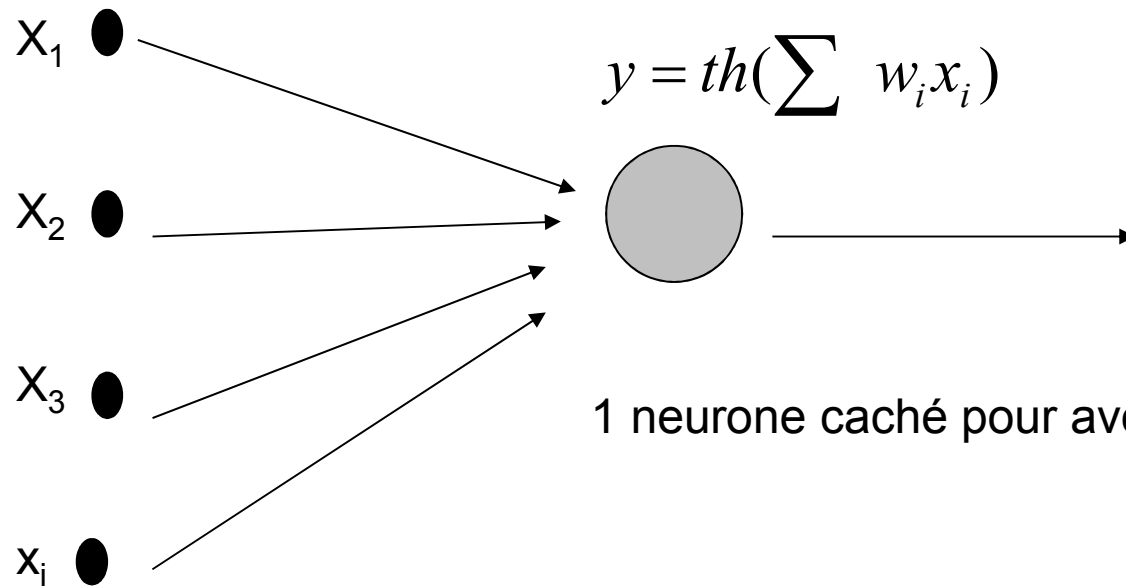


Perceptron Multicouches

$$Y = w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4$$

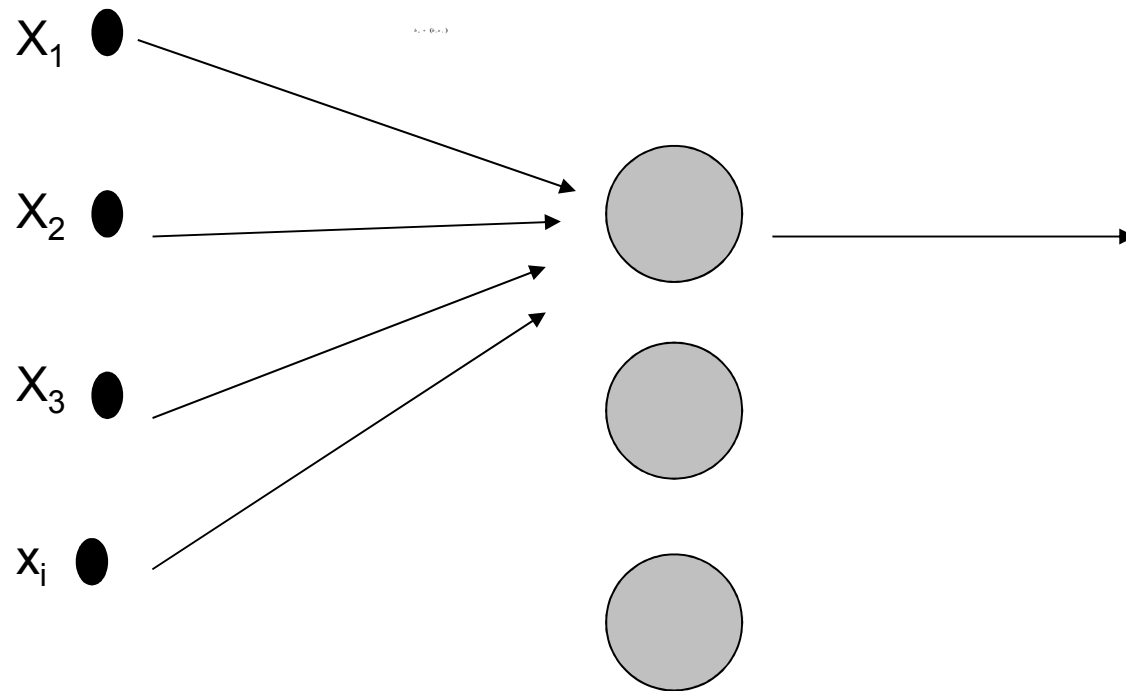


Perceptron Multicouches



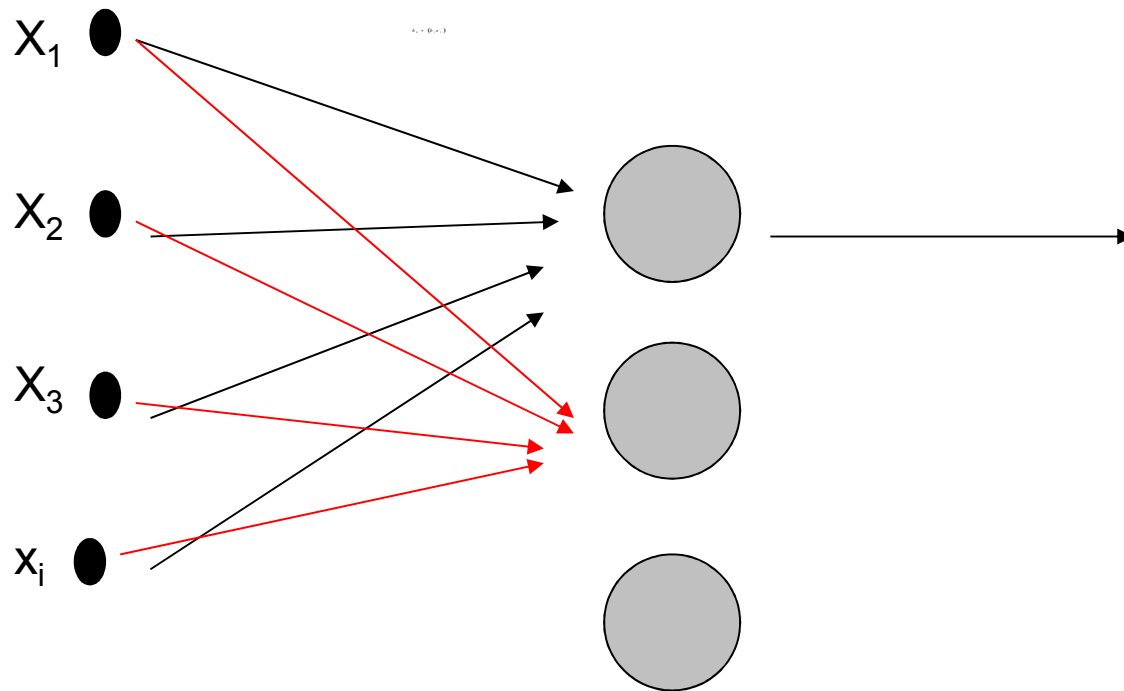
1 neurone caché pour avoir un modèle non-linéaire

Perceptron Multicouches

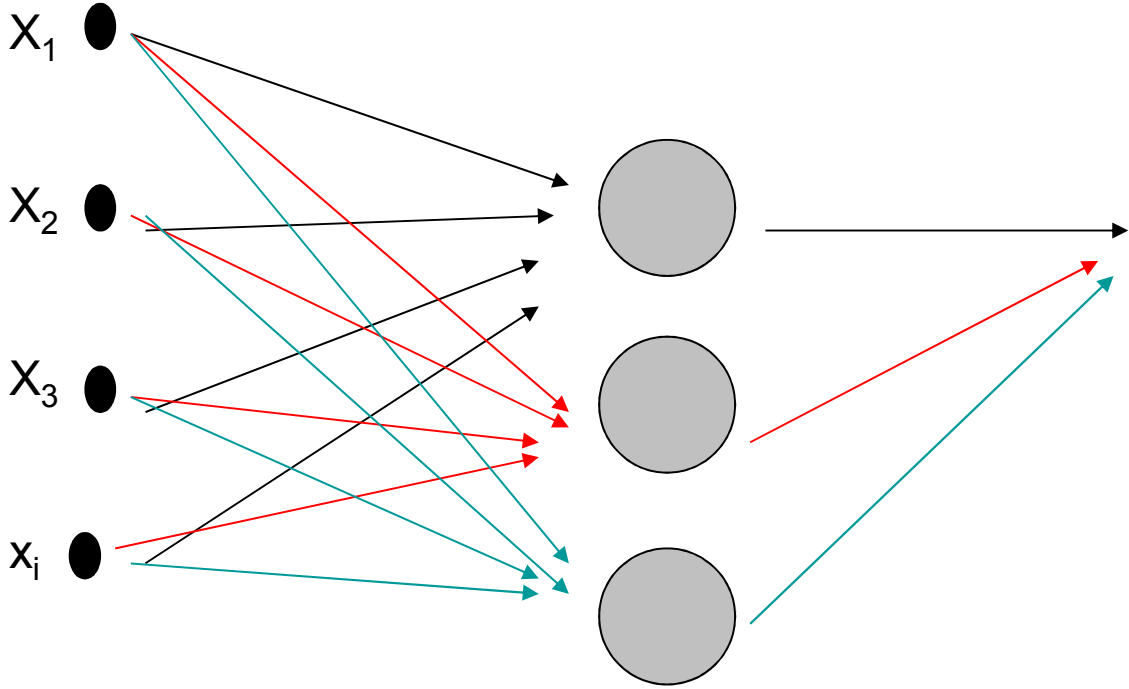


1 couche de trois neurones cachés pour avoir un modèle non-linéaire avec la même fonction d'activation

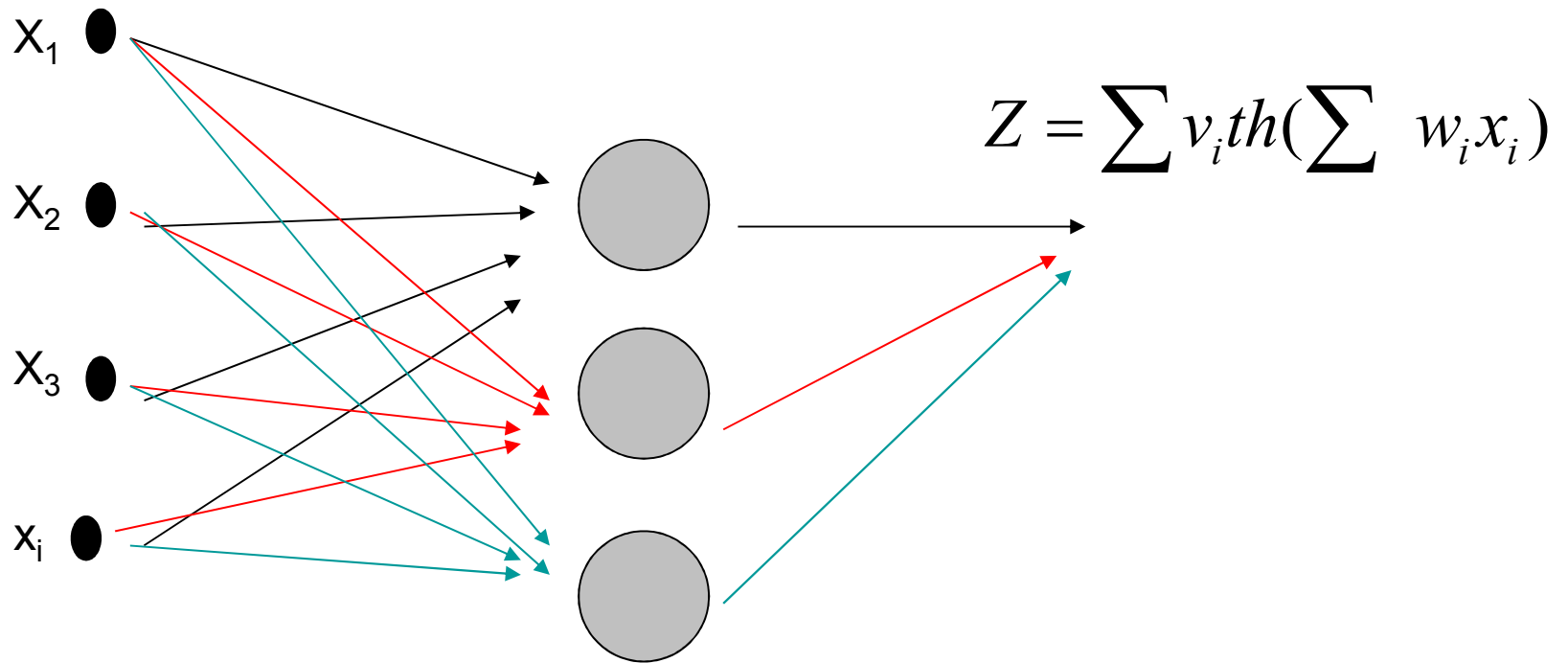
Perceptron Multicouches



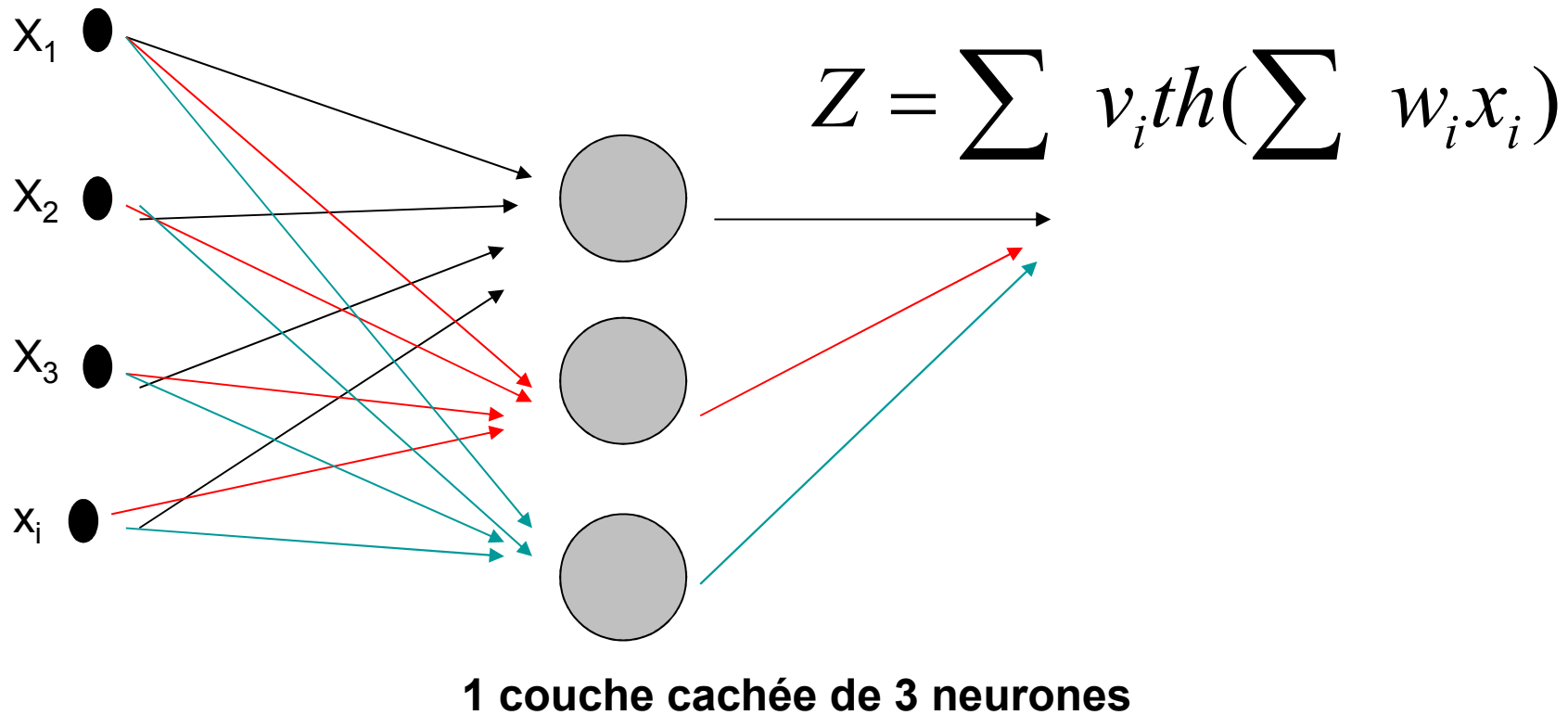
Perceptron Multicouches



Perceptron Multicouches



Perceptron Multicouches



Réseau Neurons Artificiels

Perceptron Multicouches

Découpe des données en 3 sous ensembles:

Réseau Neurons Artificiels

Perceptron Multicouches

Découpe des données en 3 sous ensembles:

- 1) Données de base \Rightarrow estimation des b_i
minimisation fonction erreur
optimisation du minimum**
- 2) Données en aveugle \Rightarrow validation du modèle**
- 3) Données de test \Rightarrow pour tester la performance du modèle retenu**

Réseau Neurons Artificiels

Exemple: application pour la prévision de l'ozone à un endroit précis sur un horizon de 24 heures

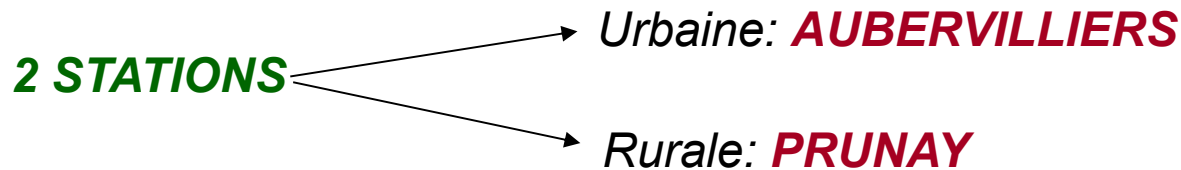
Réseau Neurons Artificiels

PREVISION DE L4OZONE APPROCHÉE PAR DES RÉSEAUX NEURONAUX

LOCALEMENT SUR UNE STATION DE MESURE

Base de données utilisée (concentrations d'ozone):

août 2000 – juillet 2001



Données météorologiques fournies par la station MONTsouris (PARIS):

- *Température*
- *Humidité relative*
- *Durée d'ensoleillement*
- *Rayonnement global*
- *Vitesse et direction du vent*

Horizon de prédiction: 24 heures

Réseau Neurons Artificiels

Fonction d'activation (neurones couche cachée): tangente hyperbolique

Rétro-propagation + Arrêt prématuré → 3 ensembles *apprentissage validation test*

Méthodes d'optimisation du gradient conjugué:

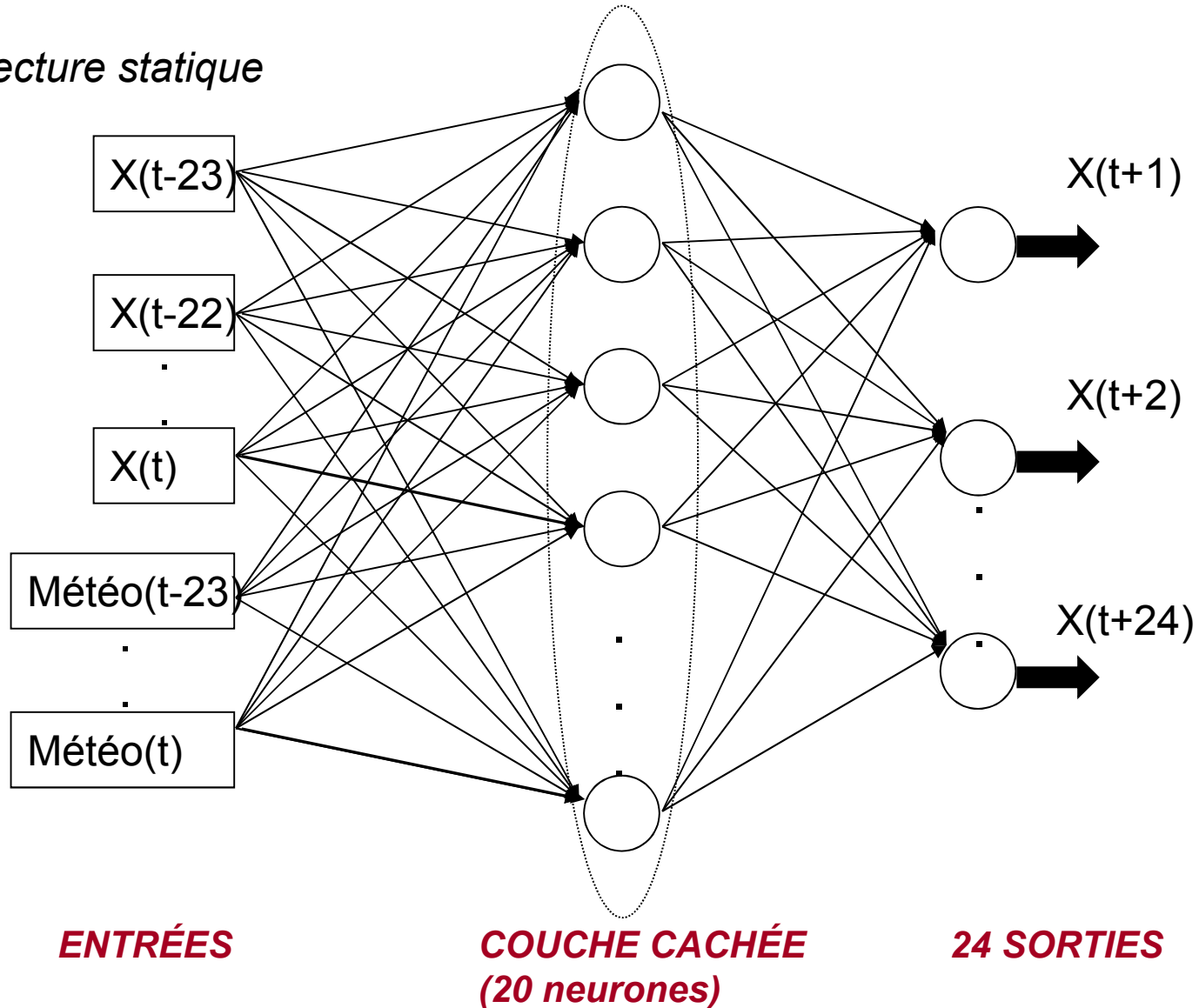
- **SCG** (Möller, 1993)
- **BFGS** (Shanno, 1978)

6 INDICES DE PERFORMANCE
calculés sur l'ensemble de test

RMSE
MAE
MBE
 d_2
 R^2
SI

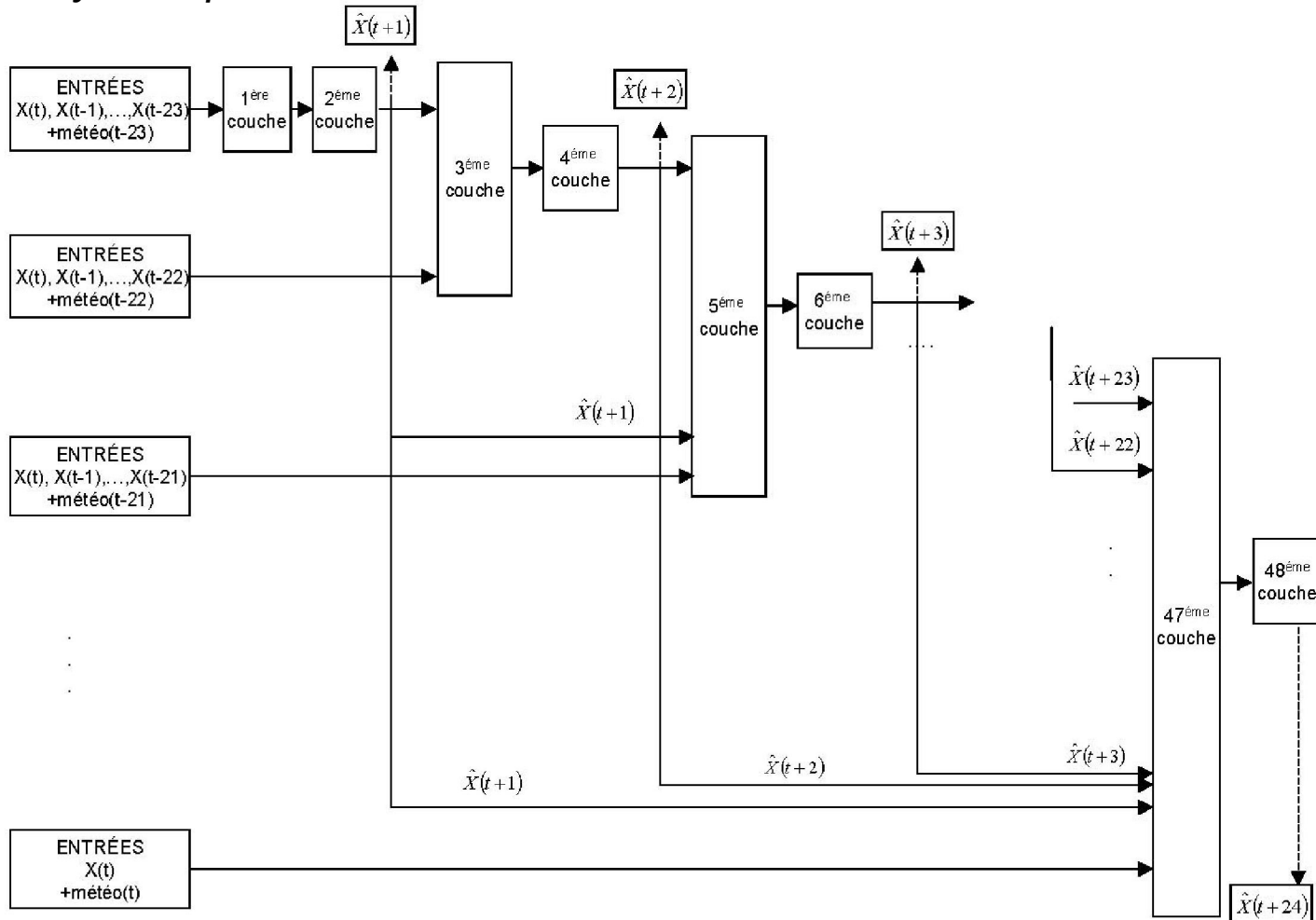
Réseau Neurons Artificiels

Architecture statique



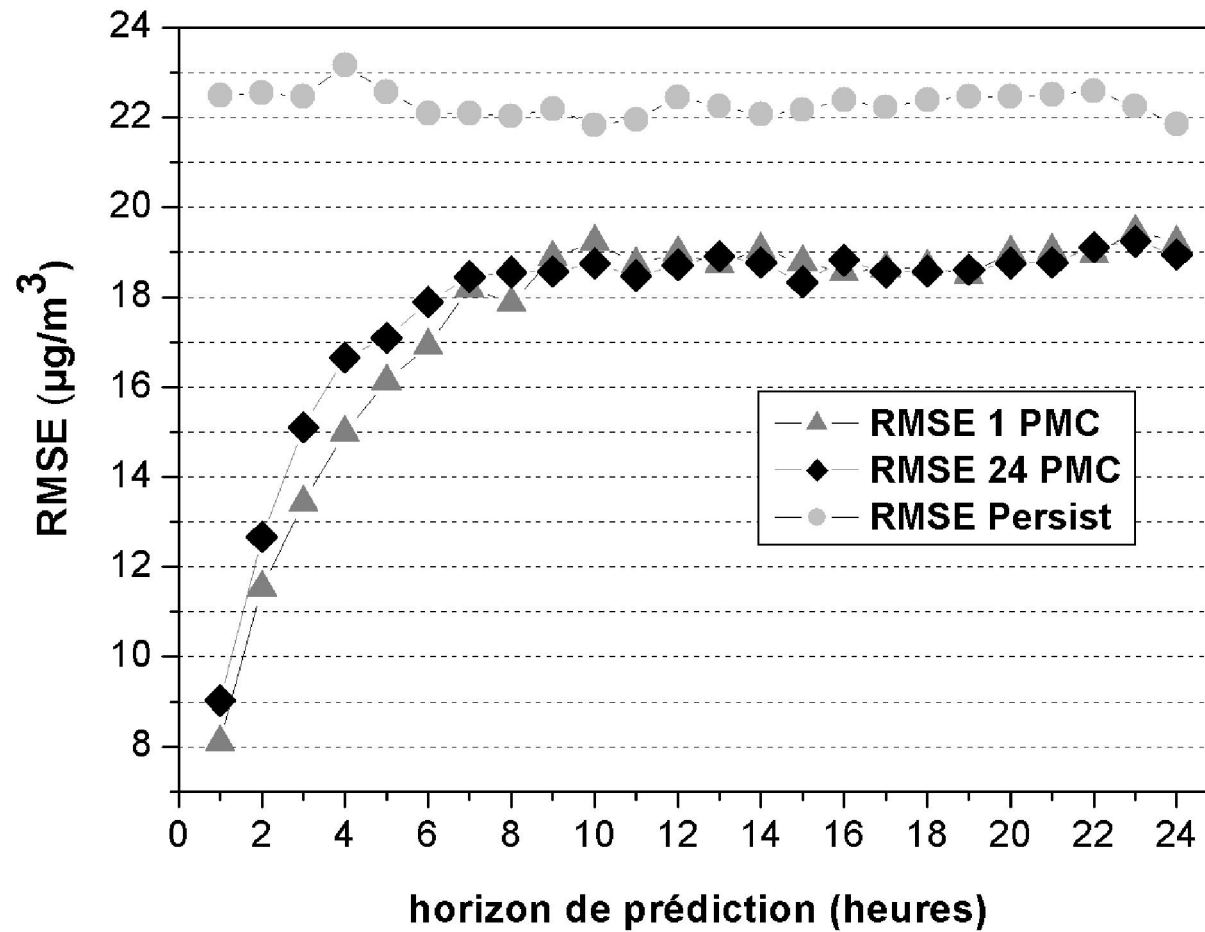
Réseau Neurons Artificiels

Architecture dynamique



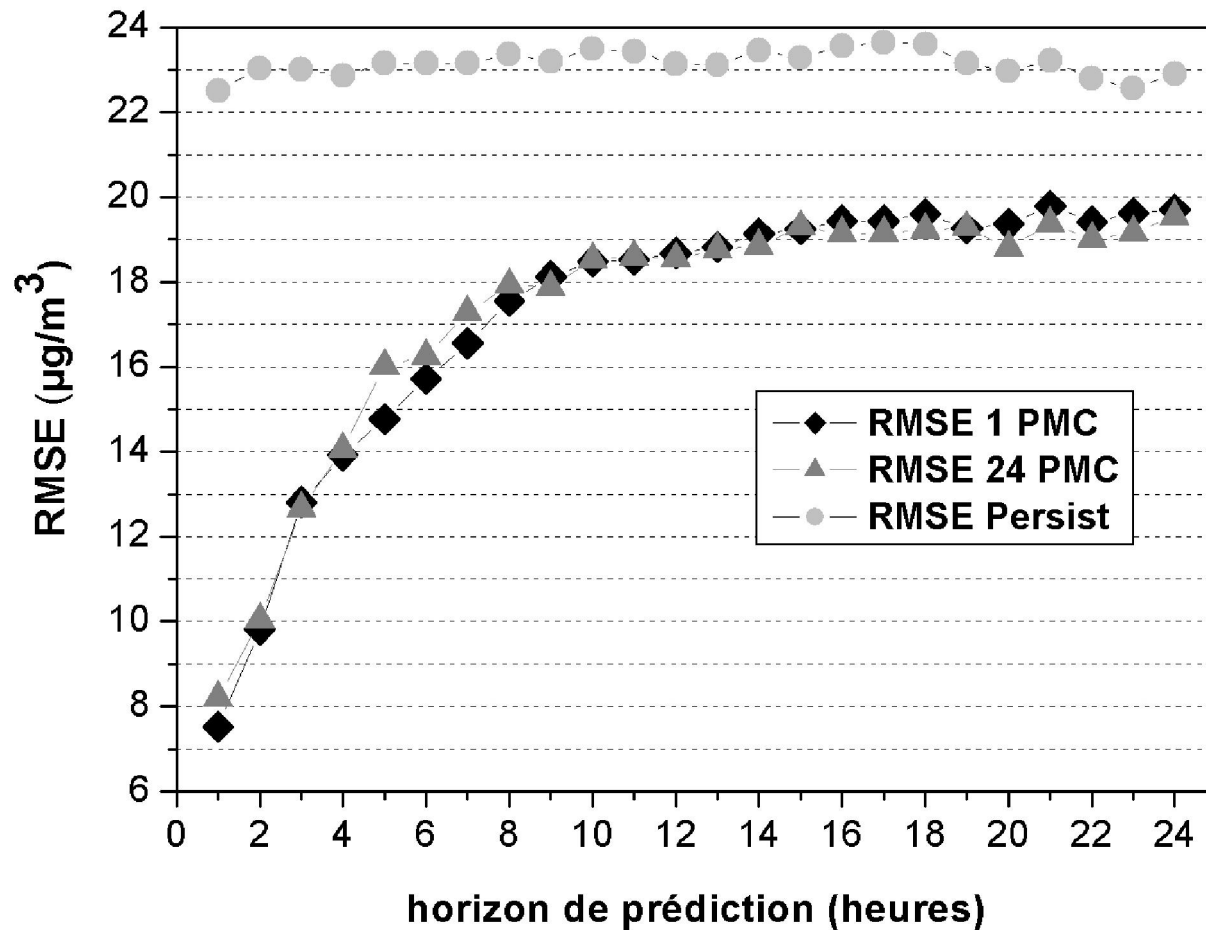
Réseau Neurons Artificiels

AUBERVILLIERS



Réseau Neurons Artificiels

PRUNAY



Puente-Lelièvre, Céline

La qualité de l'air en milieu aéroportuaire : étude sur l'aéroport Paris-Charles de Gaulle, *Thèse de doctorat de l'Université Paris 12, 2009.*

$$\text{NO}_x = f(\text{T}_{\text{air}}, \text{P}, \text{DV}, \text{X}, \text{Y}, \text{Z})$$

Afif, Charbel

L'acide nitreux atmosphérique : implications dans la chimie en milieu urbain, *Thèse de doctorat de l'Université Paris 12, 2008.*

ACP stations de NO₂ à Beyrouth

Rude, Julien

Développement d'un modèle statistique neuronal pour la description fine de la pollution par le dioxyde d'azote: application à la région parisienne, *Thèse de doctorat de l'Université Paris 12, 2008.*

$$\text{NO}_2 = f_{\text{neuronal}}(\text{DV}, \text{Orientation rues}, \text{I}_{\text{rues}}, \text{L}_{\text{rues}}, \text{Emissions}_{\text{rues}})$$

Rude, Julien

Développement d'un modèle statistique neuronal pour la description k_{OH} en fonction des caractéristiques moléculaires, *Atm. Env.*, 2008.

COV + OH \rightarrow Produits secondaires

k_{OH}

$k_{OH} = f_{\text{neuronal}}$ (types liaisons, fonctions, radicaux ...)

Chervier, Frédéric

Modélisation des variations basse fréquence des émissions de COVB à l'aide d'un réseau de neurones artificiels, *Thèse de doctorat de l'Université Paris VII*, 2005.

COV = f_{neuronal} ($T_{\text{air,sol}}$, P, HR, x, y, z)

Algorithmes de minimisation

De nombreux problèmes nécessitent de minimiser une fonction :

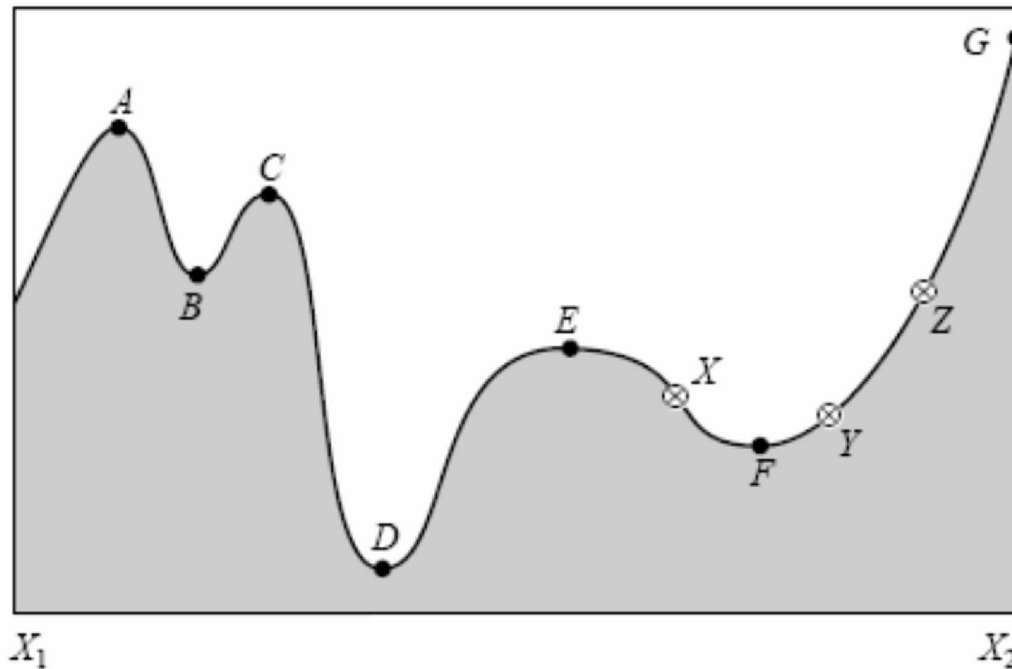
- Minimiser la distance (XHI2) entre des points de mesures et une courbe
 - Trouver l'état d'équilibre d'un système mécanique (Minimiser E_{pot})
 - Trouver l'état d'équilibre d'un gaz, d'un mélange (Maximiser Entropie)
 - Tous problèmes d'optimisation (minimiser le coût, certaines fonctions etc...)
- Etc....

Tous ces problèmes entrent dans une grande catégorie : L'optimisation

Remarque : Maximiser $F(x)$ = Minimiser $-F(x)$

Donc minimiser et maximiser sont en fait le même problème !

Une fois de plus, c'est un problème mal posé :
Il n'y a pas de méthode idéale.
Il est facile de trouver un minimum local.
Il est difficile de trouver un minimum absolu



B, D, F sont des minimums locaux

D est le minimum absolu

Cette recherche est d'autant plus longue et difficile si la fonction dépend de plusieurs variables => minimisation en N dimensions

Certains problèmes sont dit : « contraints » quand on doit minimiser une fonction et en plus, respecter certaines conditions sur les variables:

Ex 1 : Réaction chimique à l'équilibre :

Pour trouver la composition à l'équilibre on doit minimiser le potentiel chimique du système MAIS on doit respecter le fait que la masse totale du système est conservée...

Ex 2 : Calculer la forme d'un récipient de surface minimale qui contient un volume fixé... (ex : boîtes de conserve)

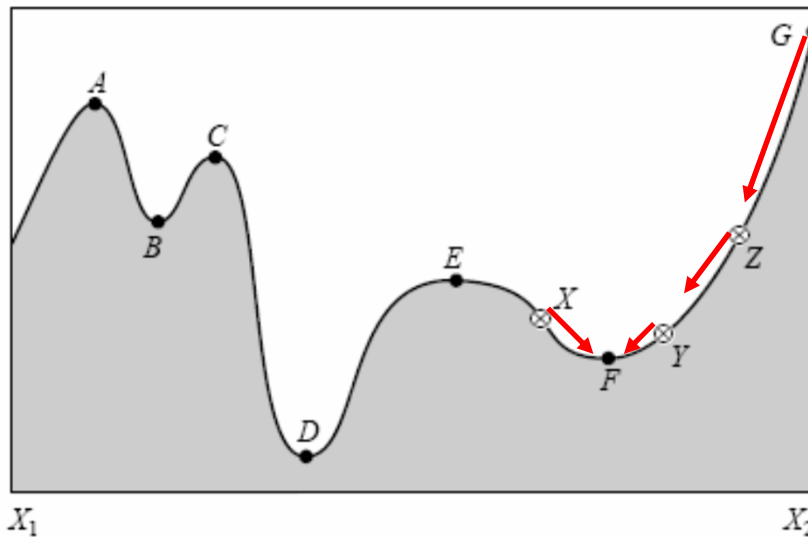
Etc....

Il existe aussi des méthodes spécifiques pour ces cas.

La méthode classique de minisation est la descente de gradient :

On part d'un point X_0 donné par l'utilisateur

On descend le long de la plus grande pente locale



C'est pourquoi la plus-part des algorithmes trouvent des *minimums locaux* et non le *minimum absolu*

Une grand partie du calcule numérique consiste à trouver la dérivée $F'(x)$

IMPORTANCE DE $F'(X)$

Soit $F(x)$ à minimiser (x peut être un vecteur)

En général, si vous connaissez $F'(X)$ (le gradient) cela pourra vous faciliter la tâche.

Si vous utiliser un algorithme de minimisation cela signifie que vous ne pouvez calculer analytiquement les zéros de la fonction $F'(X)$

- Soit par ce que vous ne connaissez pas $F'(x)$
- Soit par ce que résoudre $F'(X)=0$ est trop compliqué.

D'une manière générale, si vous pouvez calculer $F'(X)$ faites le, cela facilitera grandement la tâche de l'algorithme de minimisation.

Minimisation à 1D en utilisant le gradient (dérivée)

En 1 dimension, en fait le gradient n'apporte qu'une information assez limitée...

Pourquoi ?

A 1D le domaine de recherche est assez étroit, donc on gagne peu.

3 possibilités

1

On utilisera la méthode d'encadrement à trois points $[a,b,c]$, et la dérivée au point c , $f'(x)$ nous permettra de choisir plus rapidement la direction du prochain point (dans $[a,b]$ ou dans $[b,c]$)

2 *

On descend le long du gradient à pas fixe

3 *

Si on connaît la dérivée seconde, on peut descendre le gradient avec un pas optimal

- ***Les méthodes 2 et 3 sont intéressantes pour les fonctions à plusieurs dimensions***

MINIMISATION AVEC DESCENTE DE GRADIENT, On connaît la dérivée

On connaît F et $F'(X)$

On part d'un point de départ X_0

On calcule la suite

$$X_{k+1} = X_k + d_k g_k$$

Où g_k est la direction de descente et d_k est le pas de descente

Le choix de g_k est d_k se fera de sorte que $F(X_{k+1}) < F(X_k)$

$$g_k = -\left. \frac{df}{dx} \right|_{X_k}$$

Dans la méthode la plus simple de descente de gradient, g est simplement – le gradient de F .

Minimisation multi-dimensionnelles

Minimiser $F(X)$ où $X=(x_1 \dots , x_n)$

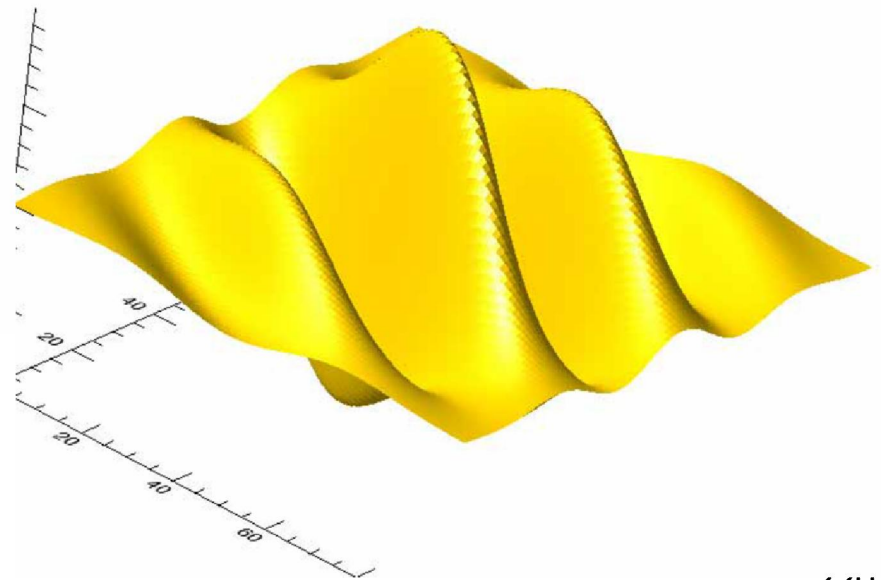
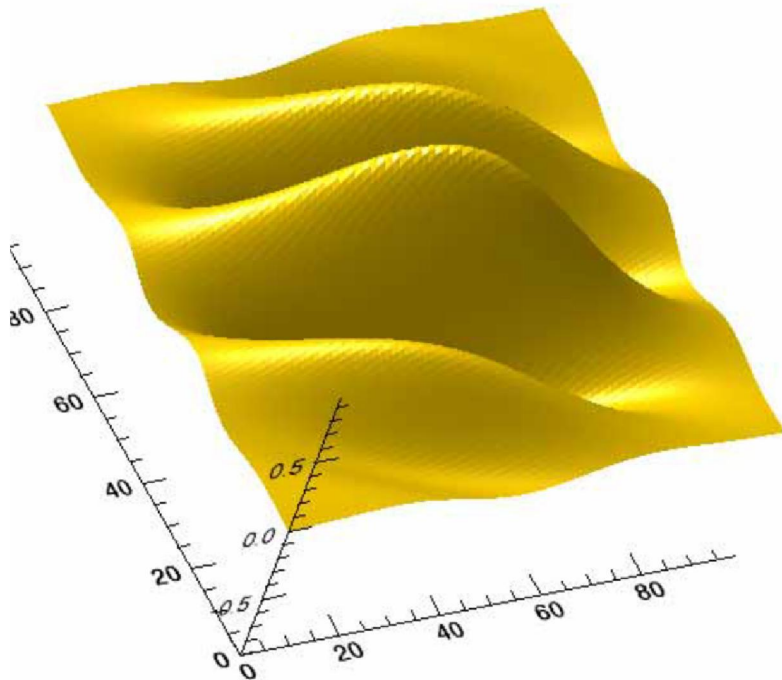
Le problème est beaucoup plus retord ...et beaucoup plus long à résoudre

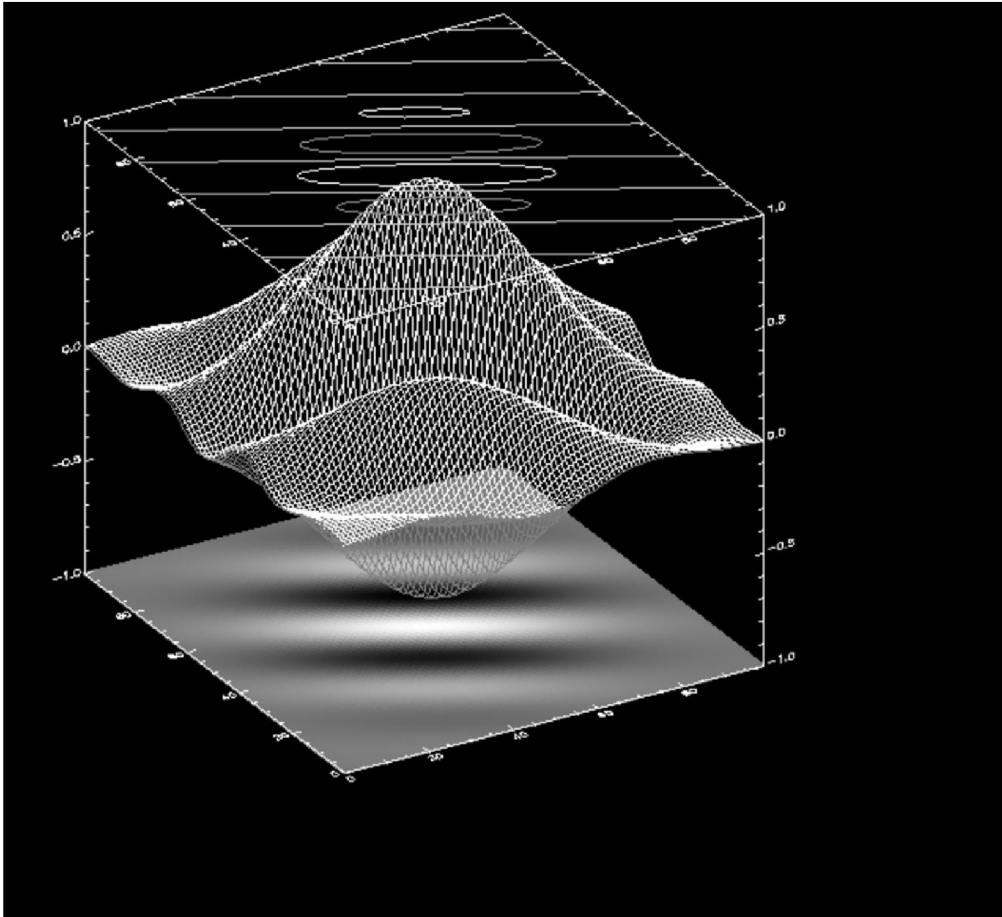
Le grand problème est souvent le CHOIX de la direction de descente

Une fois que l'on sait dans quelle direction on descend, l'autre problème est de savoir la *longueur* de la descente.

En N dimensions, ces choix sont cruciaux.

Exemple de fonction à minimiser, en 2D seulement $F(x,y)$





Ici : plusieurs minimums,
Existence de « vallées » étroites,
Maximums et minimums locaux
Etc...

Plusieurs techniques pour attaquer un tel problème. Aucune n'est universellement efficace.

TOUTES nécessitent un point de départ choisi par l'utilisateur X_0

=> Nécessité de partir d'un point X_0 pas très éloigné du minimum

La plupart utilisent le DL de la fonction à plusieurs variables (N variables)

$X = P + [x_1, \dots, x_n]$ où P est un vecteur

$$f(\mathbf{x}) = f(\mathbf{P}) + \sum_i \frac{\partial f}{\partial x_i} x_i + \frac{1}{2} \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j} x_i x_j + \dots$$
$$\approx c - \mathbf{b} \cdot \mathbf{x} + \frac{1}{2} \mathbf{x} \cdot \mathbf{A} \cdot \mathbf{x}$$

$$c \equiv f(\mathbf{P}) \quad \mathbf{b} \equiv -\nabla f|_{\mathbf{P}} \quad [\mathbf{A}]_{ij} \equiv \left. \frac{\partial^2 f}{\partial x_i \partial x_j} \right|_{\mathbf{P}}$$

\mathbf{b} est le gradient

\mathbf{A} est appelé le « Hessien » de $f(x)$. C'est simplement la matrice des dérivées secondes au point P .

Les méthodes CLASSIQUES descendent les pentes.
Elles se divisent en deux grandes familles :

Méthodes à 1 point : utilisent le gradient , ou le Hessien (si on les connaît)

- Descente de gradient à pas fixe (Gradient)
- Descente de gradient à pas optimal (Gradient + Hessien) ou Gradients conjugués (Gradient + Hessien)

Méthodes à plusieurs points : pas de gradient ni de Hessien

- Méthode Simplex (aussi appelée « Amibe », « Amoeba » en anglais)

Méthodes mixtes :

- Relaxation

Toutes ces méthodes convergent vers un minimum local

Méthode pour trouver un minimum global :
« Simulated Annealing » en anglais
« Recuit simulé » en français
Appelé aussi « relaxation lente »

Que faire pour trouver un minimum global ?

La méthode la plus utilisée aujourd'hui est celle du « Recuit Simulé »

Cette méthode, très populaire ces dernières années a permis de résoudre des problèmes très complexes du type « voyageur de commerce » où les méthodes déterministes sont rapidement piégées dans des minima locaux.

Nous la décrirons brièvement ici.

C'est une méthode de type « Monte Carlo » donc nécessite des nombres aléatoires.

Elle s'inspire de la thermodynamique...

Méthode élégante.

Principe :

Les systèmes naturels évoluent spontanément vers un minimum d'énergie ou maximum d'entropie global :

Exemple : gaz (maximum d'entropie)

conformation spatiale d'une molécule (minimum d'énergie)

cristallisation (maximum d'entropie)

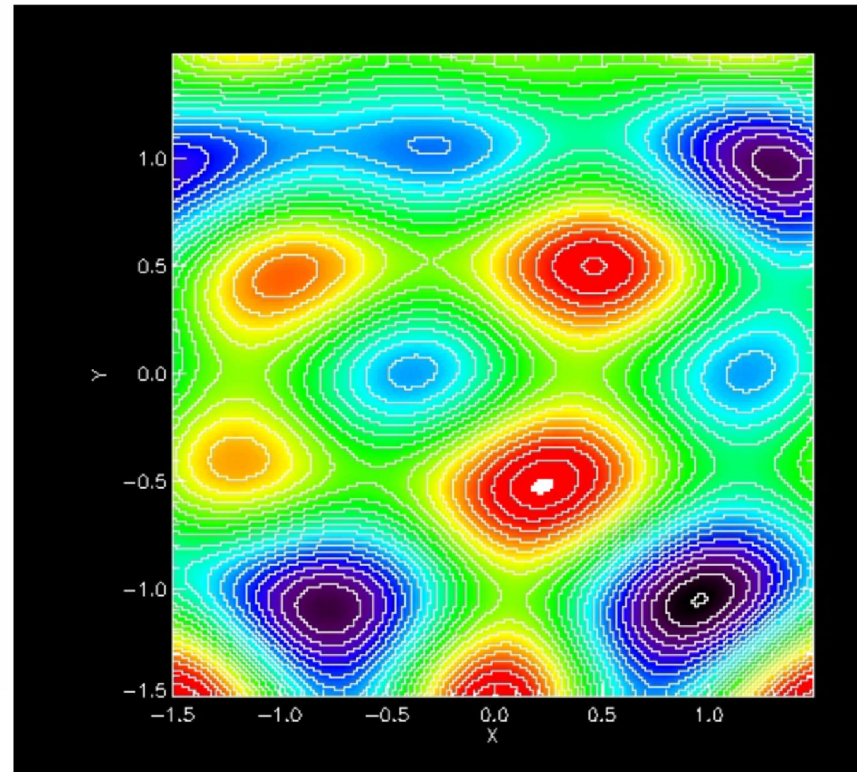
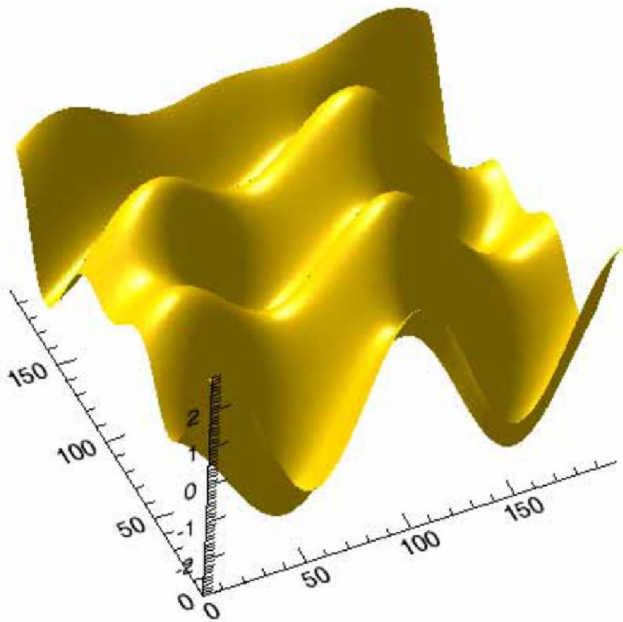
etc....

Le système explore toutes les possibilités pendant un temps (relaxation lente) puis se stabilise dans un minimum. C'est un processus de refroidissement. Plus ce refroidissement est lent, plus le système sera proche du minimum absolu.

La fonction $F(X)$ à minimiser s'appellera la fonction de « coût ».

Il faut donc minimiser le « coût »

Exemple : Voici une fonction $F(X,Y)$ dont il est très difficile de déterminer le minimum absolu. Une méthode déterministe vous donnera juste le minimum local le plus proche



La situation peut-être BIEN pire si votre fonction est en $N > 2$ dimensions !!

Maintenant inspirons nous de la thermodynamique :
En thermodynamique la probabilité d'un état d'énergie E est

$$P(E) \propto e^{-E / kT}$$

Exprimant l'idée que si un système est en équilibre thermique à la température T, alors ses états d'énergie possibles sont distribués de manière probabiliste avec une loi de probabilité qui décroît exponentiellement avec l'énergie.

Donc même si l'état de moindre énergie est le plus probable, il est toujours possible de faire des « sauts » en énergie (avec un faible probabilité) permettant de passer aux travers de minima locaux d'énergie, pour converger finalement vers le minimum global d'énergie.

Donc le système remonte parfois vers des énergies plus grandes pour redescendre vers des énergies plus faibles.

C'est la FORCE de cette approche. Une méthode directe ne peut faire cela

$$P(E) \propto e^{-E/kT}$$

La méthode du recuit-simulé reprend cette probabilité .

- **E sera la fonction de coût F : on veut minimiser F**
- **T sera un paramètre de contrôle fictif : On diminuera lentement la température T du système.**
- **La constante de Boltzmann K sera remplacée par une constante arbitraire pour que P(E) reste dans des bornes numériques accessibles par la machine.**

La probabilité de passer d'un état E_k à l'état E_{k+1} sera :

$$p(E_k \rightarrow E_{k+1}) = e^{-\frac{(E_{k+1}-E_k)}{kT}}$$

Si $E_{k+1} < E_k$ alors $P \geq 1 \Rightarrow$ la transition sera toujours acceptée. (on minimise !!!)

Si $E_{k+1} > E_k$ alors $0 < P < 1 \Rightarrow$ on tire un chiffre X entre 0 et 1. Si $X < P$ alors la transition est acceptée

En pratique, la méthode sera la suivante :

Opération
répétée N
fois à T
constant

- 1- On part d'un état numéro K , d'énergie E_k , et d'une température T
- 2- On génère aléatoirement un nouvel état, $K+1$, proche de l'état K , d'énergie E_{k+1} ($\Delta E \sim KT$)
- 3- On regarde maintenant si on accepte l'état $K+1$
Pour cela on calcule la probabilité P de passe de K à $K+1$:
On tire un chiffre X entre 0 et 1 (distrib. Uniforme)
Si $X > P$ alors on accepte le pas , sinon on le rejette
- 4- Quand E ne décroît plus, on diminue T et on recommence en 1

Remarque : KT est de l'ordre du ΔE acceptable. Donc plus on diminue KT plus les variations possibles de E seront faibles.

Dans l'étape 2, le générateur aléatoire doit être capable de générer de nouvelles configurations dont le ΔE typique est de l'ordre de KT ...

C'est dans l'étape 2 que se cache l'efficacité de la méthode.

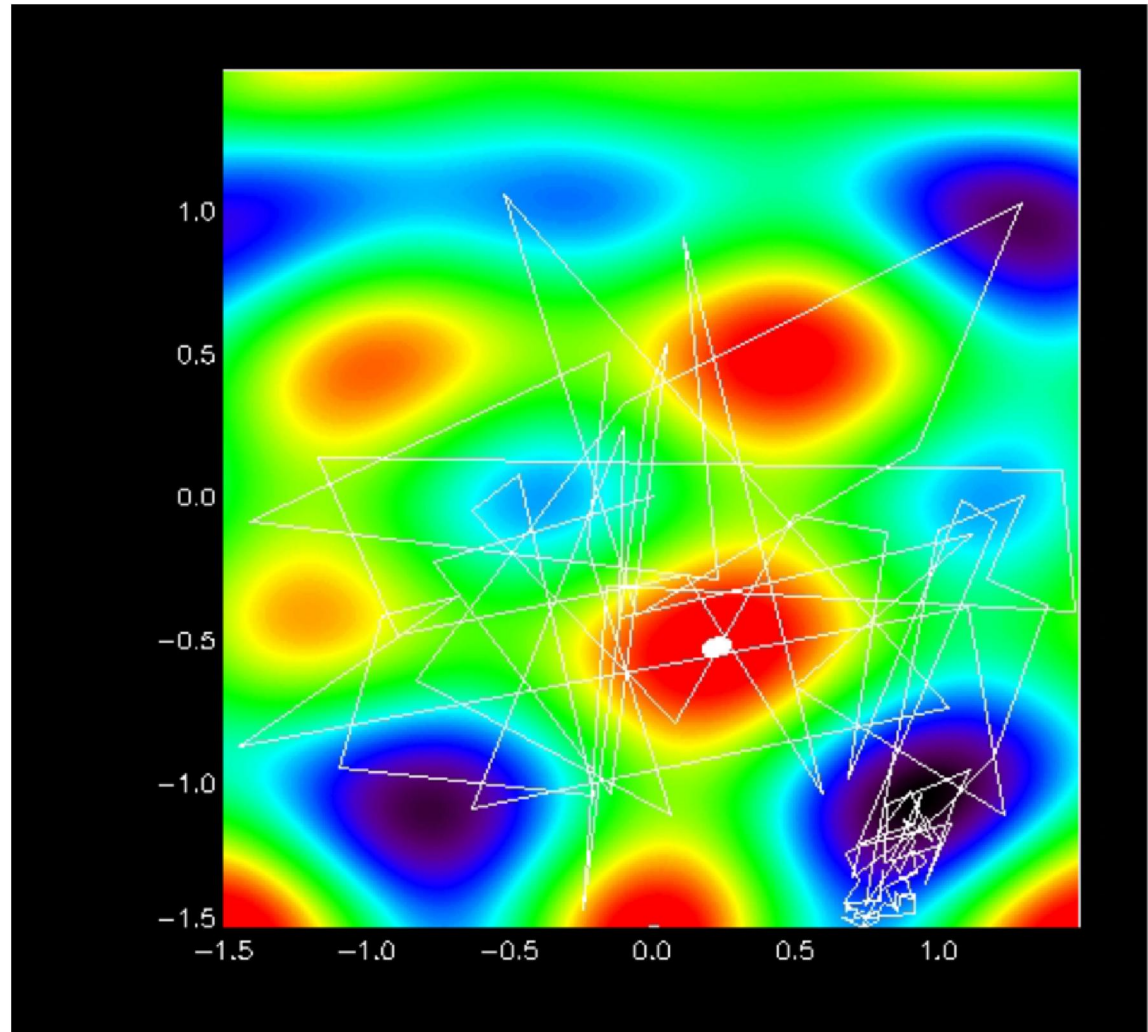
Exemple :

Point de départ:
[0,0]

Point d'arrivée :
[0.68, -1.42]

105 itérations

310 changements
acceptés



Conclusion sur le recuit simulé :

Méthode qui peut s'avérer efficace quand on cherche un minimum absolu.
Le résultat dépend **FORTEMENT** de la méthode utilisée pour générer un nouvel état aléatoire.

Très coûteuse en temps calcul => A n'utiliser que si les autres méthodes sont inefficaces.

Le méthode du « recuit simulé » est à la base d'autres méthodes de simulations physiques de type « Monté-Carlo », appelées méthodes de « Métropolis » pour simuler un gaz ou un crystal.

Alternative au recuit simulé : Algorithmes génétiques.

Conclusion sur la minimisation sans contrainte

Le problème est de minimiser une fonction $F(X)$ à N dimensions sans contrainte sur les X .

- Toutes les méthodes nécessitent un point de départ X_0 .
- Les méthodes déterministes convergent vers le *minimum local* le plus proche.
- Plus vous en saurez sur la fonction (Gradient, Hessien) plus la minimisation sera efficace. (gradient conjugué, Fletcher-Reeves)
- Les méthodes à N dim. essaient TOUTES de ramener le pb à 1 dimension : on minimise successivement sur des directions données

Si la fonction est « parasitée » de nombreux minima locaux... alors il faut utiliser une méthode de type « relaxation lente » telle que le « Simulated Annealing ». C'est une méthode Monte Carlo.

Mais le prix à payer est ...

- (1) un temps de calcul très grand
- (2) Il faut trouver un générateur de nouveaux états efficace et là il n'y a de « recette pratique »

En résumé : Comment s'attaquer à un problème de minimisation

Soit à Minimiser la fonction $F(x_1, \dots, X_n)$ à N variables.

1. Le problème est-il contraint ou non ?

Si oui : Multiplicateurs de Lagrange ou autre

2. Le problèmes est-il simple ? (idée du minimum, absence de nombreux minima parasites)

Si oui alors on peut utiliser toutes les méthodes déterministes

3. Problème simple : Connaissez vous le gradient de F ?

Si oui, vous êtes dans le meilleur cas : utilisez un gradient conjugué par exemple (Hessien), ou Fletcher Reeves

Si non et $N=1$: Utilisez une méthode de trisection ou de quadrature

Si non et $N>1$: Seule la méthode « Amibe » (« Amoeba ») non étudiée ici peut vous aider...

4. Si le problème est complexe (Pas d'idée du minimum, nombreux minima parasites etc...)

Vous êtes dans le cas le plus difficile. Tentez une méthode Monte-Carlo telle que le recuit-simulé (« Simulated Annealing »)