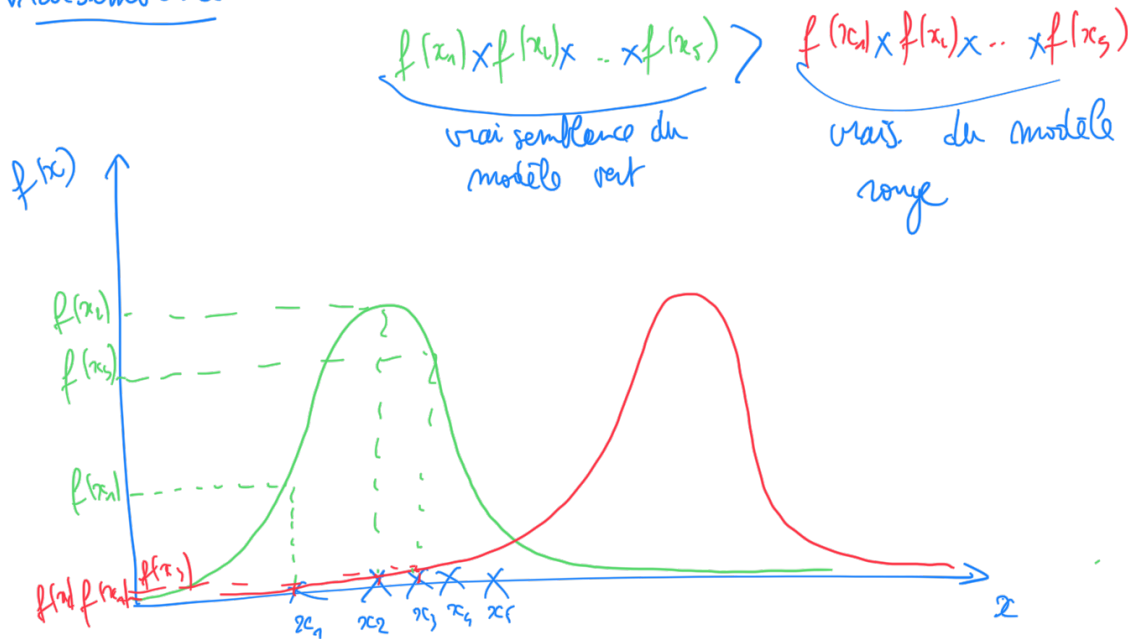


Vraisemblance



Régression (Y, X)

$(x_i, y_i)_{i=1,20}$

$$M_1: y_i = ax_i + b + \epsilon_i$$

- ⊖ mauvais ajustement, biais important $(|y_i - \hat{y}_i|)$
- ⊖ peu sensible aux fluctuations d'échantillonnage: variance

$$M_2: y_i = a_{20}x_i^{20} + a_{19}x_i^{19} + \dots + a_1x_i + b + \epsilon_i$$

😊 : biais faible

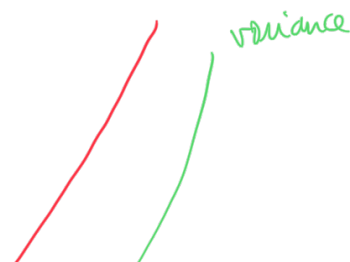
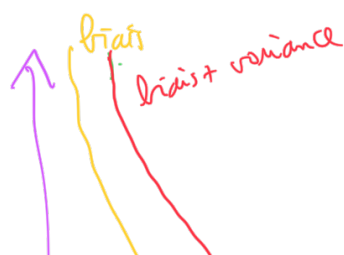
⊖ : variance trop grande
OVER-FITTING

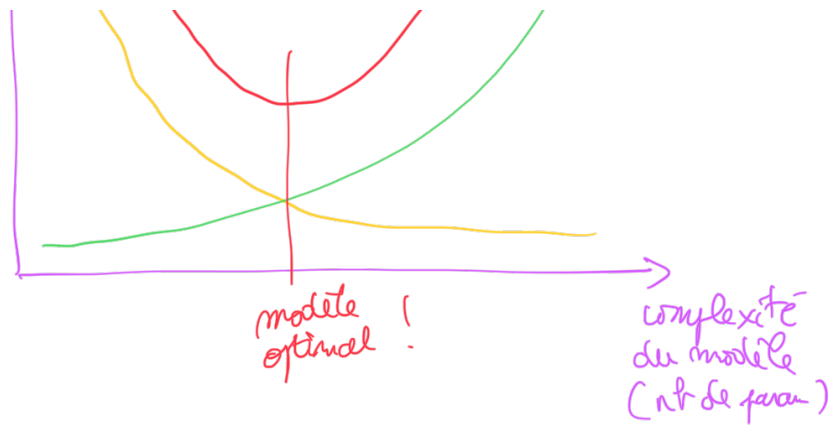
$$M_3: y_i = a_2x_i^2 + a_1x_i + b + \epsilon_i$$

😊 biais + variance faible.



On cherchera toujours un modèle avec le meilleur compromis biais variance





En pratique, pour choisir le modèle optimal, on regarde :

→ R^2 ajusté (Δ R^2 décroît avec la complexité du modèle)

→ vraisemblance pénalisée : AIC, BIC, ...
(Δ vraisemblance croît avec la complexité...)

→ erreur quadratique $RSS = \sum (y_i - \hat{y}_i)^2$ sur des données indépendantes (test, validation) des données d'entraînement (celles qui servent à estimer les paramètres des modèles)

(Δ sur les données d'entraînement, RSS croît avec la complexité, ...)

Régression linéaire

$y_i, x_{i1}, \dots, x_{ip}$
 \uparrow réponse variable explicative

modèle : $y_i = \alpha + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i$

résidu $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$

paramètres du modèle : $\alpha, \beta = (\beta_1, \dots, \beta_p), \sigma^2$

$\sim \mathcal{N}(\alpha + \beta_1 x_{i1} + \dots + \beta_p x_{ip}, \sigma^2)$

$$\Rightarrow y_i \sim \mathcal{N}(\alpha + x_{i1}\beta_1 + \dots + x_{ip}\beta_p, \sigma^2)$$

On a observé $(y_i, x_{i1}, \dots, x_{ip})$ $i=1, n$ *supposé indep.*

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \alpha \\ \vdots \\ \alpha \end{pmatrix} + \underbrace{\begin{pmatrix} x_{11} & \dots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & \dots & x_{np} \end{pmatrix}}_{X} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$$Y = \alpha \mathbb{1}_n + X \beta + \varepsilon$$

\uparrow vecteur de 1 de taille n
 \uparrow $(\beta_1, \dots, \beta_p)$

le modèle s'écrit matriciellement

$$Y = \alpha \mathbb{1}_n + X \beta + \varepsilon$$

$$Y | \alpha, \beta, \sigma^2 \sim \mathcal{N}_n \left(\underbrace{\alpha \mathbb{1}_n + X \beta}_{\text{}} , \sigma^2 \underbrace{\begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}}_{I_n} \right)$$

Q2

- La statistique bayésienne permet d'introduire de l'information dont on dispose sur le phénomène étudié en plus de celle contenue dans les données

oui ☐ non ☐

- Je peux choisir le prior que je souhaite en fonction du résultat que je veux obtenir

oui ☐ non ☐

lorsque la taille d'échantillon est grande, l'influence

- Lorsque la connaissance a priori de θ est faible
oui ☐ non ☐

- Je peux utiliser une approche bayésienne sans connaissance a priori
oui ☐ non ☐

- En statistique bayésienne, je compare des modèles pour tester mes hypothèses
oui ☐ non ☐

- Le terme $p(\theta|x) = \frac{L(x|\theta)p(\theta)}{p(x)}$ est :

la vraisemblance ☐ la loi a priori ☐ la loi a posteriori ☐

- Pour choisir le meilleur parmi deux modèles de complexité (nombre de paramètres) différente, je peux utiliser

la vraisemblance ☐

l'erreur de prédiction sur un échantillon test ☐

le critère AIC ☐

le critère BIC ☐

le critère R^2 ☐

le R^2 ajusté ☐