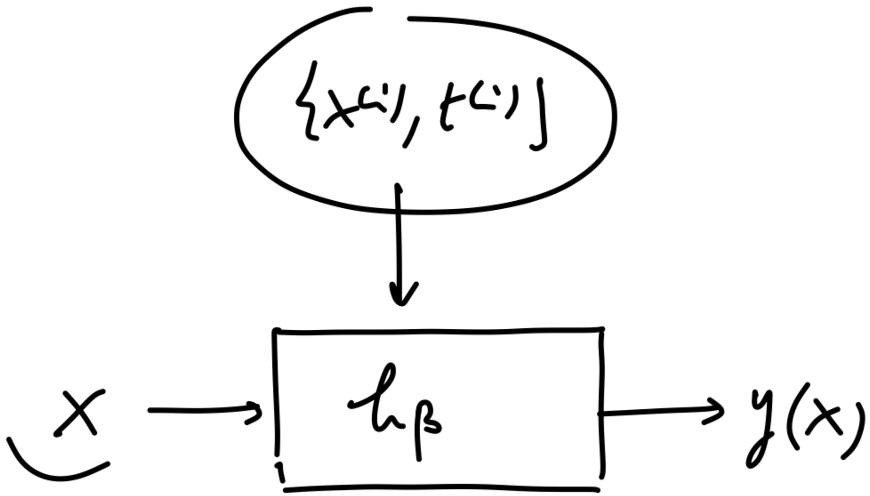
Apprentissage supervisé

$\{x^{(i)}, t^{(i)}\}_{i=1}^{N}$ $\qquad x^{(i)} \in \mathbb{R}^D \leftarrow$

$x^{(i)} = [x_1, x_2, x_3, \dots x_D]$

Apprentissage



$h_\beta(x^{(i)}) \approx t^{(i)}$

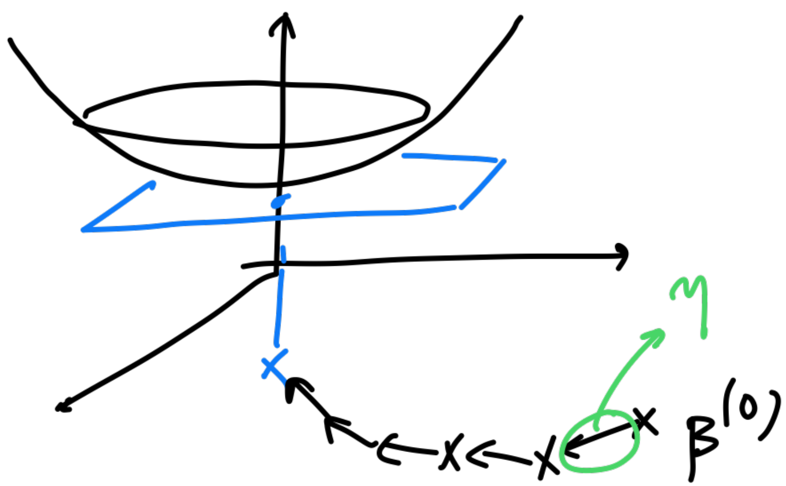
$$\ell(\beta) = \frac{1}{N} \sum_{i=1}^{N} \left(t^{(i)} - h_\beta(x^{(i)}) \right)^2$$

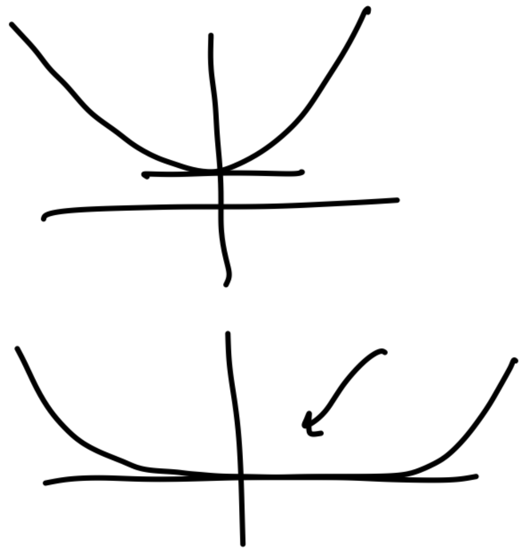
Approche classique : Combinaison linéaire / Regression linéaire

$$h_\beta(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_D x_D$$

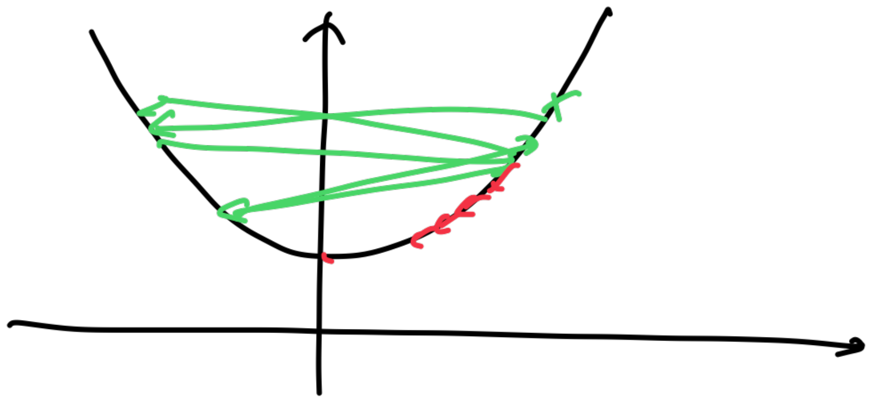
$$\ell(\beta) = \frac{1}{N} \sum_{i=1}^{N} \left(t^{(i)} - \left(\beta_0 + \beta_1 x_1^{(i)} + \dots + \beta_D x_D^{(i)} \right) \right)^2$$

$$\beta^* = \arg\min \frac{1}{N} \sum_{i=1}^{N} \left(t^{(i)} - \left(\beta_0 + \beta_1 x_1^{(i)} + \dots + \beta_D x_D^{(i)} \right) \right)^2$$





take $\beta_i^{(0)}$ aléatoire ;



$iter = 0$

while $iter < maxIter$

$$\beta \leftarrow \beta - \eta \,\text{gradient}_\beta$$

$iter{+}{+}$

$$l(\beta) = \frac{1}{N} \sum_{i=1}^{N} \left(t^{(i)} - \left(\beta_0 + \beta_1 x_1^{(i)} + \dots + \beta_D x_D^{(i)} \right) \right)^2$$

$$\text{gradient}_\beta = \left[\frac{\partial l}{\partial \beta_0}, \dots, \frac{\partial l}{\partial \beta_D} \right]$$

$$\frac{\partial l}{\partial \beta_0} = \frac{2}{N} \sum_{i=1}^{N} \left(t^{(i)} - \left(\beta_0 + \beta_1 x_1^{(i)} + \dots + \beta_D x_D^{(i)} \right) \right) (-1)$$

$$\frac{\partial l}{\partial \beta_j} = \frac{2}{N} \sum_{i=1}^{N} \left(t^{(i)} - \left(\beta_0 + \beta_1 x_1^{(i)} + \dots + \beta_D x_D^{(i)} \right) \right) \left(-x_j^{(i)} \right)$$

$$\mathrm{grad}_\beta = \left[\frac{\partial l}{\partial \beta_0}, \frac{\partial l}{\partial \beta_1}, \dots, \frac{\partial l}{\partial \beta_D} \right]$$

$$\tilde{x} = [1, \bar{x}]$$

$$h_\beta(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_D x_D$$
$$= \underline{\vec{\beta}^T \tilde{x}}$$

$$\mathrm{grad} = \frac{2}{N} \sum_{i=1}^{N} \left(t^{(i)} - \beta^T \tilde{x}^{(i)} \right) \cdot \left(-\tilde{x}^{(i)} \right) \leftarrow$$

$$\vec{t} = \left.\begin{bmatrix} t^{(1)} \\ \vdots \\ t^{(N)} \end{bmatrix}\right\} N \qquad \underset{=}{\tilde{X}} = \underbrace{\left.\begin{bmatrix} 1 & \text{---} & \vec{x}^{(1)} & \text{---} \\ \vdots & \text{---} & \vec{x}^{(2)} & \text{---} \\ \vdots & & \vdots & \\ 1 & \text{---} & x^{(N)} & \text{---} \end{bmatrix}\right\} N}_{D+1} \qquad \vec{\beta} = \left.\begin{bmatrix} \beta_0 \\ \vdots \\ \beta_D \end{bmatrix}\right) D+1$$

$$\vec{e} = \vec{t} - \underset{=}{\tilde{X}} \vec{\beta} \qquad \underbrace{e_i = t^{(i)} - \vec{\beta}^T x^{(i)}}$$

$$(AB)^T = B^T A^T$$

$$\ell = \frac{1}{N} \vec{e}^T \vec{e} \qquad \vec{e}^T \vec{e} = \sum_{i=1}^{N} e_i^2$$

$$= \frac{1}{N} (\vec{t} - \tilde{X}\vec{\beta})^T (\vec{t} - \tilde{X}\vec{\beta})$$

$$= \frac{1}{N} \left(\cancelto{0}{\vec{t}^T \vec{t}} + \vec{\beta}^T \tilde{X}^T \tilde{X} \beta - \overbrace{\underbrace{\vec{t}^T \tilde{X} \beta}} - \overbrace{\underbrace{\vec{\beta}^T \tilde{X}^T \vec{t}}}^{\vec{t}^T \tilde{X} \vec{\beta}} \right)$$

$$\frac{\partial}{\partial \beta}\Big(\underbrace{\vec{\beta}^T \tilde{X}^T \tilde{X} \vec{\beta}}_{2\tilde{X}^T\tilde{X}\beta} \; \underbrace{- 2 \vec{t}^T \tilde{X} \beta}_{-2\tilde{X}^T\vec{t}} \Big)$$

$$\frac{\partial}{\partial w}\left(v^T w\right) = \left[\frac{\partial}{\partial w_0}, \dots \frac{\partial}{\partial w_D}\right] \qquad v^T w = v_0 w_0 + v_1 w_1 + \dots + v_D w_D$$

$$= \left[v_0, \dots, v_D\right]$$

$$= \vec{v}$$

$$\frac{\partial}{\partial w}\left[w^T A w\right] = \left[w_1 \; w_2\right]\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = a_{11} w_1^2 + a_{22} w_2^2 + 2a_{12} w_1 w_2$$

$$\frac{\partial}{\partial w_1} = 2a_{11} w_1 + 2a_{12} w_2 \qquad \frac{\partial}{\partial w_2} = 2a_{22} w_2 + 2a_{12} w_1$$

$$\begin{bmatrix} \frac{\partial}{\partial w_1} \\ \frac{\partial}{\partial w_2} \end{bmatrix} = 2 \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

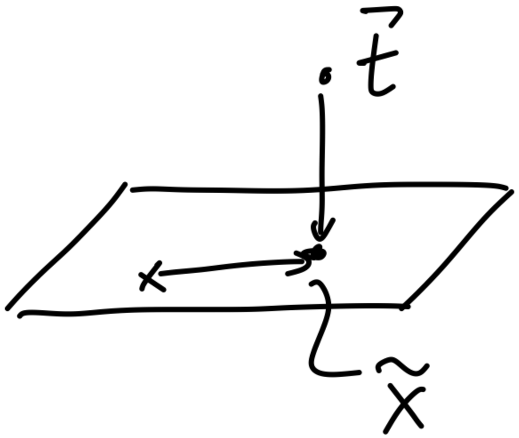
$$= 2Aw$$

$$2\tilde{X}^T\tilde{X}\beta - 2\tilde{X}^T\vec{t} = 0$$

$$\tilde{X}^T\tilde{X}\beta = \tilde{X}^T\vec{t} \quad \text{(Forme close)}$$

$$\beta = (\tilde{X}^T\tilde{X})^{-1}\tilde{X}^T\vec{t}$$

$$\ell(\beta) = \| \vec{t} - \tilde{X}\vec{\beta} \|_2^2 = \| \vec{e} \|^2 = \sum_{i=1}^{N} e_i^2 \qquad \|$$



$\rightarrow$ Orthogonalisation successive (Gram Schmidt)

Start z_0 = première colonne de $\tilde{X}$ (première caractéristique)

For $j = 1, \dots, D$

Compute the coefficients $\gamma_{\ell j} = \dfrac{\langle c_j, z_\ell \rangle}{\langle z_\ell, z_\ell \rangle}$

for $\ell = 0, \dots, j-1$

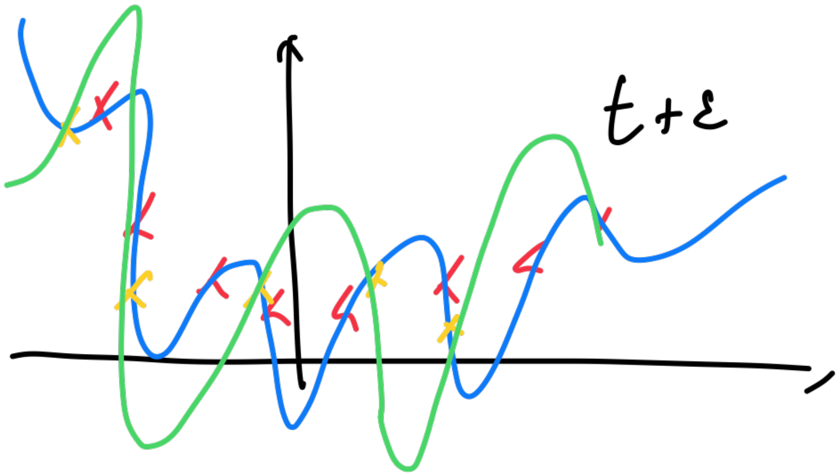
Define the new z_j as

$$z_j = c_j - \sum_{k=0}^{j-1} \gamma_{kj} z_k$$

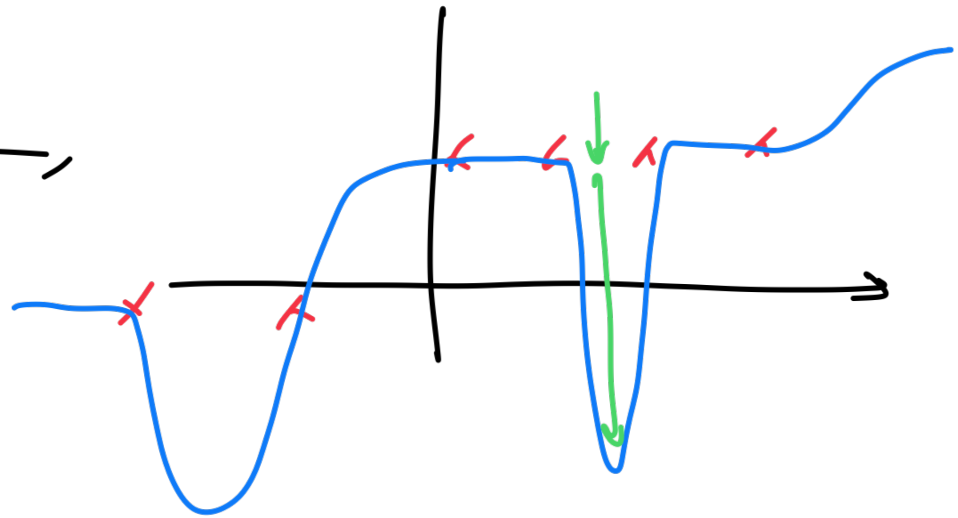
$c_D = x_D$

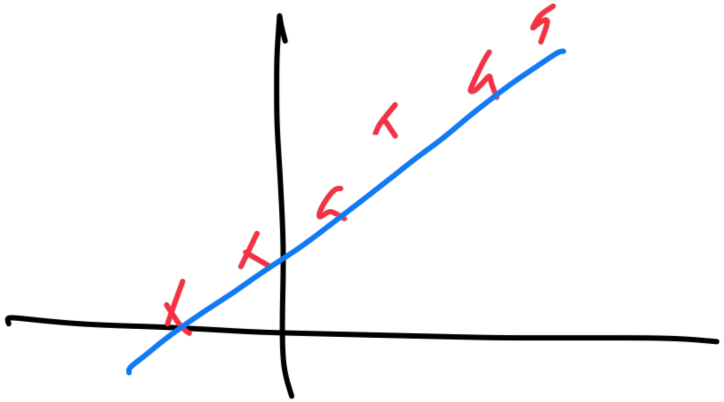
α $\quad y = \sum \alpha_k z_k \qquad \alpha_k = \dfrac{\langle t, z_k \rangle}{\langle z_k, z_k \rangle}$

$$\beta_D = \frac{\langle t, z_D \rangle}{\langle z_D, z_D \rangle}$$

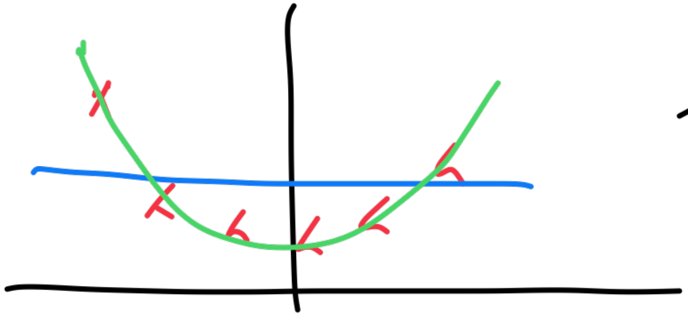


$$h_\beta(x) = \beta_0 + \beta_1 x_1 \dots + \underbrace{\beta_D x_D} + \varepsilon$$





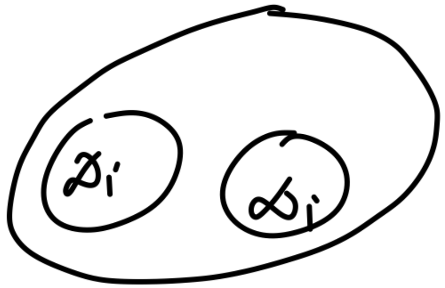
$$h_\beta(x) = \beta_0 + \beta_1 x$$



$$h_\beta(x) = \beta_0 + \beta_1 x + \beta_2 x^2$$

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2$$

$$MSE(x) = \mathbb{E}_{\mathcal{D}_i}\left\{\left(t(x) - h_\beta(x;\mathcal{D}_i)\right)^2\right\}$$



$$= \mathbb{E}_{\mathcal{D}_i}\left\{\left(t(x) - \mathbb{E}_{\mathcal{D}_i} h_\beta(x;\mathcal{D}_i) + \mathbb{E}_{\mathcal{D}_i} h_\beta(x;\mathcal{D}_i) - h_\beta(x;\mathcal{D}_i)\right)^2\right\}$$

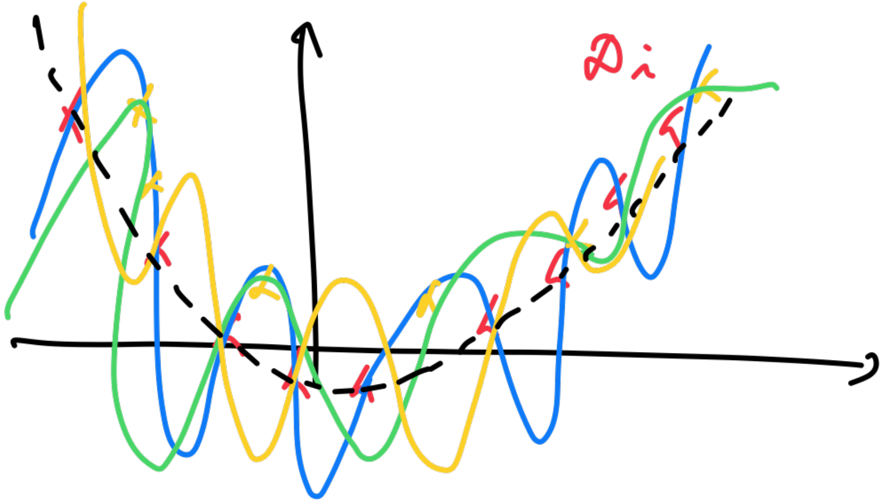
bias^2

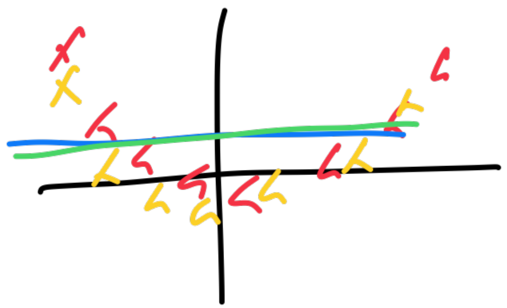
$$= \mathbb{E}_{\mathcal{D}_i}\left\{\left(t(x) - \mathbb{E}_{\mathcal{D}_i} h_\beta(x;\mathcal{D}_i)\right)^2\right\}$$

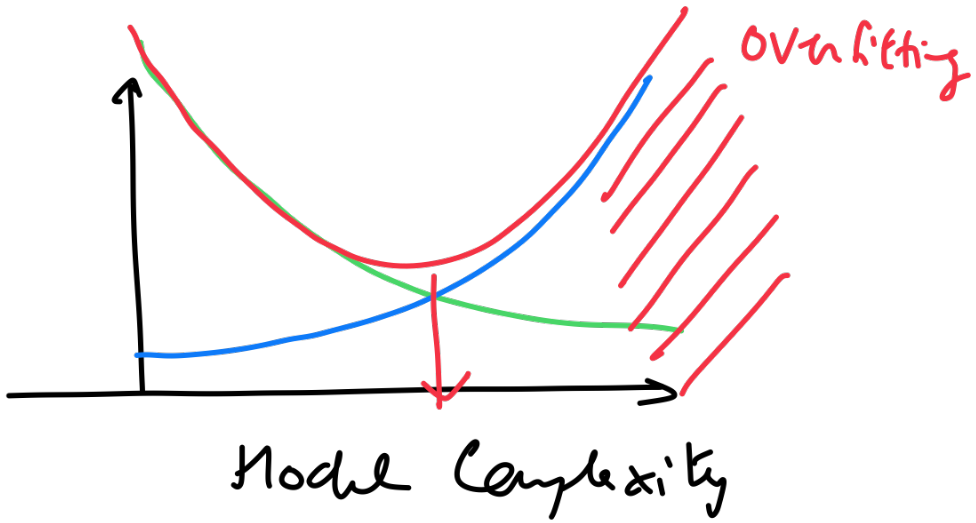
Variance

$$+ \mathbb{E}_{\mathcal{D}_i}\left\{\left(h_\beta(x;\mathcal{D}_i) - \mathbb{E}_{\mathcal{D}_i} h_\beta(x;\mathcal{D}_i)\right)^2\right\}$$

$$+ 2\,\mathbb{E}_{\mathcal{D}_i}\left\{\cancelto{0}{\left(t(x) - \mathbb{E}_{\mathcal{D}_i} h_\beta(x;\mathcal{D}_i)\right)\left(\mathbb{E}_{\mathcal{D}_i} h_\beta(x;\mathcal{D}_i) - h_\beta(x;\mathcal{D}_i)\right)}\right\}$$


Di




overfitting
Model Complexity