Optimisation stochastique

Problème 1

On veut résoudre :

$$\inf_{A\in\mathbb{R}^{MxN}} F(A) \tag{1}$$

avec :

$$F(A) = \frac{1}{2}\sum_{n=1}^{N}||Ax_n - y_n||^2 = \sum_{n=1}^{N} F_n(A) \qquad F_n(A) = \frac{1}{2}||Ax_n - y_n||^2$$

$$\begin{aligned}
F_n(A+H) &= \frac{1}{2}||(A+H)x_n - y_n||^2 \\
&= \frac{1}{2}\langle (A+H)x_n - y_n, (A+H)x_n - y_n\rangle \\
&= \frac{1}{2}\langle Ax_n - y_n, Ax_n - y_n\rangle + \langle Hx_n, Ax_n - y_n\rangle + \frac{1}{2}\langle Hx_n, Hx_n\rangle \\
&= F_n(A) + \langle Hx_n, Ax_n - y_n\rangle + \frac{1}{2}\langle Hx_n, Hx_n\rangle \\
&= F_n(A) + \langle Hx_n, Ax_n - y_n\rangle + \frac{1}{2}||Hx_n||^2
\end{aligned}$$

Donc :

$$\begin{aligned}
D_{F_n}(A)(H) &= \langle Hx_n, Ax_n - y_n\rangle_{\mathbb{R}^{MxN}} \\
&= \langle (Hx_n)^T, (Ax_n - y_n)^T\rangle_{\mathbb{R}^{NxM}} \\
&= \langle x_n^T H^T, (Ax_n - y_n)^T\rangle_{\mathbb{R}^{NxM}} \\
&= \langle H^T, x_n(Ax_n - y_n)^T\rangle_{\mathbb{R}^{NxM}} \\
&= \langle H, (Ax_n - y_n)x_n^T\rangle_{\mathbb{R}^{MxN}}
\end{aligned}$$

Donc :

$$\nabla F_n(A) = (Ax_n - y_n)x_n^T$$

Chapitre 2

$$h^* = \underset{h:\mathcal{X}\to\mathcal{Y}}{\text{argmin}}\ E(h) \qquad \text{idéal} \tag{2}$$

$$h^*_{\mathcal{H}_D} = \underset{h\in\mathcal{H}_D}{\text{argmin}}\ E(h) \qquad \text{famille}$$

$$h^*_n = \underset{h\in\mathcal{H}_D}{\text{argmin}}\ E_n(h) \qquad \text{Risque empirique}$$

<u>Optimisation :</u> $\tilde{h_n} \approx h_n^*$
On note ε les erreurs d'apprentissage :

$$\begin{aligned}
\varepsilon &= E(\tilde{h_n}) - E(h_n^*) \geq 0 \\
&= E(\tilde{h_n}) - E(h_n^*) + E(h_n^*) - E(h^*_{\mathcal{H}_D}) + E(h^*_{\mathcal{H}_D}) - E(h^*)
\end{aligned}$$

On pose :

$$\begin{aligned}\varepsilon opt &= E(\tilde{h_n}) - E(h_n^*)\\ \varepsilon_{est} &= E(h_n^*) - E(h_{\mathcal{H}_D}^*)\\ \varepsilon_{app} &= E(h_{\mathcal{H}_D}^*) - E(h^*)\end{aligned}$$

On a : $\varepsilon_{opt} \searrow$ avec N et $\varepsilon_{app} \searrow$ avec D

Erreurs d'approximation

Les erreurs d'approximation caractérisent la capacitéde la famille $\mathcal{H}_D$ à approcher le prédicteur idéal h^*.

Le choix de la famille $\mathcal{H}_D$ demande souvent un énorme travail d'analyse et de modélisation !

Largeur de Kolmogorov

Une méthode populaire pour construire des prédicteurs consiste à considérer un sous-espace vectoriel $\mathcal{H}_D$.

$$h \in \mathcal{H}_D \Leftrightarrow h = \sum_{d=1}^{D} \omega_d \psi_d$$

avec :

$$\psi_d : \mathcal{X} \to \mathcal{Y} \qquad \text{applications pré-définies}$$

Exemples :

- polynomes
- polynomes trigonométriques
- ondelettes
- splines

Supposons qu'on sache par avance que $h^* \in \mathcal{F}$ où $\mathcal{F}$ est une famille de fonctions connues.

La largeur de Kolmogorov de $\mathcal{F}$ permet de quantifier à quel point la famille $\mathcal{F}$ peut être approchée par un sous-espace vectoriel de dimension D.

<u>Définition :</u> La largeur de Kolmogorov de $\mathcal{F}$ est définie par :

$$\delta_d(\mathcal{F}, ||.||) = \inf_{\dim(\mathcal{H}_D)=d} \sup_{f \in \mathcal{F}} \inf_{h \in \mathcal{H}_D} ||h - f||$$

<u>Proposition :</u>

- $\forall \mathcal{F}, \delta_d(\mathcal{F}, ||.||) \searrow$ avec d
- $\forall \alpha, \delta_d(\alpha\mathcal{F}, ||.||) = |\alpha| \delta_d(\mathcal{F}, ||.||)$
- $\delta_d(\mathcal{F}, ||.||) = \delta_d(\text{conv}(F), ||.||)$ où $\text{conv}(F)$ est l'enveloppe convexe de $\mathcal{F}$
- $\mathcal{F}$ compact $\Rightarrow \delta_d(\mathcal{F}, ||.||) \underset{d \to +\infty}{\to} 0$ et $\mathcal{F}$ bornée.

Définition : (Espace de Soboléo)
On note :

$$\mathcal{W}_p^r([0,1]) = \{h \in \mathcal{C}^{r-1}([0,1]) \text{ tel que } h^{(r-1)} \text{absolument continue et } h^{(r)} \in L^p([0,1])\}$$

- $\mathcal{W}_2^0 = L^2$
- $\mathcal{W}_p^0 = L^p$
- $\mathcal{W}_\infty^1 =$ fonctions lipschitziennes

Théorème :

$$\delta_d(\mathcal{B}_p^r) \propto d^{-r} \text{ pour } 1 \leq p \leq +\infty$$

avec :

$$\mathcal{B}_p^r = \{h \in \mathcal{W}_p^r([0,1]) \text{ tel que } ||h^{(r)}||_{L^p} \leq 1\}$$

Les sous-espaces optimaux $\mathcal{H}_D$ sont :