

Qualité de prévision, risque et estimation du risque

Manon MAHEO - Valentin PENISSON

18/10/2021

Introduction à l'estimation de l'erreur de prévision

La performance du modèle statistique ou algorithme statistique s'évalue par un **risque** ou une **erreur de prévision**, dite encore **erreur de généralisation** dans le cas de la régression et de la classification. La mesure de cette performance est importante dans le sens où elle guide dans le choix de la méthode en comparant chacun des modèles. Une mesure de la qualité ou de la performance du modèle permet aussi de considérer la confiance que l'on peut accorder à la prévision du modèle.

On considère que l'on dispose d'un **échantillon de données observées de type entrée-sortie** de taille n : $d_1^n = \{(x_1, y_1), \dots, (x_n, y_n)\}$ avec $x_i \in \mathcal{X}$ quelconque (souvent égal à \mathbb{R}^d), $y_i \in \mathcal{Y}$ pour $i = 1 \dots n$.

Les différentes techniques de ré-échantillonnage

Approche de validation croisée hold-out

Approche de validation croisée p out

Approche de validation croisée K fold

Algorithme de Leave-One-Out bootstrap

Conclusion