

TP2 Projet hélicoptère : construction de modèles de krigage

Mines Saint-Étienne, majeure Data Science, 2016 - 2017

Le but de ce TP est de construire un modèle statistique du temps de chute à partir des expériences qui ont été réalisées. Le meilleur modèle obtenu sera utilisé pour les séances suivantes.

Réflexions préliminaires

Avant de vous lancer dans la création de modèles, effectuez une première analyse des données. De plus, prenez le temps de vous poser des questions telles que :

- Un pré-traitement des données est-il nécessaire (données aberrantes, transformation log, ...)?
- Est-il pertinent de changer la paramétrisation du problème?
- ...

Construction de modèles de krigage (1h30)

Le package `DiceKriging` vous permettra de gagner un temps précieux. La construction de modèles se fait alors à l'aide de la fonction `km`, et les prédictions à l'aide de la fonction `predict`. De plus, ce package possède déjà des outils de diagnostic des modèles (`plot(model)`, `summary(model)`) et il permet d'estimer les paramètres des noyaux par maximum de vraisemblance ou par validation croisée. Par ailleurs, le package `DiceView` peut être utile pour obtenir des représentations graphiques de modèles.

Comparez différents modèles en essayant différents noyaux et différentes tendances. Ces modèles permettent-ils d'obtenir de meilleurs résultats qu'un simple modèle de régression linéaire?

Choisissez le modèle qui vous semble le mieux adapté et donnez des arguments vous ayant amenés à faire ce choix. Attention, cela a-t-il du sens d'appliquer des méthodes *leave-one-out* pour juger de la qualité d'un modèle? Si le temps le permet, analysez la robustesse des estimations des paramètres des noyaux.

Interprétez les valeurs des paramètres du modèle retenu (bruit d'observation, portées, ...). Si vous choisissez d'utiliser des termes de tendance, peuvent-ils s'interpréter?

Analyse de sensibilité (1h30)

Appliquez les méthodes vues en cours d'analyse de sensibilité pour représenter les effets principaux et calculez les indices de Sobol. D'après le modèle, quelles sont les variables (et les interactions) qui sont influentes? Cela peut-il s'interpréter physiquement?

Région contenant l'optimum

À partir du modèle dont vous disposez, êtes-vous capable d'identifier une région contenant *a priori* l'optimum? Quelle méthode pouvez-vous proposer pour enrichir le plan d'expériences dans cette région? Si le temps le permet, réalisez 10 nouvelles expériences pour affiner le modèle dans cette région d'intérêt.