

SY19 – A18

TP 4: Splines, modèles additifs généralisés

Exercice 1

Les données `mcycle` du package `MASS` contiennent des mesures d'accélération de la tête d'un mannequin en fonction du temps, au cours d'une simulation d'accident de moto.

1. Estimer la valeur moyenne de l'accélération en fonction du temps par régression polynomiale d'ordre p . Représenter graphiquement les données et la fonction de régression estimée pour différentes valeurs de p . Quelles valeurs semblent visuellement satisfaisantes ?
2. Déterminer la valeur optimale de p par validation croisée (faire $K = 5$ groupes).
3. Mêmes questions en utilisant des splines cubiques naturelles. On expérimentera cette fois différents nombres de degrés de liberté (paramètre `df` de la fonction `ns()`), puis on déterminera la valeur optimale par validation croisée.
4. Utiliser cette fois les splines de lissage (fonction `smooth.spline()`). Trouver la valeur optimale du paramètre `df` par la méthode leave-one-out, puis faire varier ce paramètre autour de la valeur optimale, et estimer les erreurs par validation croisée avec le même découpage que dans les deux questions précédentes.
5. Représentez sur un même graphique les fonctions de régression estimées par les trois méthodes. Laquelle donne les meilleurs résultats sur ces données ? Justifiez votre réponse.

Exercice 2

Les données `kyphosis` du package `gam` concernent 81 enfants ayant subi une opération correctrice de la colonne vertébrale. Les variables sont :

- **Kyphosis** : présence ou absence d'une cyphose (un type of déformation) après l'opération.
- **Age** : âge en mois
- **Number** : nombre de vertèbres impliquées

- **Start** : numéro de la première vertèbre opérée (en partant du haut) .
1. A l'aide de statistiques descriptives, explorer la relation entre la variable **Kyphosis** à expliquer, et les trois variables explicatives.
 2. Estimer différents modèles additifs généralisés sur ces données. Représenter graphiquement et interpréter les résultats.
 3. Estimer les probabilités d'erreur des différents modèles par la méthode *leave-one-out*.