Assignment

Statistical Methods for Bioinformatics

In this assignment you will analyse data from a study into the relation between retinol/carotene intake from food and their levels in blood. The levels of retinol and carotene in blood have been linked to cancer risk. To understand how to control these values by nutrition a study has been done, linking nutrients to the blood levels of these chemicals.

1. Study and describe the data. Do you see indications of potential issues when statistically modelling the data? Explain.
2. Study and model appropriately the relation between alcohol intake and carotene levels in the blood.
3. *All other things being equal*, does gender influence blood carotene levels?
4. Make your best model for predicting carotene levels in the blood. Make and compare a ridge regression model, a linear model and a generalized additive model (GAM).
5. Summarize your analyses. At least discuss which is the most important dietary contribution.

Annotation to the datafile:

Variable Names in order from left to right:

age

gender

smoking Smoking habit

weightOverHeightSQ weight/(height^2)

vitaSuppl Vitamin Use

cal Dietary energy intake (calories per day).

fat Dietary fat (g per day).

fiber Dietary fiber (g per day).

alcohol Alcoholic drinks (number per week).

chol Dietary cholesterol (mg per day).

carodiet Dietary beta-carotene (mcg per day).

retinoldiet Dietary retinol (mcg per day)

caroBlood: Blood beta-carotene (ng/ml)

retinolBlood Blood Retinol (ng/ml)