1. Use the data from the iod.csv file you analyzed in the previous homework to create a model for GFR that considers nonlinear transformations (polynomials, step functions, splines, generalized additive models) for continuous predictors. Try each type of nonlinear function and give statistical support (e.g, by using a hypothesis test) to your final choice. If the function is nonparametric or hard to interpret, try to represent it using a function that can be interpreted straightforwardly.

You can either divide your data into a training set and a test set to validate or use cross-validation. If you are using cross-validation, you will need to automate your model fitting process in order to be able to replicate it on the various folds.

As always, explain the methods you are using, provide useful tables and figures with accompanying text and elucidate the implications and conclusions of your modeling process

Biostat PhD problems (Extra Credit for everyone else):

2. Consider the truncated power series representation for cubic splines with K interior knots

$$f(x) = \sum_{j=0}^{3} \beta_j X^3 + \sum_{k=1}^{K} \theta_k (X - \xi_k)_+^3$$

Show that the truncated power basis functions $1, X, X^2, X^3, (X - \xi_1)_+^3, (X - \xi_2)_+^3$ where ξ_1 and ξ_2 are two interior knots form a basis for a cubic spline with two knots. In other words, the piecewise cubic polynomials are continuous and have continuous first and second derivatives.

3. Using the same truncated power series representation, show that the boundary conditions for natural cubic splines (that they are linear beyond the boundary knots) implies the following linear constraints on the coefficients:

$$\beta_2 = 0, \beta_3 = 0, \sum_{k=1}^{K} \theta_k = 0, \sum_{k=1}^{K} \xi_k \theta_k = 0$$

Hence derive the basis functions

$$N_1(X) = 1$$

$$N_2(X) = X$$

$$N_{k+2}(X) = d_k(X) - d_{K-1}(X)$$

where

$$d_k(X) = \frac{(X - \xi_k)_+^3 - (X - \xi_K)_+^3}{\xi_K - \xi_k}$$

for $k = 1, 2, \ldots, K - 2$. Note that each of these basis functions has zero second and third derivatives for $X \geq \xi_K$.