7.1: Generalized Additive Models (GAM)

Dr. Bean - Stat 5100

When using smoothing splines, why is it recommended to set a random number seed in proc gampl?

Smoothing splines, by default, use cross validation to determine the effective degrees of freedom of the spline. Changes in the grouping of the observations for cross validation can change the recommended model smoothness.

Can you apply a smoothing spline to a categorical variable? Explain why or why not.

Not appropriately. There are only two unique values of X (0 and 1) which makes it impossible to define knots for the smoothing splines. In addition, there is no point to have a smooth transition between $x = 0$ and $x = 1$ as it is impossible to observe a value between 0 and 1.

GAMs and Regression Trees are both considered non-parametric models. Name one fundamental difference between regression trees and smoothing splines.

- Regression trees give discrete predictions, GAMs give smooth predictions

- Regression trees handle high ordered interactions, GAMs do not handle interactions.

- Effects of explanatory variables can be visualized holding all other predictors constant in GAMs, not so for Regression Trees.

Why might piecewise cubic polynomials be preferred to piecewise linear models?

Piece-wise linear models may approximate the average well, but the model will not be differentiable i.e. there will be sharp corners at each one of the spline knots.