Practice Problems

Smoothing Splines

* TRUE/FALSE
  + Smoothing splines contain knots at every single data point.
  + As lambda increases, the overall fit of a smoothing spline will become less smooth.
  + When fitting smoothing splines in R, is it *always* good to cross-validate.
* Use the Boston dataset to predict crime rate “crim” using the predictor “lstat.”
  + Fit a smoothing spline with specified degrees of freedom of 30.
  + Next, fit a smoothing spline with a cross-validated smoothing parameter (hint: there is a “cv” argument within the smooth.spline function).
  + Display number of degrees of freedom in the cross-validated smoothing spline.
  + Plot the data and both smoothing splines. In a comment line, pick which fit would be most favorable and explain your reasoning.

Local Regression

* TRUE/FALSE
  + Increasing the span increases the variance in the model.
  + When weighting the points, points outside of the span have a weight of 0.
  + When weighting the points, the furthest point(s) of the span have a weight of 0.
  + Local regression is not a memory based model.
  + *Lowess* is used more frequently than the *loess* model.
* Using the Boston dataset from the MASS package, construct a local regression to predict the crime rate “crim,” using the predictor “lstat.”
  + Build a model that has a span of 0.25 and find the MSE.
  + Now build a model with a span of 0.5 and find the MSE.
  + Plot the two models together.
  + Which model is more accurate? Why?
* Using the Boston data set, write a *for* loop to find the span that minimizes the MSE.
  + Use cross validation to find the optimal span by setting the number of folds to 10, then 20, then 50.
  + How does increasing the number of folds change the optimal span?