

STAT/CS 5525 Homework 4

(Due on November 16th, 2023)

1. Analyze the “motor cycle data” (use “library(MASS)”, then load “data(mcycle)”, the data are $x=\text{times}$, $y=\text{accel}$). Data can also be found from the following link,

https://www.dropbox.com/scl/fi/nosez2q9y6btuoq1ybgqb/data_mcycle.txt?rlkey=na5c1r410t5yogo2lbv7yob7j&dl=0

Use smoothing splines to fit the data. Try different df's in $[5, 20]$. Find the optimal df in $[5, 10]$ according the cross-validation criterion (in the function “smooth.spline”, specify “cv=T”). What is the λ and cross-validation error of the best fit? Return the following three plots:

- (a) The observation points and the optimal smoothing spline fit.
- (b) The observation points and the three smoothing splines with $\text{df}=5, 10, 15$ (three different colored curves). Then you should also add a “legend” to denote these lines.
- (c) Plot the cross validation errors against different df's from 5 to 20 (show both points and lines). The step of df's is 0.5. (Hint: from this plot you can find the optimal df.)