Problem 1

Suppose we have n pairs of observations (x_i, y_i) with both $x_i \in [0,1]$, $y_i \in \mathbb{R}$. Imagine we would like to solve the smoothing spline problem

$$\hat{f} \leftarrow \text{argmin}_f \sum_{i=1}^{n} (y_i - f(x_i))^2 + \lambda \int_0^1 (f^{''}(x))^2 dx$$

Suppose rather than optimizing over all possible twice differentiable f, we instead preselect a set of basis functions $\psi_j(x)$, $j = 1, \ldots, J$ (for example, perhaps $\psi_j(x) = x^j$ might be used), and would like to solve the smoothing spline problem where f is restricted to be a linear combination of these basis functions. In other words, we would like to solve

$$\hat{\beta} \leftarrow \text{argmin}_\beta \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{J} \beta_j \psi_j(x_i) \right)^2 + \lambda \int_0^1 \left[\frac{\partial^2}{\partial x^2} \left(\sum_{j=1}^{J} \beta_j \psi_j(x) \right) \right]^2 dx \tag{1}$$

with $\hat{f} \leftarrow \sum_{j=1}^{p} \hat{\beta}_j \psi_j$.

(a) Show that we can rewrite (1), in matrix/vector form as

$$\hat{\beta} \leftarrow \text{argmin}_\beta \|\underline{y} - \Psi\beta\|_2^2 + \lambda \beta^\top \Omega \beta \tag{2}$$

for a properly chosen Ψ and Ω (be explicit about what these matrices are!)

(b) Show that the solution to (2) is given by

$$\hat{\beta} = \left(\Psi^\top \Psi + \lambda \Omega \right)^{-1} \Psi^\top \underline{y}$$

(c) Roughly how many basis vectors (J) do you think should be used? (as a function of n). What happens if we use a large number of basis vectors (eg. $J > n$?). How does this compare to using a projection estimator without penalization? (do we need to cross-validate over both penalty parameter and number of basis vectors? Which do we use to control the bias/variance tradeoff?)

Problem 2

Take the simulation study you conducted for HW 1, and add local polynomial regression to it (you can use the loess function in R). Be intentional in how you choose the bandwidth. As a reminder, your simulation study should include *multiple* simulated datasets/replicates per scenario/sample-size (don't just simulate one dataset and try to draw conclusions... that is not how a simulation study works!)