Problem 1

In this problem, we empirically explore the curse of dimensionality.

Suppose we observe n iid triples $\mathbf{x} = (x_1, x_2, x_3)$ drawn uniformly from $[0, 1]^3$, with

$$y_i = f(\mathbf{x}_i) + \epsilon_i$$

where $\epsilon_i \overset{iid}{\sim} N(0, 1)$. Consider the following three functions:

(a) $f(x_1, x_2, x_3) = x_1 + x_2 + x_3$

(b) $f(x_1, x_2, x_3) = sin(4x_1) + 2\sqrt{x_2} + e^{x_3}$

(c) $f(x_1, x_2, x_3) = (x_1 x_2 x_3)^{1/3}$

Design and implement a simulation experiment to compare the performance of

- The multivariate linear model estimate;
- An additive model estimate of your choice (eg. basis expansion with a growing number of elements)
- A general nonparametric estimator of your choice (eg. Nadaraya-Watson or local polynomial estimation).

Briefly plot/discuss the results. How do you expect this to change if we were in > 3 dimensions?

Problem 2

Work on your project!