Assignment1

Tiffany Woodley

5/25/2020

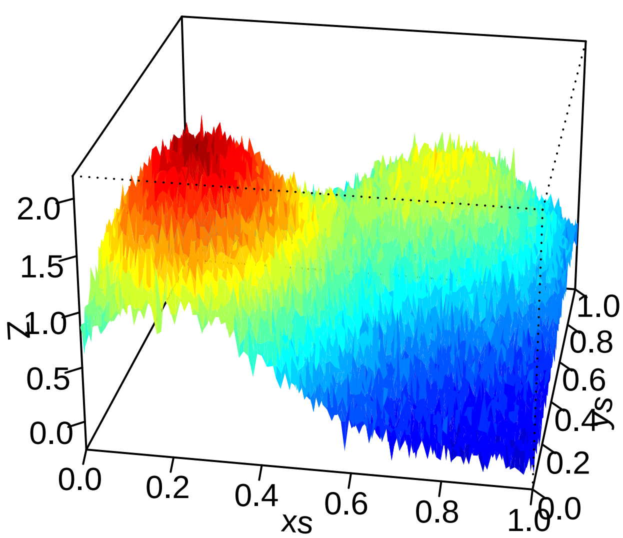
Question 2: Thin-plate Splines

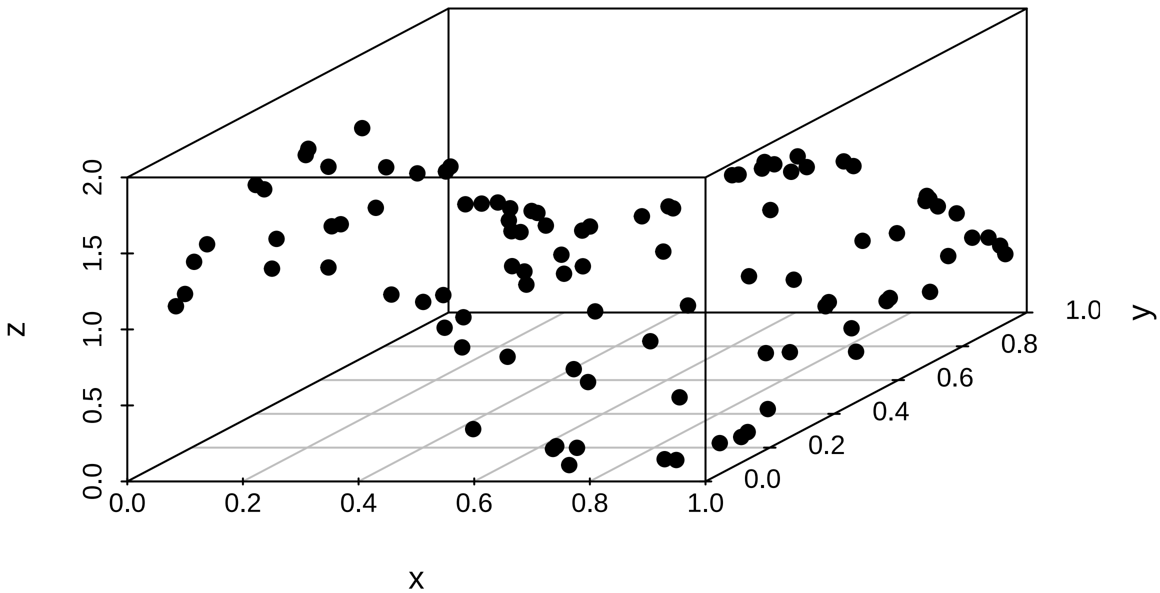
In this question two methods are compared for creating a reduced thin-plate regression spline basis. Both constructed basis' are then fit to a function and the surfaces are compared. The first method uses 16 evenly spaced knots to create 16 basis functions. The second method uses all the points as knots to create basis functions and then uses eigen value decomposition to reduce this to 16 basis functions which contain the most variation.

2.1

100 points were simulated from the following function on a unit square with added noise, $e \sim N(0, 0.1^2)$. The function surface can be seen below and the 100 points can be seen in the scatterplot.

$$f(x,z) = \frac{0.75}{\pi\sigma_x\sigma_z}\exp\{-(x-0.2)^2/\sigma_x^2-(z-0.3)^2/\sigma_z^2\}+\frac{0.45}{\pi\sigma_x\sigma_z}\exp\{-(x-0.7)^2/\sigma_x^2-(z-0.8)^2/\sigma_z^2\};\;\;\sigma_x = 0.3\;\;\sigma_z = 0.4$$





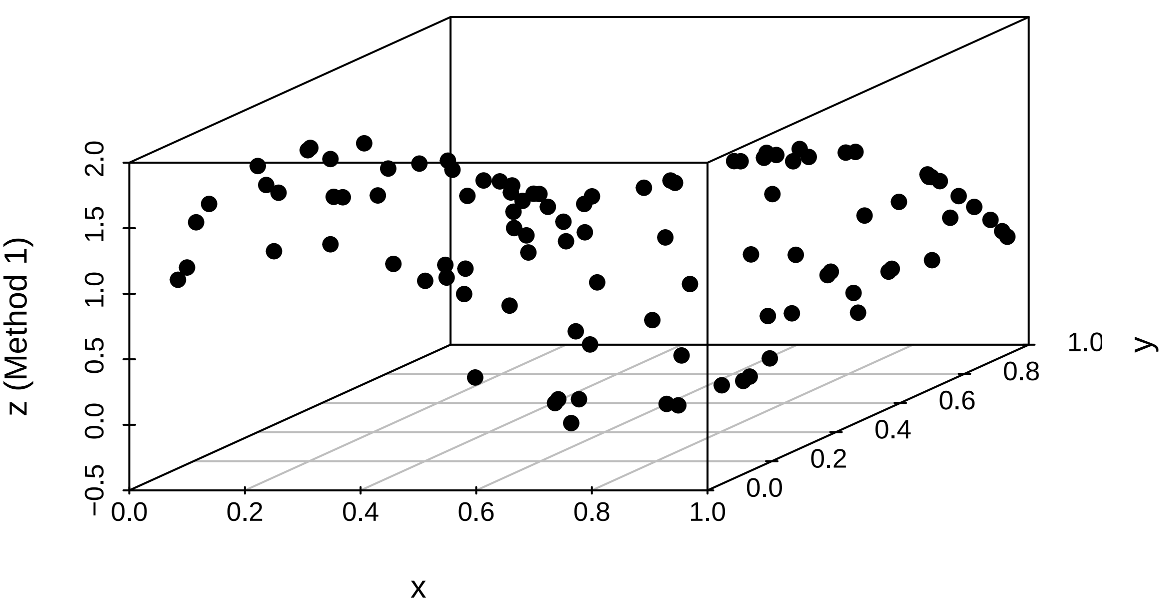
2.2

16 evenly spaced knots were used to construct a thin-plate spline basis. This basis was then used to fit the data.

The thin spline basis functions are created using a radial basis, each basis function j is calculated by the following equation with $u = |x - x_j|^2$ being the distance between the points and the jth knot.

$$r(u) = u^2 log(u)$$

The basis is then fitted to the simulated points, the resultant fitted points can be seen in the scatterplot below.



2.3.

A thin spline basis was created using all 100 simulated points as knots. It was then reduced using Singular Value Decomposition, finding the 16 basis functions which accounted for the greatest variation. This reduced basis was then fit to the data.

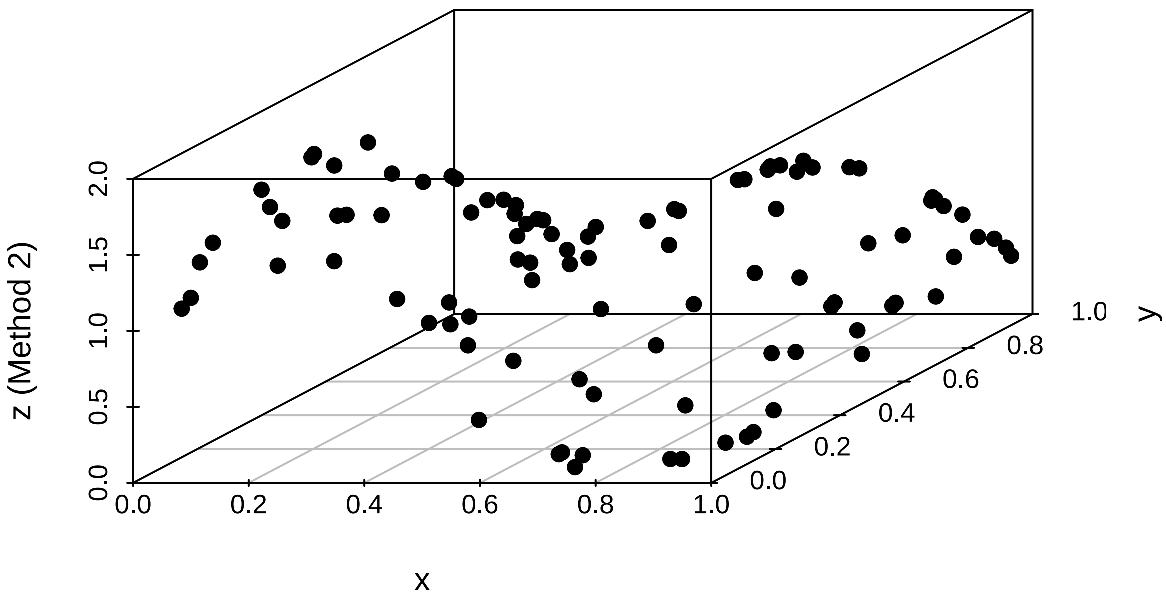
The process of creating the reduced basis is outlined in *Thin plate regression splines* (Simon N. Wood, 2003.). The steps taken are shown in the appendix.

The EE matrix is created by creating a function for each of the knots resulting in a 100 by 100 matrix.

T contains basis functions for constant and linear terms: $M = \binom{m+d-1}{m}$ functions are linearly independent polynomials spanning the space of polynomials with degree less than m (pg. 97).

The T matrix was constructed with $m = 2$, and the steps in the appendix were followed with the exception of using Singular Value Decomposition instead of Eigen Value Decomposition as when creating the basis for the surface in the following question the EE matrix became non-singular.

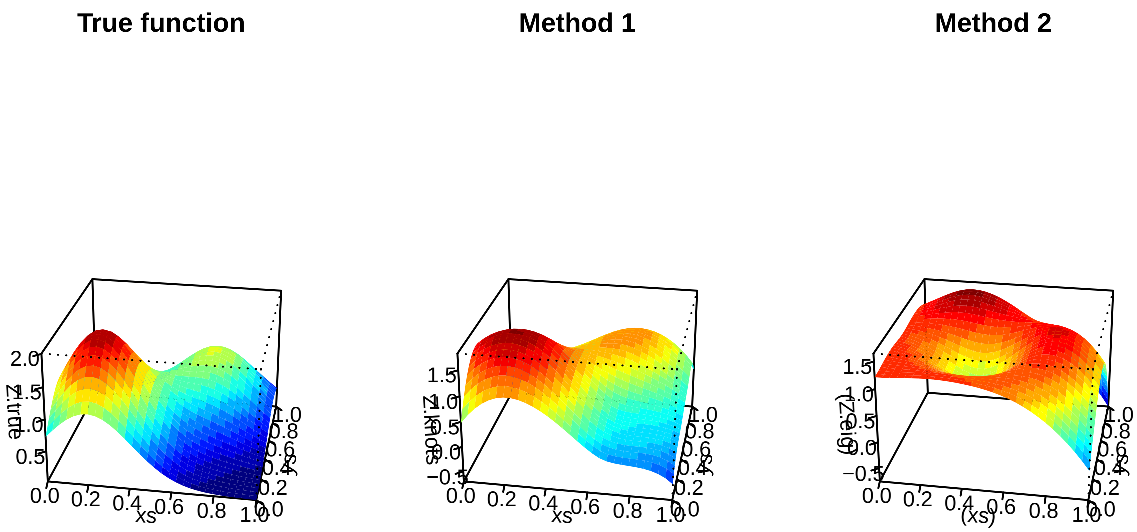
The basis was then fit to the simulates points, the fitted points can be seen in the scatterplot below.

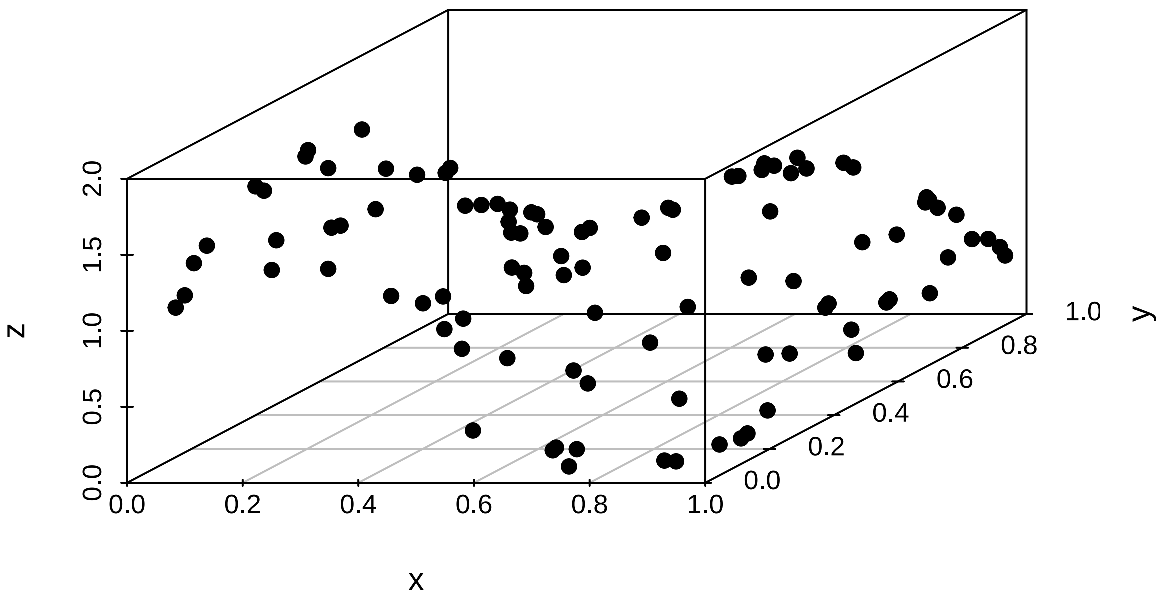


2.4

Using the same knots, basis functions for a 30 by 30 grid were created and predicted from the previous fitting models, these plotted surfaces can be seen below next to the true function.

Method 1 captures the general trend of the true function, forming less sharp bumps than the true function. Method 2 also attempts to capture the two bumps, but has a large divit which the true function doesn't. When looking at the simulated data the divit occurs where there is a large gap between data points and method 2 struggled to fill in the gaps.





References

1. Simon N. Wood, 2003. "Thin plate regression splines," Journal of the Royal Statistical Society Series B, Royal Statistical Society, vol. 65(1), pages 95-114, February.

2. Erni, Birgit, 2020, *Chapter 2: Multidimensional Splines*, lecture notes, Advanced Analytics STA5057W, University of Cape Town, May 2020