STATG006: Solutions to Exercise Sheet #7

The exercises in this sheet focus on nonlinear regression. As before, we provide solutions, sometimes detailed, sometimes a sketch that should point you to the complete solution. Skeetches should not be taken at face value as the level of detail required for an exam answer.

1. (a) It is clear that, if $x \leq \xi$, then $(x - \xi)_+ = 0$, so that the polynomial is $f_1(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$.

   (b) Now $(x-\xi)_+^3 = (x-\xi)^3$, so that, among other things, $f_2(x)$ will have $\beta_3 + \beta_4$ as the coefficient of x^3 and $\beta_0 - \xi^3$ as the intercept. You should finish the details, it is just a matter of rearranging $f_2(x) = f_1(x) + (x - \xi)^3$.

   (c) $f_2(\xi) = f_1(\xi) + (\xi - \xi)^3 = f_1(\xi)$.

   (d) Since $f_2(x) = f_1(x) + (x - \xi)^3$, we have that $f_2'(x) = f_1'(x) + 3(x - \xi)^2$, so $f_2'(\xi) = f_1'(\xi)$.

   (e) Analogous to (d).

2. (a) If $g(\cdot)$ is continuous, then it needs to be zero everywhere so that the resulting error is just the sum of the squares of the outcome variable. But if no assumption of continuity is required, then it can be zero everywhere *but* the training points, where we would be allowed to set $g(x^{(i)}) = y^{(i)}$ and get zero training error!

   (b) (This now assumes that at least the first derivative exists everywhere, so it must also be continuous everywhere. We will in general not be allowed to play with $g(x^{(i)}) = y^{(i)}$). In this case, the first derivative must be zero, so that the function must be constant. The best constant function that minimises RSS is $f(x) = \bar{y}$.

   (c) Analogously, now the function must be linear. This recovers linear least-squares.

   (d) The function must be quadratic. So it is the the least squares solution to $Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon$.

   (e) There is no penalty. This interpolates the data ($g(x^{(i)}) = y^{(i)}$) but it is not identifiable outside the training data (except for the assumption that the third derivative exists).

3. This is similar to the next question, so I will present the solution to the next question only.

4. Sketching this functions gives insights on the notion of piecewise functions. In an exam, you must be able to sketch this by hand. Here is some R code so that you can verify your sketch.

```
> x <- seq(-2, 2, 0.01)
> beta_0 <- 1; beta_1 <- 1; beta_2 <- 3
> f <- beta_0 * (x >= 0 & x <= 2) +
+ beta_1 * (x - 1) * (x >= 1 & x <= 2) +
+ beta_2 * ((x - 3) * (x >= 3 & x <= 4) + (x >= 4 & x <= 5))
> plot(x, f, type = "l")
```

   The main thing to pay attention to is how the range of the input is split. For instance, you should be able to identify that each region $[-2, 0], [0, 1], [1, 2], [2, 3], [3, 4][4, 5]$ may need to be assessed separately. For instance, what happens between 0 and 1?

5. (a) As it should be clear from Question 2, we get more flexibility by increasing m. So $\hat{g}_2$ will have the smaller RSS.
   (b) Same old question, you have seen this thousands of times by now with different guises.
   (c) As in Question 2, they will interpolate the training data perfectly (zero training error for both) and be unidentifiable elsewhere. If we pick arbitrary functions, $\hat{g}_1$ may be smoother, but who knows what this means in terms of test error.

6. See EX7.R.

7. See EX7.R.

8. See EX7.R.

9. See EX7.R.

10. See EX7.R.