STAT 4/579: STATISTICAL LEARNING Quiz 3 Name: _____

The purpose of this exercise is for you to explore the methods for nonparametric regression. The predictor variable x in not uniformly sampled. In the model

$$y_i = m(x_i) + e_i$$

the function $m(x)$ is quite complicated. (I know what it is.) But to make your lives simpler I have actually given you the values of $m(x)$.

```
> x
 [1]  1.0  1.3  1.6  1.9  2.2  2.5  2.8  3.1  3.4  3.7  4.0  4.0  4.5  5.0  5.0
[16]  9.0 13.0 17.0 21.0 25.0 29.0 33.0 37.0 41.0 45.0 49.0 53.0 57.0 61.0 65.0
[31] 69.0 73.0 77.0 81.0 85.0
> y
 [1]  9.966161 10.508760 10.666419 10.972181 11.331496 11.460036 11.673590
 [8] 11.372626 10.829283 11.601400 12.218124 11.631192 11.933786 11.433394
[15] 11.841649 11.752662 11.639633 11.599726 10.905967 10.557000  9.455023
[22]  8.783393  8.365466  8.550886  8.505317  7.751282  7.376445  7.021851
[29]  6.880376  7.034911  5.997330  6.986051  5.838446  5.460420  5.606327

> m(x)
 [1] 10.131332 10.431434 10.688689 10.909840 11.100308 11.264507 11.406063
 [8] 11.527985 11.632792 11.722605 11.799226 11.799226 11.901694 11.977566
[15] 11.977566 12.029649 11.656773 11.164844 10.651344 10.152012  9.680119
[22]  9.239795  8.831233  8.452942  8.102763  7.778353  7.477409  7.197768
[29]  6.937451  6.694663  6.467797  6.255411  6.056219  5.869071  5.692940
```

For your own edification you should plot both y versus x and $m(x)$ versus x. We will be using these same data and **these same models** in Quiz 5.

1. What is the smallest polynomial that fits well? Give R^2 for that model.

2. What is the smallest cosine model that fits well? Give R^2 for that model.

3. Fit a Haar wavelet (or step function) with 8 equal sized partition sets and give R^2.

4. What is R^2 from fitting separate lines using partition points at 10 and 50?

5. What is R^2 from fitting linear splines using knots at 10 and 50?

6. What is R^2 from fitting cubic splines using knots at 10 and 50?

7. Intuitively, where should you put the knots/partition-points?