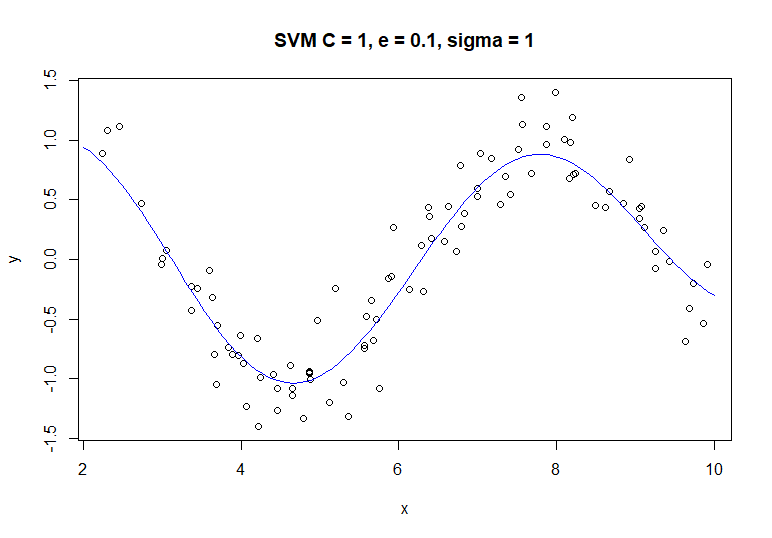
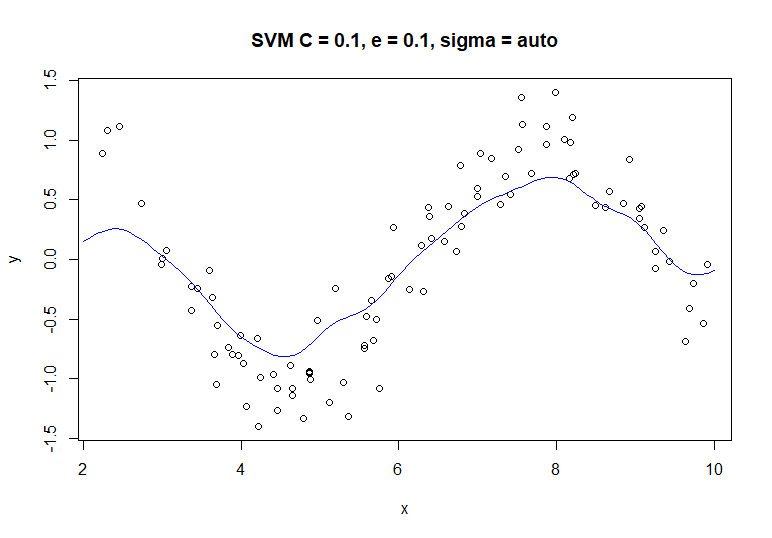
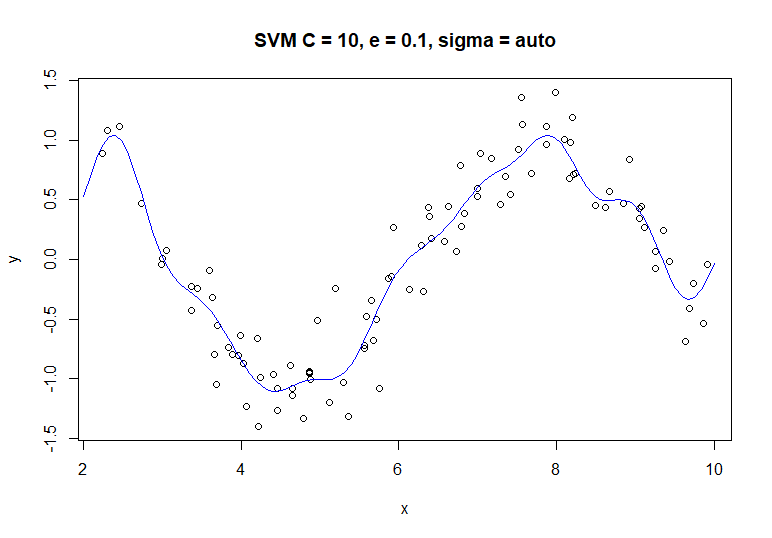
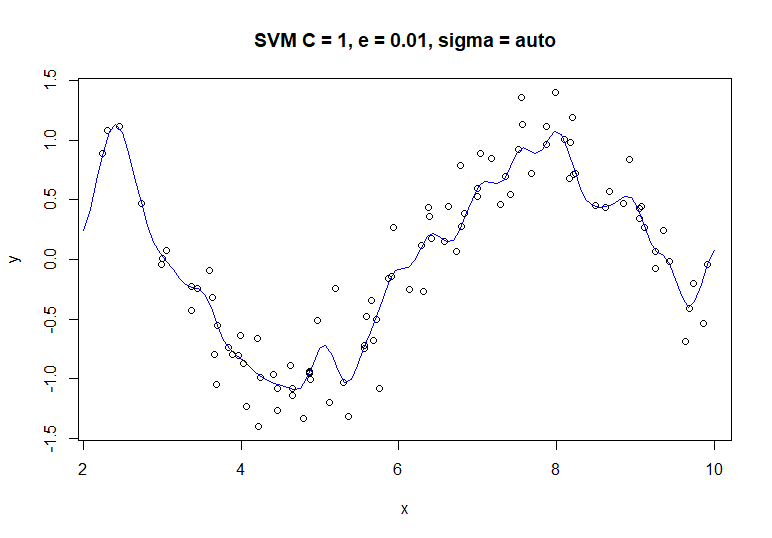
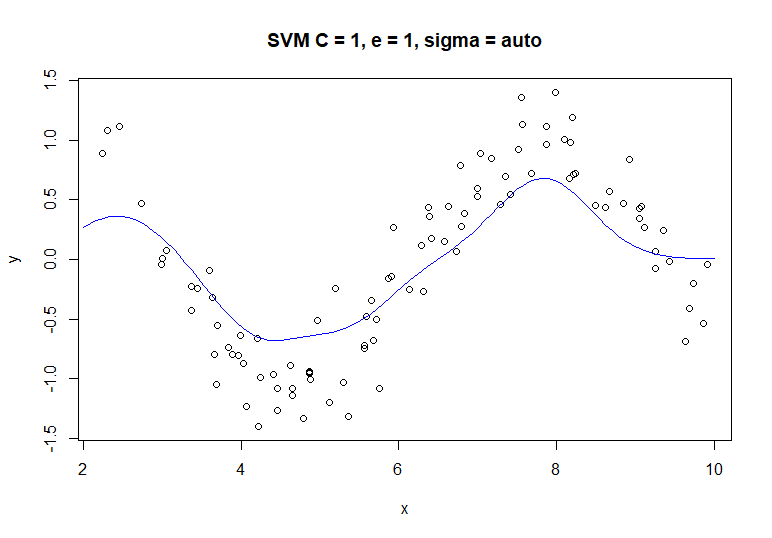
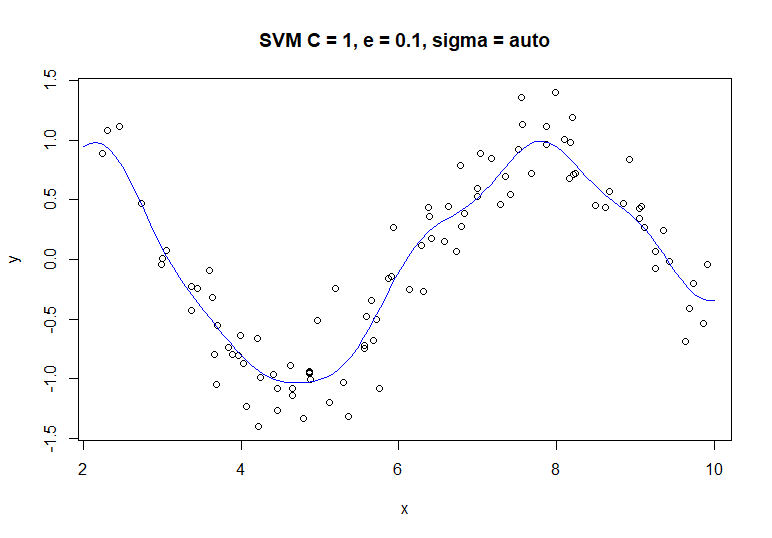
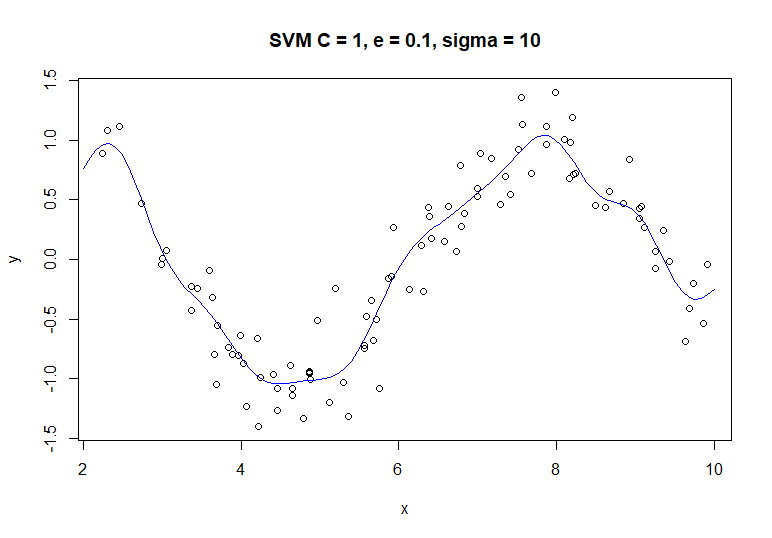
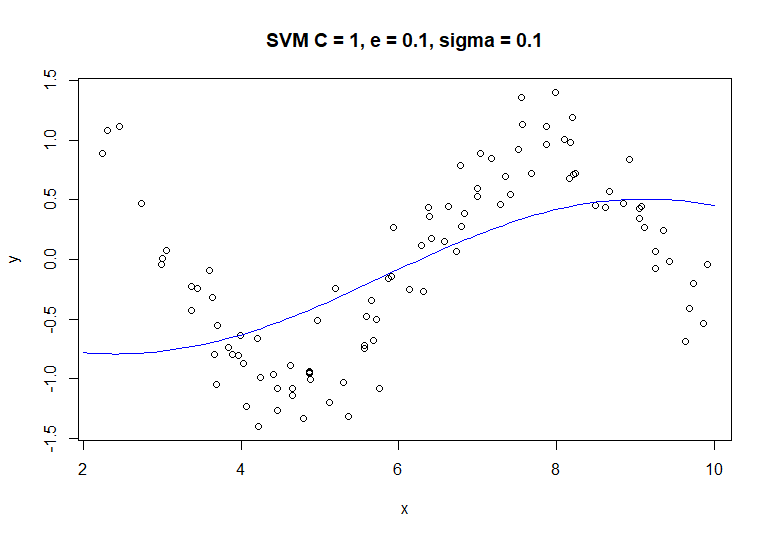
Deanna Springgay

Ch 7 Homework

7.1





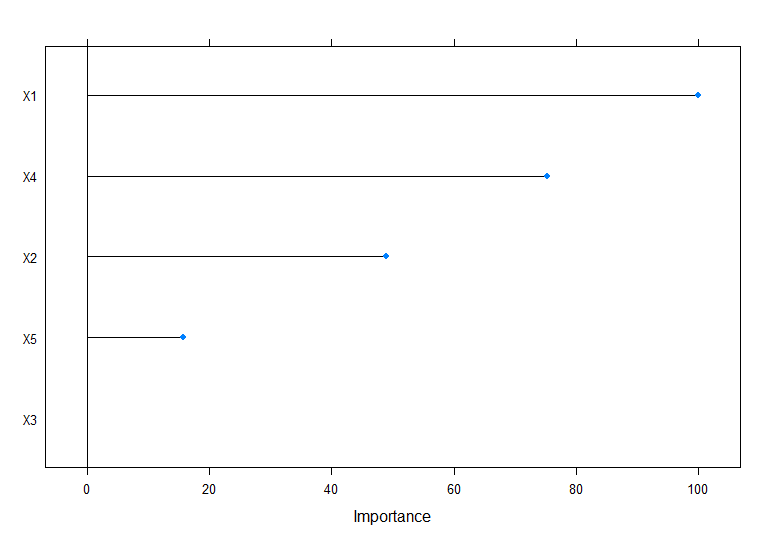
Based on the previous plots, the cost parameter effects penalties on large residuals: a higher cost value leads to less penalties, higher complexity, and likely to over-fit. ε is the margin of tolerance for penalties, a higher ε leads to a smaller margin, though has less of an effect on penalties than the cost parameter. σ affects the variance and bias in the model: a larger σ leads to less variance and more bias, which can lead to overfitting.

7.2

MARS appears to give a better performance than KNN as it has a lower RMSE value on the testing data as shown in the table below.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **Best Tuning Parameter** | **Training** | | **Testing** | |
| **RMSE** | **R²** | **RMSE** | **R²** |
| KNN | k = 17 | 3.349428 | 0.5452823 | 3.2040595 | 0.6819919 |
| MARS | nprune = 13  degree = 2 | 1.261130 | 0.9357469 | 1.2803060 | 0.9335241 |

MARS does not select all the informative predictors, seeing as X3 has an importance of zero as shown by the following plot:



7.5

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **Best Tuning Parameter** | **Training** | | **Testing** | |
| **RMSE** | **R²** | **RMSE** | **R²** |
| Neural Network | size = 8  decay = 0.1  bag = FALSE | 0.7520576 | 0.4805693 | 0.24868980 | 0.93915322 |
| MARS | nprune = 7  degree = 1 | 0.6600063 | 0.5686642 | 0.6296545 | 0.6023783 |
| SVM | sigma = 0.01906204  C = 4 | 0.6495206 | 0.6135844 | 0.3166915 | 0.9035918 |
| KNN | k = 4 | 0.6735640 | 0.5716295 | 0.5544946 | 0.7096414 |

The neural network is the best model out of the group, and the top predictors are a mix of manufacturing and biological processes. The linear model has primarily manufacturing processes as the most important predictors, so there is some difference between the two models.

