STAT-721 Homework (Figure 1.4 reproduction)

Segbehoe, Lawrence Sethor

September 8, 2018

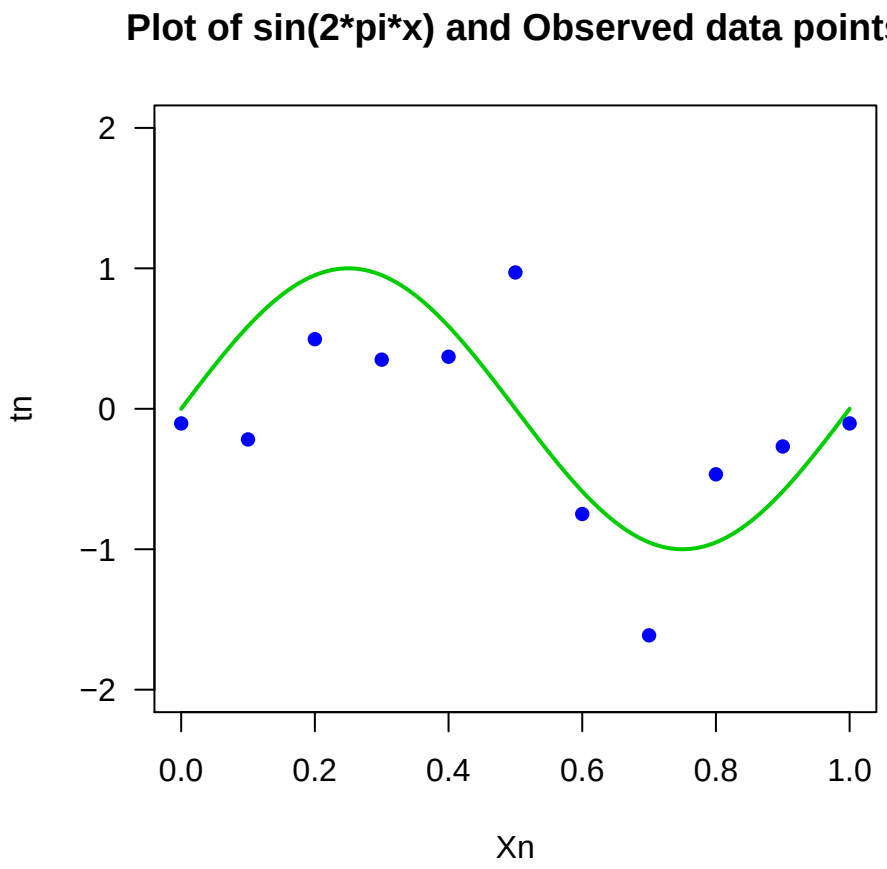
Reproducing the figure 1.4 from the text book

Using Base R Plots

```
## Creating a function for sin(2*pi*x)
f <- function(x){sin(2*pi*x)}
## Creating input variable for the function
Xn <- seq(0,1, 0.001)
## making a dataframe for the two sets of data above
data_sinx <- data.frame(Xn, tn = f(Xn))
## setting  a seed to avoid changing random samples from rnorm()
set.seed(589)
## space the training data set out on the range[0,1]
X_training <- seq(0,1,.1)

## Creating the target variable
## Varying the variance parameter in the rnorm function show how far
## the blue points are away from the green curve
sigma_squared <- .6
t_target = f(X_training) + rnorm(10, 0, sigma_squared)
## making a dataframe form the observed (training and target) dataset
(datframe = data.frame(X_training, t_target))
```

```
##    X_training   t_target
## 1         0.0 -0.1041727
## 2         0.1 -0.2181944
## 3         0.2  0.4958160
## 4         0.3  0.3498535
## 5         0.4  0.3709094
## 6         0.5  0.9708989
## 7         0.6 -0.7491750
## 8         0.7 -1.6132434
## 9         0.8 -0.4664768
## 10        0.9 -0.2682152
## 11        1.0 -0.1041727
```

```
plot(tn~Xn, data = data_sinx, col = 3, type = "l", las = 1, lwd = 2,
     main = "Plot of sin(2*pi*x) and Observed data points", ylim = c(-2,2))
points(t_target~X_training, data = datframe, col = 4, pch = 16)
```



Fitting Polynomial Curve for Observed (training and target) dataset

When the order = 0

```
## When M = 0
Model_order_0 <- glm(t_target ~ 1, data = datframe,
                     family = gaussian(link = "identity"))
B0 <- Model_order_0$coefficients
```

When the order = 1

```
## When M = 1
Model_order_1 <- glm(t_target~ X_training, data = datframe,
                     family = gaussian(link = "identity"))
B1 <- Model_order_1$coefficients
```

When the order = 3

```
## When M = 3
Model_order_3 <- glm(t_target~X_training +I(X_training^2) +I(X_training^3),
                 data = datframe, family = gaussian(link = "identity"))
B3 <- Model_order_3$coefficients
```

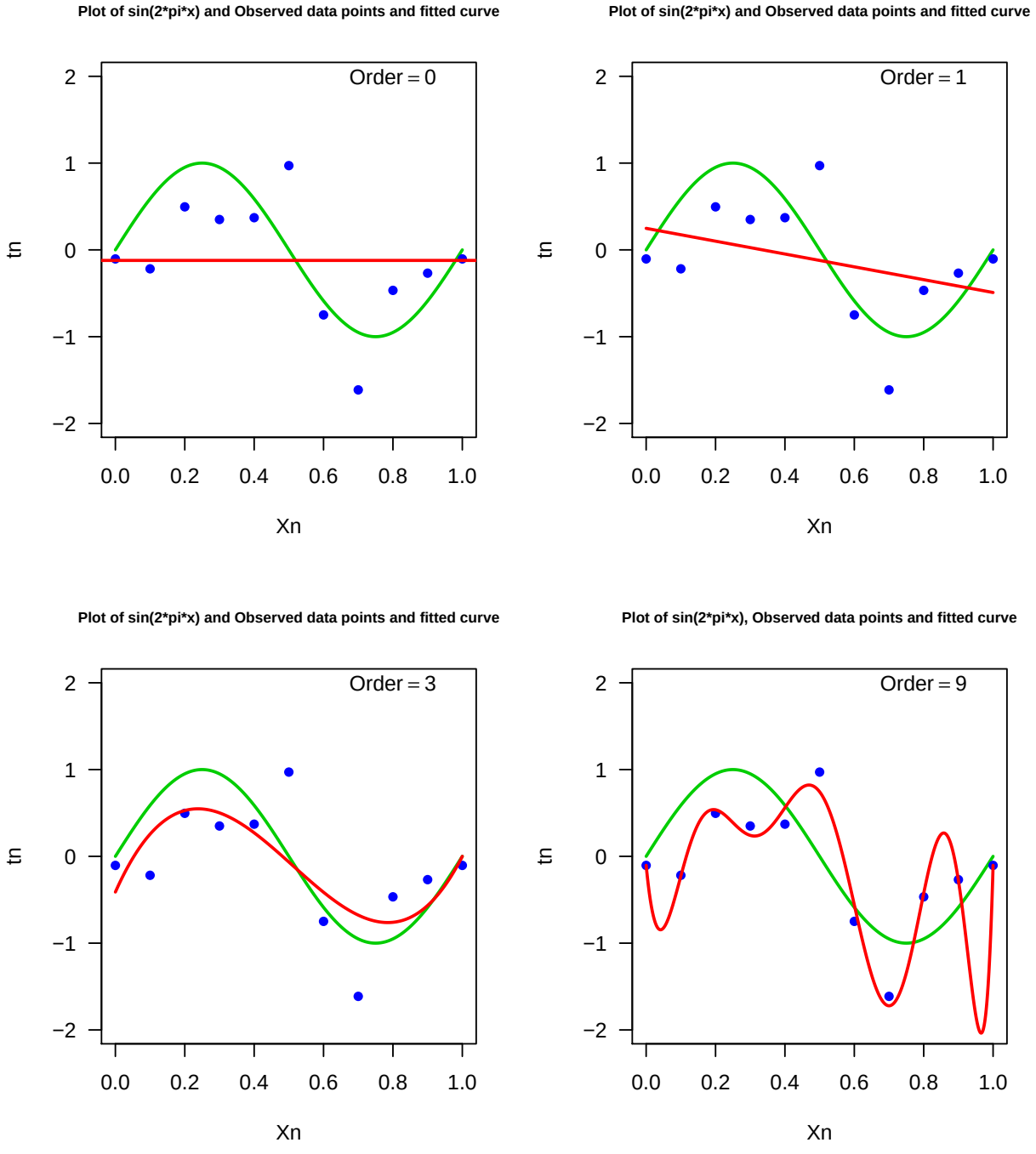
When the order = 9

```
Model_order_9 <- glm(t_target~X_training + I(X_training^2) + I(X_training^3)
                 + I(X_training^4) + I(X_training^5) + I(X_training^6)
                 + I(X_training^7) + I(X_training^8) + I(X_training^9),
                 data = datframe, family = gaussian(link = "identity"))
B9 <- Model_order_9$coefficients
```

Plotting all the fitted polynomial curves to the on the scatter plot of the observed (training and target) dataset.

```
## demacation of the plotting area
layout(matrix(1:4, ncol = 2, byrow = T ))
## Plot with of the polynomial with order = 0
plot(tn~Xn, data = data_sinx, col = 3, type = "l", las = 1, lwd = 2,
     main = "Plot of sin(2*pi*x) and Observed data points and fitted curve",
     cex.main = 0.7, ylim = c(-2,2))
text(.8,2, expression( Order == 0))
points(t_target~X_training, data = datframe, col = 4, pch = 16)
abline(Model_order_0,col = 2, lwd = 2)

## Plot with of the polynomial with order = 1
plot(tn~Xn, data = data_sinx, col = 3, type = "l", las = 1, lwd = 2,
     main = "Plot of sin(2*pi*x) and Observed data points and fitted curve",
     cex.main = 0.7, ylim = c(-2,2))
text(.8,2, expression( Order == 1))
points(t_target~X_training, data = datframe, col = 4, pch = 16)
lines(Xn,B1[1] + B1[2]*Xn, col = 2, lwd = 2)

## Plot with of the polynomial with order = 3
plot(tn~Xn, data = data_sinx, col = 3, type = "l", las = 1, lwd = 2,
     main = "Plot of sin(2*pi*x) and Observed data points and fitted curve",
     cex.main = 0.7, ylim = c(-2,2))
text(.8,2, expression( Order == 3))
points(t_target~X_training, data = datframe, col = 4, pch = 16)
lines(Xn,B3[1] + B3[2]*Xn + B3[3]*Xn*Xn + B3[4]*Xn*Xn*Xn, col = 2, lwd = 2)

plot(tn~Xn, data = data_sinx, col = 3, type = "l", las = 1, lwd = 2,
     main = "Plot of sin(2*pi*x), Observed data points and fitted curve",
     cex.main = 0.7,ylim = c(-2,2))
text(.8,2, expression( Order == 9))
points(t_target~X_training, data = datframe, col = 4, pch = 16)
lines(Xn,B9[1]+B9[2]*Xn + B9[3]*I(Xn^2)+ B9[4]*I(Xn^3)+ B9[5]*I(Xn^4)
```

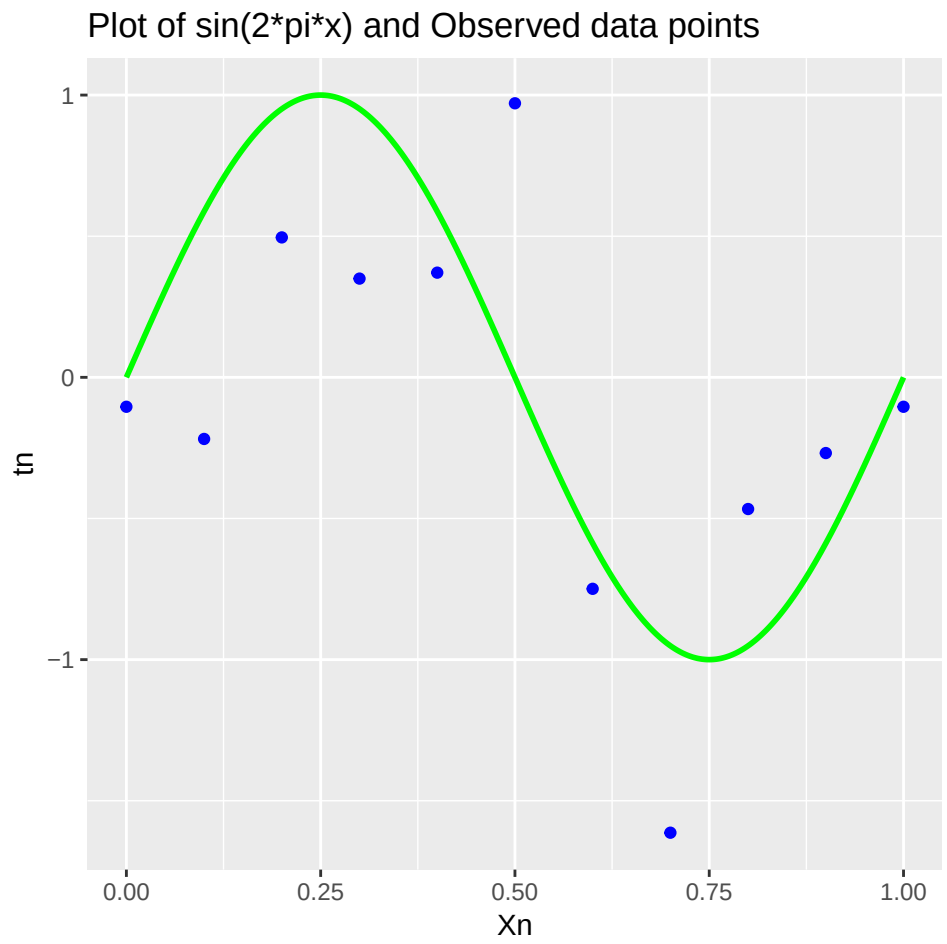
```
      + B9[6]*I(Xn^5)+ B9[7]*I(Xn^6)+ B9[8]*I(Xn^7)+ B9[9]*I(Xn^8)
      + B9[10]*I(Xn^9), col = 2, lwd = 2)
```




```
layout(matrix(1:1, ncol = 2, byrow = T ))
```

Using ggplot2

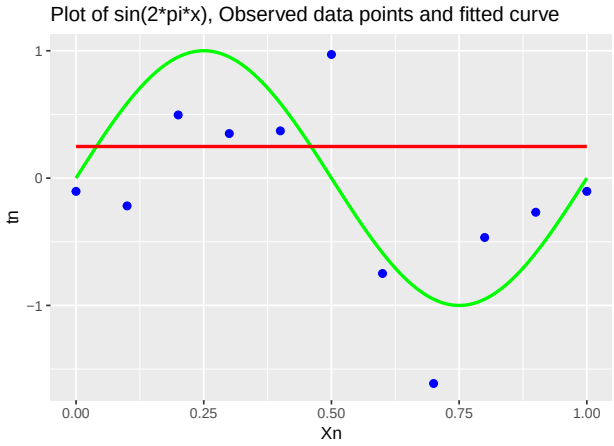
```
library(ggplot2)
library(gridExtra)
## Using ggplot
ggplot() +  geom_line(data = data_sinx, aes(Xn,tn), colour = "green", size = 1) +
  ggtitle("Plot of sin(2*pi*x) and Observed data points") +
  geom_point(data = datframe, aes(X_training,t_target), colour = "blue")
```

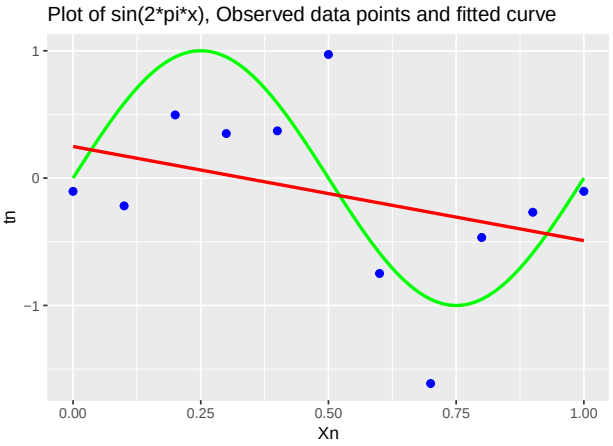


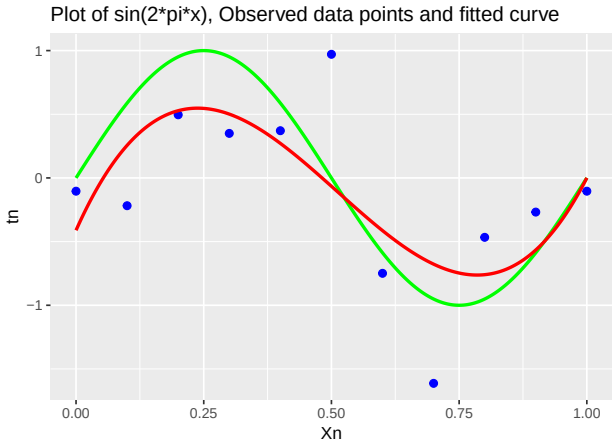
Plotting all the fitted polynomial curves to the on the scatter plot of the observed (training and target) dataset.

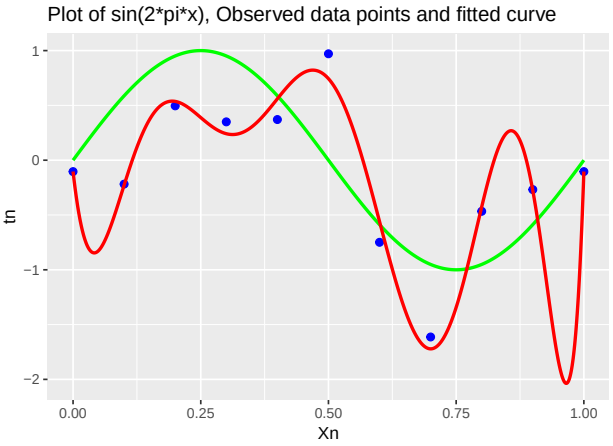
```
## Plot with of the polynomial with order = 0
dat0 <- data.frame(Xn, tn0 = B1[1])

p1=ggplot() +  geom_line(data = data_sinx, aes(Xn,tn), colour = "green", size = 1) +
  ggtitle("Plot of sin(2*pi*x), Observed data points and fitted curve") +
  geom_point(data = datframe, aes(X_training,t_target),colour="blue",size = 2)+
  geom_line(data = dat0, aes(Xn,tn0), colour = "red", size = 1)

## Plot with of the polynomial with order = 1
```

```
dat1 <- data.frame(Xn, tn1 = B1[1] + B1[2]*Xn)

p2=ggplot() +  geom_line(data = data_sinx, aes(Xn,tn), colour = "green", size = 1) +
  ggtitle("Plot of sin(2*pi*x), Observed data points and fitted curve") +
  geom_point(data = datframe, aes(X_training,t_target),colour="blue",size = 2)+
  geom_line(data = dat1, aes(Xn,tn1), colour = "red", size = 1)

## Plot with of the polynomial with order = 3
dat3 <- data.frame(Xn, tn3 = B3[1] + B3[2]*Xn + B3[3]*Xn*Xn + B3[4]*Xn*Xn*Xn)

p3=ggplot() +  geom_line(data = data_sinx, aes(Xn,tn), colour = "green", size = 1) +
  ggtitle("Plot of sin(2*pi*x), Observed data points and fitted curve") +
  geom_point(data = datframe, aes(X_training,t_target),colour="blue",size = 2)+
  geom_line(data = dat3, aes(Xn,tn3), colour = "red", size = 1)


## Plot with of the polynomial with order = 9
dat9 <- data.frame(Xn, tn9 = B9[1]+B9[2]*Xn + B9[3]*I(Xn^2)+ B9[4]*I(Xn^3)+
                     B9[5]*I(Xn^4) + B9[6]*I(Xn^5)+ B9[7]*I(Xn^6)+ B9[8]*I(Xn^7)+
                     B9[9]*I(Xn^8) + B9[10]*I(Xn^9))

p4=ggplot() +  geom_line(data = data_sinx, aes(Xn,tn), colour = "green", size = 1) +
  ggtitle("Plot of sin(2*pi*x), Observed data points and fitted curve") +
  geom_point(data = datframe, aes(X_training,t_target),colour="blue",size = 2)+
  geom_line(data = dat9, aes(Xn,tn9), colour = "red", size = 1)

grid.arrange(p1, p2, p3,p4 , nrow = 2, ncol = 2)
```


Plot of sin(2*pi*x), Observed data points and fitted curve
tn
Xn
1
0
−1
0.00
0.25
0.50
0.75
1.00


Plot of sin(2*pi*x), Observed data points and fitted curve
tn
Xn
1
0
−1
0.00
0.25
0.50
0.75
1.00


Plot of sin(2*pi*x), Observed data points and fitted curve
tn
Xn
1
0
−1
0.00
0.25
0.50
0.75
1.00


Plot of sin(2*pi*x), Observed data points and fitted curve
tn
Xn
1
0
−1
−2
0.00
0.25
0.50
0.75
1.00