Chapter2 Assignment3

Matt Richey

2/15/2020

```
library(tidyverse)
```

```
## -- Attaching packages -------- tidyverse 1.2.1 --

## v ggplot2 3.2.1     v purrr   0.3.3
## v tibble  2.1.3     v dplyr   0.8.4
## v tidyr   1.0.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.3.0

## -- Conflicts ----------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Directions

Use the ideas from the Bias-Variance Tradeoff RMarkdown to create a graphic demonstrating how bias, variance, and MSE change as a function of both the input (x0) and the flexibililty

Specifically, use three different underlying models: linear, quadratic, and a cubic. Pick a specific value of the noise (sigma), say around 1.5.-2.5. For each, compute the bias, variance, and total MSE for values of x0 ranging from -1 to +1 is steps of, say, 0.1. Use these to put together a 3x3 graphic that has a facet for each combination of error type (mse, bias, variance) and flexibility (1,2,3).

Your final dataframe could be something like allTheData.df which has columns x, errorValue, errorType, and Flexibility.

Then the graphic could be generated via:

allTheData.df %>% ggplot()+ geom_line(aes(x,errValue))+ facet_wrap(errorType~Flexibility)+ labs(title=Bias, Variance, MSE and Flexibility“, subtitle=”-1 <=x <= +1“)

Conclude with a comment on what you observe about bias, variance, and MSE from this graphic. Call this file “BiasVarianceGraphic.Rmd”

Setup

Use the function biasVarTO3.lm(form1,sizeDS,numReps,x0)

```
biasVarTO3.lm <- function(form,sizeDS,numDS,x0){
  allVals <- matrix(ncol=2,nrow=numDS)
  for(m in 1:numDS){
    ##the
    mod <- lm(formula(form),buildData(f,sizeDS,sig))
    pred <- predict(mod,newdata=data.frame(x=x0))
    allVals[m,1] <- pred
  }
  ##Same as above from here on.
```

```
  allVals[,2] <- f(x0)+rnorm(numDS,0,sig)
  allVals.df <- data.frame(pred=allVals[,1],true=allVals[,2])
  mse <- with(allVals.df,mean((pred-true)^2))
  var0 <- with(allVals.df,var(pred))
  bias2 <- with(allVals.df,mean(pred-true))^2
  noise <- sig^2
  c(mse,var0,bias2,noise)
}
```

And the buildData function.

```
buildData <- function(func,sizeDS,sig,xMin = -1, xMax = 1){
  ##predictor
  x<-runif(sizeDS,xMin, xMax) # inputs
  ## Repsonse
  y<-func(x)+rnorm(sizeDS,0,sig) #realized values f(x)+noise
  ## Put in a data frame
  data.frame(x,y)
}
```

Variables

We need all these variables

```
f1 <- function(x) 1+x
f2 <- function(x) (x-1)*(x+1)
f3 <- function(x) x*(x-1)*(x+1)
form1 <- "y~x"
form2 <- "y~x+I(x^2)"
form3 <- "y~x+I(x^2)+I(x^3)"

sizeDS <- 50
numReps <- 100
sig <- 1.5
```

Set up the range of x values.

```
K <- 100
xVals0 <- seq(-1,1,length=K)
```

Build the data

There are a lot of slick ways to do this. Let's keep it simple

I'll use a quadratic model

```
f <- f2
```

Flexibility 1

```
resVals1 <- matrix(nrow=K,ncol=4)
for(k in  1:K){
  resVals1[k,] <- biasVarTO3.lm(form1,sizeDS,numReps,xVals0[k])
}
```

Flexibility 2

```
resVals2 <- matrix(nrow=K,ncol=4)
for(k in  1:K){
  resVals2[k,] <- biasVarTO3.lm(form2,sizeDS,numReps,xVals0[k])
}
```

Flexibility 3

```
resVals3 <- matrix(nrow=K,ncol=4)
for(k in  1:K){
  resVals3[k,] <- biasVarTO3.lm(form3,sizeDS,numReps,xVals0[k])
}
```

Pack it all together

Flex 1

```
data1.df <- data.frame(resVals1,deg=1)
names(data1.df) <- c("mse","var","bias2","noise","deg")
```

Flex 2

```
data2.df <- data.frame(resVals2,deg=2)
names(data2.df) <- c("mse","var","bias2","noise","deg")
```

Flex 3

```
data3.df <- data.frame(resVals3,deg=3)
names(data3.df) <- c("mse","var","bias2","noise","deg")
```

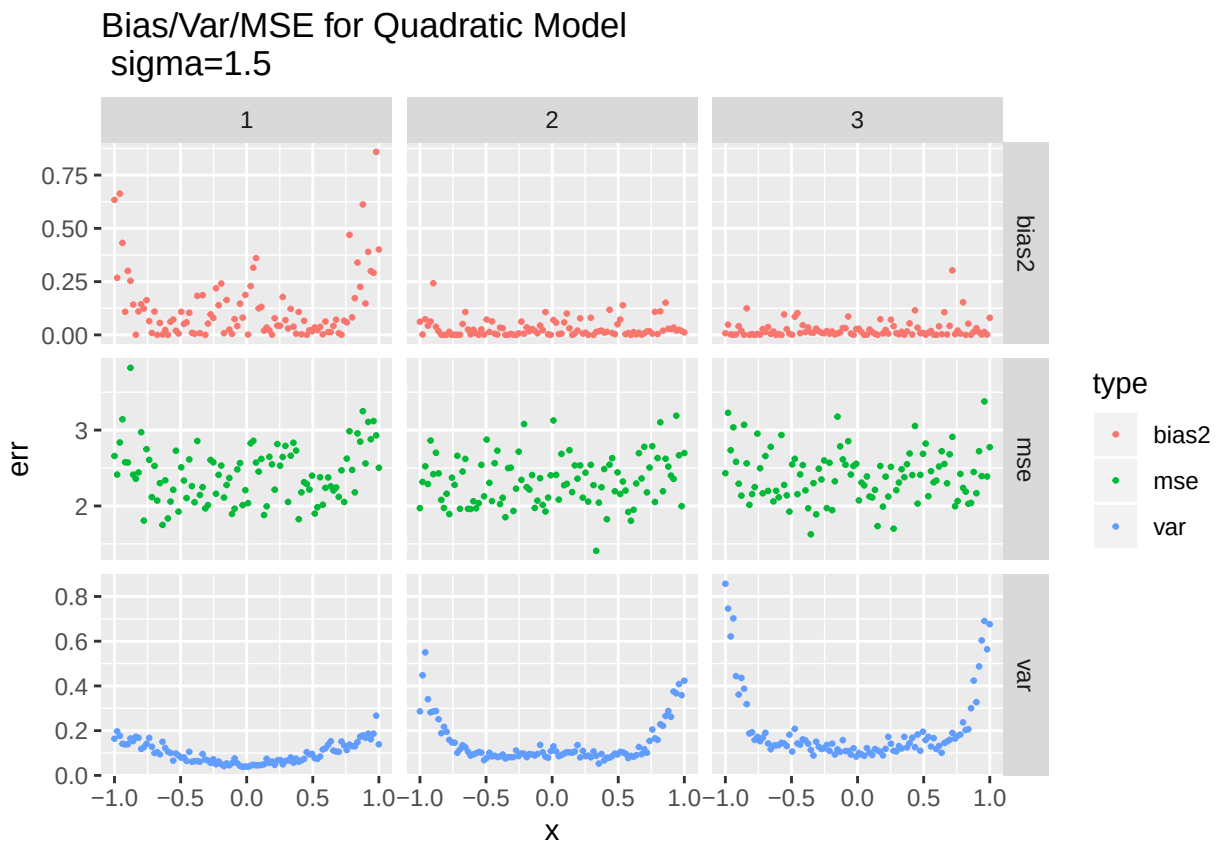
The whole shebang

```
data.df <- bind_rows(data1.df,data2.df,data3.df)
##Add  x values..note the repetition  (one set for each flex value)
data.df$x <- rep(xVals0,3)
```

The Plot

Here it is.

```
data.df %>%
  select(-noise) %>%
  gather(type,err,c(mse,var,bias2)) %>%
  ggplot()+
  geom_point(aes(x,err,color=type),size=.5)+
  facet_grid(type~deg,scale="free_y")+ ##free_y allows different scales
  labs(title=sprintf("Bias/Var/MSE for Quadratic Model\n sigma=%s",sig))
```



The Bias^2, Variance, and MSE all increase significantly near the endpoints of the interval at Flexibility -1. This trend continues for the variance across all the flexibility values.