Homework 2

Sohaib Syed

2023-02-15

Problem 2 Exercise 3.2

Part 1

Given X and Y, assuming Y is the response variable and X are the input.

We want to fit f(x)= B0X^0 + B1X + B2X^2 + B3X^3

We want fitted curve so need a plot.

For this first part need 95% CI at each x0, which is

b_hat +/- 1.96 * sqrt(vj * sigma^2)

where vj is diagonal of (X_t*X)^-1

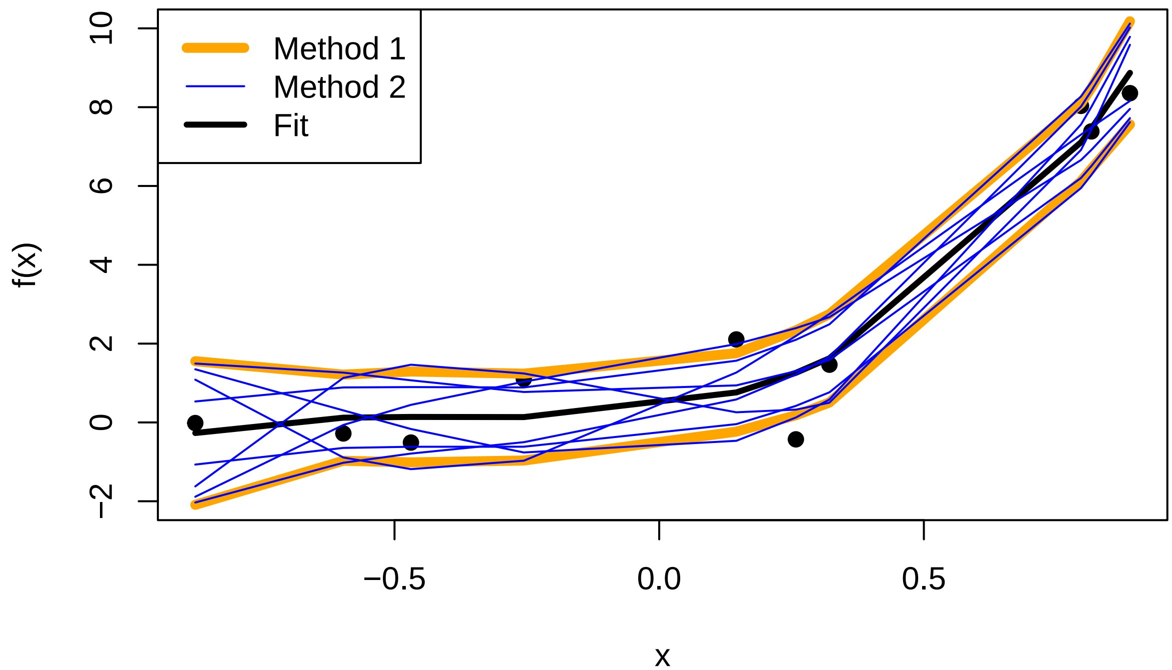
We have to find Var(a^T * B)= a^T * Var(B) * a = sigma^2 X_t (X_t * X) ^-1 * X

```
set.seed(1)
N <- 10
X <- runif(N,-1,1)
Y <- 1+(2*X)+(3*X^2)+(4*X^3) +rnorm(N)
fit <- lm(Y ~ X + I(X^2) + I(X^3))
yhat <- fit$fitted.values
sigma_square <- sum(fit$residuals^2) / (N - 2)

X_combined <- cbind(1, X, X^2, X^3)

# Point wise (local) confidence intervals
Xt_X <- t(X_combined) %*% X_combined
invXX <- solve(Xt_X)
vj_sigmasquare <- apply(X_combined, 1, function(x) t(x) %*% invXX %*% x) * sigma_square

idx <- order(X)
```

Part 2

For the second approach we are trying to solve for (Bhat-B) which will result in 4 upper bound lines and 4 lower bound lines for each Beta rather than for each point as in approach 1. This will result in a wider confidence CI since it is a global approach to the problem rather than local. The derivation is done in paper submission and follows a SVD on X. This way solving for (Bhat-B) becomes simpler.

```r
s <- svd(X_combined) # we get U D and V'

# essentially the variance of this confidence band
var_Cb<-sqrt(sigma_square * qchisq(.95, 4))
Var_over_diagonalofX<-var_Cb/s$d
# put back into matrix form since D needs to be pxp
diagonal_entriesofvar<-diag(Var_over_diagonalofX)

# we need to solve this problem :
#  Ax=B, where x is (betahat-beta), A is V' and B is the diagonal entries in pxp dimensions
betachange<-solve(t(s$v),diagonal_entriesofvar)

store_upper <- list()
store_lower <- list()
for( i in 1:ncol(betachange)){
  store_upper <- append(store_upper, X_combined[idx,] %*% (fit$coefficients + betachange[,i]))
  store_lower <- append(store_lower, X_combined[idx,] %*% (fit$coefficients - betachange[,i]))
}

plot(X, Y, xlab = "x", ylab = "f(x)",ylim = c(-2,10),pch=19)
lines(X[idx],yhat[idx],lwd=3) # the fit line

lines(X[idx], yhat[idx] + qnorm(.975) * sqrt(vj_sigmasquare[idx]), col = "orange",lwd =5) # upper bound
lines(X[idx], yhat[idx] - qnorm(.975) * sqrt(vj_sigmasquare[idx]), col = "orange",lwd=5) # lower bound

lines(X[idx],  (unlist(store_upper[1:10])), col = "blue") # upper bound b0
lines(X[idx],  (unlist(store_upper[11:20])), col = "blue") # upper bound b1
lines(X[idx],  (unlist(store_upper[21:30])), col = "blue") #upper bound b2
lines(X[idx],  (unlist(store_upper[31:40])), col = "blue") #upper bound b3
lines(X[idx],  (unlist(store_lower[1:10])), col = "blue") # lower bound b0
lines(X[idx],  (unlist(store_lower[11:20])), col = "blue") #lower bound b1
lines(X[idx],  (unlist(store_lower[21:30])), col = "blue") #lower bound b2
lines(X[idx],  (unlist(store_lower[31:40])), col = "blue") #lower bound b3


legend("topleft", c("Method 1", "Method 2", "Fit"), col= c("orange", "blue", "black"), lwd = c(5, 1, 3)
```



As we see the two methods plotted, we see that the approaches are certainly different. We have stated as before that approach 2 is a global approach to confidence interval of Betas and approach 1 is a local approach for betas using the points x0. The graph confirms that approach 2 has wider interval as we see at certain points the blue lines go out of the area of the orange lines.