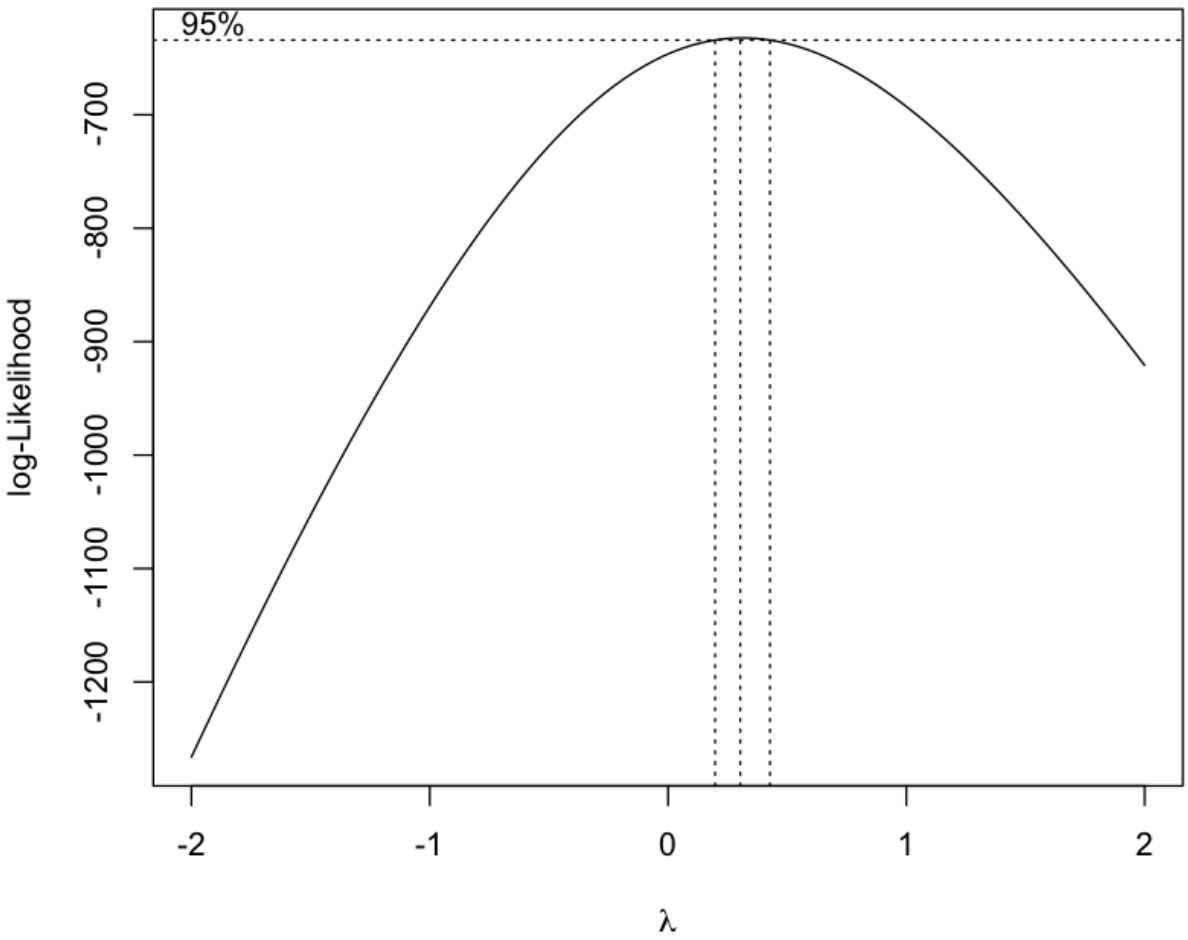
CS 498 HW6 Report

1. Indexes of removed points:

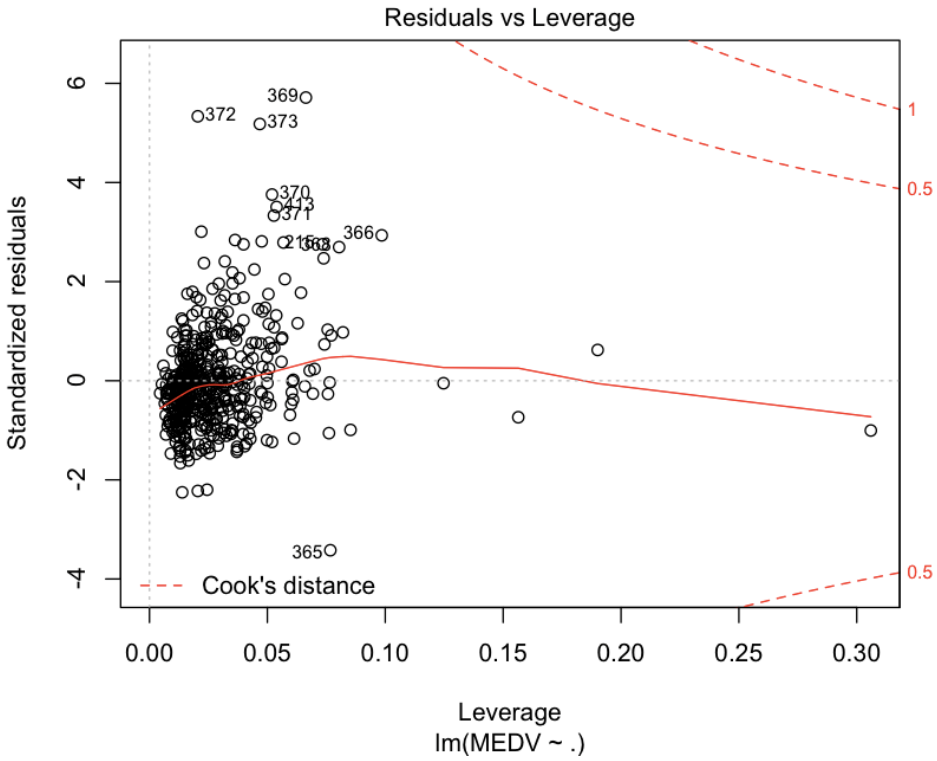
   413, 366, 369, 372, 373, 370, 371, 365, 368

2. Log-likelihood vs. λ:

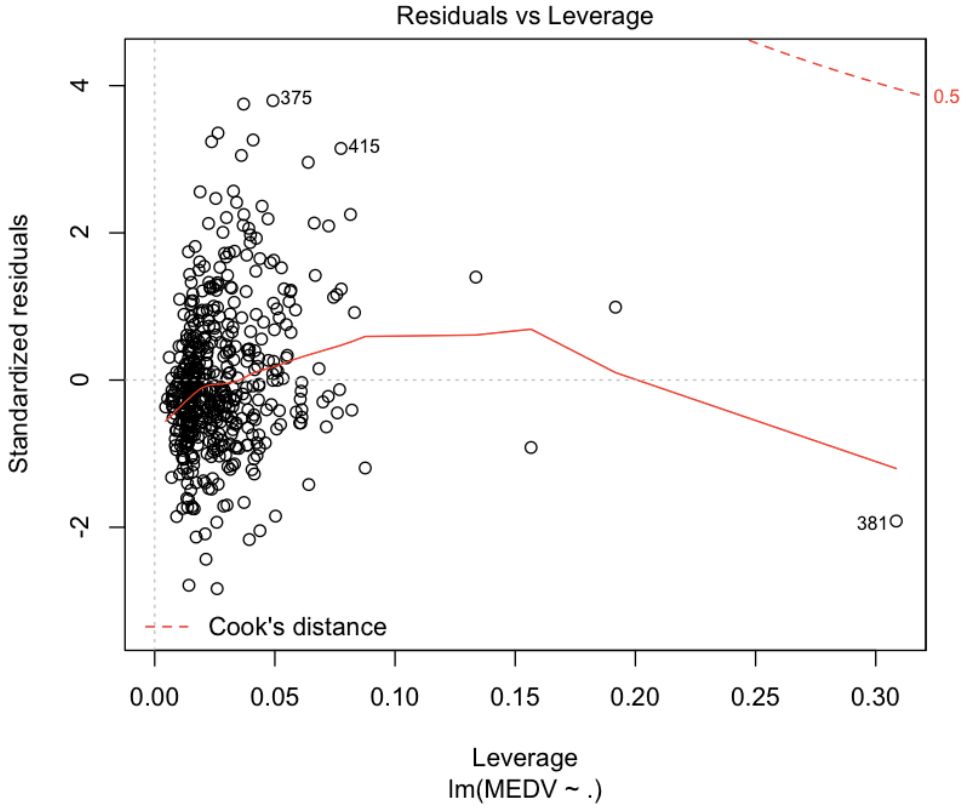


The best lambda we found is 0.3030303.

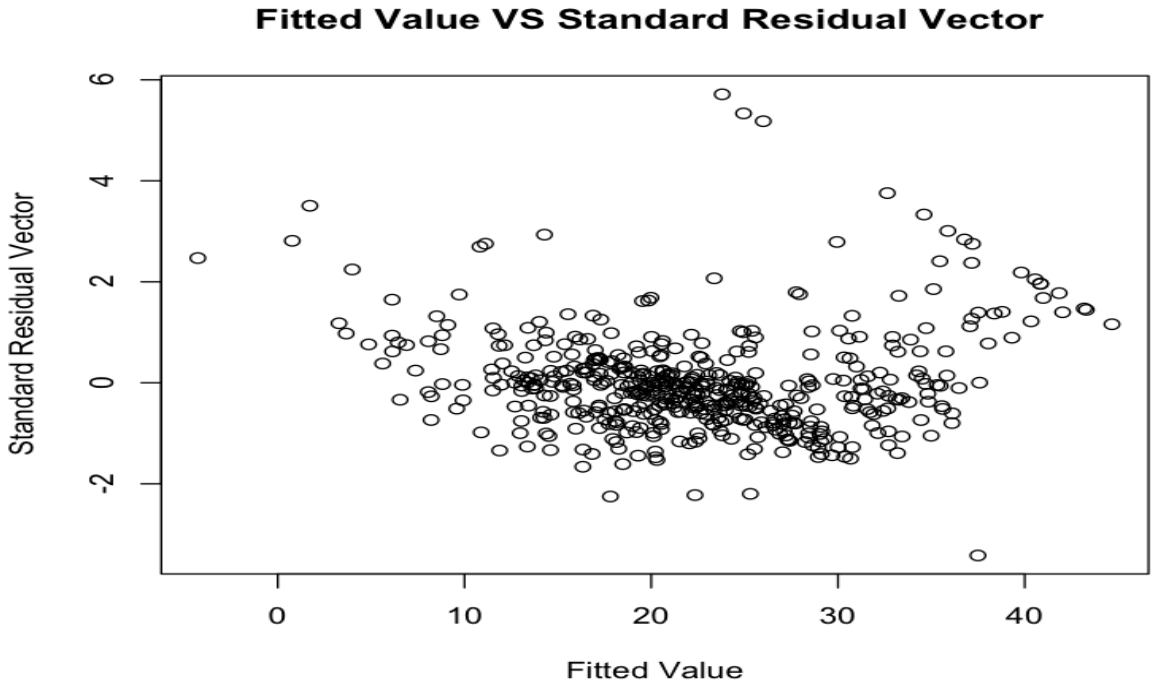
3. Diagnostic Plot



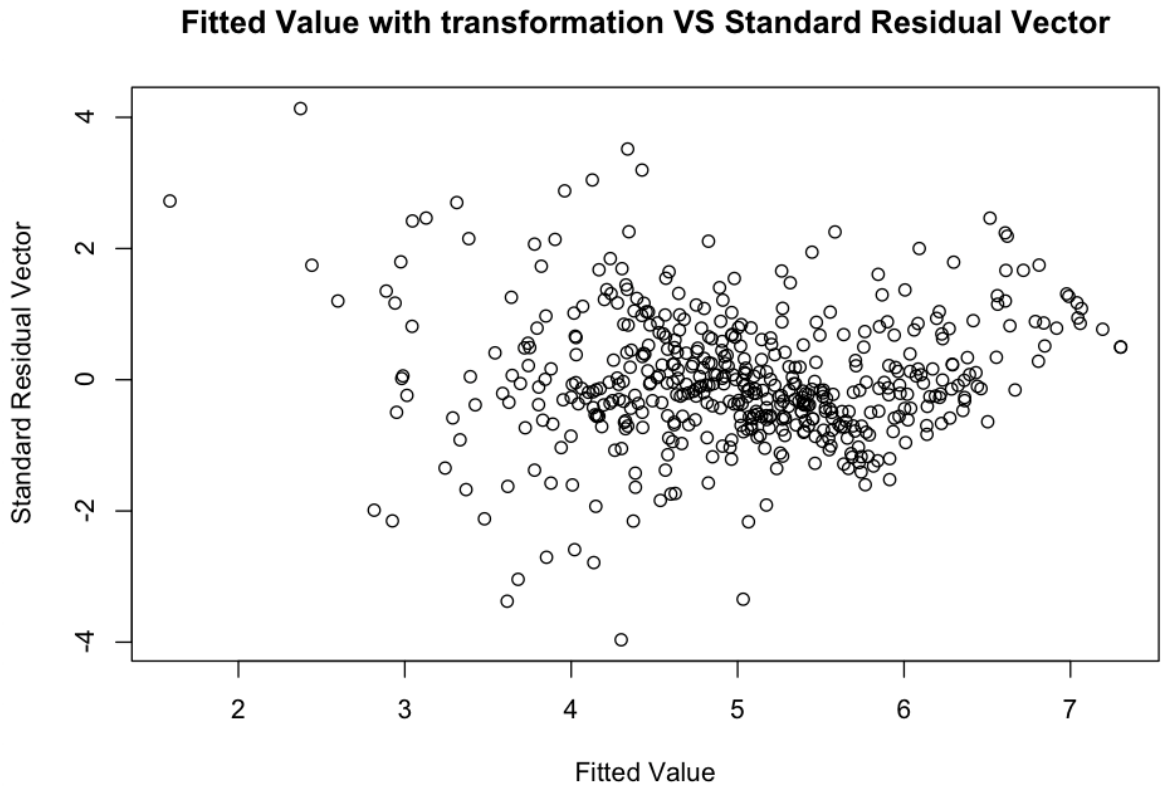
Final Diagnostic Plot



4. Plot of Standardized Residuals vs. Fitted Value for linear regression

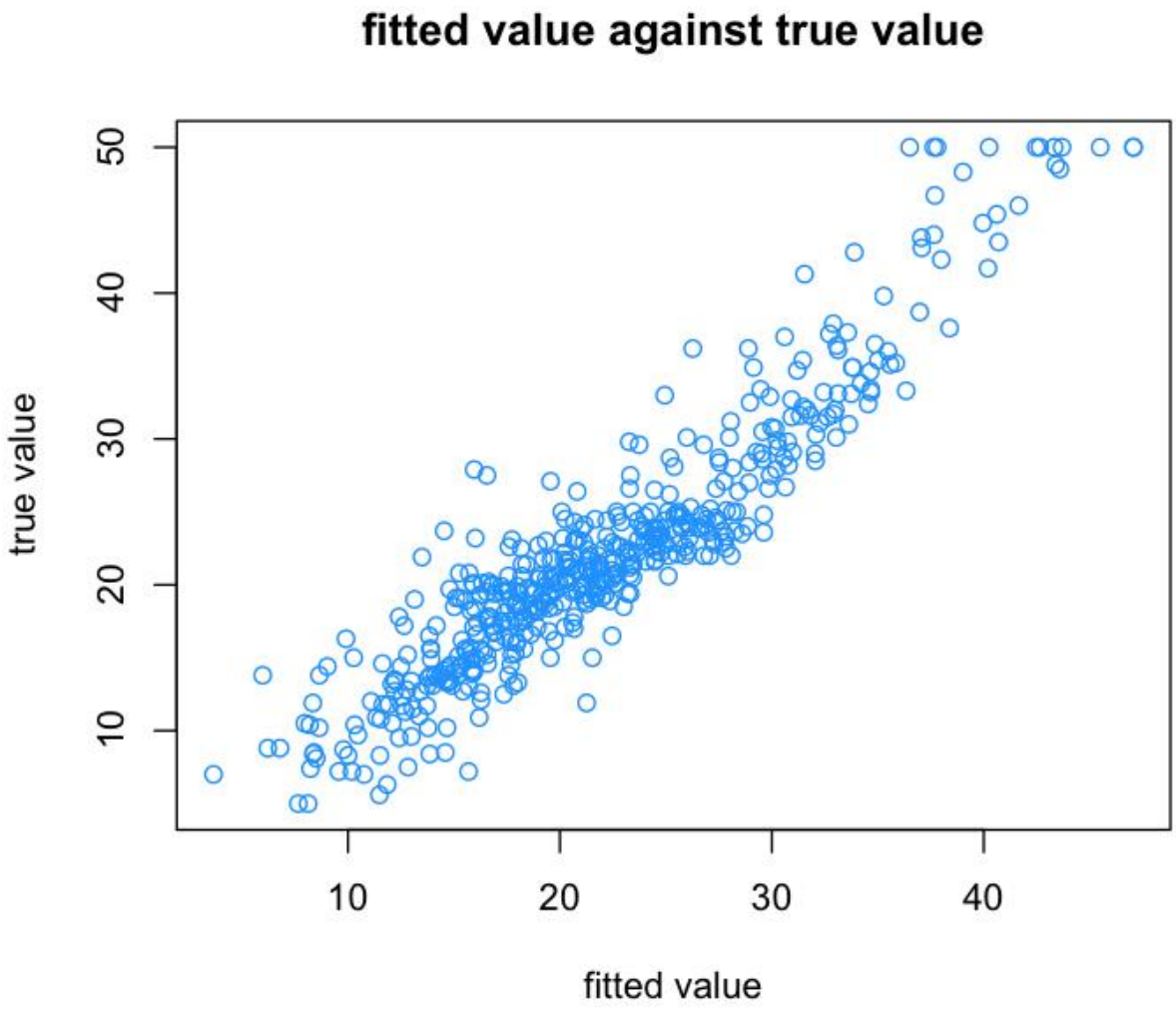


5. Plot of Standardized Residuals vs. Fitted Value for final regression



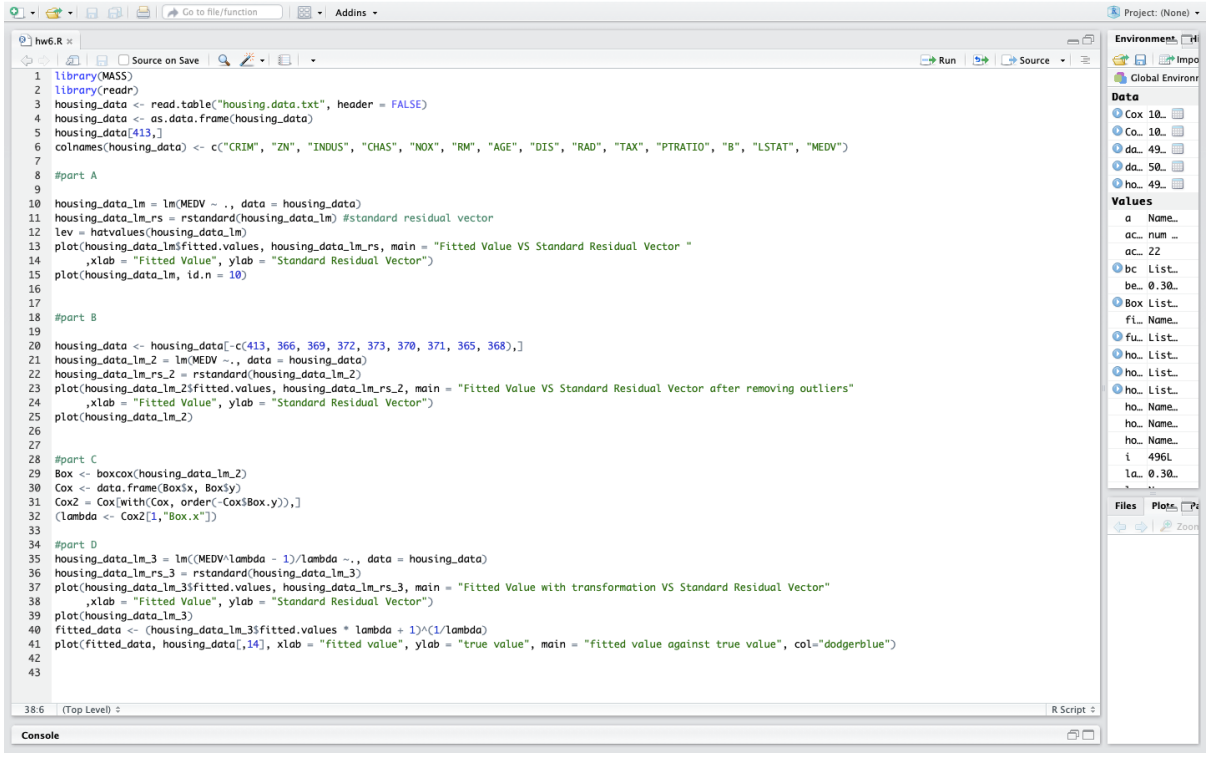
6. Observation: Before transformation, the plot does not look like random normalized. There are some points with extreme standard residuals accumulated around the origin. After the transformation, the plot looks like a random normal distribution, and the residuals are more scattered around the origin.

7. Final Plot of Fitted House Price vs. True House Price



Observation: The sample points are distributed around a linear line intercepting y-axis at 0. This outcome implies that fitted values are mostly close to the true values. Therefore, the model is effective and accurate.

8. Code Screenshot



```
library(MASS)
library(readr)
housing_data <- read.table("housing.data.txt", header = FALSE)
housing_data <- as.data.frame(housing_data)
housing_data[413,]
colnames(housing_data) <- c("CRIM", "ZN", "INDUS", "CHAS", "NOX", "RM", "AGE", "DIS", "RAD", "TAX", "PTRATIO", "B", "LSTAT", "MEDV")

#part A

housing_data_lm = lm(MEDV ~ ., data = housing_data)
housing_data_lm_rs = rstandard(housing_data_lm) #standard residual vector
lev = hatvalues(housing_data_lm)
plot(housing_data_lm$fitted.values, housing_data_lm_rs, main = "Fitted Value VS Standard Residual Vector "
     ,xlab = "Fitted Value", ylab = "Standard Residual Vector")
plot(housing_data_lm, id.n = 10)


#part B

housing_data <- housing_data[-c(413, 366, 369, 372, 373, 370, 371, 365, 368),]
housing_data_lm_2 = lm(MEDV ~., data = housing_data)
housing_data_lm_rs_2 = rstandard(housing_data_lm_2)
plot(housing_data_lm_2$fitted.values, housing_data_lm_rs_2, main = "Fitted Value VS Standard Residual Vector after removing outliers"
     ,xlab = "Fitted Value", ylab = "Standard Residual Vector")
plot(housing_data_lm_2)


#part C
Box <- boxcox(housing_data_lm_2)
Cox <- data.frame(Box$x, Box$y)
Cox2 = Cox[with(Cox, order(-Cox$Box.y)),]
(lambda <- Cox2[1,"Box.x"])

#part D
housing_data_lm_3 = lm((MEDV^lambda - 1)/lambda ~., data = housing_data)
housing_data_lm_rs_3 = rstandard(housing_data_lm_3)
plot(housing_data_lm_3$fitted.values, housing_data_lm_rs_3, main = "Fitted Value with transformation VS Standard Residual Vector"
     ,xlab = "Fitted Value", ylab = "Standard Residual Vector")
plot(housing_data_lm_3)
fitted_data <- (housing_data_lm_3$fitted.values * lambda + 1)^(1/lambda)
plot(fitted_data, housing_data[,14], xlab = "fitted value", ylab = "true value", main = "fitted value against true value", col="dodgerblue")
```