

Group Members: Edgar Pino and Gassan Soukaev  
CS498 Applied Machine Learning  
**CS498 AMO**

### **Code for regression and resulting model**

```
orig_model = lm(MEDV ~ ., data = housing_data)
```

```
par(mfrow=c(2,2))
```

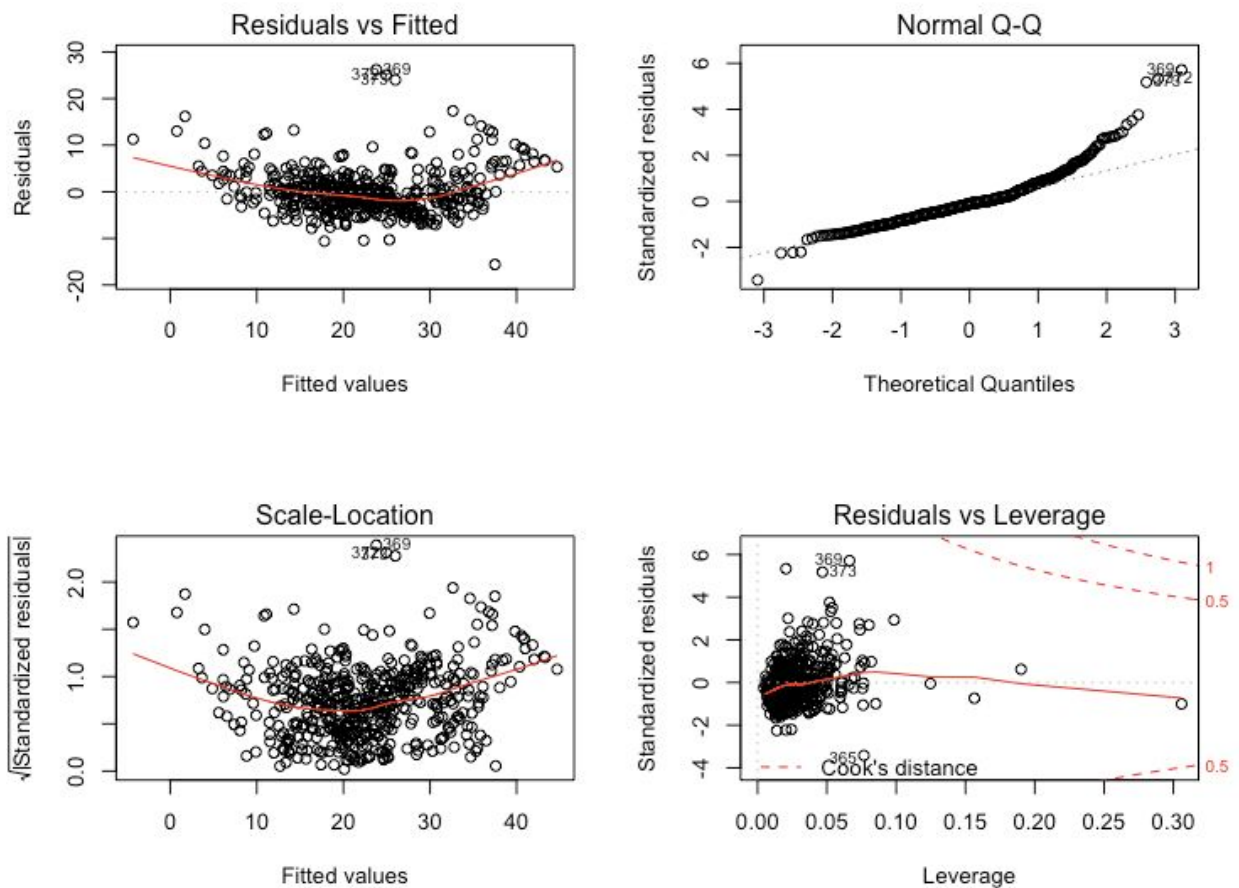
```
plot(orig_model)
```

```
par(mfrow=c(1,1))
```

```
plot_fitted_resid(orig_model)
```

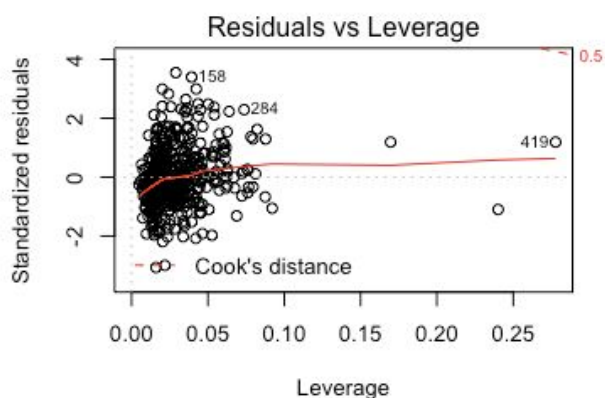
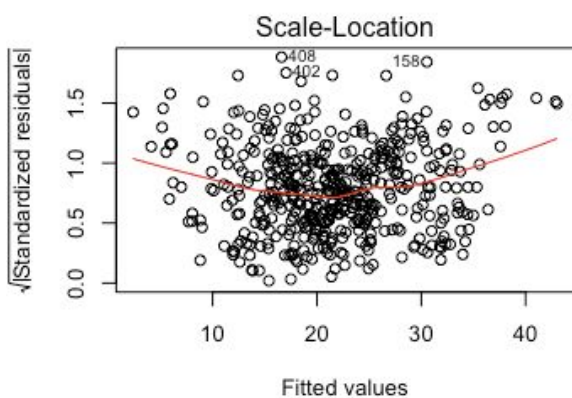
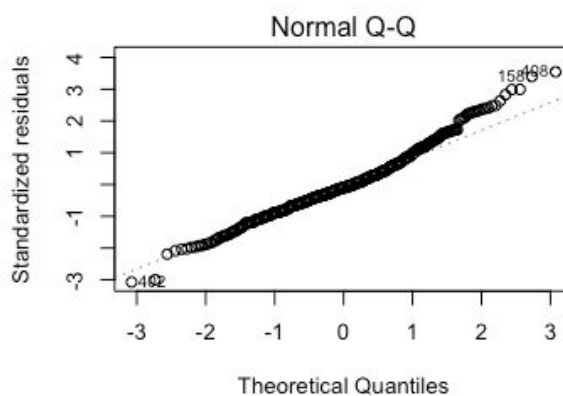
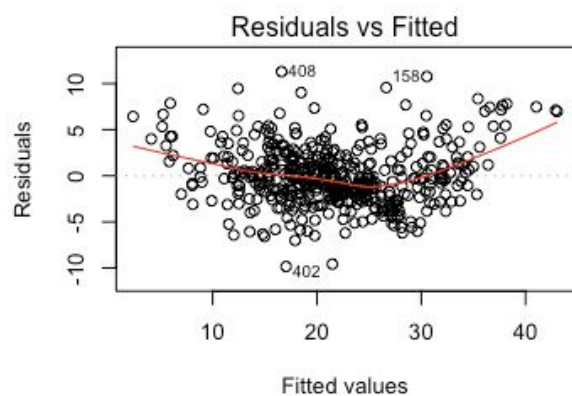
```
possible_outliers = as.numeric(names(resid(orig_model)[cooks.distance(orig_model) > 10 /  
length(cooks.distance(orig_model))]))
```

## Diagnostic Plot



We decided to remove anything with a cooks distance of 4. This is because most of the data points were below that. This are the points being removed:  
 65,142,149,162,163,164,167,187,196,205,215,226,229,234,254,263,268,365,366,368,369,370,  
 371,372,373,375,376,381,413,415.

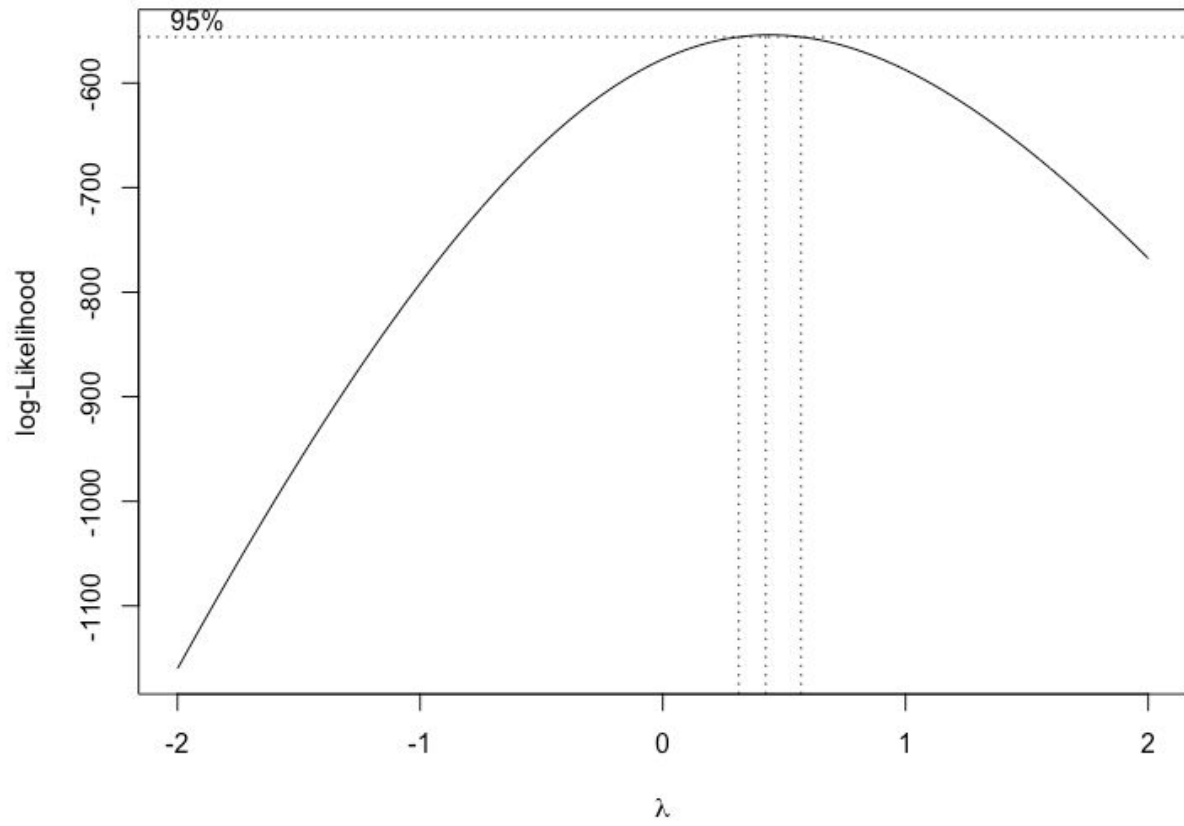
## New diagnostic plot



## Code for subproblem 2

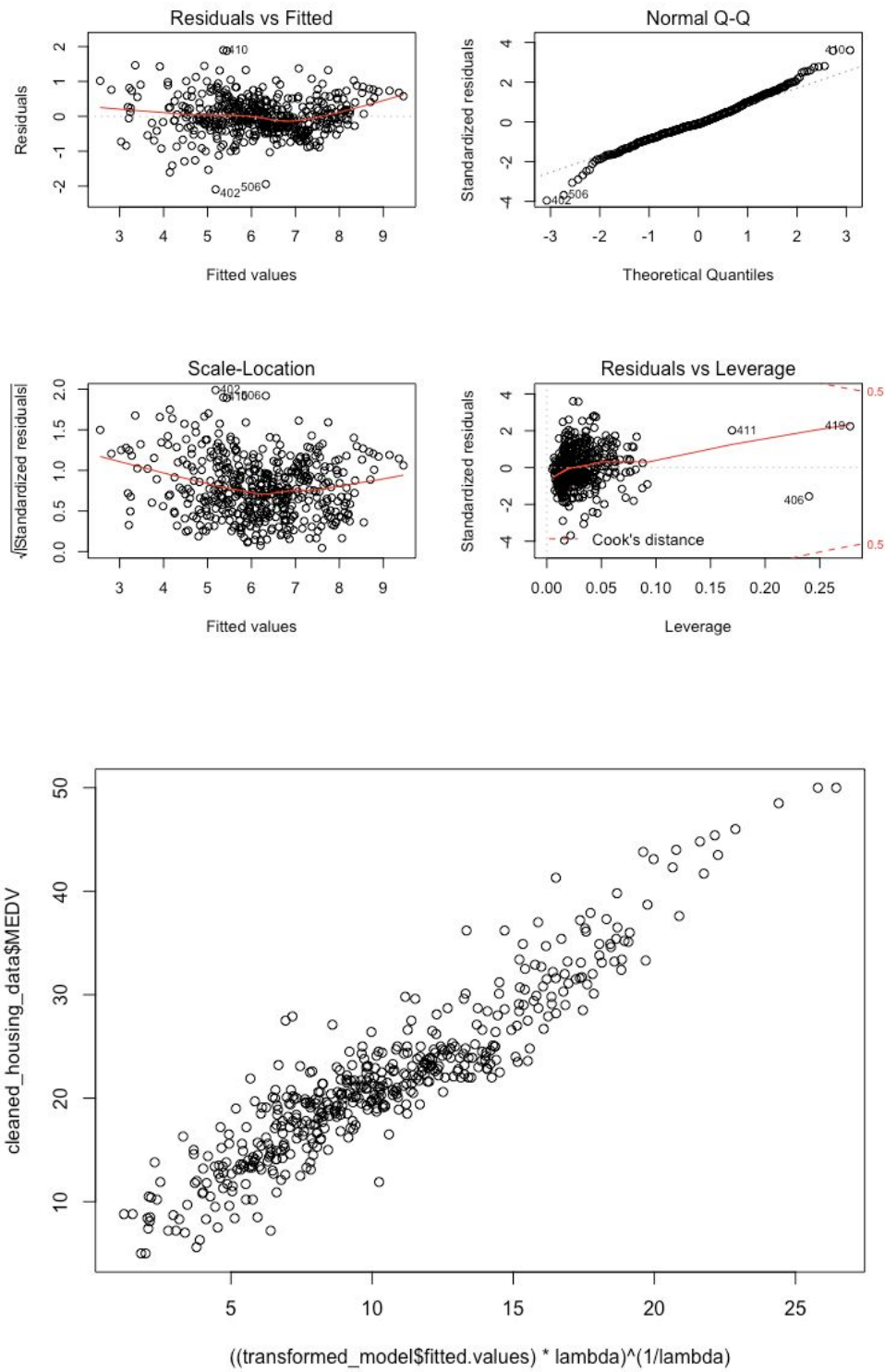
```
require(MASS)
bc = boxcox(new_model)
lambda = bc$x[which.max(bc$y)]
```

Box-Cox transformation plot



Best Value: 0.4242424

## Results after Box-Cox



### Code for subproblems 3 and 4

```
transformed_model = lm(((MEDV ^ lambda - 1)/lambda) ~ ., data =  
cleaned_housing_data)
```

```
par(mfrow=c(2,2))  
plot(transformed_model)  
par(mfrow=c(1,1))
```

```
plot_fitted_resid(transformed_model)
```

```
plot(((transformed_model$fitted.values)*lambda)^(1/lambda),  
cleaned_housing_data$MEDV)
```