Brian Holliday

Professor Li

Intro to Data Mining

4 May 2020

Project 6b

1.

Find the linear regression function for model:

$X3 = \beta1X1 + \beta2X2 + \beta4X4 + \beta5X5.$

Figure 1: Multilinear Model Summary

```
Call:
lm(formula = x3 ~ x1 + x2 + x4 + x5 - 1, data = dat, na.action = na.omit)

Residuals:
    Min      1Q  Median      3Q     Max 
-30.642 -18.932  -6.257   7.463  40.550 

Coefficients:
   Estimate Std. Error t value Pr(>|t|)    
x1  -1.4748     0.7327  -2.013   0.0477 *  
x2  48.1178     2.6601  18.088  < 2e-16 ***
x4  -8.8958     0.9940  -8.949 1.67e-13 ***
x5  -0.1826     0.2833  -0.644   0.5213    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 17.49 on 76 degrees of freedom
Multiple R-squared:  0.9023,	Adjusted R-squared:  0.8972 
F-statistic: 175.5 on 4 and 76 DF,  p-value: < 2.2e-16
```

We can see that for our model, we do have a model. We reject the null hypothesis for all our variables except for x5. The null hypothesis being that the coefficient of the variable is equal to zero. We have an R^2 value of about 90 percent, meaning that we can account for 90 percent of the variation of the data. Our equation is:

$$X3 = -1.4748*X1 + 48.1178*X2 + -8.8958*X4 + -0.1826*X5$$

2. Compute the root mean square error (RMSE) for your model.

Figure 2: Multilinear RMSE

```
> y <- na.omit(dat$x3)
>
> y_pred <- predict(multireg_model, newdata = dat)
>
> RMSE <- sqrt(sum((y_pred - dat$x3)^2))/sqrt(length(y))
> RMSE
[1] 17.05075
```

From the RMSE, we can see how far in total the data points of our model are from the actual dat a. So, we can say that our data is about 17.05 points away from the actual data.