Brian Holliday

Professor Li

Intro to Data Mining

11 May 2020

Project 6c

1. Fit the given data for model

$$X3 = \beta1 \sin(X2) + \beta2X\ 3\ 2 + \beta4X4 + \beta5 \ln(X5) + C.$$

Figure 1: Variation Multilinear Model Summary

```
Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)  -8.2825     3.2568  -2.543   0.0130 *
b1            8.7047     3.1088   2.800   0.0065 **
b2            5.3452     0.1124  47.568  < 2e-16 ***
b4           -2.7565     0.3691  -7.468 1.22e-10 ***
b5            3.2023     0.3878   8.258 3.86e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.869 on 75 degrees of freedom
Multiple R-squared:  0.998,   Adjusted R-squared:  0.9979
F-statistic:  9526 on 4 and 75 DF,  p-value: < 2.2e-16
```

We can see that for our model, we do have a model. All of our coefficents are statistically significant. have an R^2 value of about 99 percent, meaning that we can account for 99 percent of the variation of the data. We almost have a perfect fit for our data and our equation is:

$$x3 = 8.7047*\sin(x2) + 5.3452*(x2^3) + -2.7565*x4 + 3.2023*\ln(x5) – 8.2825$$

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

Figure 2: Variation Multilinear RMS

```
> y_pred <- predict(var_model, newdata = data.frame(Amatrix))
>
> RMSE <- sqrt(sum((y_pred - y)^2))/sqrt(length(y))
> RMSE
[1] 1.810087
```

We can see that the RMSE is 1.810087. We can say that we are an average of 1.81 points away from our actual data with our model.

3.

Use the above model to estimate the value of X3 when X2 = 1.5, X4 = 5 and X5 = 10.

Figure 3: Prediction Model

```
> pred <- predict(var_model, newdata=data.frame(b1 = sin(1.5),
> pred <- predict(var_model, newdata=data.frame(b1 = sin(1.5),
+                                               b2 = (1.5)^3, b4 = 5,
+                                               b5 = log(10)))
> pred
      1
12.0316
>
>
>
> exact <- 8.7047*sin(1.5) + 5.3452*(1.5)^3 + -2.7565*5 + 3.2023*log(10) -8.2
825
>
> abs(exact - pred)
           1
8.784121e-05
```

Using our prediction model, we get that 12.0316 is the value if x2 = 1.5, x4 = 5 and x5 = 10.

This is 8.784121e-05 away from our actual value. This shows us that our prediction values are ve ry close to our actual value. This fact combined with a 0.998 R^2 value tells us that we have a ne ar perfect model for our data.