Homework 05

Anika Cartas

Wednesday, April 08, 2015

### 0. Read Data

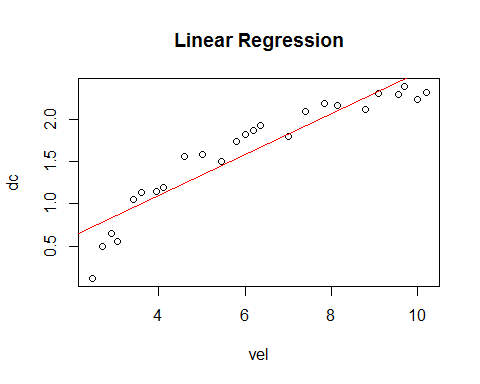
The table lists measurements of DC output at different wind velocity.

* Dependent variable: DC
* Independent variable: Wind Velocity

data <- data.frame(  
 vel = c(5,6,3.4,2.7,10,9.7,9.55,3.05,8.15,6.2,2.9,6.35,4.6,5.8,7.4,3.6,7.85,8.8,7,5.45,9.1,10.2,4.1,3.95,2.45),  
 dc = c(1.582,1.822,1.057,0.5,2.236,2.386,2.294,0.558,2.166,1.866,0.653,1.93,1.562,1.737,2.088,1.137,2.179,2.112,1.8,1.501,2.303,2.310,1.194,1.144,0.123))  
  
data <- data[with(data,order(vel)),]  
  
attach(data)

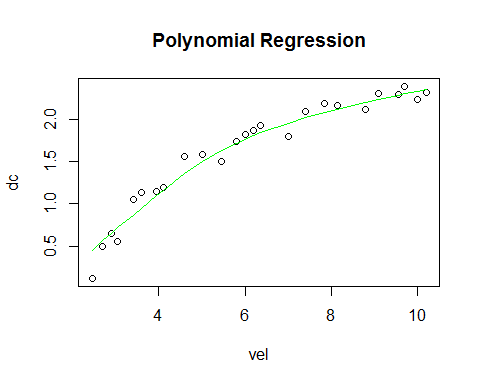
### 1. Linear Regression

reg.linear <- lm(dc~vel)  
  
plot(dc~vel, main="Linear Regression")  
abline(reg.linear, col="red")



### 2. Polynomial Regression

plot(dc~vel, main="Polynomial Regression")  
lines(lowess(dc~vel), col="green")



Using lowess, we visualize this to be a 4th degree polynomial. However, let's see if we can get more details.

fit2 <- lm(dc ~ poly(vel, 2, raw=TRUE))  
fit3 <- lm(dc ~ poly(vel, 3, raw=TRUE))  
fit4 <- lm(dc ~ poly(vel, 4, raw=TRUE))  
fit5 <- lm(dc ~ poly(vel, 5, raw=TRUE))  
summary(fit2)

##   
## Call:  
## lm(formula = dc ~ poly(vel, 2, raw = TRUE))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.26347 -0.02537 0.01264 0.03908 0.19903   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.155898 0.174650 -6.618 1.18e-06 \*\*\*  
## poly(vel, 2, raw = TRUE)1 0.722936 0.061425 11.769 5.77e-11 \*\*\*  
## poly(vel, 2, raw = TRUE)2 -0.038121 0.004797 -7.947 6.59e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1227 on 22 degrees of freedom  
## Multiple R-squared: 0.9676, Adjusted R-squared: 0.9646   
## F-statistic: 328.3 on 2 and 22 DF, p-value: < 2.2e-16

summary(fit3)

##   
## Call:  
## lm(formula = dc ~ poly(vel, 3, raw = TRUE))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.199016 -0.054316 0.008116 0.064812 0.164210   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.358550 0.438464 -5.379 2.46e-05 \*\*\*  
## poly(vel, 3, raw = TRUE)1 1.426866 0.246758 5.782 9.70e-06 \*\*\*  
## poly(vel, 3, raw = TRUE)2 -0.160366 0.042056 -3.813 0.00101 \*\*   
## poly(vel, 3, raw = TRUE)3 0.006458 0.002211 2.921 0.00817 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1059 on 21 degrees of freedom  
## Multiple R-squared: 0.9769, Adjusted R-squared: 0.9737   
## F-statistic: 296.6 on 3 and 21 DF, p-value: < 2.2e-16

summary(fit4)

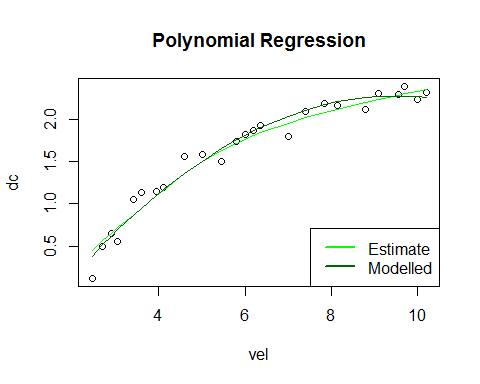
##   
## Call:  
## lm(formula = dc ~ poly(vel, 4, raw = TRUE))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.17091 -0.06373 0.02672 0.07621 0.11434   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -4.7151534 1.0997243 -4.288 0.000359 \*\*\*  
## poly(vel, 4, raw = TRUE)1 3.2773039 0.8353441 3.923 0.000842 \*\*\*  
## poly(vel, 4, raw = TRUE)2 -0.6620619 0.2214626 -2.989 0.007245 \*\*   
## poly(vel, 4, raw = TRUE)3 0.0627544 0.0245587 2.555 0.018861 \*   
## poly(vel, 4, raw = TRUE)4 -0.0022306 0.0009698 -2.300 0.032342 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.09648 on 20 degrees of freedom  
## Multiple R-squared: 0.9818, Adjusted R-squared: 0.9781   
## F-statistic: 269.2 on 4 and 20 DF, p-value: < 2.2e-16

summary(fit5)

##   
## Call:  
## lm(formula = dc ~ poly(vel, 5, raw = TRUE))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.16744 -0.06573 0.02651 0.06511 0.11238   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -6.4743544 3.1589991 -2.049 0.0545 .  
## poly(vel, 5, raw = TRUE)1 5.0279731 3.0604169 1.643 0.1168   
## poly(vel, 5, raw = TRUE)2 -1.3154225 1.1201753 -1.174 0.2548   
## poly(vel, 5, raw = TRUE)3 0.1775419 0.1943954 0.913 0.3725   
## poly(vel, 5, raw = TRUE)4 -0.0117847 0.0160763 -0.733 0.4725   
## poly(vel, 5, raw = TRUE)5 0.0003034 0.0005096 0.595 0.5586   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.09808 on 19 degrees of freedom  
## Multiple R-squared: 0.9821, Adjusted R-squared: 0.9774   
## F-statistic: 208.5 on 5 and 19 DF, p-value: 6.523e-16

Depending on how specific we want to get, it looks like we are adding significant variables up to 4th degree, though maybe just 2nd or 3rd is fine. I'm not sure how to decide what the cut-off point is, but based on recommendations am going to continue with 2nd degree polynomial.

plot(dc~vel, main="Polynomial Regression")  
lines(lowess(dc~vel), col="green")  
reg.poly <- lm(dc ~ poly(vel, 2, raw=TRUE))  
lines(vel, predict(reg.poly),col="darkgreen")   
  
legend("bottomright", c("Estimate","Modelled"), col=c("green","darkgreen"), lwd=2)



### 3. Transformation Based Regression

The Box-Cox Power Transformation identifies an appropriate exponent (lambda) to transform the data into a normal shape. The Lambda value indicates the power to which all data should be raised.

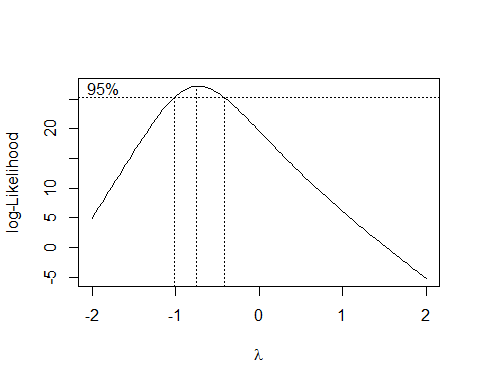
When lambda is plotted against log likelihood, the best lambda is the one that maximizes the log likelihood.

Changed this to be vel ~ dc following advice from article sent out by Danielle, but I'm not sure why it works this way. Originally I had it as dc ~ vel, which led to lambda = 2.

library(MASS)

## Warning: package 'MASS' was built under R version 3.1.3

bc <-boxcox(vel ~ dc)



(maxll <- max(bc$y))

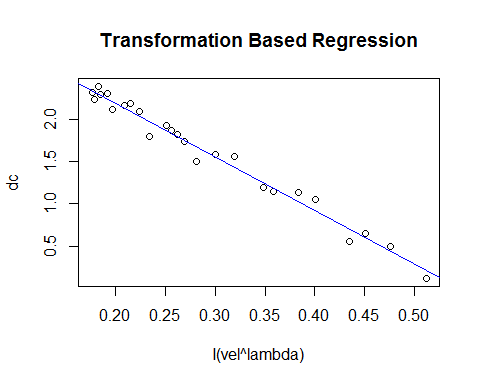
## [1] 27.17658

(lambda <- bc$x[which(bc$y == maxll)])

## [1] -0.7474747

Maximum log likelihood of 27.17 is achieved when lamda is equal to -0.75.

reg.trans<-lm(dc ~ I(vel^lambda))  
plot(dc ~ I(vel^lambda), main="Transformation Based Regression")  
abline(reg.trans, col="blue")



### 4. Nonlinear Regression

I will attempt to solve using least squares approach, following directions from Simran's tutorial. As of class time, I am unsure of how to determine the model I should be using.

### 5. Discussion

Discuss which models are best for this dataset and the physical processes described by your models.