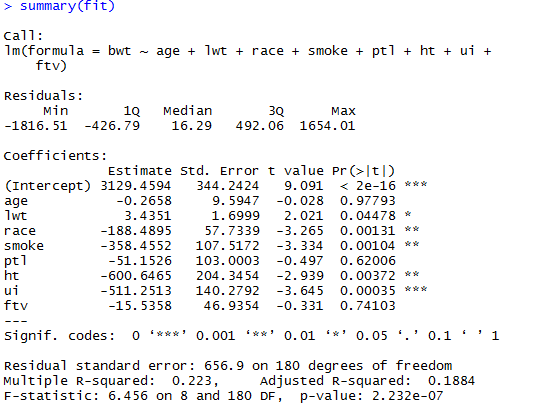
Xuan Zhang(xz2463)

ADA HW4

#1

(i)



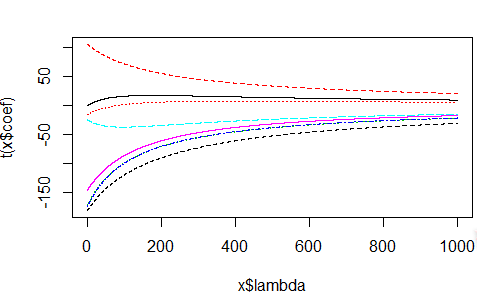
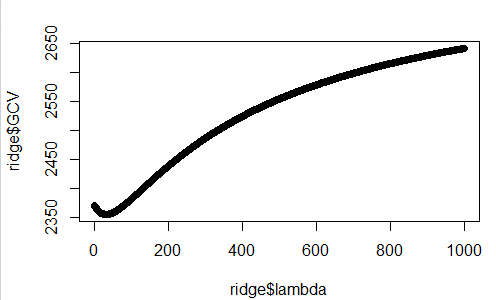
C:\Users\think\AppData\Local\Temp\2CB2.tmp.png

C:\Users\think\AppData\Local\Temp\EA4.tmp.png

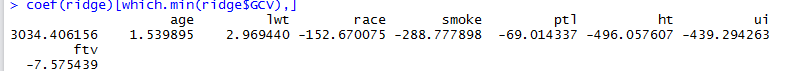
C:\Users\think\AppData\Local\Temp\C062.tmp.png

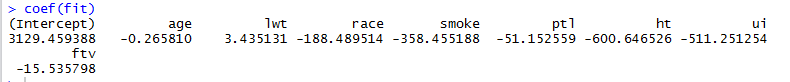
The VIFs for all the variables are not greater than 10, but the mean of VIF =1.138795 >1, and the condition number for multiple linear regression is 683.5843, which is greater than 30. So, >30, so multi-collinearity probably exist.

(ii)

C:\Users\think\AppData\Local\Temp\BDA3.tmp.png

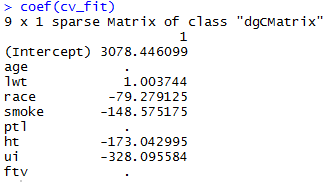




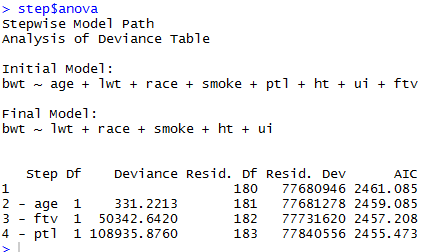
The coefficients for ridge regression and multiple linear regression are pretty similar.

#2

Lasso:



Stepwise(AIC):



We can see from above that the final model selected using LASSO and a stepwise procedure is same, with variable *age*, *ftv* and *ptl* deleted.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | OLS | Ridge | lasso | Elastic Net | |
| Performance when p >> n (high dimensional feature space) | 3 | 2 | 2 | 1 |
| Performance under multicollinearity | 3 | 2 | 1 | 1 | |
| Unbiased estimators | 1 | 2 | 2 | 3 | |
| Model selection capability | 3 | 2 | 1 | 1 | |
| Simplicity: Computation, Inference, Interpretation | 1 | 2 | 3 | 3 | |

#1(i)

library(MASS)

attach(birthwt)

fit<-lm(bwt~age+lwt+race+smoke+ptl+ht+ui+ftv)

summary(fit)

library(car)

vif<-vif(fit)

vif\_bar<-sum(vif)/length(vif)

kappa(fit)

#(ii)

ridge<-lm.ridge(bwt~age+lwt+race+smoke+ptl+ht+ui+ftv,lambda = seq(0,1000,0.001))

select(ridge)

plot(ridge)

coef(ridge)[which.min(ridge$GCV),]

coef(fit)

ridge$lambda[which.min(ridge$GCV)]

plot(ridge$lambda, ridge$GCV)

#2

library(glmnet)

x<-as.matrix(birthwt[,2:9])

y<-as.matrix(birthwt[,10])

cv\_fit<-cv.glmnet(x,y)

coef(cv\_fit)

step <- stepAIC(fit, direction="both")

step$anova