**Assignment (Session 16-20)**

**Problem Statement**

Task 1:

1. Use the below-given data set

Data Set

2. Perform the below-given activities:

a. Predict the no of comments in next H hrs

Note:-

1. Use LASSO, Elastic Net and Ridge and other regression techniques that are covered in the module

2. Report the training accuracy and test accuracy

3. Compare with linear models and report the accuracy

4. Create a graph displaying the accuracy of all models

ANS-

|  |  |
| --- | --- |
|  | crime<-read.csv("uscrime.txt",header = TRUE, sep = "") |
|  | # Data summary |
|  | print(summary(crime)) |
|  | #================================================= |
|  | # Stepwise regression using MASS package and AIC |
|  | #================================================= |
|  | library(MASS) |
|  | # Fit the full model |
|  | full.model <- lm(Crime ~., data = crime) |
|  | # Stepwise regression model |
|  | step.model1 <- stepAIC(full.model, direction = "both", |
|  | trace = TRUE) |
|  | # Full Model summary |
|  | print(summary(full.model)) |
|  |  |
|  | # Stepwise regression model summary |
|  | print(summary(step.model1)) |
|  |  |
|  | # Plotting the model |
|  | par(mfrow=c(2,2)) |
|  | plot(step.model1) |
|  |  |
|  | # ANOVA of Deviance showing how the final model was obtained |
|  | step.model1$anova |
|  |  |
|  | #================================================== |
|  | # Stepwise regression using Caret package and RMSE |
|  | #================================================== |
|  | library(caret) |
|  | # Set up repeated k-fold cross-validation |
|  | train.control <- trainControl(method = "cv", number = 10) |
|  | # Train the model |
|  | step.model2 <- train(Crime ~., data = crime, |
|  | method = "leapSeq", |
|  | tuneGrid = data.frame(nvmax = 1:8), |
|  | trControl = train.control |
|  | ) |
|  |  |
|  | # Result of the stepwise regression |
|  | print(step.model2$results) |
|  |  |
|  | # Best model |
|  | print(step.model2$bestTune) |
|  |  |
|  | # Coefficients of best model |
|  | print(coef(step.model2$finalModel, 6)) |
|  |  |
|  | # Plotting the RMSE with number of variables |
|  | par(mfrow=c(1,1)) |
|  | plot(step.model2,lwd=3,cex=1.5,cex.lab=1.25,cex.axis=2) |
|  |  |
|  | #====================================== |
|  | # LASSO and Elastic Net using glmnet |
|  | #====================================== |
|  | library(glmnet) |
|  | # Convert dataframe to matrix format |
|  | x <- as.matrix(crime[,1:15]) |
|  | y<-as.matrix(crime[,16]) |
|  |  |
|  | #============================== |
|  | # LASSO regression with alpha=1 |
|  | #============================== |
|  | lasso.model<-glmnet(x,y,alpha=1,family="gaussian") |
|  | plot(lasso.model) |
|  |  |
|  | #Cross-validation with glmnet |
|  | cv.lasso.model<-cv.glmnet(x,y,alpha=1,family="gaussian") |
|  | plot(cv.lasso.model) |
|  |  |
|  | #============================== |
|  | # Ridge regression with alpha=0 |
|  | #============================== |
|  | ridge.model<-glmnet(x,y,alpha=0,family="gaussian") |
|  | plot(ridge.model) |
|  |  |
|  | #Cross-validation with glmnet |
|  | cv.ridge.model<-cv.glmnet(x,y,alpha=0,family="gaussian") |
|  | plot(cv.ridge.model) |
|  |  |
|  | #======================================== |
|  | # Elastic Net regression with alpha=0.5 |
|  | #======================================== |
|  | elastic.model<-glmnet(x,y,alpha=0.5,family="gaussian") |
|  | plot(elastic.model) |
|  |  |
|  | #Cross-validation with glmnet |
|  | cv.elastic.model<-cv.glmnet(x,y,alpha=0.5,family="gaussian") |
|  | plot(cv.elastic.model) |
|  |  |
|  | # Lambda for lowest MSE error |
|  | cv.elastic.model$lambda.min |
|  | # Lambda for error is within one standard error of the minimum |
|  | cv.elastic.model$lambda.1se |