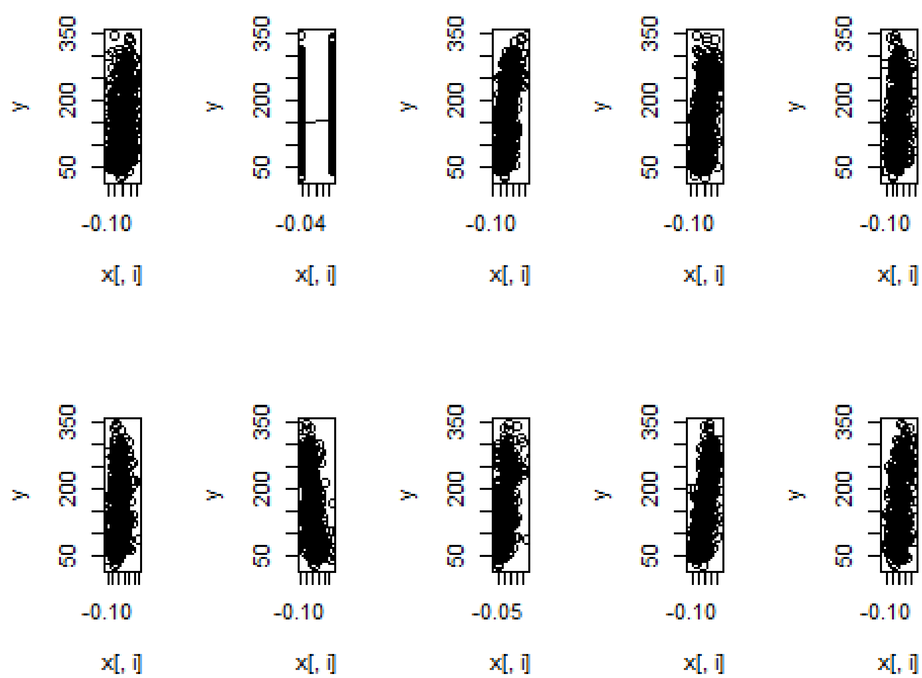
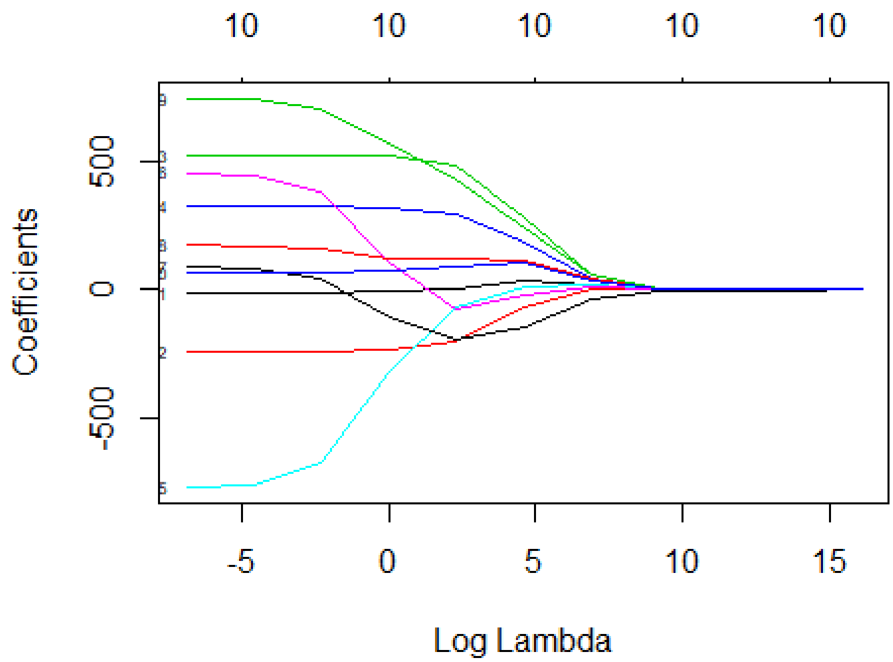
```r
library(lars)

## Loaded lars 1.2

library(glmnet)

## Loading required package: Matrix

## Loaded glmnet 3.0-2

library(caret)

## Loading required package: lattice

## Loading required package: ggplot2

data(diabetes)
attach(diabetes)
set.seed(1234)
par(mfrow=c(2,5))
for(i in 1:10){   plot(x[,i], y)
abline(lm(y~x[,i])) }
```



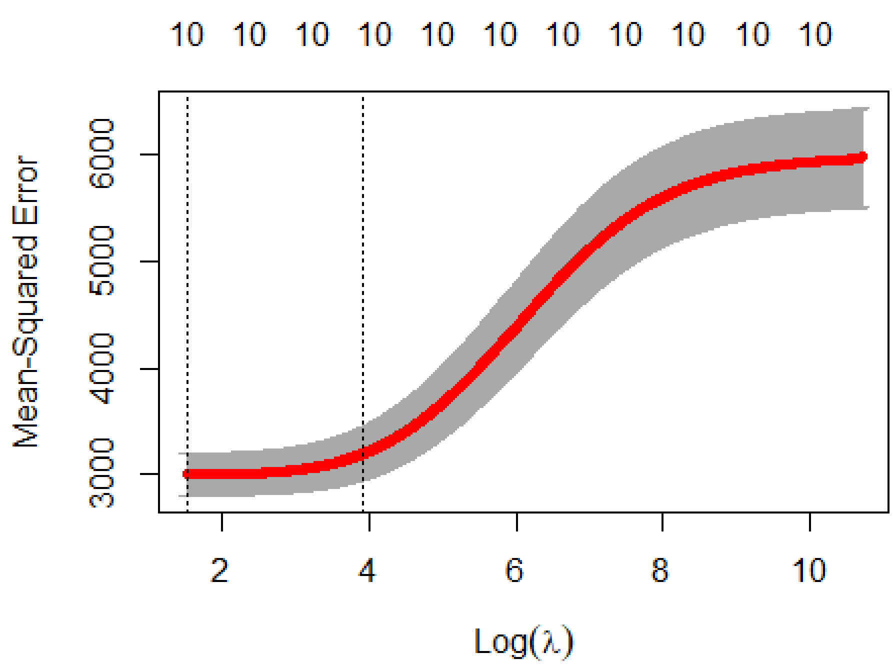
```r
layout(1)

model_ols <- lm(y ~ x)
summary(model_ols)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -155.829  -38.534   -0.227   37.806  151.355
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  152.133      2.576  59.061  < 2e-16 ***
## xage         -10.012     59.749  -0.168 0.867000
## xsex        -239.819     61.222  -3.917 0.000104 ***
## xbmi         519.840     66.534   7.813 4.30e-14 ***
## xmap         324.390     65.422   4.958 1.02e-06 ***
## xtc         -792.184    416.684  -1.901 0.057947 .
## xldl         476.746    339.035   1.406 0.160389
## xhdl         101.045    212.533   0.475 0.634721
## xtch         177.064    161.476   1.097 0.273456
## xltg         751.279    171.902   4.370 1.56e-05 ***
## xglu          67.625     65.984   1.025 0.305998
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 54.15 on 431 degrees of freedom
## Multiple R-squared:  0.5177, Adjusted R-squared:  0.5066
## F-statistic: 46.27 on 10 and 431 DF,  p-value: < 2.2e-16
```

```r
#ridge regression #2
lambdas <- 10^seq(7, -3)
model_ridge <- glmnet(x, y, alpha = 0, lambda = lambdas)
plot(model_ridge, xvar = "lambda", label = TRUE)
```




```
#3
cv_fit <- cv.glmnet(x=x, y=y, alpha = 0, nlambda = 1000)
plot(cv_fit)
```



```r
cv_fit$lambda.min

## [1] 4.516004

#4
fit <- glmnet(x=x, y=y, alpha = 0, lambda=cv_fit$lambda.min)
fit$beta

## 10 x 1 sparse Matrix of class "dgCMatrix"
##                 s0
## age   -1.962198
## sex -218.721144
## bmi  504.449471
## map  309.708401
## tc  -120.010727
## ldl  -49.525779
## hdl -180.549976
## tch  113.395195
## ltg  472.947829
## glu   80.612740

#5
fit <- glmnet(x=x, y=y, alpha = 0, lambda=cv_fit$lambda.1se)
fit$beta

## 10 x 1 sparse Matrix of class "dgCMatrix"
##                 s0
## age   23.935931
## sex -113.200049
## bmi  355.815095
## map  229.598721
## tc    -6.723554
## ldl  -47.972296
## hdl -167.389042
## tch  121.092669
## ltg  304.387870
## glu  112.675894

intrain <- createDataPartition(y=diabetes$y,p = 0.8,list = FALSE)
training <- diabetes[intrain,]
testing <- diabetes[-intrain,]
cv_ridge <- cv.glmnet(x=training$x, y=training$y,alpha = 0, nlambda = 1000)
ridge_reg <- glmnet(x=training$x, y=training$y,alpha = 0, lambda=cv_ridge$lam
bda.min)
ridge_reg$beta

## 10 x 1 sparse Matrix of class "dgCMatrix"
##                s0
## age   31.33299
## sex -210.53163
## bmi  423.74733
```

```
## map  337.34705
## tc   -99.82516
## ldl  -46.24579
## hdl -241.00979
## tch   67.98749
## ltg  432.93107
## glu  118.84876
```

```
ridge_reg <- glmnet(x=training$x, y=training$y,alpha = 0, lambda=cv_ridge$lam
bda.1se)
ridge_reg$beta
```

```
## 10 x 1 sparse Matrix of class "dgCMatrix"
##              s0
## age   43.80905
## sex -111.99653
## bmi  322.61859
## map  252.51659
## tc   -17.89400
## ldl  -43.95097
## hdl -180.01382
## tch  104.13127
## ltg  300.13712
## glu  127.46322
```

```
#8
ridge_reg <- glmnet(x=training$x, y=training$y,alpha = 0, lambda=cv_ridge$lam
bda.min)
ridge_pred <- predict.glmnet(ridge_reg,s = cv_ridge$lambda.min, newx = testin
g$x)
sd((ridge_pred - testing$y)^2)/sqrt(length(testing$y))
```

```
## [1] 354.8159
```

```
ridge_reg <- glmnet(x=training$x, y=training$y,alpha = 0, lambda=cv_ridge$lam
bda.1se)
ridge_pred <- predict.glmnet(ridge_reg,s = cv_ridge$lambda.1se, newx = testin
g$x)
sd((ridge_pred - testing$y)^2)/sqrt(length(testing$y))
```

```
## [1] 380.1932
```

```
#9
ols_reg <- lm(y ~ x, data = training)
summary(ols_reg)
```

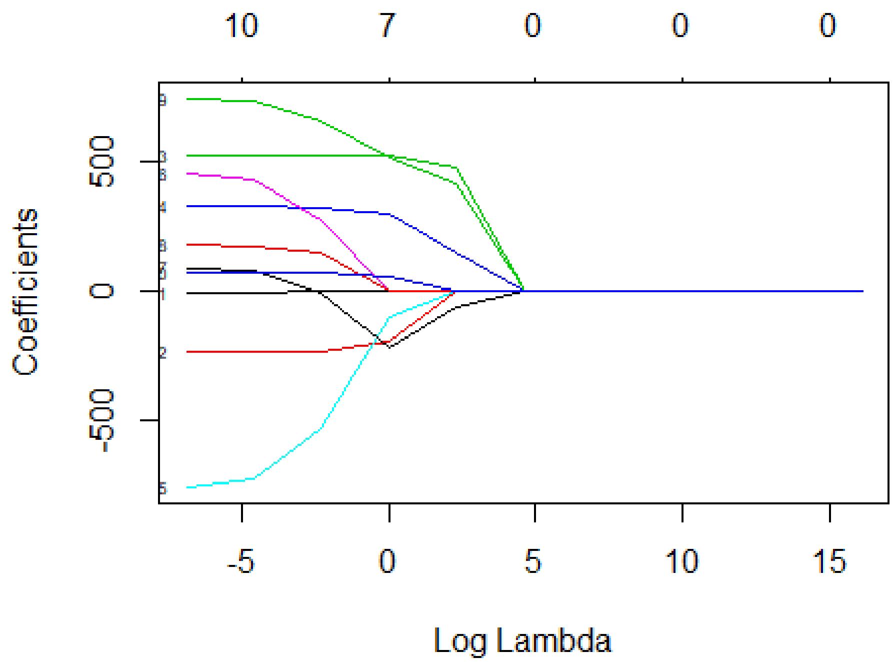
```
##
## Call:
## lm(formula = y ~ x, data = training)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -151.967  -39.604   -2.989   40.180  156.881
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  153.3004     2.9338  52.253  < 2e-16 ***
## xage          23.1119    67.9522   0.340 0.733974
## xsex        -243.0285    68.8648  -3.529 0.000474 ***
## xbmi         432.5835    77.2865   5.597 4.46e-08 ***
## xmap         367.4892    75.8545   4.845 1.92e-06 ***
## xtc         -807.5205   501.4664  -1.610 0.108246
## xldl         534.1182   408.4487   1.308 0.191857
## xhdl          -0.1645   249.3326  -0.001 0.999474
## xtch          71.4788   190.6822   0.375 0.707997
## xltg         738.4486   202.1615   3.653 0.000300 ***
## xglu         107.4243    74.4419   1.443 0.149913
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 55.19 on 344 degrees of freedom
## Multiple R-squared:  0.4963, Adjusted R-squared:  0.4817
## F-statistic:  33.9 on 10 and 344 DF,  p-value: < 2.2e-16
```

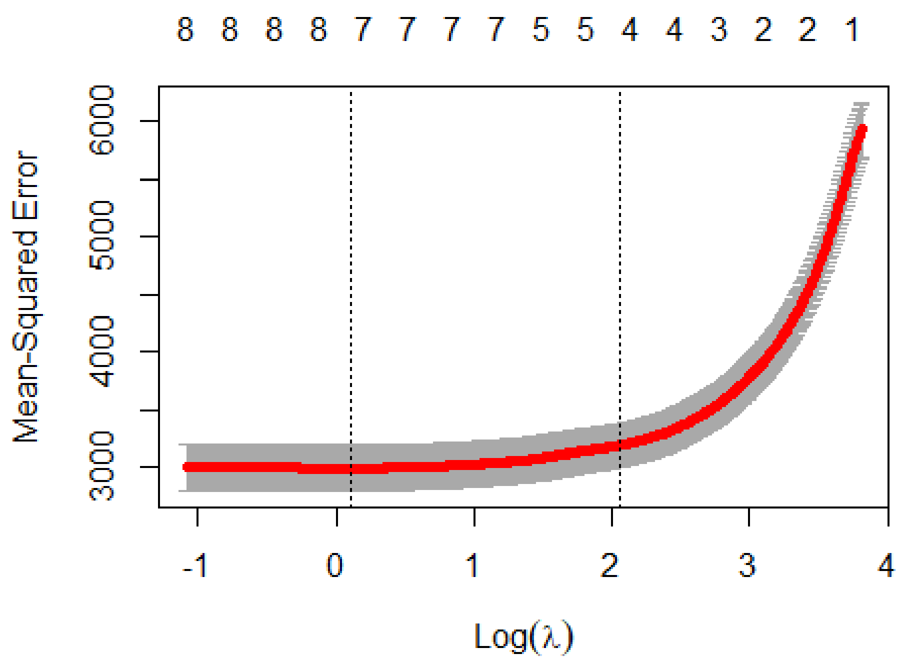
```r
#10
ols_pred <- predict(ols_reg, newdata=testing$x, type = "response")
sd((ols_pred - testing$y)^2)/sqrt(length(testing$y))
```

```
## [1] 358.3017
```

```r
#least squares prediction error is higher.

#lasso regression
#2
llambdas <- 10^seq(7, -3)
model_ridgel <- glmnet(x, y, alpha = 1, lambda = llambdas)
plot(model_ridgel, xvar = "lambda", label = TRUE)
```




```
#3
cv_fitl <- cv.glmnet(x=x, y=y, alpha = 1, nlambda = 1000)
plot(cv_fitl)
```



```
#4
cv_fitl$lambda.min

## [1] 1.109544

fitl <- glmnet(x=x, y=y, alpha = 1, lambda=cv_fitl$lambda.min)
fitl$beta

## 10 x 1 sparse Matrix of class "dgCMatrix"
##               s0
## age    .
## sex -192.03578
## bmi  521.39363
## map  294.15488
## tc   -96.74435
## ldl    .
## hdl -222.17067
## tch    .
## ltg  510.46339
## glu   51.86951

#5
fitl <- glmnet(x=x, y=y, alpha = 1, lambda=cv_fitl$lambda.1se)
fitl$beta

## 10 x 1 sparse Matrix of class "dgCMatrix"
##             s0
## age   .
## sex   .
## bmi 492.5053
## map 170.4232
## tc    .
## ldl   .
## hdl -92.8612
## tch   .
## ltg 427.6289
## glu   .

#6
intrainl <- createDataPartition(y=diabetes$y, p = 0.8,list = FALSE)
trainingl <- diabetes[intrainl,]
testingl <- diabetes[-intrainl,]
#7
cv_lasso <- cv.glmnet(x=trainingl$x, y=trainingl$y,alpha = 1, nlambda = 1000)
lasso_reg <- glmnet(x=trainingl$x, y=trainingl$y,alpha = 1, lambda=cv_lasso$l
ambda.min)
lasso_reg$beta

## 10 x 1 sparse Matrix of class "dgCMatrix"
##               s0
## age  -19.33869
```

```
## sex -204.69511
## bmi  571.29778
## map  314.57239
## tc  -527.94256
## ldl  231.36446
## hdl    .
## tch  163.05944
## ltg  630.97242
## glu   40.36855
```

```
lasso_reg <- glmnet(x=trainingl$x, y=trainingl$y,alpha = 1, lambda=cv_lasso$l
ambda.1se)
lasso_reg$beta
```

```
## 10 x 1 sparse Matrix of class "dgCMatrix"
##            s0
## age   .
## sex   .
## bmi 523.53483
## map 148.78442
## tc    .
## ldl   .
## hdl -79.69628
## tch   .
## ltg 382.45308
## glu   .
```

```
#8
lasso_reg <- glmnet(x=trainingl$x, y=trainingl$y,alpha = 1, lambda=cv_lasso$l
ambda.min)
lasso_pred <- predict.glmnet(lasso_reg,s = cv_lasso$lambda.min, newx = testin
gl$x)
sd((lasso_pred- testingl$y)^2)/sqrt(length(testingl$y))
```

```
## [1] 465.1801
```

```
lasso_reg <- glmnet(x=trainingl$x, y=trainingl$y, alpha = 1, lambda=cv_lasso$
lambda.1se)
lasso_pred <- predict.glmnet(lasso_reg,s = cv_lasso$lambda.1se, newx = testin
gl$x)
sd((lasso_pred - testingl$y)^2)/sqrt(length(testingl$y))
```

```
## [1] 413.3923
```