Introduction to Machine Learning - Lab 2

Gustav Sternelöv

Thursday, October 29, 2015

Assignment 1

1.1

The function *ridgereg_nfoldCV* which performs ridge regression by using n-fold cross validation is implemented at the first step of assignment 1. The R-code used to create the function can be seen below.

```
ridgereg_nfoldCV <- function(x, y, lambda, nfolds){
  # Create the folds and initialize CV vector
  n <- length(y)
  seques <- floor(n/nfolds)
  reps <- n%%nfolds
  groups <- rep(seq(1,nfolds, 1), seques)
  end_values <- rep(nfolds, reps)
  folds <- c(groups, end_values)
  folds <- sort(folds)

  x <- cbind(rep(1, nrow(x)), x)
  CV <- 0
  for (i in 1:nfolds){
    testIndexes <- which(folds==i,arr.ind=TRUE)
    testData_x <- x[testIndexes, ]
    testData_y <- y[testIndexes]
    trainData_x <- x[-testIndexes, ]
    trainData_y <- y[-testIndexes]

    # Perform ridge regression on train data
    x_t <- t(trainData_x)
    I <- diag(ncol(trainData_x))
    BetaRidge <- solve(x_t %*% trainData_x + lambda * I) %*% x_t %*% trainData_y
    # Test regression on test data and compare with true values for test data
    y_hat_test <- testData_x %*% BetaRidge
    # Calculates CV
    CV[i] <- sum((testData_y - y_hat_test)^2)
  }
  CV_score <- (1/nfolds * sum(CV))
  return(CV_score)
}
```

1.2

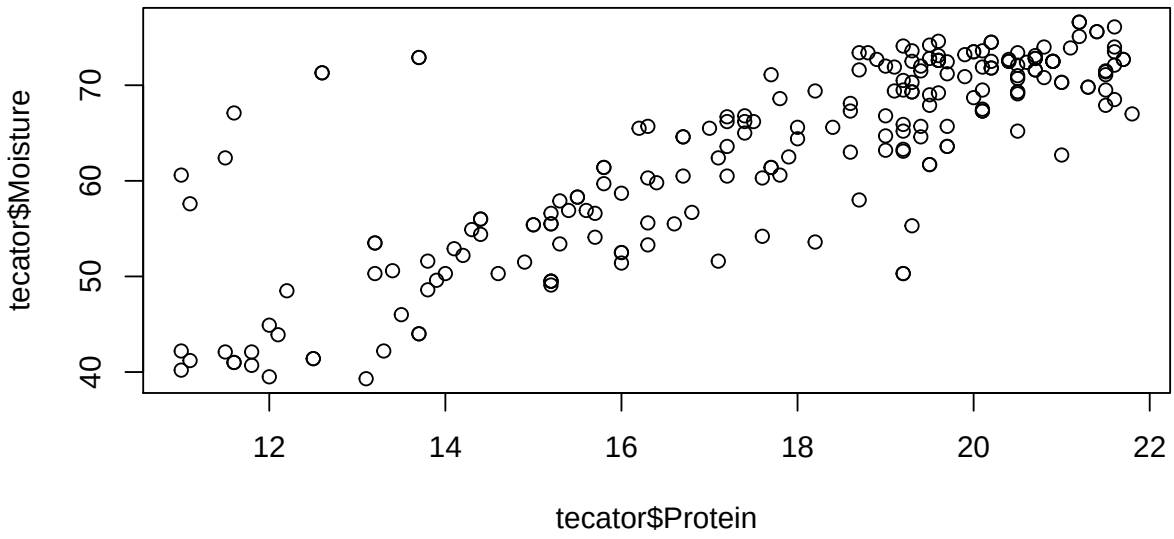
The function implemented in 1.2 is tested with the longley data set. The predictor variables has been centered, but not scaled. Number of folds is set to 10 and the value for lambda goes from 1 to 7, by 1. The given values for the CV score is shown in the following table and it can be seen that the CV score increases for higher values of lambda. The conclusion from this result is that the best model is given with lambda set to 1.

```
## [1] 1220.437
## [1] 2921.322
## [1] 4750.991
## [1] 6569.879
## [1] 8316.805
## [1] 9966.241
## [1] 11510.01
```

Assignment 2

In the second assignment the data material called *tecator* is used. This file contains information from a study where...

2.1



To use a linear model to describe the levels of moisture with the levels of protein may be an appopriate approach by the look of the plot.

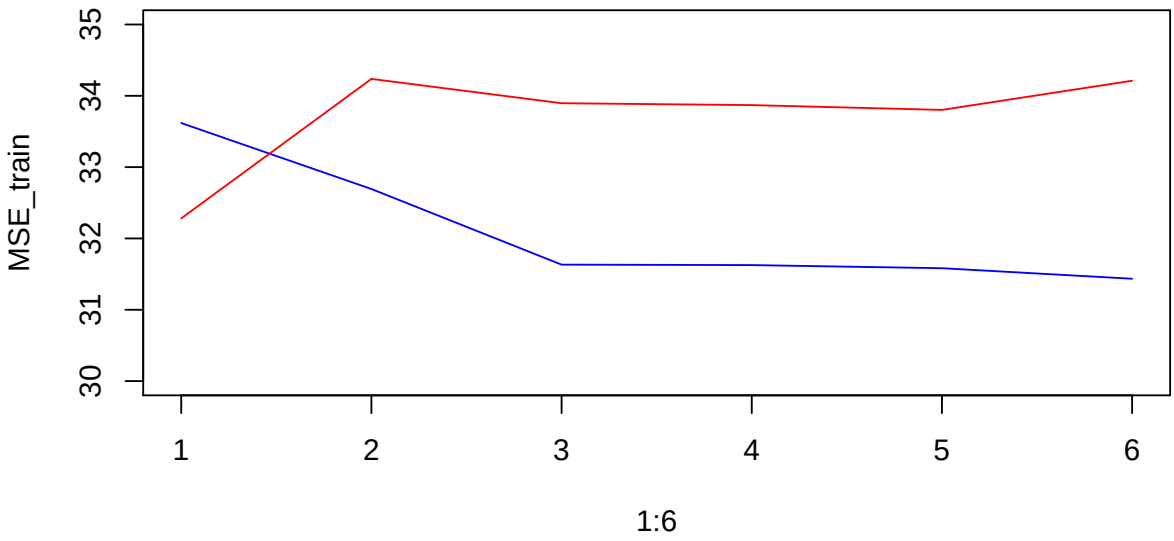
2.2

A Probabilistic model...

2.3

For exercise 2.3 the data set is divided into a training set and a validation set where each set contains 50 percent of the observations.
For models with polynomials of higher power the MSE for test data in general increases, and for training data it decreases. The best model for training data in terms of MSE is the, probably overfitted, model with polynomial terms up to a power of 6. This can be compared to the the test data which has the lowest MSE

for a simple linear model. In terms of the bias/variance trade-off the plot an interpretation of the plot can be that...



2.4

Now the whole data set is used to construct the same models. AIC values is computed to evaluate the models and decide which model that is the best one. The criterion used for comparing the models is that the AIC value should be as low as possible. With that criterion in mind is it concluded that the model with polynomial terms up to a power of 3 is the preferred model.

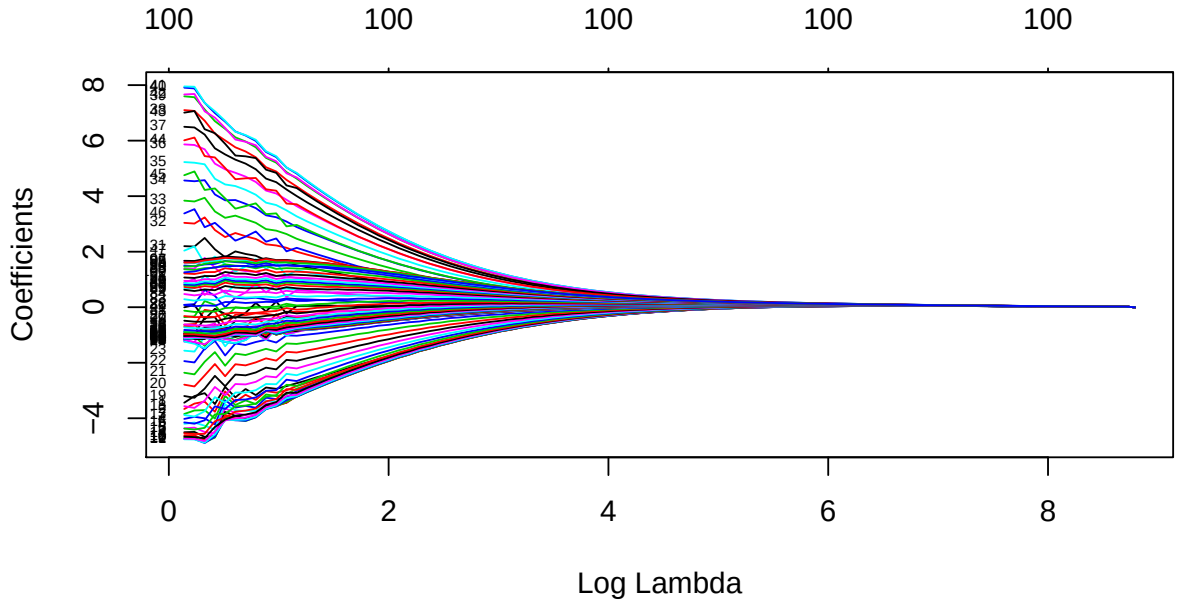
```
## [1] 754.7710 756.4501 753.6657 755.3104 757.1727 759.1538
```

2.5

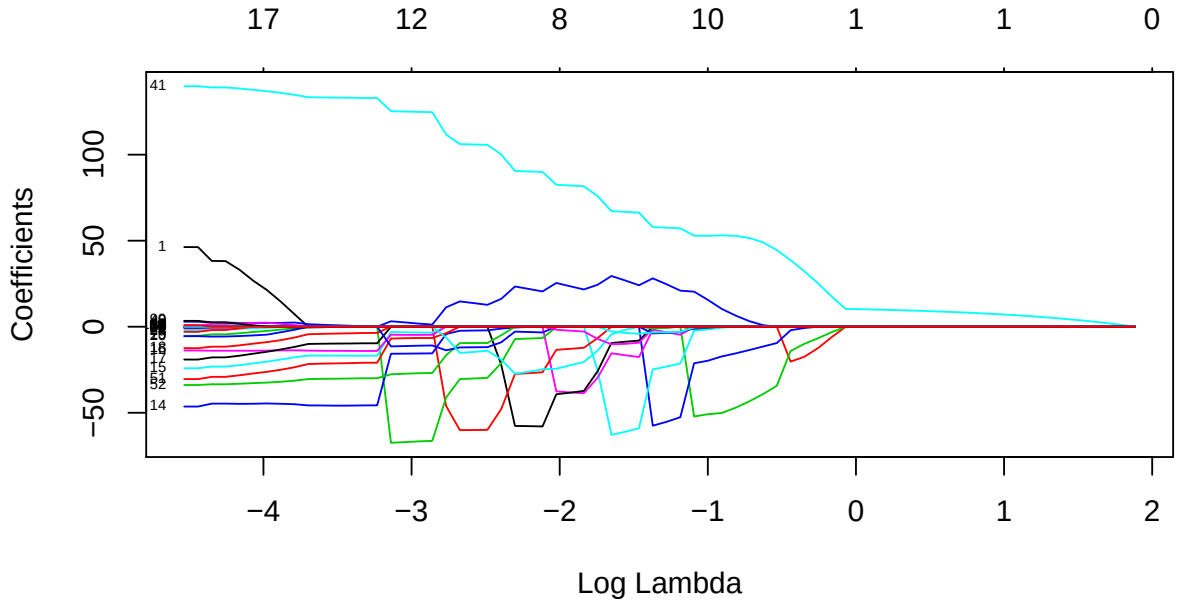
Another area where the AIC criterion can be useful is in the variable selection. Here the function *stepAIC* starts with a linear model that has *fat* as response variable and *channel1-channel100* as predictors. After running the variable selection procedure it is concluded that 63 of the predictors are selected.

2.6 - 2.7

With the same response and predictor variables as in 2.5 a ridge regression model is fitted. The difference between a ridge regression model and a linear model is the use of a penalty factor λ in the former model. The value of λ affects the coefficients in such a way that they shrinks.
How the values of the coefficients in the ridge regression model depends on the log of λ is illustrated by plotting these values.



When fitting a LASSO model instead the dependence between the coefficients and log of λ changes.

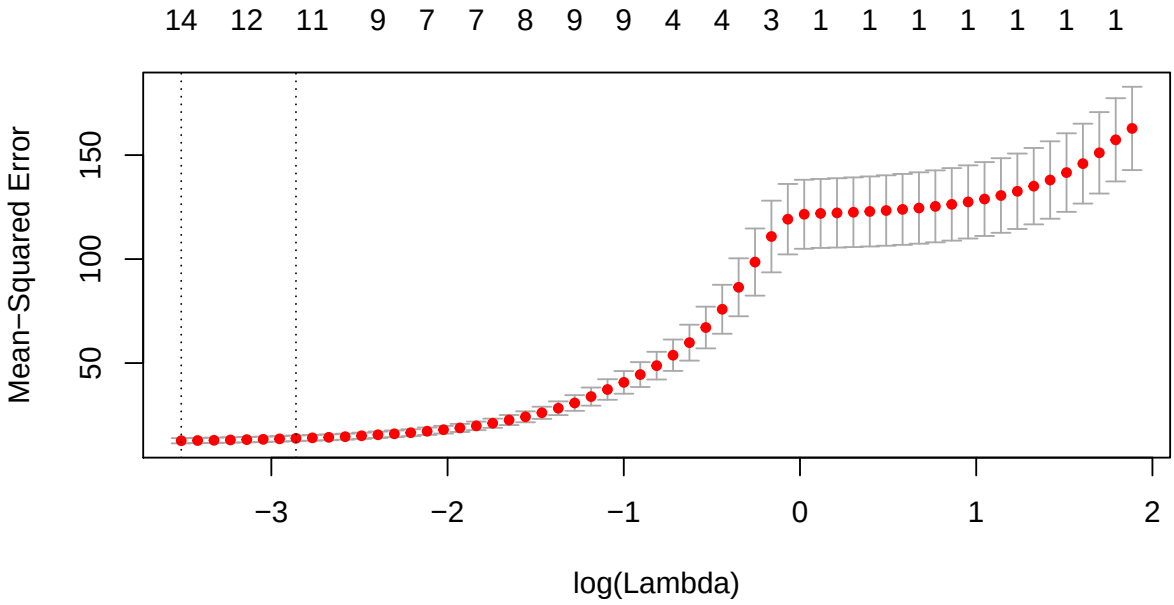


Compare the plots, conclusions The difference between a Ridge regression model and a LASSO model...

2.8

To find a optimal LASSO model a possible approach is to use cross-validation. This results in a model with 0.03 as the optimal value of lambda and 14 variables.

How the CV score depend on the log of λ



2.9

The variable selection done with *stepAIC* and the selection done with the cross-validated LASSO model choose a significantly different amount of variables. With *stepAIC* a model with 63 variables were obtained and from the cross-validated LASSO computations a model with 14 variables.