Exercise 11

Homework exercise

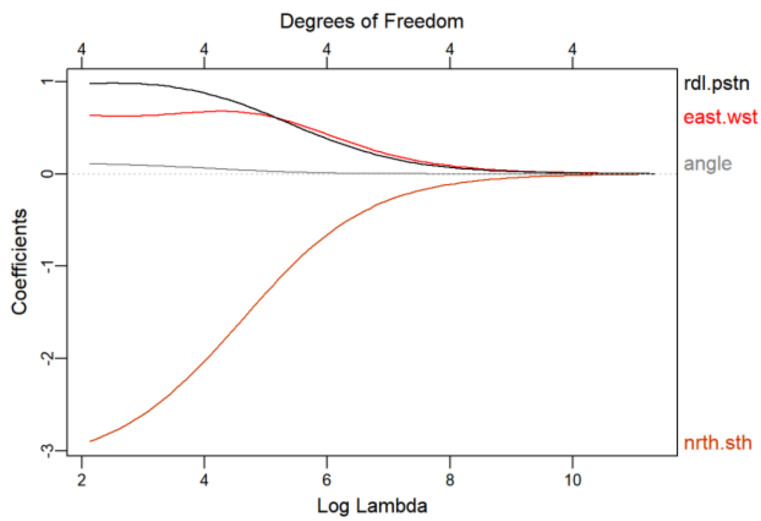
To be solved at home before the exercise session.

---

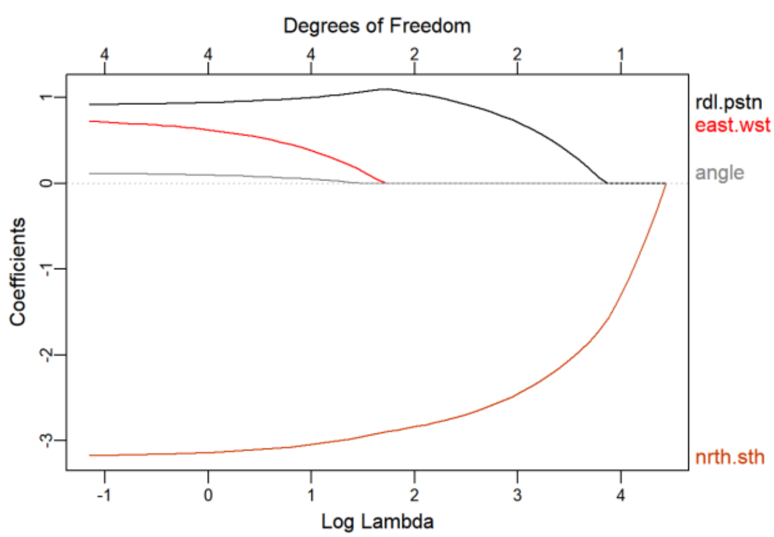
1. a. The data set `galaxy` from the package `ElemStatLearn` contains measurements on the position and radial velocity (the response) of the galaxy NGC7531. The following two plots show the ridge and LASSO coefficient profiles of the four explanatory variables. Compare the two plots and interpret them (for example, deduce which of the explanatory variables are the most ``important''?)

```r
library(ElemStatLearn)
library(glmnet)
library(plotmo)

ridge_galaxy <- glmnet(as.matrix(galaxy[, 1:4]), as.matrix(galaxy[, 5]), alpha = 0)
plot_glmnet(ridge_galaxy, xvar = "lambda", label = TRUE)
```




```r
lasso_galaxy <- glmnet(as.matrix(galaxy[, 1:4]), as.matrix(galaxy[, 5]), alpha = 1)
plot_glmnet(lasso_galaxy, xvar = "lambda", label = TRUE)
```



The order of importance for the four explanatory variables are in the order as follows:
nrth.sth > rdl.pstn > east.wst > angle

Justifications:

- For the Ridge Regression, we know that the coefficients are never reduced to 0, so the degree of freedom is all 4 along the row as all four variables are included in the model. For the first range of log lambda 2 to 4, we can see that nrth.sth value is furthest from 0, then rdl.pstn and east.wst => nrth.sth is more important than rdl.pstn, which in turn is more important than east.wst. Angle at the very start is already close to 0, suggesting that it is the least important variable. When log-lambda = 6, nrth.sth, rdl.pstn and east.wst are roughly equal as the effects of log-lambda starts to be dominant in the regression model.
- For the Lasso Regression, we know that the coefficients can be shrunk to exactly 0, so the degree of freedom is all 4 along log lambda -1 to 1, but gradually reduce to 2 and 1 as some of the variables get shrunk to 0 and vanished out of the model. We can know which variable is more important by checking if its coefficients shrunk later as log lambda increases. As we can see, angle shrunk earliest to 0, then east.wst, rdl.ptsn and finally nrth.sth, suggesting that the order of importance nrth.sth > rdl.pstn > east.wst > angle is indeed correct.

b. Consider a regression model where the response `Y` is explained using the covariates `X1`, `X2` and `X3`. The p-values corresponding to the models of all possible combinations of the covariates are listed below. Use them to perform variable selection with both backward and forward selection with the p-value cutoff $\alpha_0 = 0.05$.

```
##     X1
## 0.0071
```

```
##     X2
## 0.4221
```

```
##     X3
## 0.0014
```

```
##     X1     X2
## 0.0055 0.2809
```

```
##     X1     X3
## 0.0021 0.0004
```

```
##     X2     X3
## 0.1267 0.0006
```

```
##     X1     X2     X3
## 0.0010 0.0516 0.0001
```

The backward selection works by selecting a p-value cutoff α_0 (e.g. 0.05) and proceeding as follows:

1. Estimate the model using all predictors.
2. Remove the predictor with the highest p-value greater than or equal to α_0 and estimate the new model.
3. Repeat step 2 until all predictors have p-values less than α_0.

That is, backward selection begins with a full model and one-by-one removes the variables that are the least "important", until we are left with the subset of "most important" variables.

Backward selection steps from the model above is as follows:
Estimate the model using all predictors

```
## X1       X2       X3
## 0.0010   0.0516   0.0001
```

X2 p-value is greater than alpha = 0.05 => X2 is omitted

=> New model: X1 and X3

```
## X1       X3
## 0.0021   0.0004
```

No variables whose p-value is smaller than alpha = 0.05. Backward selection stops. X1 and X3 will be chosen as the explanatory variables

The forward selection works by selecting a p-value cutoff α_0 (e.g. 0.05) and proceeding as follows:

1. Start with a model with no predictors at all.
2. For each predictor one at a time, check what their p-value would be if they were added to the model and add the one with the smallest p-value below α_0 to the model.
3. Repeat step 2 until no new predictors with p-values less than α_0 can be added.

That is, forward selection begins with an empty model and one-by-one adds the variables that are the most "important", until no more "important" variables are left to be added.

Forward selection steps from the model above is as follows:

```
## X1
## 0.0071

## X2
## 0.4221

## X3
## 0.0014
```

For each predictor one at a time, check what their p-value would be if they were added to the model and add the one with the smallest p-value below α0 to the model. Since X3 has the smallest p-value under alpha = 0.05, it will first be added to the model.
The current model: X3

Now we add X1 to X3, and X2 to X3 and estimate the two models:

```
## X1       X3
## 0.0021   0.0004

## X2       X3
## 0.1267   0.0006
```

Now since X1 is smaller than alpha = 0.05 and X2 is larger than alpha = 0.05 => X1 is added to the model
The current model: X1, X3

Finally we add X2 to the model

```
## X1      X2      X3
## 0.0010  0.0516  0.0001
```

Since X2 is larger than alpha = 0.05, it will not be added to the model. Forward selection stops.
X1 and X3 will be chosen as the explanatory variables