Compulsory exercise 2: Group XX

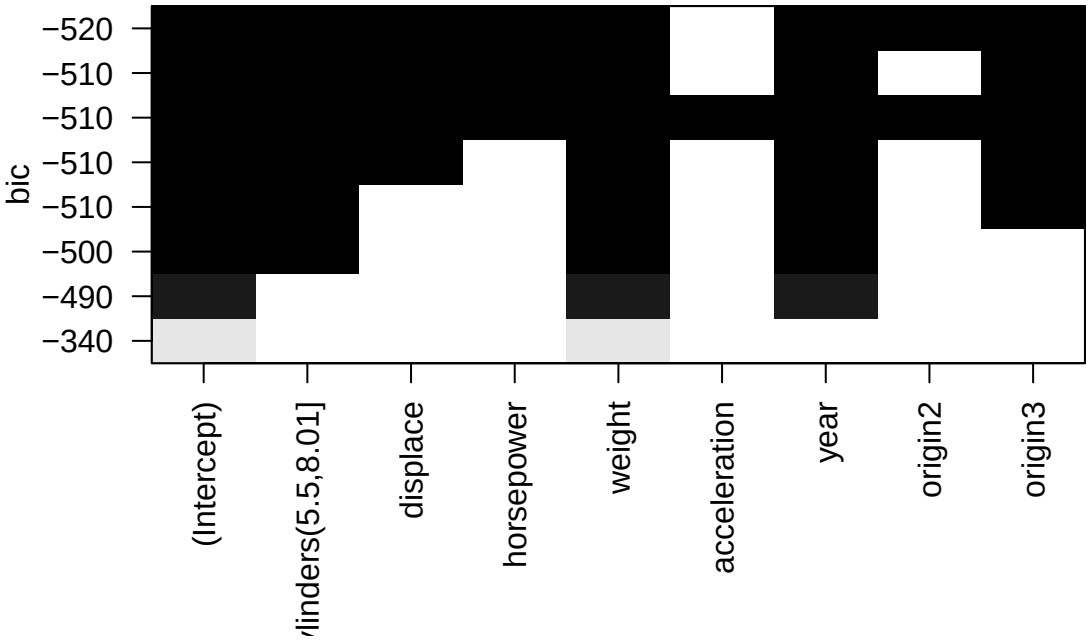
TMA4268 Statistical Learning V2018

NN1, NN2 and NN3

06 mars, 2018

1a)

- Q1: How many possible models?
- Q2: How to choose best model



1b)

- Q3: How best model chosen for each model complexity + best for 2 covariates.
- Q4: Choose model and evaluate.

```
# R code to fit your final model.
```

- Q5: Use this model fit to predict new values for `ourAutoTest` and report the MSE.

1c)

- Q6. Explain how k-fold cross-validation is performed.
- Q7. Why may k-fold cross-validation be preferred to leave-one-out cross-validation?

1d@9

- Q8. R-code for 10-fold CV.

```
library(caret)
library(leaps)
# and so on
```

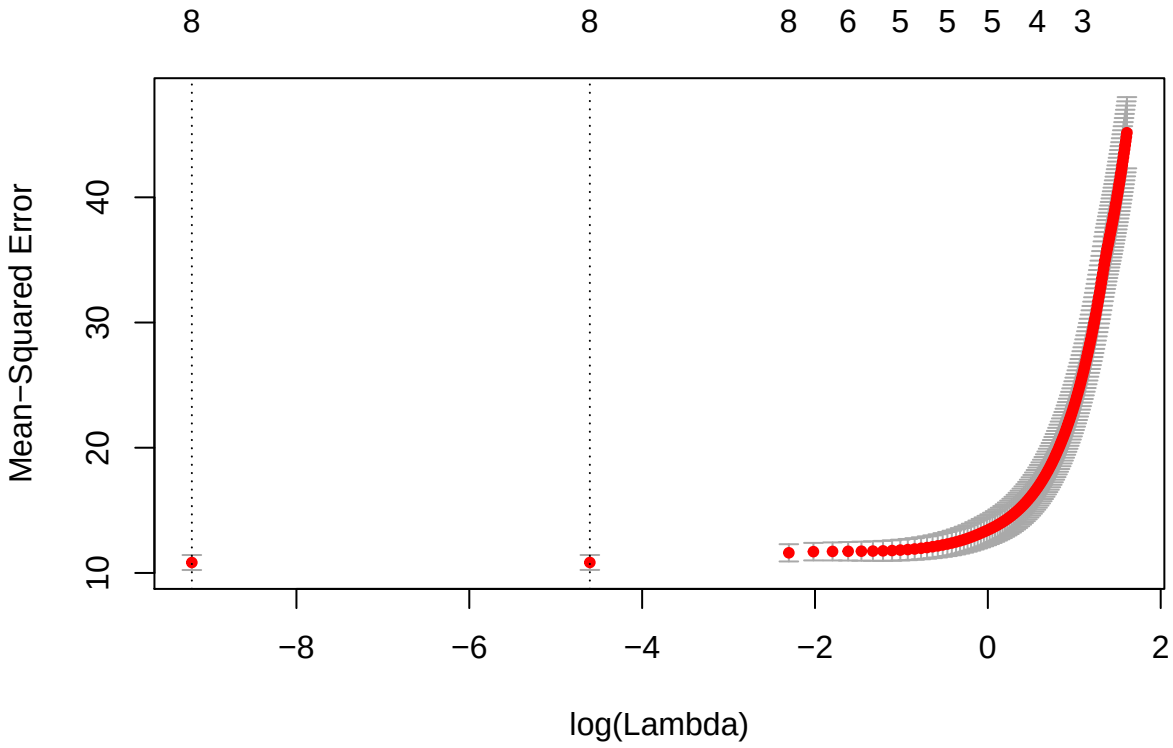
- Q9. What is the optimal model complexity (number of parameters) in your regression?

```
# MSE on test set
```

- Q10. Evaluate best model (or refer to Q4 and Q5).

2a) Explain figures

- Q11: Which figure (1 or 2) corresponds to ridge and which figure corresponds to lasso?
- Q12. Use the two figures and the above formulas to explain...
- Q13. Can you use lasso and/or ridge regression to perform model selection similar to what you did in Problem 1?



2b) Finding the optimal λ

- Q14: Explain what the function `cv.glmnet` does.
- Q15. Explain what we see in the above plot.
- Q16: Finding the optimal lambda:

```
# need some R code here
```

3c) Prediction

- Q17: Fit model, coefficients,...

```
# fit the lasso
```

```
# 0 for cylinder, displace, horsepower, weight, acceleration, year, 0 for origin2 and 0 for origin3
newx=matrix(c(0,150,100,3000,10,82,0,0),nrow=1)
# then do the prediction
```

- Q18: Predicted value:

3a)

- Q19: Fitting the specified `gam`

```
library(gam)
# write R code
```

- Q20: The cubic spline basis.