

DATS 7860 Statistical and Machine Learning for Big Data

Homework 4 (Chapter 6)

1. In class we mentioned that the lasso is implemented with the *glmnet* package. Please read the vignette of the *glmnet* package (uploaded to D2L can be found on the webpage of the package <https://cran.r-project.org/web/packages/glmnet/index.html>) and answer this question: what other types of regression models can this package be applied for (in addition to the regular linear regression model) and briefly explain the types of data for which each model is modelling? You don't need to figure out all details of this package, and my purpose is to let you know which models and what types of data this package can be applied for.

Optional task: We have mentioned several extensions to the lasso / ridge regression in class, including the elastic net, the group lasso, the fussed lasso and the bridge regression. Please read the Wikipedia page on the lasso ([https://en.wikipedia.org/wiki/Lasso_\(statistics\)](https://en.wikipedia.org/wiki/Lasso_(statistics))). Also, for those of you who are interested in the underlining methodologies and have time, I suggest reading the original papers on those aforementioned extended approaches (uploaded in the "Optional Readings" folder in Week 7 on D2L).

Problems below are all from Exercises 6.6 in Chapter 6 in the textbook (ISL):

2. Problem 8 (Page 285).

- Please submit your complete R codes for this problem.

3. Problem 9 (Page 286).

Some specific notes for this problem:

- Please submit your complete R codes for this problem.
- The College dataset is available from the R package ISLR2. You can use the following R codes to load the dataset:

```
require(ISLR2)
```

```
data(College)
```

- In this problem, please use the variable `Apps` (i.e. number of applications received) as the response variable.

4. Problem 10 (Page 287).

Some specific notes for this problem:

- Depending on how you simulate the data, the answers will not be the same. I attached an example FYI. Please do not just copy and paste it, but write your own code and simulations.

- Please submit your complete R codes for this problem.

5. Problem 11 (Page 287).

Some specific notes for this problem:

- Please submit your complete R codes for this problem.
- Please at least try the following methods: the best subset selection, the lasso, ridge regression and PCR. You are also encouraged but not required to try other approaches, such as standard linear regression, PLS, etc.
- The Boston dataset is available from the R package MASS. You can use the following R codes to load the dataset and those methods you will try:

```
require(leaps) # for best subset selection
```

```
require(glmnet) # for ridge and lasso
```

```
library(pls) # for PCR
```

```
require(MASS) # Boston data set
```

```
data(Boston)
```