**WEEK 5 HOMEWORK­­**

**INSTRUCTIONS**

* Every learner should submit his/her own homework solutions. However, you are allowed to discuss the homework with each other (in fact, I encourage you to form groups and/or use the forums) – but everyone must submit his/her own solution; you may not copy someone else’s solution.
* The homework will be peer-graded. In analytics modeling, there are often lots of different approaches that work well, and I want you to see not just your own, but also others.
* The homework grading scale reflects the fact that the primary purpose of homework is learning:

|  |  |  |
| --- | --- | --- |
| **Rating** | **Meaning** | **Point value (out of 100)** |
| 4 | All correct (perhaps except a few details) with a deeper solution than expected | 100 |
| 3 | Most or all correct | 90 |
| 2 | Not correct, but a reasonable attempt | 75 |
| 1 | Not correct, insufficient effort | 50 |
| 0 | Not submitted | 0 |

**Question 11.1**

Using the crime data set uscrime.txt from Questions 8.2, 9.1, and 10.1, build a regression model using:

# Stepwise regression

1. Lasso
2. Elastic net

For Parts 2 and 3, remember to scale the data first – otherwise, the regression coefficients will be on different scales and the constraint won’t have the desired effect.

For Parts 2 and 3, use the glmnet function in R.

Notes on R:

* For the elastic net model, what we called λ in the videos, glmnet calls “alpha”; you can get a range of results by varying alpha from 1 (lasso) to 0 (ridge regression) [and, of course, other values of alpha in between].
* In a function call like glmnet(x,y,family=”mgaussian”,alpha=1) the predictors x need to be in R’s matrix format, rather than data frame format. You can convert a data frame to a matrix using as.matrix – for example, x <- as.matrix(data[,1:n-1])
* Rather than specifying a value of T, glmnet returns models for a variety of values of T.

**Question 12.1**

Describe a situation or problem from your job, everyday life, current events, etc., for which a design of experiments approach would be appropriate.

**Question 12.2**

To determine the value of 10 different yes/no features to the market value of a house (large yard, solar roof, etc.), a real estate agent plans to survey 50 potential buyers, showing a fictitious house with different combinations of features. To reduce the survey size, the agent wants to show just 16 fictitious houses. Use R’s FrF2 function (in the FrF2 package) to find a fractional factorial design for this experiment: what set of features should each of the 16 fictitious houses have? Note: the output of FrF2 is “1” (include) or “-1” (don’t include) for each feature.

**Question 13.1**

For each of the following distributions, give an example of data that you would expect to follow this distribution (besides the examples already discussed in class).

# Binomial

# Geometric

# Poisson

# Exponential

# Weibull