

Final Exam

– Due Date : By 6:00pm on December, 16th –

The objective of this exam is to take the methods that you study in this course and to apply them to data from a scientific research project. The goal is to build multiple models for prediction to compare them to each other and to propose the optimal model to analyze the data.

The dataset “exam.Rdata” consists of the training set (`x.train`, `y.train`) and the test set `x.test`. Develop an R function to compute the predicted value of y given by the test set data, and then upload your R script file (“exam.r”) to the course website. Your script file needs to contain

1. A list of `library(...)` to call necessary packages which are covered in class.
2. A main function which has a form of

```
main <- function(x.train, y.train, x.test) {  
    ....  
    ....  
    return(y)  
}
```

3. Some sub-functions if necessary.

Note that

- You should not include a `set.seed()` function.
- Make sure that the processing time of the `main` function does not take too long.
- Do not include any R packages and functions which have not been studied in class.
- You should not use the test set to build your prediction model.
- You don't need to include all R commands used for selecting your final prediction model after comparing multiple statistical learning methods.

Evaluation (a total score: 40 points)

1. Creativity and validity (10 points) : Good (10), Ordinary (5), Bad (0)
2. Accuracy (30 points) : A winner of students with the smallest test error (30)