MTH210: Assignment 4

Kaggle-like competion

In your repositories, there is a csv file with a dataset `assign4_train.csv`.
This dataset has the following attributes:

- The first column `y`, is the response

- The rest of the columns are your covariates/predictors including a
  column for the intercept. There are 100 columns.

- There are `n = 500` observations

Your goal is submit a function, `predict.y()` that returns the predicted
value of y for any numeric input $\mathbf{x} \in \mathbb{R}^{100}$, where $\mathbf{x} = (1, x_1, x_2, ..., x_{100})$.
The function should look like:

```
# x = vector of length 100
predict.y <- function(x)
{
  load("fit_params.Rdata")
  ....
  f.x <- ....

  return(f.x)
}
```

Submission instructions

- All your cleaned and error-free code for this problem must be uploaded in the `assign4_script.R` file. This includes are cross-validation.

- In this `assign4_script.R` file you can save the final fitted model parameters in an object called `fit_params.Rdata` using the `save()` function. This `save()` line should be the last line in your file.

- If an `assign4_script.R` file is missing/empty, you get 0/10.

- The `predict.y` should be in the `assign4_prediction.R` file. Only this function should be in the file.

- The `fit_params.Rdata` should then be loaded in the `predict.y` function.

- DO NOT fit the data in the function `predict.y`. This should be a very fast and clean function. If you fit the data in this function, you will get 0/10.

Evaluation

- I will apply your `predict.y()` function on a test dataset that I have kept.

- I will use the **squared error loss** to calculate overall test error for your submitted model.

- Your final marks will be based on how your test error compares with my test error:

  - If you beat or match my test error, you get 10/10 (highly unlikely).

  - The farther you are from my test error, the lower your marks.

Deadline(s)

Below are the submission deadlines and instructions:

- There are two submission deadline.

  - Round 1 deadline: 1:00 pm on April 16th

  - Round 2 deadline: 5:00 pm on April 17th

- I will grade Round 1 and upload your individual test errors, and your marks, and a summary of the class test error.

- You then have a day to improve your model and resubmit for Round 2.

- To submit for any round, just push on your GitHub repositories. Whatever is the submission at 1:00pm and 5:00pm on both April 16th and 17th, respectively, will be the submission that I will download.

- In Round 1 if your `predict.y()` or `assign4_script.R` is missing or empty, you immediately get 0 in the assignment, and you are ineligible for Round 2.