Mid-term assignment

Mario Rosario GUARRACINO

•

May 3
100 points
Due Tomorrow, 11:59 PM

- **The challenge** consists in dealing with a regression problem using a parametric and a nonparametric approach.
  In particular, the former will consist in estimating a linear regression model for the train data contained in the *train_ch.csv*. The latter will consist in using KNN (properly tuned) to predict the response values for the test observations.
- **Your task** is to use the training data to build your models and to test them on the *test_ch.csv* observations.
- You will receive a training set (*train_ch.csv*) of 1000 observations and a test set (*test_ch.csv*) of 100, with nine independent variables.
- **Your submission** will consist of:
  1. A RData file whose name is formatted as: StudentRegistrationNumber_FamilyName_challenge1.Rdata
  2. The file will contain the output of lm() in *fit*, and two variables *knn_pred* and *lm_pred*, one for each method, containing the 100 predicted values for the test data.
  3. A presentation in PDF with up to 6 pages, in which you describe how you obtained the model.
  4. The R macro named *solution.R* used to obtain the results.

  Results will be raked according to RMSE test.