STA 314, Homework

Due dates: different for different parts

This homework will require you to download the files 'trainingdata.csv' and 'testpredictors.csv' from quercus and load both files into R. All questions are related to the corresponding data sets. **Submission: questions 1-4 should be submitted via quercus, a special place for that will be provided. Questions 5-7 should be submitted via kaggle.**

1. **Due November 22** (1 mark) Write your team name for the competition on quercus after you register an account on kaggle and decide your team.

2. **Due November 22** (2 marks) Run a linear model for y as response using all predictors and compute a prediction for the *predictor vector with id 37* from the test set.

3. **Due November 22** (2 marks) Which predictors does a model with two predictors selected by forward stepwise selection contain?

4. **Due November 22** (2 marks) Run a gam model of the form

$$f(X) = b_0 + b_1 X15 + g_1(X10) + g_2(X102)$$

   where g_1 is a natural cubic spline with knots $-1, 0, 1$ and g_2 is a polynomial of degree 4. What is your estimated value for the coefficient b_1?

5. **Due November 25** (3 marks) Submit a prediction that has a score < 4.5 on the public leaderboard.

6. **Due November 25** (3 marks) Submit a prediction that has a score < 4.19 on the public leaderboard.

7. **Due November 25** (2 marks) Submit a prediction that has a score < 4.0 on the public leaderboard.