# Statistical and Machine Learning (Spring 2018)

## Mini Project 1

---

**Instructions:**

- Due date: Jan 25, 2018.
- Total points = 20.
- Submit a typed report.
- Submit only one report per group, and include a description of the contribution of each member.
- It is OK to discuss the project with other students in the class (even those who are not in your group), but each group must write its own code and answers. If the submitted report (including code and answer) is similar (either partially or fully) to someone else's, this will be considered evidence of academic dishonesty, and you will referred to appropriate university authorities.
- Do a good job.
- You must use the following template for your report:

  Mini Project #
  Name
  Names of group members
  Contribution of each group member
  Section 1. Answers to the specific questions asked
  Section 2: R code. Your code must be annotated. No points may be given if a brief look at the code does not tell us what it is doing.

---

1. Consider the training and test data posted on eLearning in the files `1-tranining-data.csv` and `1-test-data.csv`, respectively, for a classification problem with two classes.

   (a) Fit KNN with $K = 1, 2, \ldots, 30, 35, \ldots, 100$.

   (b) Plot training and test error rates against $K$. Explain what you observe. Is it consistent with what you expect from the class?

   (c) What is the optimal value of $K$? What are the training and test error rates associated with the optimal $K$?

   (d) Make a plot of the training data that also shows the decision boundary for the optimal $K$. Comment on what you observe. Does the decision boundary seem sensible?