# Homework 1: Model Building and Model Selection/Fitting

## Overview

**Due Sunday by 5 pm**

For each of the following prompts, produce responses *with* code in-line. Only single PDF submissions will be graded.

## Fork the `problem-set-1` repository

## Building models

**Deviant aggressive behavior**

Consider four well-known and widely believed theories of socially deviant aggressive behavior (e.g. criminal behavior, revolutionary behavior, rude behavior, etc.):

- **Theory I**: Deviant aggressive behavior is learned from experience. Individuals in a society learn to do those things for which they receive rewards and to avoid those things for which they receive punishment.
- **Theory II**: Deviant aggressive behavior is a symbolic expression of hostility toward personal authority figures. When an individual is frustrated in his personal life, he becomes angry toward parents, bosses, or public officials. He will express this anger by deviant aggressive behavior.
- **Theory III**: Deviant aggressive behavior is the rational action of oppressed individuals. Social rules systematically discriminate among people. People who are most hurt by the rules are least likely to profit from conforming to them and thus do conform less.
- **Theory IV**: Deviant aggressive behavior is a social role. Individuals are socialized into the role through contact with a deviant subculture.

Answer the following question in 500-800 words:

1. What social policy might be appropriate to reduce deviant aggressive behavior if Theory I were correct? Theory II? Theory III? Theory IV?

**Waiting until the last minute**

> People often do things at the last minute (students turning in papers, professors grading exams, and so on).

a. Ask yourself **why** the observation might be true and write down your explanations.
b. Generalize the explanatory model – that is, induce the most general, abstract model you can produce that still has the original observation as a consequence.
c. Induce an alternative model that also has the original observation as a consequence.
d. For each of the two general models produced in (b) and (c), derive two interesting predictions (four predictions in total). Be sure the logical connection between your model and your predictions is explicitly stated and that any assumed facts concerning the world are made explicit.

## Selecting and fitting a model

1. For each part, indicate whether we would generally expect the performance of a flexible statistical learning method to be better or worse than an inflexible method. Justify your answer.
   a. The sample size $n$ is extremely large, and the number of predictors $p$ is small.
   b. The number of predictors $p$ is extremely large, and the number of observations $n$ is small.
   c. The relationship between the predictors and response is highly non-linear.

d. The variance of the error terms $\sigma^2 = \text{Var}(\epsilon)$ is extremely high.

2. Bias-variance: Think about the graph including bias, variance, training error, test error, and irreducible error curves, moving from less flexible statistical learning methods towards more flexible approaches.
   a. Explain why each of the five curves has the shape it has.