In class exercise week 11
Topic: prediction models

Problem 1 (Regression to the mean)

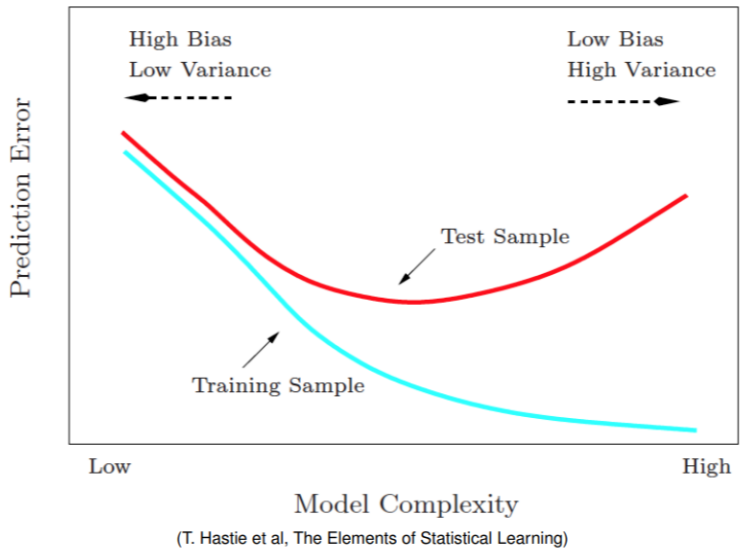
Read the first two parts of the article "Three things that every medical writer should know about statistics" by Stephan Senn. Then, read the description of the following study and answer the questions:

A randomized controlled trial (RCT) was conducted in order to examine if praise or criticism is helpful for medical students when learning to take blood samples from patients. Therefore, students were randomly assigned to one of two groups. In the first group, students who performed bad in taking blood samples were criticized. In group 2, teachers praised students performing very well. Next time, when the students were taking blood samples from the patients, the following observations were made: On average, the criticized students from group 1 improved while the praised students from group two worsened.

a) Based on this study, can you conclude that criticizing is beneficial to praising students when learning new task like taking blood?
b) How could you design the study when you want to conclude that praise respectively criticism is helpful?

Problem 2 (Bias Variance Tradeoff)

Consider the following figure:



(T. Hastie et al, The Elements of Statistical Learning)

a) What indicates that models with high complexity lead to overfitting?
b) Based on which criterion can you find the optimal model complexity?
c) Why are low complexity models correlated with high bias?