Problem Set 1

Hans Trautlein

January 25, 2016

Flexible - more likely to overfit (follow error/noise more than a good model would) but you also might increase the explanatory power of f in the process.

On 1, ask for clarification about 1 and 2, what would be “flexible” vs. “inflexble.” on 3, make sure non-linear relationships can still be fit in a flexible way

Book Problems

1. Indicate whether we would generally expect the performance of a flexible statistical learning method to be better or worse than an inflexible method. Justify my answer.[1]
   (a) A flexible method (to the degree that a model with a small amount of available predictors can be flexible) might be most appropriate, assuming that the
   (b) You would want an inflexible method in this circumstance. If you have a small amount of variables you are more likely to have a model that is difficult to fit exactly, and the large amount of predictors might allow you to accidentally overfit the model.[2]
   (c) This depends on whether or not there is some relationship at all betweent the predictors and the observations if you can map them to the function. Perhaps there is a non-linear relationship that explains the variables well, in which case a flexible relationship might be best.
   (d) If the variance of the error terms $\sigma^2 = Var(\epsilon$ is extremely high I would expect a less flexible model to be worse than an inflexible method because you might be more likely to overfit your model. In Nate Silver parlance your model would end up confusing the “signal” for the “noise.”
2. (a) $n = 500$, $p = 4$. This a regression problem that is most interested in inference.
   (b) $n = 20$, $p = 14$. This is a classification problem that is most interested in prediction.
   (c) $n = 52$ (because of weekly data), $p = 4$. This is a regression problem that is most interested in prediction.
4. (a) Real-life applications of *classification*. Describe the response, as well as the predictors. Is the goal of each application inference or prediction? Explain.
      i.
      ii.
      iii.
   (b) Real-life applications of *regression*. Describe the response, as well as the predictors. Is the goal of each application inference or prediction? Explain.
      i.
      ii.
      iii.
   (c) Real-life applications of *cluster analysis*.

[1] Every answer in this Problem Set has referred frequently to Chapter 2 in James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning*. New York: Springer, 2013. I have also gotten help from

[2] Wonderful examples of what this can lead to are available on the Tyler Virgin's Spurious Correlations website. Did you know that the US's per capita consumption of mozzarella cheese highly correlates with people who die falling down the stairs? This is a good example of correlation not equaling causation.

i.
ii.
iii.

5. A very flexible model for regression and classification has both advantages and disadvantages

   A very inflexible model for regression and classification has both advantages and disadvantages as well, as it often is easier to make inferences from more inflexible methods compared to flexible methods. For example, if inference is the overall goal then you might prefer a less flexible model, like "least squares" or "subset selection lasso" as opposed to "support vector machines." The restriction inherent in more inflexible models makes inference easier.

   A more flexible approach might be preferred when prediction is weighed much heavier than inference in the model's goals. This is a slight oversimplification, but then the "why" takes a backseat to the "what," that is, the output that the model creator cares about is less a great understanding

   A less flexible model might be preferred when

6. The differences between a parametric and a non-parametric statistical learning approach are

   The advantages of a parametric approach as opposed to a non-parametric approach are

   The disadvantages of a parametric approach are

7. (a)
   (b) My prediction when $K = 1$ is
   (c) My prediction when $K = 3$ is
   (d)

Additional exercise

Using the notation standards described at the end of chapter one, please provide notation for the following objects:

Input: The 10 photos that we looked at on the first day, as if they were scanned at 64 x 64 pixel resolution.

Transformed Input: The 10 photos, after a small number of features have been identified.

Output: The associated actual ages of those 10 photos.

Model: Provide a guess at what f might look like (there is no single right answer here).

need to understand irreducible error better