STATISTICS 452/652: Statistical Learning and Prediction

October 4, 2020

Lecture 12a: Projection Pursuit

(Reading: not in book)

1 Goals of lecture

- We learned about Partial Least Squares
  - Create new components that are linear combinations of the original explanatory variables
  $$Z = \sum_{j=1}^{p} \phi_j X_j$$
  - Choose ϕ weights to try to create a Z that might be strongly related to Y
  - Repeat M times and fit linear regression model to these M components
- Result of PLS is still a linear model in the original X_j, $j = 1, \ldots, p$
  - Let $Z_1 = \phi_{11}X_1 + \phi_{12}X_2$ and $Z_2 = \phi_{21}X_1 + \phi_{22}X_2$
  - Fit model $f(X) = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2$
  - Then $f(X) = \beta_0 + \beta_1(\phi_{11}X_1 + \phi_{12}X_2) + \beta_2(\phi_{21}X_1 + \phi_{22}X_2) = \beta_0 + [\beta_1\phi_{11} + \beta_2\phi_{21}]X_1 + [\beta_1\phi_{12} + \beta_2\phi_{22}]X_2$
- So PLS is just fancy linear regression
- Meanwhile, GAM allows fitting of nonlinear models
  - But practically restricted to “additive” (non-interactive) models
- We would like something smooth and curvy that can adapt to the surface suggested by Y automatically
- We will quickly learn a technique that can do this, but is not used much...
- ...because there is something similar that does the same thing, only better!

2 Projection Pursuit Regression (PPR)

Note: This method is not used very much, but represents (1) an interesting idea, and (2) a very direct link to *much* more complicated but useful methods that we will cover next.

1. PLS:
   (a) Make components that are linear combinations of X_j's in directions suggested by Y
   (b) Do a linear regression on these components
2. GAM:
   (a) Create a function in each X_j direction.
   (b) Make a linear combination of the functions
3. PROJECTION PURSUIT REGRESSION (PPR) is sort of like combining PLS with GAM:
   (a) Make a linear combination of the X_j's

   $$Z_1^{PPR} = \sum_{j=1}^{p} \omega_{1j} X_j$$

   i. This represents a "direction" in X space,
   ii. In 3 dimensions, think of it as "spinning the cube" to look at how Y changes with X_1 and X_2 from different angles
      A. **SEE EXAMPLE**
   iii. Goal is to find the direction in which Y seems to have the strongest-looking relationship with X (not restricted to linear)
   iv. Use that direction to define the ω weights

   (b) Fit a spline function $f_1(Z_1^{PPR})$ to model the relationship between Y and Z
   i. Use this function to predict Y in the given direction

   (c) Now repeat this M times
   i. Create M sets of linear combination weights $\omega_1, \omega_2 \ldots, \omega_M$
      A. Each weight finds the direction of greatest Y-variation, using residuals from previous projections.
   ii. Create M functions $f_1, \ldots, f_M$

   (d) Combine these into one prediction

   $$f(X) = \sum_{m=1}^{M} \beta_m f_m(Z_m^{PPR})$$

   (e) Hope is to select $M < p$, so that you can describe the surface using a smaller number of functions
   i. But there are extra parameters in the ω_m's
   ii. M is a tuning parameter. can be selected by CV (or stepwise, stopping when the next term represents little improvement).

2.1 Notes

- PPR can adapt to interactions because the splines are fit in directions that are not necessarily aligned with the axes
  - Adding several of these together can create interesting shapes
- The process of finding the optimal location is a little complicated, but not too bad
- This method is not used much because there is a more accurate way to find "best" directions and combine them into predictions.

Example: PPR on the Prostate Data (L12 - PPR Prostate.R)

The `ppr()` function in the `stats` package fits the models. I start with fitting a models to just the 2-variable model with `lcavol` and `pgg45`. This lets us see how it works. You can fit models with any number of terms and then backward-eliminate them to a smaller number.

See the results.

3 Exercises

Application

Refer to the Air Quality data described previously, and the analyses we have done with `Ozone` as the response variable, and the five explanatory variables (including the two engineered features).

1. Use PPR to model the relationship between `Ozone` and all five explanatories (you don't need to use the `scale()` function in the formula). Use `max.terms=5`. Also use the `gcv.spline` smoothing method as shown in the example.

   (a) Specify `nterms=1`.

      i. **Show the plot** of the spline for the selected projection

      ii. Report the training SSE from the summary

   (b) Repeat with `nterms=2`.

      i. **Show the plot** of the spline for the selected projection

      ii. Report the training SSE from the summary

2. Now we must figure out which number of terms to use in a final prediction. We need to tune this parameter. Use 10-fold cross-validation to train models and compute MSPE for values of `nterms` from among 1, 2, 3, 4, and 5, maintaining `max.terms=5`. Be sure to train each version of the model on each fold so that the comparison across the tuning parameters is easy.

   (a) **Report the matrix of MSPEs from CV.** (There should be 10 rows and 5 columns)

      i. Comment on any consistent patterns you see in the comparison among numbers of terms. Specifically, are there one or more values that seem much better than others?

   (b) **Create and show the side-by-side boxplots of these 10 MSPEs for each number of terms (5 boxes)**

   (c) **Repeat using relative MSPE**

   (d) Based on what you have seen, **how many terms would you use?** If there is no clear winner, then choose the least complicated model than is among the top models.

3. Add "tuned PPR" to the CV comparison that has been used for other methods. Use the same folds as were used for the other methods. Arrange it so that in each fold, the best PPR model is chosen exactly as above and is used to produce PPR's predictions for that fold. *This means that you will need to tune PPR exactly as in the previous problem, separately within each fold!!! (You can use 5-fold CV for tuning if you want to save a little time.* Separately save out the number of terms used in the best model for each fold.

(a) **Add the tuned PPR to the boxplots. Present the plots and write a sentence describing how well tuned PPR performs compared to other methods we have used thus far.**

(b) **Repeat this using relative MSPE.**

(c) **List the optimal tuning parameters that were selected for the tuned PPR in each of the 10 folds**