Homework Assignment 8: Red Brain, Blue Brain

36-402, Advanced Data Analysis, Spring 2012

Due at 10:30 am on Tuesday, 3 April 2012

We continue to work with the n90_pol.csv data set from last week's assignment. Recall that the three variables are `orientation`, an ordinal measure of political orientation running from 1 (very conservative) to 5 (very liberal); `amygdala`, which measures the volume of that region of the brain after adjusting for body size and the like; and `acc`, which measures the volume of a different region called the anterior cingulate cortex. Last time, we saw that `orientation` is positively correlated with `amygdala` and negatively correlated with `acc`. We also looked at the distribution of the brain volumes conditional on `orientation`. This time, we will try to predict political orientation from the brain volumes. We saw last time that when `orientation` is 3 or larger, there isn't much difference in the distribution of brain volumes, so we will concentrate on predicting whether `orientation` is less than or equal to 2.

1. *Creating a binary response variable* (15 total)
   (a) (5) Create a vector, `conservative`, which indicates whether each subject has `orientation` ≤ 2 (in which case the corresponding value in `conservative` should be 1) or has `orientation` ≥ 3 (in which case `conservative` should be 0).
   (b) (5) Check that your `conservative` vector has the proper values, *without* manually examining all 90 entries.
   (c) (5) Add `conservative` to your data frame. (Creating a new data frame with a new name will only get you partial credit.)
2. *Logistic regression* (15 total)
   (a) (5) Fit a logistic regression of `conservative` (not `orientation`) on `amygdala` and `acc`. Report the coefficients to no more than three significant digits. Explain what the coefficients mean.
   (b) (10) Using case resampling, give bootstrap standard errors and 95% confidence intervals for the coefficients. Was the restriction to three significant digits reasonable?
3. *Generalized additive model.* (10) Fit a generalized additive model for `conservative` on `amygdala` and `acc`. Make sure you are using a logistic link function. Report the intercept with reasonable precision. Plot the partial response functions, and explain what they mean (be careful!).

4. *Kernel conditional probability estimation.* (15 total)

   (a) (10) Using `npcdens`, find the conditional probability of `conservative` given `amygdala` and `acc`. Make sure `npcdens` treats `conservative` as a categorical variable and not a continuous one. Report the bandwidths.

   (b) (3) Plot the estimated conditional probability that `conservative` is 1, with `acc` set to its median value and `amygdala` running over the range $[-0.07, 0.09]$.

   (c) (2) Plot the estimated conditional probability that `conservative` is 1, with `amygdala` set to its median value and `acc` running over the range $[-0.04, 0.06]$.

5. *Probability surfaces* (15) For each of the three models, create a plot showing the estimated probability that `conservative` is 1, given `amygdala` or `acc`. The range for `amygdala` should be $[-0.07, 0.09]$, and the range for `acc` should be $[-0.04, 0.06]$. Compare and contrast the three plots.

   Contour, wireframe and heatmap/level-plot plots are all acceptable, but all access must be clearly labeled with numerical scales, and, if you use color, there must be a color key.

   Hint: use `predict`; be careful that you are predicting the right thing.

6. *Calibration* (15) Make calibration plots for each of the three models, as in chapter 13 of the notes. Which models (if any) seem reasonably calibrated? Explain with reference to your plots.

7. *Classification* (15) All three of our models predict probabilities for `conservative`. If we have to make a point prediction of whether someone is conservative or not, we should predict 1 if the probability is ≥ 0.5 and 0 otherwise. Find such predictions for each subject, under each of the three models. What fraction of subjects are mis-classified? What fraction would be mis-classified by "predicting" that none of them are conservative?