More Issues with Logistic Regression

STA303/STA1002: Methods of Data Analysis II, Summer 2016

Michael Guerzhoy

Perfectly Fitting the Data

- Likelihood: $\prod_{i=1}^{n} \pi_i^{y_i}(1-\pi_i)^{1-y_i}$
- $\pi_i = \frac{1}{1+e^{-\beta_0-\beta_1 x_1^{(i)}-\beta_2 x_2^{(i)}-\cdots}}$
- Want π_i as close as possible to y_i
  - If $y_i = 1$, want $-\beta_0 - \beta_1 x_1^{(i)} - \beta_2 x_2^{(i)} - \cdots$ as negative as possible
  - If $y_i = 0$, want $-\beta_0 - \beta_1 x_1^{(i)} - \beta_2 x_2^{(i)} - \cdots$ as positive as possible
- Suppose $y_i = x_1^{(i)}$ always
  - What's the β_1 that maximizes the likelihood?

Perfectly Fitting the Data

- Suppose $y_i = x_1^{(i)}$ always
  - What's the β_1 that maximizes the likelihood?
  - The smaller β_1, the better
- In general, suppose
  - $e^{-\beta_0-\beta_1 x_1^{(i)}-\beta_2 x_2^{(i)}-\cdots-\beta_p x_p^{(i)}} < 1 \Leftrightarrow y_i = 1$
- How can we then increase the likelihood by changing the β's?

- In general, suppose
  - $e^{-\beta_0-\beta_1 x_1^{(i)}-\beta_2 x_2^{(i)}-\cdots-\beta_p x_p^{(i)}} < 1 \Leftrightarrow y_i = 1$
- How can we then increase the likelihood by changing the β's?
  - Multiply all of them by a constant factor

Perfect Separation

- $e^{-\beta_0-\beta_1 x_1^{(i)}-\beta_2 x_2^{(i)}-\cdots-\beta_p x_p^{(i)}} < 1 \Leftrightarrow y_i = 1$
- $\beta_0 + \beta_1 x_1^{(i)} + \beta_2 x_2^{(i)} + \cdots + \beta_p x_p^{(i)} > 0 \Leftrightarrow y_i = 1$
- With two covariates:
  - (in R)
  - The line $\beta_0 + \beta_1 x_1^{(i)} + \beta_2 x_2^{(i)}$ separates the points belonging to the two classes
- With multiple covariates:
  - The hyperplane separates the points belonging to the two classes

Residuals

- The Deviance Residual for point *i* is such that the sum of squared residuals adds up to the deviance

$$sign(y_i - \pi_i)\sqrt{2\{y_i \log\left(\frac{y_i}{\pi_i}\right) + (1 - y_i)\log(\frac{1-y_i}{1-\pi_i})\}}$$

- (Note: there is no constant here, since we've explicitly chosen how to define it here)
  - $const = 2\sum_i \ (y_i \log y_i + (1 - y_i)\log(1 - y_i)$
  - Reminder: $deviance =$

$$const - 2\log \mathrm{P}(\mathrm{y}|\beta) =$$

$$const - 2\sum_i (y_i \log \pi_i + (1 - y_i)\log(1 - \pi_i))$$

- If the residual for any group of people look unusual, examine the data

Extra-Binomial Variation

- So far, we assumed that $Var(Y_i|x,\beta) = \pi_i(1-\pi_i)$
- That is not true when the observations Y_i are not conditionally independent
- That is also not true when in fact the true probabilities π_i and π_j are not the same even though $x^{(i)} = x^{(j)}$
  - Suppose we omit a covariate (e.g., the class of the passenger in the Titanic dataset)
  - The probability of drowning for women in first class is different from the probability of drowning for women in second class

Quasilikelihood Approach

- Estimate the overdispersion parameter
  - One estimate: $\hat{\psi} = \frac{Deviance}{Degrees\ of\ Freed}$
  - Use $SE_{\hat{\psi}}(\beta) = \sqrt{\hat{\psi}} SE_{est}(\beta)$
  - Overdispersion increases the uncertainty about the parameters
  - Note: this is called “Quasilikelihood” because we are not using our model anymore – we’re adjusting the Standard Error
  - Justification: hopefully the estimate for β are not biased