P9120 Homework 2

Shan Jiang

October 16, 2019

Problem 1

Ex. 4.5

Consider a two-class logistic regression problem with $x \in R$. Characterize the maximum-likelihood estimates of the slope and intercept parameter if the sample x_i for the two classes are separated by a point $x_o \in R$. Generalize this result to (a) $x \in R_p$ and (b) more than two classes.

First we assume that $x_0 = 0$ and y = 1 for $x_i > 0$ and y = 0 for $x_i < 0$.

$$p(x;\beta) = \frac{\exp(\beta_x + \beta_0)}{1 + \exp(\beta_x + \beta_0)}$$

$$1 - p(x;\beta) = 1 - \frac{\exp(\beta_x + \beta_0)}{1 + \exp(\beta_x + \beta_0)} = \frac{1}{1 + exp(\beta_x + \beta_0)}$$

Given the condition that $x_0 = 0$ is the boundary, then $p(x_0) = 1 - p(x_0)$, $\beta_0 = 0$, so the above function can be simplified as:

$$p(x;\beta) = \frac{\exp(\beta_x)}{1 + \exp(\beta_x)}$$

$$1 - p(x;\beta) = 1 - \frac{1}{1 + \exp(\beta_x)}$$

Therefore we derive the likelihood function as:

$$L(\beta; y, x) = \prod_{i=1}^{N} p(x_i;\beta)^{y_i}[1 - p(x_i;\beta)]^{1-y_i}$$

$$= \prod_{i=1}^{N} [\frac{p(x_{i;}\beta)}{1 - p(x_{i;}\beta)}]^{y_i}[1 - p(x_i;\beta)]$$

$$= \prod_{i=1}^{N} [\exp(\beta x_i)]^{y_i}[1 - p(x_i;\beta)]$$

Then we take log of both sides:

$$logL(\beta; y, x) = \Sigma_{i=1}^{N} y_i[\beta x_i] - log[1 + exp\{\beta x_i\}]$$

After then we take the first-order derivatives w.r.t to β, and use the condition for y as an replacement.

$$\frac{dlogL(\beta; y, x)}{d\beta} = \Sigma_{i=1}^{N} x_i(y_i - \frac{exp(\beta x_i)}{1 + exp(\beta x_i)})$$

$$= \Sigma_{X_i>0}^N x_i(1 - \frac{exp(\beta x_i)}{1+exp(\beta x_i)}) - \Sigma_{x_i<0}^N x_i(\frac{exp(\beta x_i)}{1+exp(\beta x_i)})$$

$$= \Sigma_{X_i>0}^N x_i - \Sigma_{X_i>0}^N x_i(\frac{exp(\beta x_i)}{1+exp(\beta x_i)}) - \Sigma_{x_i<0}^N x_i(\frac{exp(\beta x_i)}{1+exp(\beta x_i)}).$$

$$set \frac{dlogL(\beta; y, x)}{d\beta} = 0 \Leftrightarrow \Sigma_{X_i>0}^N x_i = \Sigma_{x_i=1}^N x_i(\frac{exp(\beta x_i)}{1+exp(\beta x_i)})$$

Thus, we can infer that for any dataset $\Sigma_{x_i=1}^N x_i$, only when $\beta \to \infty$,the first order derivative can be set to 0 and holds.

(b) Suppose that there are K classes, with x_k sperates between K-1 class and K class, and -$\infty < x_0 < x_1 < x_2 < x_3 < \cdots < x_{k-1} < x_k = \infty$.

Each probability is defined as:

$$P_1(x;\beta) = \frac{exp(\beta_1 x + \beta_{01})}{1 + \Sigma_{j=1}^{K-1} exp(\beta_j x + \beta_{0j})}$$

$$P_2(x;\beta) = \frac{exp(\beta_2 x + \beta_{02})}{1 + \Sigma_{j=1}^{K-1} exp(\beta_j x + \beta_{0j})}$$

$$\vdots$$

$$P_{k-1}(x;\beta) = \frac{1}{1 + \Sigma_{j=1}^{K-1} exp(\beta_j x + \beta_{0j})}$$

Given the assumption, for y is y_i= 1 if $x_{j-1} < x_i < x_j$ and $y_i = 0$ otherwise for observation i = 1, . . . , N and class j = 1, . . . , K. we can take a new likelihood function:

$$L(\beta; y, x) = \prod_{j=1}^{K} \prod_{i=1}^{N_j} [p_j(x_i;\beta)^{y_i}]^{y_i},$$

where N_j is the number of obs. in class j, then we take the log-likelihood function:

$$logL(\beta; y, x) = \Sigma_{j=1}^{K-1} \Sigma_{i=1}^{N_j} log[\frac{exp(\beta_j x_i + \beta_{0j})}{1 + \Sigma_{j=1}^{K-1} exp(\beta_j x + \beta_{0j})}] + \Sigma_{i=1}^{N_k} y_i log[\frac{1}{1 + \Sigma_{j=1}^{k-1} exp(\beta_j x_i + \beta_{0j})}]$$

$$= \Sigma_{j=1}^{K-1} \Sigma_{i=1}^{N_j} y_i[\beta_j x_i + \beta_{0j}] - \Sigma_{j=1}^{K} \Sigma_{i=1}^{N_j} y_i log[1 + \Sigma_{j=1}^{k-1} exp(\beta_j x_i + \beta_{0j})]$$

Then we take the derivative wrt to $\beta = (\beta_1, \beta_2, \cdots, \beta_{k-1})$, given that for $x_{j-1} < x < x_j$, $p(x; \beta_j) = \frac{exp(\beta_j x + \beta_{0j})}{1 + \Sigma_{j=1}^{k-1} exp(\beta_j x_i + \beta_{0j})}$, where $\beta_{0j} = log[exp(\beta_j x_{j-1}) - exp(\beta_j x_j)]$

The first-order derivative is then written as:

$$\frac{dlogL(\beta; y, x)}{d\beta_j} = \Sigma_{i=1}^{N_j} x_i + \Sigma_{i=1}^{N_j} \frac{exp(\beta_j x_{j-1}) x_{j-1} - exp(\beta_j x_j) x_j}{exp(\beta_j x_{j-1}) - exp(\beta_j x_j)}$$

$$-\Sigma_{i=1}^{N} [X_i + \frac{exp(\beta_j x_{j-1} - exp(\beta_j x_j) x_j}{exp(\beta_j x_{j-1}) - exp(\beta_j x_j)}](\frac{exp(\beta_j x_i - \beta_{0j})}{1 + \Sigma_{j=1}^{K-1} exp(\beta_j x_i + \beta_{0j})})$$

Set the $\frac{dlogL(\beta;y,x)}{d\beta_j} = 0$, Since the question listed two scenarios:

(a) Now, suppose that there are two classes in which x $\in R_p$. Suppose that $\mathbf{X_1}$ and $\mathbf{X_2}$ are two vectors, we have that $\boldsymbol{\beta}(\mathbf{X_1}$ -$\mathbf{X_2}) = 0$.

$$p(\mathbf{x};\beta) = \frac{exp(\beta' x + \beta_0)}{1 + exp(\beta' x + \beta_0)}$$

As we know that the X_0 is the seperating parameter, we can simplify the above equation as:

$$1 - p(\mathbf{x};\beta) = \frac{1}{1 + exp((x - x_0))}$$

This is similar to the univariate case in that once taking derivatives of the log-likelihood function wrt to $\beta = (\beta_1, \beta_2, \cdots, \beta_{k-1})$, and setting them equal to zero.

In conclusion, the Generalized form is that when $||\beta|| \rightarrow \infty$, the maximum likelihood estimator is attained, note that β is a vector.

Problem 2

Ex. 5.1 Show that the truncated power basis functions in (5.3) represent a basis for a cubic spline with the two knots as indicated.

$$h_1(X) = 1, h_3(X) = X^2, h_5(X) = (X - \xi_1)^3_+$$

$$h_2(X) = X, h_4(X) = X^3, h_5(X) = (X - \xi_2)^3_+$$

The proof requires a fulfillment of Condition C_2 leads to C_1,

The C_1 requires truncated power basis functions; While for C_2, it required a collection of cubic functions.

To prove from cubic function $C_2 \rightarrow C_1$

The cubic polynomial can be expressed by:

$$f(x) = \Sigma^6_{m=1}\beta_m h_m(x)$$

Then we need to show the continuity of the f(x) at knots $\xi_1$and ξ_2, and the first and second order derivatives' continuity.

1. The continuity of f(x):

By proving the left limit = right limit at ξ_1,the continuity can be achieved:

The Left-limit:

$$f(\xi_1 - k) = \beta_1 + \beta_2(\xi_1 - k) + \beta_3(\xi_1 - k)^2 + \beta_4(\xi_1 - k)^3 + \beta_5(\xi_1 - k - \xi_1)^3_+ + \beta_6(\xi_1 - k - \xi_2)^3_+, k > 0$$

$$= \beta_1 + \beta_2(\xi_1 - k) + \beta_3(\xi_1 - k)^2 + \beta_4(\xi_1 - k)^3 + 0 + 0$$

So,

$$lim_{k \rightarrow 0^-} f(\xi_1 - k) = \beta_1 + \beta_2\xi_1 + \beta_3(\xi_1)^2 + \beta_4(\xi_1)^3$$

The right limit:

$$f(\xi_1 + k) = \beta_1 + \beta_2(\xi_1 + k) + \beta_3(\xi_1 + k)^2 + \beta_4(\xi_1 + k)^3 + \beta_5(\xi_1 + k - \xi_1)^3_+ + \beta_6(\xi_1 + k - \xi_2)^3_+, k > 0$$

$$lim_{k \rightarrow 0^+} f(\xi_1 + k) = \beta_1 + \beta_2\xi_1 + \beta_3(\xi_1)^2 + \beta_4(\xi_1)^3$$

Thus, the left limit is equal to the right limit at ξ_1,the continuity can be achieved.

2. Continuity of $f'(x)$

Then we take the first order derivatives at $x = \xi_1$,

$$f'(\xi_1) = lim_{k\to 0}\frac{f(\xi_1) - f(\xi_1 - k)}{k}$$

Based on the definition of $f(x), and\ f(\xi_1) = \beta_1 + \beta_2\xi_1 + \beta_3(\xi_1)^2 + \beta_4(\xi_1)^3$ then we have:

$$f(\xi_1) - f(\xi_1 - k) = \beta_2\xi_1 + 2\beta_3(\xi_1)k + 3\beta_4(\xi_1)^2k + O(k)$$

Then the lfet-side derivative w.r.t. k is written as:

$$f'_-(\xi_1) = \beta_2\xi_1 + 2\beta_3(\xi_1) + 3\beta_4(\xi_1)^2$$

The right-side derivative is written as:

$$f'_+(\xi_1) = lim_{k\to 0}\frac{f(\xi_1 + k) - f(\xi_1)}{k}$$

For the same reason like above,

$$f(\xi_1 + k) - f(\xi_1) = \beta_2\xi_1 + 2\beta_3(\xi_1)k + 3\beta_4(\xi_1)^2k$$

$$+\beta_5(\xi_1 + k - \xi_1)^3_+ + \beta_6(\xi_1 + k - \xi_2)^3_+ + O(k)$$

$$f'_+(\xi_1) = \beta_2 + 2\beta_3\xi_1 + 3\beta_4\xi_1^2$$

3. Continuity of $f''(x)$

Then we take the second order derivatives at $x = \xi_1$

For the same reason like above part 2, that the left-side limit is equal to right-side limit,

$$f''(\xi_1) = f''_-(\xi_1) = f''_+(\xi_1) = 6\beta_4\xi_1^2$$

CONCLUSION: AT ξ_1, FUNCTION, FIRST AND second order derivatives are all continuous at this knot. Similarly, we can prove it on ξ_2, THUS we can say the cubic spline function with two knots is equivalent to that of the basis function as it fufills the continuity requirement for function, first-order derivative and second-order derivative.

Problem 3

see next page R code.