

P9120 Homework 2

Shan Jiang

October 6, 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:

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

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

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

Then we take log of both sides:

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

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

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

$$\begin{aligned}
&= \sum_{X_i > 0}^N x_i \left(1 - \frac{\exp(\beta x_i)}{1 + \exp(\beta x_i)}\right) - \sum_{x_i < 0}^N x_i \left(\frac{\exp(\beta x_i)}{1 + \exp(\beta x_i)}\right) \\
&= \sum_{X_i > 0}^N x_i - \sum_{X_i > 0}^N x_i \left(\frac{\exp(\beta x_i)}{1 + \exp(\beta x_i)}\right) - \sum_{x_i < 0}^N x_i \left(\frac{\exp(\beta x_i)}{1 + \exp(\beta x_i)}\right). \\
&\text{set } \frac{d \log L(\beta; y, x)}{d\beta} = 0 \Leftrightarrow \sum_{X_i > 0}^N x_i = \sum_{x_i = 1}^N x_i \left(\frac{\exp(\beta x_i)}{1 + \exp(\beta x_i)}\right)
\end{aligned}$$

Thus, we can infer that for any dataset $\sum_{x_i=1}^N x_i$, only when $\beta \rightarrow \infty$, 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 $-\infty < x_0 < x_1 < x_2 < x_3 < \dots < x_{k-1} < x_k = \infty$.

Each probability is defined as:

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

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

\vdots

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

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:

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

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

Then we take the derivative wrt to $\beta = (\beta_1, \beta_2, \dots, \beta_{k-1})$, given that for $x_{j-1} < x < x_j$,

$$p(x; \beta_j) = \frac{\exp(\beta_j x + \beta_{0j})}{1 + \sum_{j=1}^{K-1} \exp(\beta_j x + \beta_{0j})}, \text{ where } \beta_{0j} = \log[\exp(\beta_j x_{j-1}) - \exp(\beta_j x_j)]$$

The first-order derivative is then written as:

$$\begin{aligned}
\frac{d \log L(\beta; y, x)}{d\beta_j} &= \sum_{i=1}^{N_j} x_i + \sum_{i=1}^{N_j} \frac{\exp(\beta_j x_{j-1}) x_{j-1} - \exp(\beta_j x_j) x_j}{\exp(\beta_j x_{j-1}) - \exp(\beta_j x_j)} \\
&- \sum_{i=1}^N [X_i + \frac{\exp(\beta_j x_{j-1}) - \exp(\beta_j x_j) x_j}{\exp(\beta_j x_{j-1}) - \exp(\beta_j x_j)}] \left(\frac{\exp(\beta_j x_i - \beta_{0j})}{1 + \sum_{j=1}^{K-1} \exp(\beta_j x_i + \beta_{0j})} \right)
\end{aligned}$$

Set the $\frac{d \log L(\beta; y, x)}{d \beta_j} = 0$, 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 $\beta(\mathbf{X}_1 - \mathbf{X}_2) = 0$.

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

As we know that the X_0 is the separating 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 $\beta = (\beta_1, \beta_2, \dots, \beta_{k-1})$, 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.

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

$$h_2(X) = X, h_4(X) = X^3, h_6(X) = (X - \xi_2)_+^3$$

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:

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

Then we need to show the continuity of the $f(x)$ at knots ξ_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 $x = \xi_1$,

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

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 left-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:

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

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 fulfills the continuity requirement for function, first-order derivative and second-order derivative.

Problem 3