Maximum likelihood estimation in logistic regression

AlgoNeural

March 23, 2023

We want to maximize the probability of observing our data, which is given by the following formula:

$$p(y_1|x_1)\, p(y_2|x_2) \ldots p(y_m|x_m)\ . \tag{1}$$

Even though parameters a and b are not explicitly mentioned in (1), they are implicitly included in the definition of each $p(y_i|x_i)$:

$$p(y_i|x_i) = \begin{cases} \frac{1}{1+e^{-(ax_i+b)}}, & y_i = 1\ , \\ 1 - \frac{1}{1+e^{-(ax_i+b)}}, & y_i = 0 \end{cases} \tag{2}$$

If we denote

$$h(a, b, x_i) = \frac{1}{1 + e^{-(ax_i+b)}}\ , \tag{3}$$

then we can rewrite (2) as:

$$p(y_i|x_i) = h(a, b, x_i)^{y_i}\, (1 - h(a, b, x_i))^{(1-y_i)}\ . \tag{4}$$

To find the best fitting curve, we need to maximize (1) with respect to a and b. Using (4), we can write our optimization problem as follows:

$$\max_{a,b} \prod_{i=1}^{m} h(a, b, x_i)^{y_i} (1 - h(a, b, x_i))^{1-y_i}\ . \tag{5}$$

To rewrite this expression as a negative log-likelihood, first observe that the logarithm is a monotonically increasing function, which means that optimizing $\log(f(x))$ is equivalent to optimizing $f(x)$. For the sake of convenience, we will also multiply the final result by $1/m$ (multiplying the objective function by a positive constant does not change the optimal values of a and b). Hence, (5) becomes:

$$\max_{a,b} \frac{1}{m} \sum_{i=1}^{m} \left[y_i \log h(a, b, x_i) + (1 - y_i)(1 - h(a, b, x_i))\right]\ . \tag{6}$$

Finally, by inverting the problem and using (3), we get:

$$\min_{a,b} -\frac{1}{m}\sum_{i=1}^{m}\left[y_i \log\left(\frac{1}{1+e^{-(ax_i+b)}}\right) + (1-y_i)\log\left(1-\frac{1}{1+e^{-(ax_i+b)}}\right)\right] . \tag{7}$$