SAAS MLE Workshop Note

Wenhao Pan

December 1, 2022

1 Simple Linear Regression

Here we describe the application of MLE to simple linear regression. For observed data $\{x_i, y_i \in \mathbb{R}\}_{i=1}^n$, we have the following modeling assumption

$$y_i = \alpha + \beta x_i + \epsilon_i, \quad \epsilon_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2) \tag{1}$$

where α, β are unobserved constants in $\mathbb{R}$. We can treat α, β as two *unobserved parameters* we want to estimate.

If we assume x_i are also constants, then ϵ_i's are the only sources of the randomness in our data. Since any linear transformation of a normally distributed random variable is still normally distributed, y_i for $i = 1, \ldots, n$ has the following distribution

$$y_i \sim \mathcal{N}(\alpha + \beta x_i, \sigma^2). \tag{2}$$

In other words, y_i has the following probability density function

$$f_{y_i}(y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{y - (\alpha + \beta x_i)}{\sigma}\right)^2\right). \tag{3}$$

Thus, the log-likelihood function for estimating α, β is

$$\begin{aligned}
L(\alpha, \beta) \coloneqq \log lik(\alpha, \beta) &= \log P(y_1, \ldots, y_n | \alpha, \beta) \\
&= \log\left(\prod_{i=1}^n P(y_i | \alpha, \beta)\right) \\
&= \sum_{i=1}^n \log f(y_i | \alpha, \beta) \\
&= \sum_{i=1}^n \log\left(\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{y - (\alpha + \beta x_i)}{\sigma}\right)^2\right)\right) \\
&= \sum_{i=1}^n \log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{1}{2}\left(\frac{y - (\alpha + \beta x_i)}{\sigma}\right)^2 \\
&= n \log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y - (\alpha + \beta x_i))^2
\end{aligned} \tag{4}$$

Maximizing such log-likelihood function to find $\hat{\alpha}, \hat{\beta}$ is equivalent to

$$\begin{aligned}
\hat{\alpha}, \hat{\beta} &= \arg\max_{\alpha,\beta} \; L(\alpha, \beta) \\
&= \arg\max_{\alpha,\beta} \; n \log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y - (\alpha + \beta x_i))^2 \\
&= \arg\max_{\alpha,\beta} \; -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y - (\alpha + \beta x_i))^2 \\
&= \arg\min_{\alpha,\beta} \; \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y - (\alpha + \beta x_i))^2 \\
&= \arg\min_{\alpha,\beta} \; \sum_{i=1}^{n}(y - (\alpha + \beta x_i))^2
\end{aligned} \tag{5}$$

which is minimizing the L2 loss. It is the same as the well-known method *least squares*.

2 Simple Logistic Regression

Here we describe the application of MLE to simple logistic regression for binary classification. For observed data $\{x_i \in \mathbb{R}, y_i \in \{0, 1\}\}_{i=1}^n$, we have the following modeling assumption

$$y_i \sim \text{Bernoulli}(\pi_i), \quad \pi_i = \sigma(\alpha + \beta x_i) \tag{6}$$

where $\sigma(\cdot)$ is the logistic function and α, β are unobserved constants in $\mathbb{R}$. The logistic function has the following form

$$\sigma(x) = \frac{1}{1 + \exp(-x)} \tag{7}$$

for $x \in \mathbb{R}$. We can treat α, β as two *unobserved parameters* we want to estimate. y_i has the following probability mass function

$$P(y_i = y) = \pi_i^y (1 - \pi_i)^{1-y} = \begin{cases} \pi_i & \text{if } y = 1 \\ 1 - \pi_i & \text{if } y = 0 \end{cases} \tag{8}$$

Thus, the log-likelihood function for estimating α, β is

$$\begin{aligned}
L(\alpha, \beta) \coloneqq \log lik(\alpha, \beta) &= \log P(y_1, \ldots, y_n | \alpha, \beta) \\
&= \log\left(\prod_{i=1}^{n} P(y_i | \alpha, \beta)\right) \\
&= \sum_{i=1}^{n} \log P(y_i | \alpha, \beta) \\
&= \sum_{i=1}^{n} \log\left(\pi_i^{y_i}(1 - \pi_i)^{1-y_i}\right) \\
&= \sum_{i=1}^{n} y_i \log \pi_i + (1 - y_i)\log(1 - \pi_i)
\end{aligned} \tag{9}$$

Maximizing such log-likelihood function to find $\hat{\alpha}, \hat{\beta}$ is equivalent to

$$\begin{aligned}
\hat{\alpha}, \hat{\beta} &= \arg\max_{\alpha,\beta} \; L(\alpha, \beta) \\
&= \arg\max_{\alpha,\beta} \; \sum_{i=1}^{n} y_i \log \pi_i + (1 - y_i)\log(1 - \pi_i) \\
&= \arg\min_{\alpha,\beta} \; -\sum_{i=1}^{n} y_i \log \pi_i - (1 - y_i)\log(1 - \pi_i)
\end{aligned} \tag{10}$$

which is minimizing the well-known cross-entropy loss.