Interpreting Log Error

James Chen

March 5, 2018

Let $\hat{P}$ be an estimate of the target value P. Say in the training we have obtained

$$\log(1+P) = \log\left(1+\hat{P}\right) \pm \hat{\epsilon} \tag{1}$$

where $\hat{\epsilon}$ is some estimate [1] of the test error and

$$\hat{\epsilon} \ll 1 \ll \hat{P}, \tag{2}$$

then

$$P \approx \hat{P}\left(1 \pm \hat{\epsilon}\right), \quad \hat{P} \approx P\left(1 \pm \hat{\epsilon}\right). \tag{3}$$

In other words, the error $\hat{\epsilon}$ for $\log(1+\hat{P})$ translates [2] to the percentage-error $100\hat{\epsilon}$ % for $\hat{P}$.

Proof. From (1), we have $\log\left[(1+P)/(1+\hat{P})\right] = \pm\hat{\epsilon}$, that is,

$$P = -1 + e^{\pm\hat{\epsilon}}\left(1+\hat{P}\right) \tag{4}$$

Since $\hat{\epsilon}$ is relatively small, e.g. $\hat{\epsilon} \approx 0.01$ (see (2)), by Taylor series,

$$e^{\pm\hat{\epsilon}} \approx 1 \pm \hat{\epsilon} \tag{5}$$

Substitue (5) into (4), we have

$$\begin{aligned} P &\approx \hat{P} \pm \hat{\epsilon}\hat{P} \pm \hat{\epsilon} \\ &\approx \hat{P}\left(1 \pm \hat{\epsilon}\right) \end{aligned} \tag{6}$$

Notice that in the last step leading to (6), we drop the term $\pm\hat{\epsilon}$ since $\hat{\epsilon} \ll \hat{P}$ and $\hat{\epsilon} \ll \hat{\epsilon}\hat{P}$ (see (2)). This proves the first approximation in (3).

Next, from (6), we estimate that

$$\hat{P} \approx P\left(1 \pm \hat{\epsilon}\right), \tag{7}$$

which can be derived by substituting the following two approximations into (6):

$$\begin{aligned} \frac{1}{1-\hat{\epsilon}} &= 1 + \hat{\epsilon} + \hat{\epsilon}^2 + \hat{\epsilon}^3 + \ldots \approx 1 + \hat{\epsilon}, \\ \frac{1}{1+\hat{\epsilon}} &= 1 - \hat{\epsilon} + \hat{\epsilon}^2 - \hat{\epsilon}^3 + \ldots \approx 1 - \hat{\epsilon}. \end{aligned}$$

This proves the second approximation in (3). □

[1] For example, we may take $\hat{\epsilon}$ to be the worst (largest) root mean square error from cross validations.

[2] By the way, (3) also hold for the case when $\log P = \pm\hat{\epsilon} + \log\hat{P}$ under the same assumption (2). The proof is similar.