

cross validation error (CVE)

An absolute measure of discrepancy for the predictions made by a given *model*, where smaller values indicate a better fit of the model to the data. It differs from the *mean squared error* (MSE), because it measures the *test error* instead of the *training error*, i.e. it approximates the quality of the predictions made by a model. There are different ways to compute the CVE. A simple variant is the “leave one out CV” (or LOOCV), which is computed as follows:

$$CVE = \frac{1}{n} \sum_{i=1}^n (f_i(x_i) - y_i)^2$$

where f_i is a model fitted on the data set S_i which is the original data set S with the i^{th} value left out, i.e.:

$$S_i = \{x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n\}$$

So the LOOCV uses a subset S_i of S to fit a model and then computes the *error* of the omitted element $\hat{y}_i = f_i(x_i)$, whose actual value y_i is known from the data. This process is repeated for each pair $x_i, y_i \in S$. The CVE is the mean of all of the computed errors. Because LOOCV has to fit the model f n times, it is expensive to compute.

Other means of computing the CVE are the LpOCV, which leaves out p elements instead of one, and the validation set approach, which divides the set S into a “training set” and a “test set” and then uses the training set for model fitting and the test set for computing the CVE. The latter method is biased, though, (\rightarrow *bias*) because the CVE here depends on the choice of the test set.