



Mean vs Total Squared (or Absolute) Error

A potential confusion is the following: How do we know if we should use the mean or the total squared (or absolute) error?

The total squared error is the sum of errors at each point, given by the following equation:

$$M = \sum_{i=1}^m \frac{1}{2} (y - \hat{y})^2,$$

whereas the mean squared error is the average of these errors, given by the equation, where m is the number of points:

$$T = \sum_{i=1}^m \frac{1}{2m} (y - \hat{y})^2.$$

The good news is, it doesn't really matter. As we can see, the total squared error is just a multiple of the mean squared error, since

$$M = mT.$$

Therefore, since derivatives are linear functions, the gradient of T is also m times the gradient of M .

However, the gradient descent step consists of subtracting the gradient of the error times the learning rate α . Therefore, choosing between the mean squared error and the total squared error really just amounts to picking a different learning rate.

In real life, we'll have algorithms that will help us determine a good learning rate to work with. Therefore, if we use the mean error or the total error, the algorithm will just end up picking a different learning rate.

[NEXT](#)