To fit the regression model, we choose a model form , where y is the label, f is the model function, x are the features, w are the model weights, and e is the error. The minus sign ( - ) is chosen for the error term so that a positive error implies the estimate overshoots the actual value, and a negative error indicates an estimate that undercuts the actual value.

This allows for a nicely human-interpretable metric.

Constructing an objective function for tuning parameters, consider the error terms by solving for them in the model equation. This gives , and using the errors to construct a convex loss function leads to using .

In the case of a linear regression, we choose a model where f is an affine function of the features, and the coefficients are given by the weights. This looks like for a model with n features, where b is the bias and the w are the feature weights. In the notation above, this can be represented by treating b as a special additional weight in the w vector, which then gives .

In this formulation, x is an n-dimensional vector of features, and w is an n+1 dimensional vector of parameters, where represents the n-dimensional parameters without the bias term, thus constituting the feature coefficients. The notation represents the inner product, which for our purposes means the same as adding the pairwise products of each element in the vector.

This is the same as saying .