# Bias-variance decomposition

Assume the true model is and the observation is , where and . The expected predication error of a regression fit   at the input vector using the squared-error loss . Different data sets will lead different values. Thus, the performance of a particular learning algorithm is assessed by taking the average over the ensemble of data sets [1]. In other word, we may take   as a random variable which depends on the training set . To get the expected loss for , we first solve its expectation with respect to sample noise



The last step is because . Continue our analysis from . We obtain the expectation of the loss function with respect to



The squared bias represents the extent to which the average predication over all data sets differ from the true value. The variance term measures the extent to which the predications based on individual data sets vary around their average. In summary, we have

* Very flexible models: low bias and high variance
* Simple models: high bias and low variance

# Reference

[1] Pattern Recognition and machine learning. Page 148.