

Bias and Variance

Suppose you have implemented regularized logistic regression to predict what items customers will purchase on a web shopping site. However, when you test your hypothesis on a new set of customers, you find that it makes unacceptably large errors in its predictions. Furthermore, the hypothesis performs poorly on the training set.

Which of the following might be promising steps to take? Check all that apply.

- (i) Try increasing the regularization parameter λ .

The poor performance on both the training and test sets suggests a high bias problem. Increasing regularization will decrease the fit of the hypothesis to the data, exacerbating the high bias problem.

- (ii) Try evaluating the hypothesis on a cross validation set rather than the test set.

A cross validation set is useful for choosing the optimal non-model parameters like the regularization parameter λ , but the train / test split is sufficient for debugging problems with the algorithm itself.

- (iii) Try adding polynomial features.

The poor performance on both the training and test sets suggests a high bias problem. Adding more complex features will increase the complexity of the hypothesis, thereby improving the fit to both the train and test data.

- (iv) Try to obtain and use additional features.

The poor performance on both the training and test sets suggests a high bias problem. Using additional features will increase the complexity of the hypothesis, thereby improving the fit to both the train and test data.