

Overfitting Card

The performance of your learning algorithm is excellent on the training set, i.e. the data set that was used to train the algorithm. However, when you calculate its performance on a test set, it is much worse. You've heard of this phenomenon called overfitting, but you don't remember how to correct it.

What do you do?

NB: Overfitting is a property of a learning algorithm that learns "by heart" the training data and thus has excellent performance on the training set. As with all rote learning, the understanding of the underlying phenomena is lacking and the algorithm's performance on the test is disastrous.