

## Training, Test, and Validation Datasets

The goal of supervised machine learning is to be able to predict or classify the targets on unseen examples correctly. We can misjudge the performance of our learning models if we evaluate the model performance with the same samples used to train the model as explained previously.

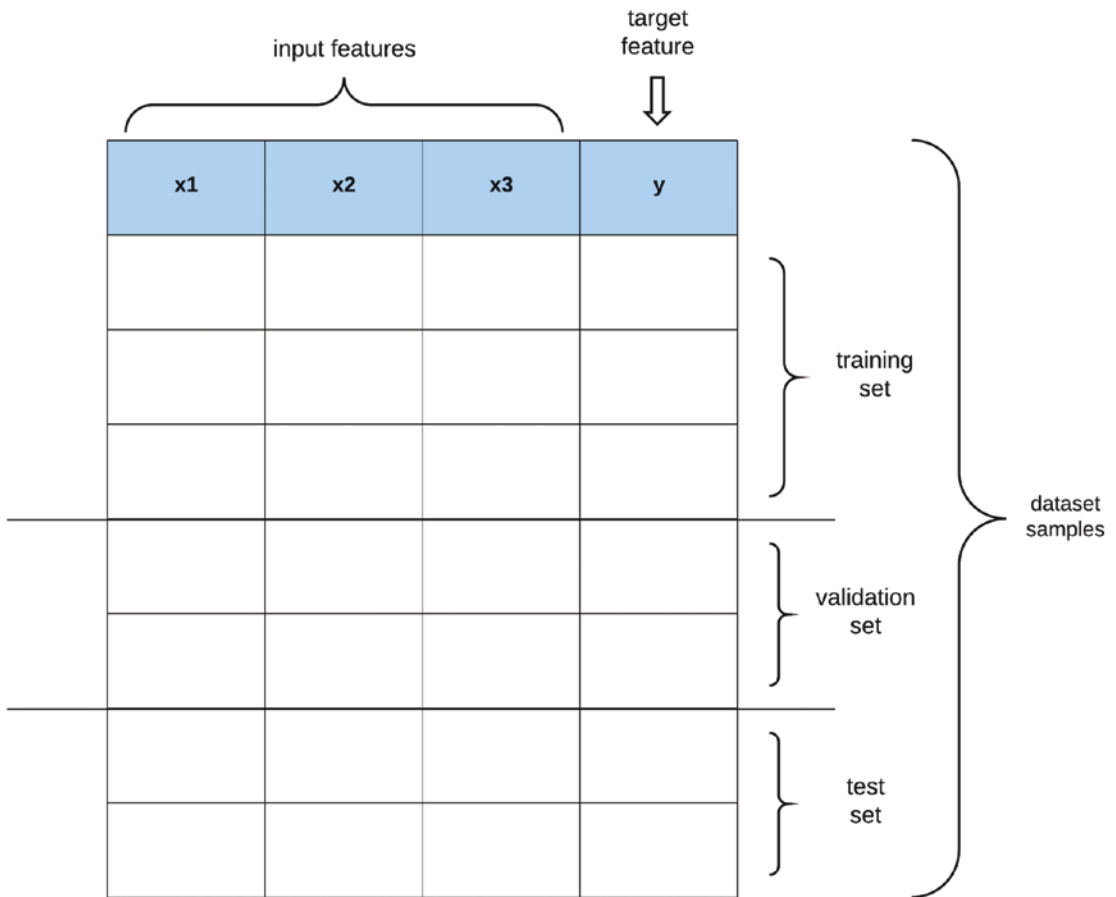
To properly evaluate the performance of a learning algorithm, we need to set aside some data for testing purposes. This held-out data is called a **test set**.

Another situation arises when we have trained the model on a dataset, and we now need to improve the performance of the model by adjusting some of the learning algorithm's parameters.

We cannot use the test set for model tuning because if we do that, the model's parameters are trained with the test dataset rendering it unusable as an unseen held-out sample for model evaluation. Hence, it is typical to divide the entire dataset into

- The training set (to train the model)
- The validation set (to tune the model)
- The test set (to evaluate the effectiveness of the model)

A common and straightforward strategy is to split 60% of the dataset for training, 20% for validation, and the final 20% for testing. This strategy is popularly known as the 60/20/20 rule. We will discuss more sophisticated methods for resampling (i.e., using subsets of available data) for machine learning later in this chapter. See Figure [14-4](#).



**Figure 14-4.** Training, test, and validation set

## Bias vs. Variance Trade-Off

The concept of bias vs. variance is central to machine learning and is critical to understanding how the model is performing, as well as in suggesting the direction in which to improve the model.

A model is said to have high **bias** when it oversimplifies the learning problem or when the model fails to accurately capture the complex relationships that exist between the input features of the dataset. High bias makes the model unable to generalize to new examples.