What is the difference between a validation set and a test set?

Authored by

stats writer

May 12, 2024

A validation set and a test set are commonly used in machine learning and data science to evaluate the performance of a model. The main difference between these two sets is their purpose and the stage at which they are used in the model development process.

A validation set is used to assess the performance of a model during the training stage. It is used to tune the hyperparameters of a model and to select the best performing model for further analysis. The validation set is usually a subset of the training data and is not used in the actual training process.

On the other hand, a test set is used to evaluate the final performance of a model. It is used once the model has been trained and validated to test its generalization capabilities on unseen data. The test set is completely independent from the training and validation data and is only used for final evaluation.

In summary, the main difference between a validation set and a test set is their purpose and the stage at which they are used. The validation set is used for model selection and hyperparameter tuning during training, while the test set is used for final evaluation after the model has been trained. Both sets are crucial in ensuring the reliability and accuracy of a model's performance.

Validation Set vs. Test Set: What's the Difference?

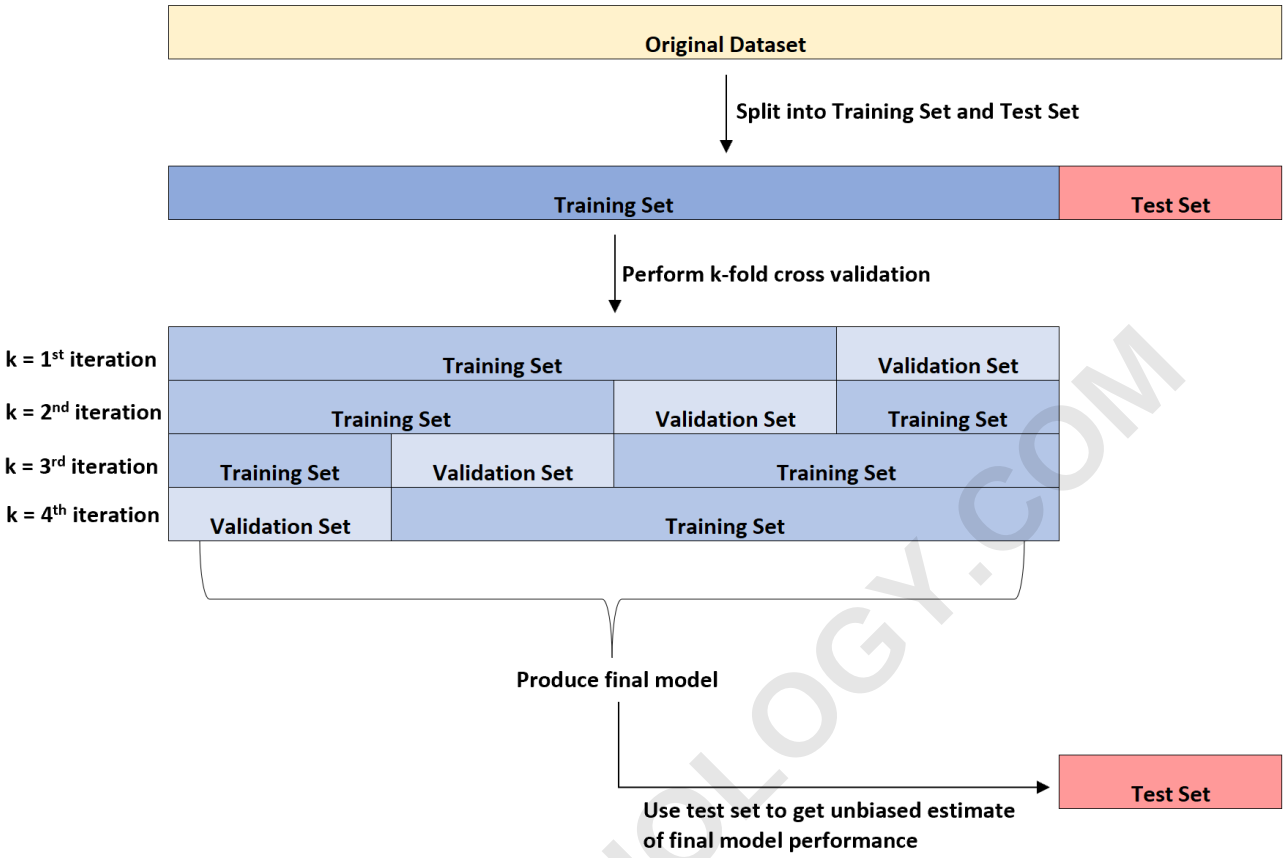
Whenever we fit a to a dataset, we typically split the dataset into three parts:

1. Training Set: Used to train the model.

2. Validation Set: Used to optimize model parameters.

3. Test Set: Used to get an unbiased estimate of the final model performance.

The following diagram provides a visual explanation of these three different types of datasets:



One point of confusion for students is the difference between the validation set and the test set.

In simple terms, the validation set is used to optimize the model parameters while the test set is used to provide an unbiased estimate of the final model.

It can be shown that the error rate as measured by k-fold cross validation tends to underestimate the true error rate once the model is applied to an unseen dataset.

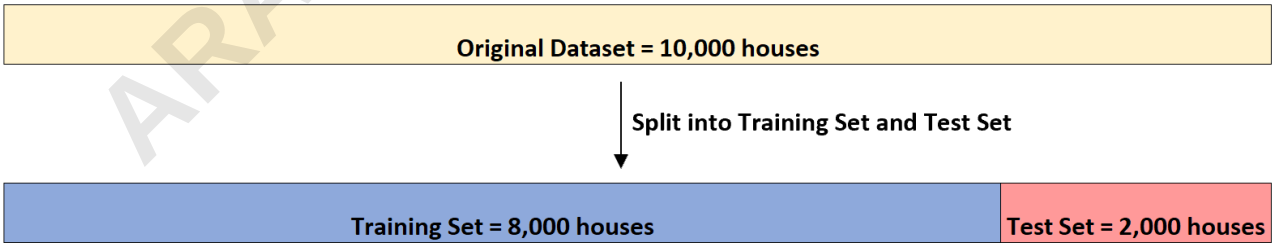
Thus, we fit the final model to the test set to get an unbiased estimate of what the true error rate will be in the real world.

The following example illustrates the difference between a validation set and a test set in practice.

Example: Understanding the Difference Between Validation Set & Test Set

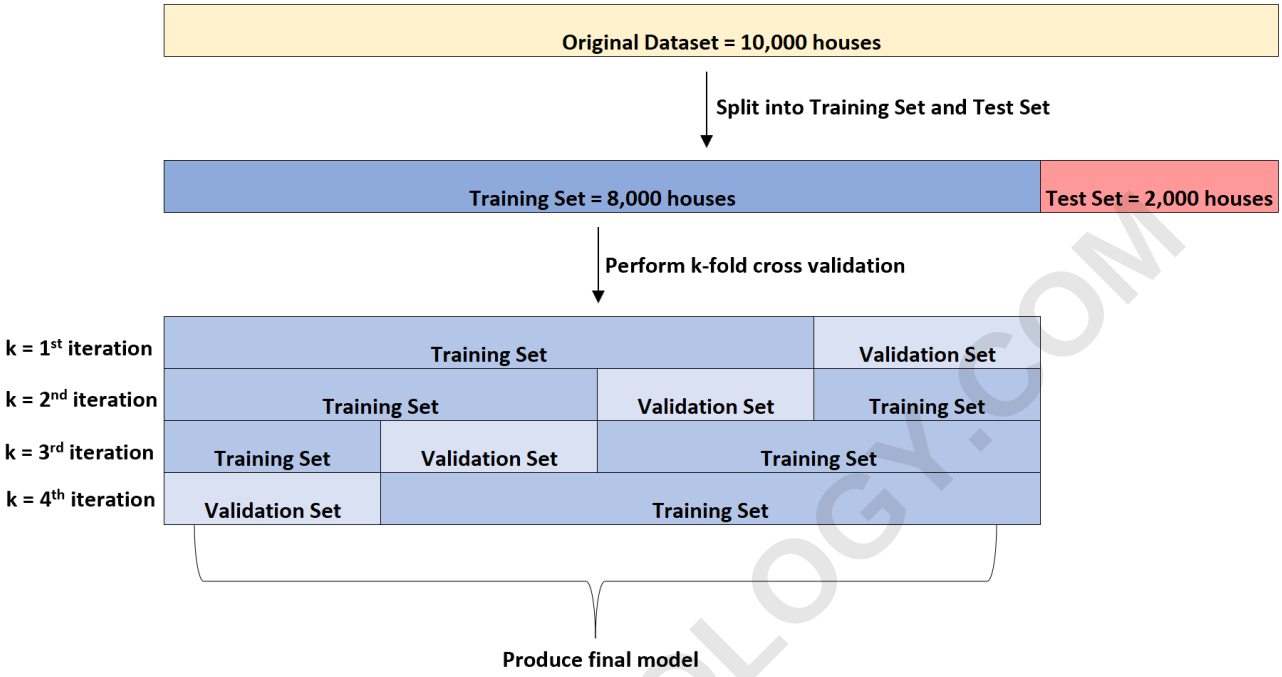
Suppose a real estate investor wants to use (1) number of bedrooms, (2) total square feet, and (3) number of bathrooms to predict the selling price of a given house.

Suppose he has a dataset with this information on 10,000 houses. First, he'll split up the dataset into a training set of 8,000 houses and a test set of 2,000 houses:



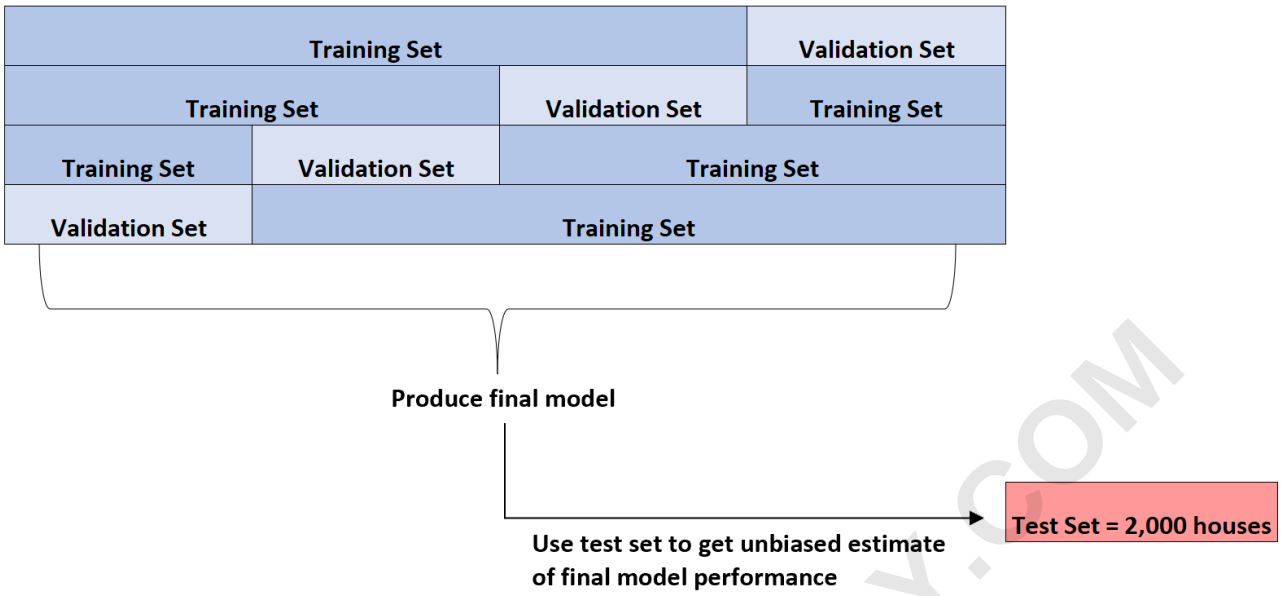
Next, he'll fit a multiple linear regression model to the dataset four times. Each time he'll use 6,000 houses for the training set and 2,000 houses for the validation set.

This is known as k-fold cross validation.



He may perform this k-fold cross validation on several different types of regression models to identify the model that has the lowest error (i.e. identify the model that fits the dataset best).

Only once he has identified the best model will he then use the test set of 2,000 houses that he held out at the beginning to get an unbiased estimate of the final model performance.



For example, he might identify a specific type of regression model that has a mean absolute error of 8,345. That is, the mean absolute difference between the predicted house price and actual house price is $8,345.

He may then fit this exact regression model to the test set of 2,000 houses that has not yet been used and find that the mean absolute error of the model is 8,847.

Thus, the unbiased estimate of the true mean absolute error for the model is $8,847.

How to Perform K-Fold Cross Validation in Python