**Choice of k in k-fold Cross Validation**

The k value must be chosen carefully for your data sample.

A poorly chosen value for k may result in a misrepresentative idea of the skill of the model, such as a score with a high variance (that may change a lot based on the data used to fit the model), or a high bias, (such as an overestimate of the skill of the model).

Three common tactics for choosing a value for k are as follows:

* **Representative**: The value for k is chosen such that each train/test group of data samples is large enough to be statistically representative of the broader dataset.
* **k=10**: The value for k is fixed to 10, a value that has been found through experimentation to generally result in a model skill estimate with low bias a modest variance.
* **k=n**: The value for k is fixed to equal n, where n is the size of the dataset to give each test sample an opportunity to be used in the hold out dataset. **This approach is called leave-one-out cross-validation.**

*The choice of k is usually 5 or 10, but there is no formal rule. As k gets larger, the difference in size between the training set and the resampling subsets gets smaller. As this difference decreases, the bias of the technique becomes smaller*

— Page 70, [Applied Predictive Modeling](http://amzn.to/2Fmrbib), 2013.

A value of k=10 is very common in the field of applied machine learning, and is recommend if you are struggling to choose a value for your dataset.

*To summarize, there is a bias-variance trade-off associated with the choice of k in k-fold cross-validation. Typically, given these considerations, one performs k-fold cross-validation using k = 5 or k = 10, as these values have been shown empirically to yield test error rate estimates that suffer neither from excessively high bias nor from very high variance.*

— Page 184, [An Introduction to Statistical Learning](http://amzn.to/2FkHqvW), 2013.

If a value for k is chosen that does not evenly split the data sample, then one group will contain a remainder of the examples. It is preferable to split the data sample into k groups with the same number of samples, such that the sample of model skill scores are all equivalent.

<https://www.youtube.com/watch?v=qOwT553oMzs>



