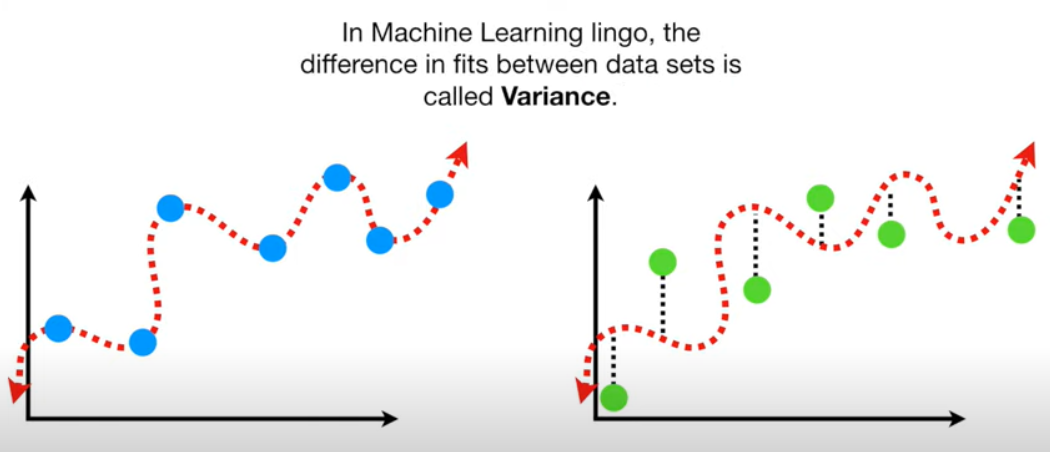
**VARIANCE AND BIAS**

Machine learning algorithms use mathematical or statistical models with inherent errors in two categories: reducible and irreducible error. Irreducible error, or inherent uncertainty, is due to natural variability within a system. In comparison, reducible error is more controllable and should be minimized to ensure higher accuracy.Bias and variance are components of reducible error. Reducing errors requires selecting models that have appropriate complexity and flexibility, as well as suitable training data. Data scientists must thoroughly understand the difference between bias and variance to reduce error and build accurate models.

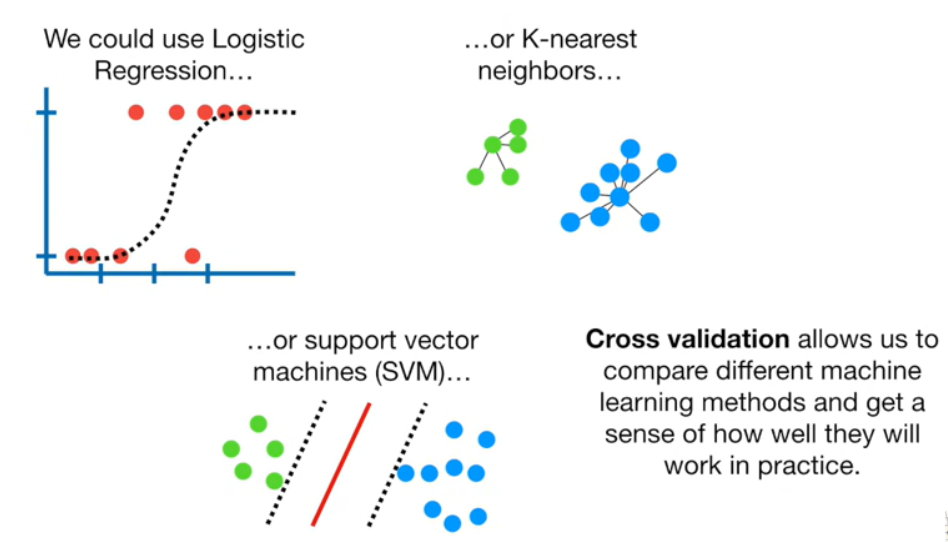


Reference Video: <https://www.youtube.com/watch?v=EuBBz3bI-aA>

**CROSS VALIDATION**

Cross-validation is a resampling procedure used to evaluate machine learning models on a limited data sample.The general procedure is as follows:

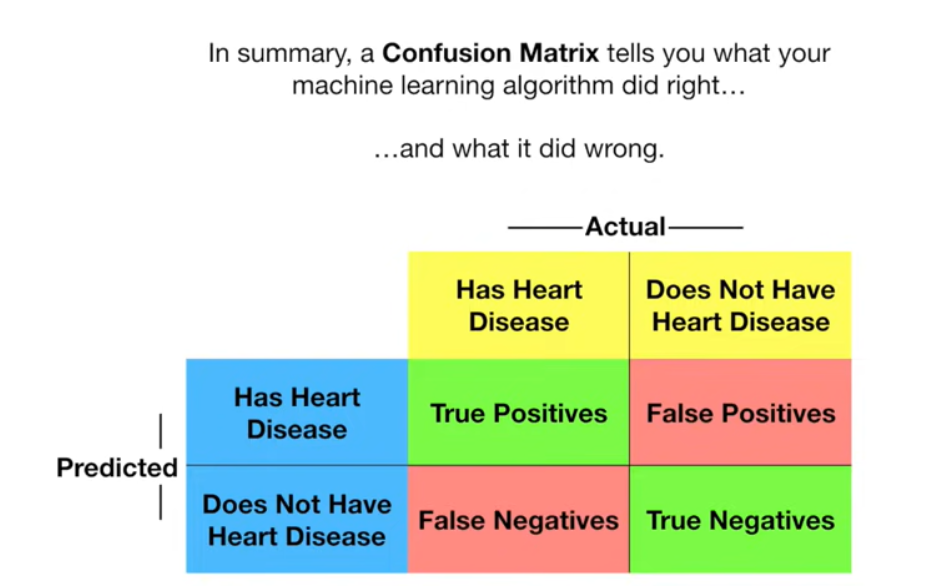
1. Shuffle the dataset randomly.
2. Split the dataset into k groups
3. For each unique group:
   1. Take the group as a hold out or test data set
   2. Take the remaining groups as a training data set
   3. Fit a model on the training set and evaluate it on the test set
   4. Retain the evaluation score and discard the model
4. Summarize the skill of the model using the sample of model evaluation scores



Reference Video: <https://www.youtube.com/watch?v=fSytzGwwBVw>

**CONFUSION MATRIX**

A Confusion matrix is an N x N matrix used for evaluating the performance of a classification model, where N is the number of target classes. The matrix compares the actual target values with those predicted by the machine learning model. This gives us a holistic view of how well our classification model is performing and what kinds of errors it is making**.**

****

Reference Video: <https://www.youtube.com/watch?v=Kdsp6soqA7o>

**REFERENCE:**

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