# Sampling Techniques

The most common approaches are sampling based, or more specifically, re-sampling based. Between up-sampling and down-sampling (also called over-sampling and under-sampling), down-sampling is a bit simpler conceptually. Fundamentally, if you have a minority class or classes that is noticeably underrepresented. In a random way you can drop some of those from the training data so that the proportions are more closely matched across classes. There are additional methods, one of which is inspired by K Nearest Neighbors (KNN) that can improve on pure random selection.

A major caveat to down-sampling is that we are not using all of our data. Over-sampling techniques also come in several flavors, from random or naive versions to classes of algorithms like the Synthetic Minority Oversampling Technique (SMOTE) [2] and the Adaptive Synthetic (ADASYN) [8] sampling method. Synthetic Minority Over-sampling Technique for Nominal and Continuous (SMOTENC) is an extension of the original SMOTE method designed to handle a mixture of categorical and continuous features. SMOTE has a number of other variants including ones that make use of Support Vector Machines and K-means clustering to improve on the synthetic samples.