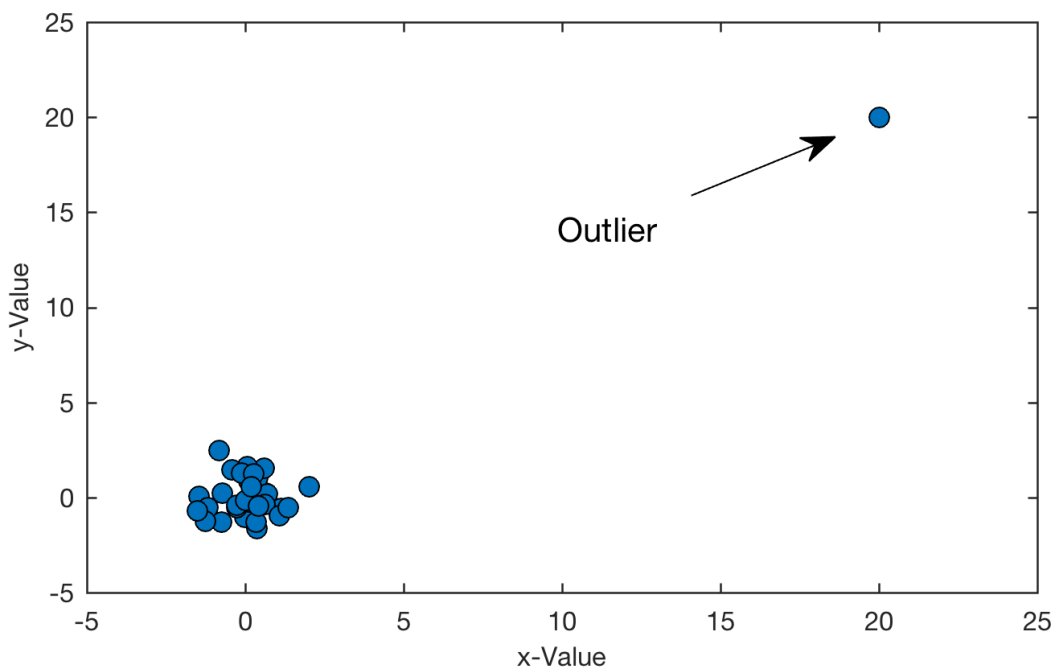


Manual: 7.1.7. Pre-processing

Before raw measurement data can be used in the context of machine learning, it is usually passed through several stages of processing. These methods are collectively called pre-processing and are essential to getting good results.

Implausible values are removed from the dataset according to the allowed ranges of values specified in the metadata. Outliers are removed next. These are data points that exhibit highly irregular or impossible features such as very large gradients. Values that are removed are filled in with the most recent prior plausible measurement.



Values are also normalized into a coordinate system so that each variable has an average of zero and standard deviation of one. This removes any natural scaling and thus removes any incompatibility of data related purely to the magnitude of the data. For example, you could measure a weight in kilograms or tons. For human understanding it does not matter much but for a computer the numbers are much larger in kilograms than in tons and so they would receive a different influence in comparison to something else like a pressure. Normalization removes any such features and pays attention only to relative variation in structure.