# Feature Selection

*“When dealing with high dimensionality data, it is often useful to reduce the dimensionality by projecting the data to a lower dimensional subspace which captures the essence of the data.”* [1, p11].

Although there are many features which can be extracted from PE files, not all of these will contribute positively to the accuracy of our solution and the inclusion of some features for clustering may, in fact, result in reduced accuracy. To increase the accuracy of our solution we will be carrying out dimensionality reduction [1, p11].

# References

[1]K. Murphy, *Machine learning*. Cambridge, Mass.: MIT Press, 2012.