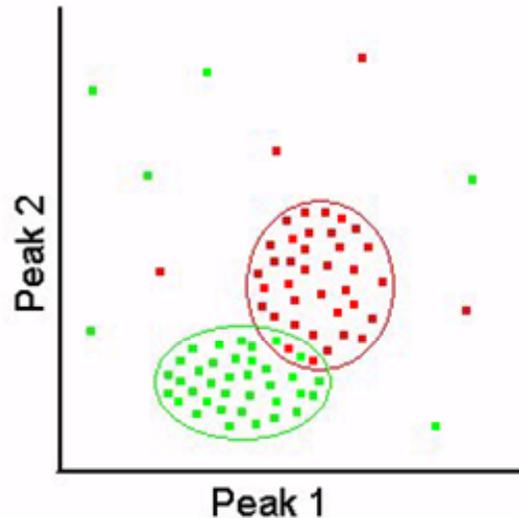


What Causes an “Undetermined” Classification?

Brian T. Luke (lukeb@ncifcrf.gov)

The figure to the right shows a hypothetical result for a classifier consisting of two peak intensities and demonstrates two different causes of an “Undetermined” classification. While most of the samples lie either in the Case region (red ellipse) or the Control region (green ellipse), a small number of samples lie in or near the intersection of these regions. These samples should receive an “Undetermined” classification since their nearest neighbors are a mixture of Cases and Controls. This was used to determine an “Undetermined” region in a classifier based on the blood concentration of complement C3a anaphylatoxin [[Hab-06](#)] to determine if an individual has colorectal cancer or polyps.

The samples in the figure that lie outside of the ellipses and far from any other sample will also [receive](#) an “Undetermined” classification. These samples are “Undetermined” because of an incomplete [coverage](#) of feature-space by the available samples. If the study assumes that the sampling of individuals is sufficient to cover the intensity ranges for each putative biomarker, a classification like this one should be strongly avoided.



(Last updated 4/28/07)