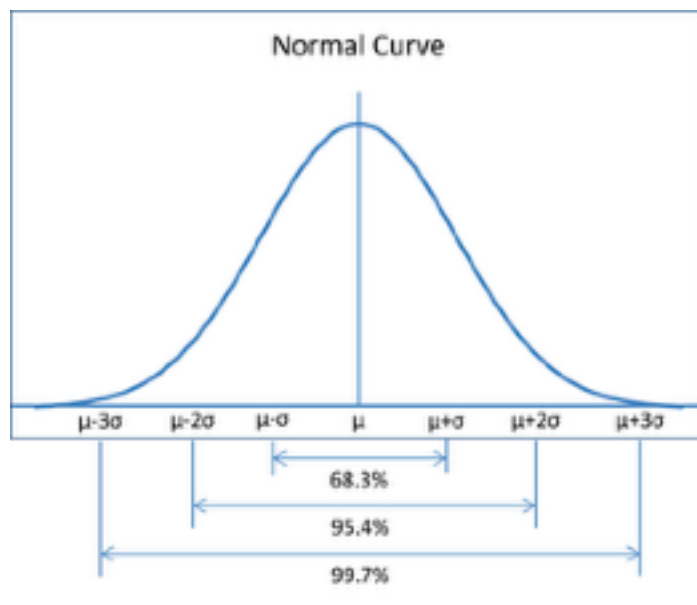


How is it usually done?

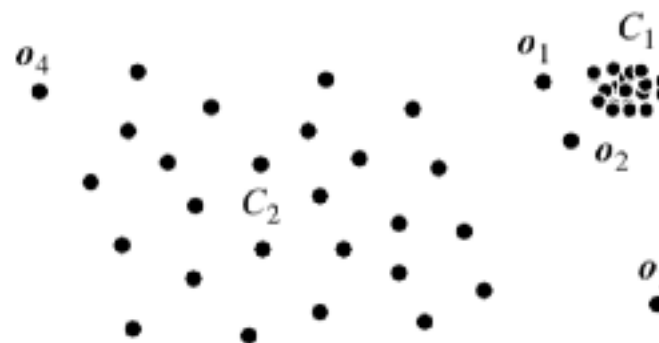
Parametric methods

Distribution based, if the observation is past a certain standard deviation it can probably be considered an outlier.



* this is by no means an extensive list of methods!

Distance based

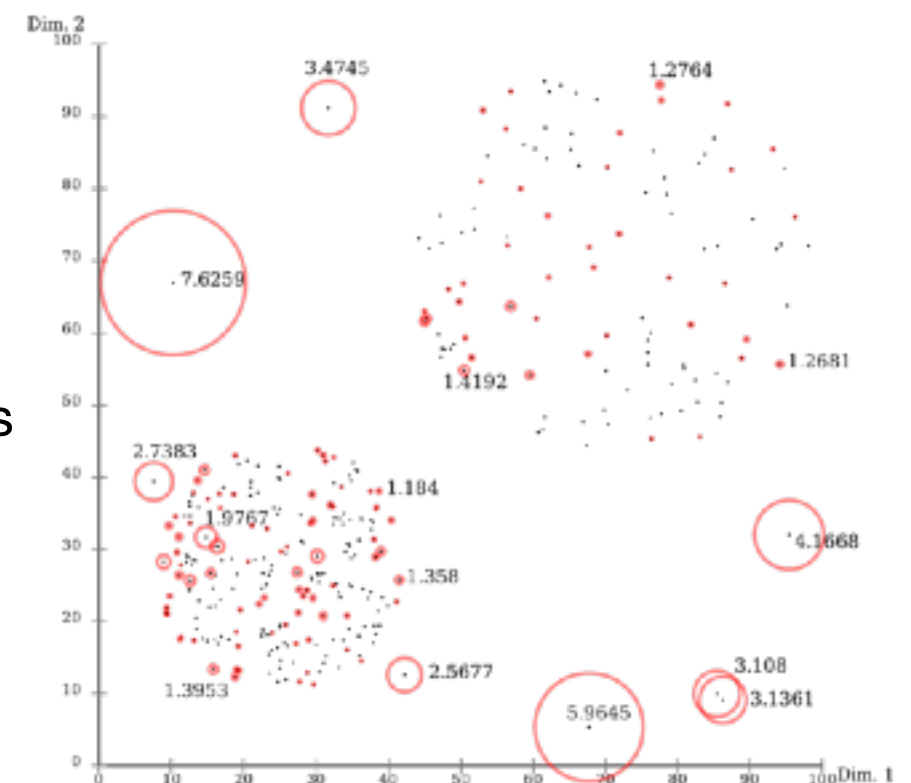


The neighbourhood around an outlier point is sparse; the observation is rather lonely.

Note: Popular, easy to understand distance based algorithms are $O(n^2)$ time-complexity.

Density based

Incorporates distance **and** density into its outlier detection process. This means not just *global lonely* points are found, but *locally lonely* points too!



Tree-based methods?

