

First use pca to reduce the number of dimension to 2
And then use grid-based outlier detection method to find outliers in a large data set

Setting $D = 100$ and $M = 100$ I find 22 outliers.

As the data is too large, I try to set $d=10$ and $M=10$, but the running time is too long to get the result before the due of the exam

Reference: https://github.com/vrzhao/COMP4331_A4