

1154-68-398

Mireille Boutin* (mboutin@purdue.edu). *Highly Likely Clusterable Data With No Cluster.*

Data generated as part of a real-life experiment is often quite organized. So much so that, in many cases, projecting the data onto a random line has a high probability of uncovering a clear division of the data into two well-separated groups. In other words, the data can be clustered with a high probability of success using a hyperplane whose normal vector direction is picked at random. We call such data “highly likely clusterable.” The clusters obtained in this fashion often do not seem compatible with a cluster structure in the original space. In fact, the data in the original space may not contain any cluster at all. This talk is about this surprising phenomenon. We will discuss empirical ways to detect it as well as how to exploit it to cluster datasets, especially datasets consisting of a small number of points in a high-dimensional space. We will also present a possible mathematical model that would explain this observed phenomenon. This is joint work with Sangchun Han, Tarun Yellamraju, and Alden Bradford. (Received September 05, 2019)