Data Mining

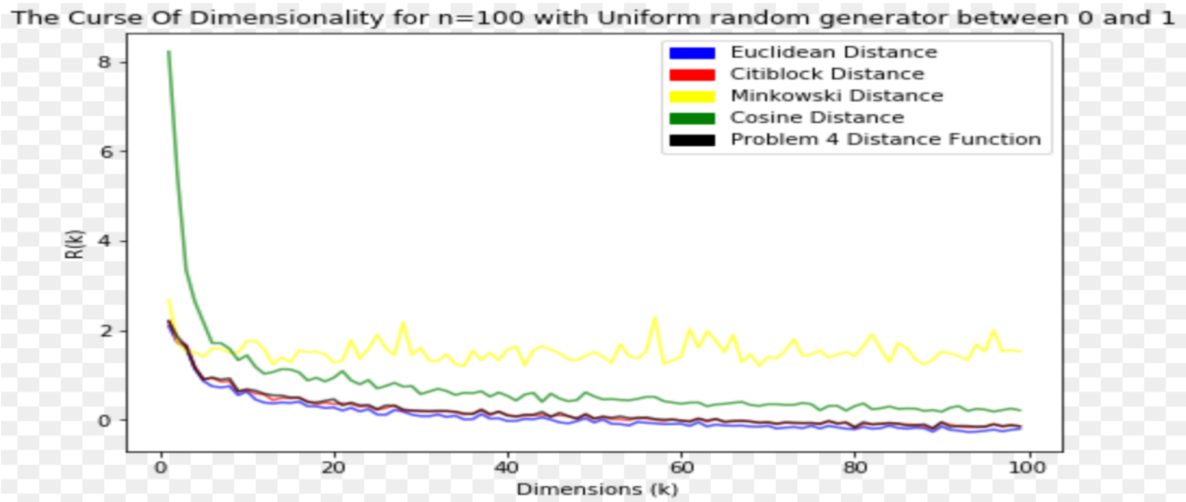
Name – Surya Prakash Sekar

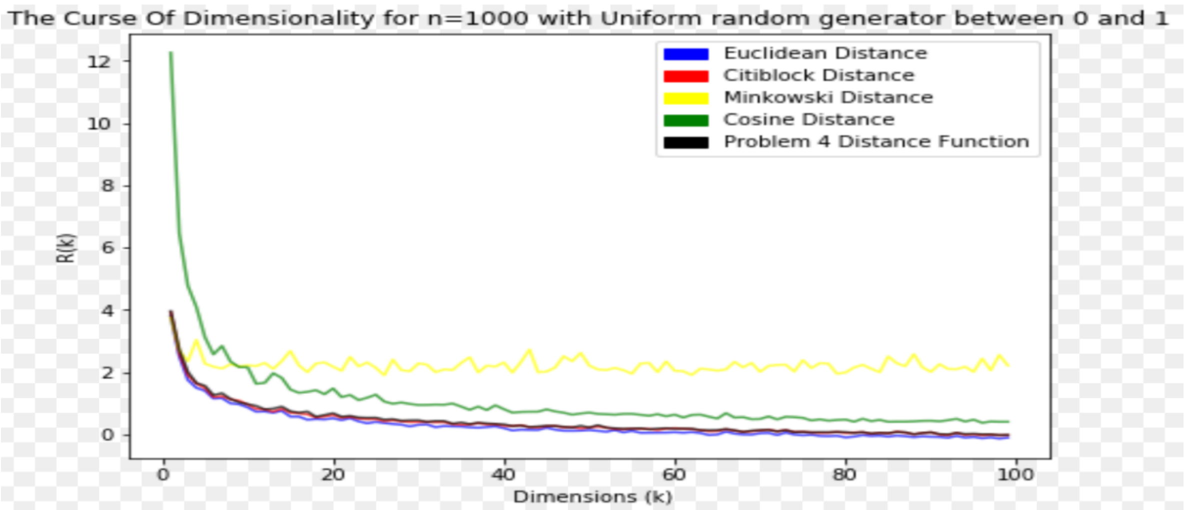
Email id – sursekar@iu.edu

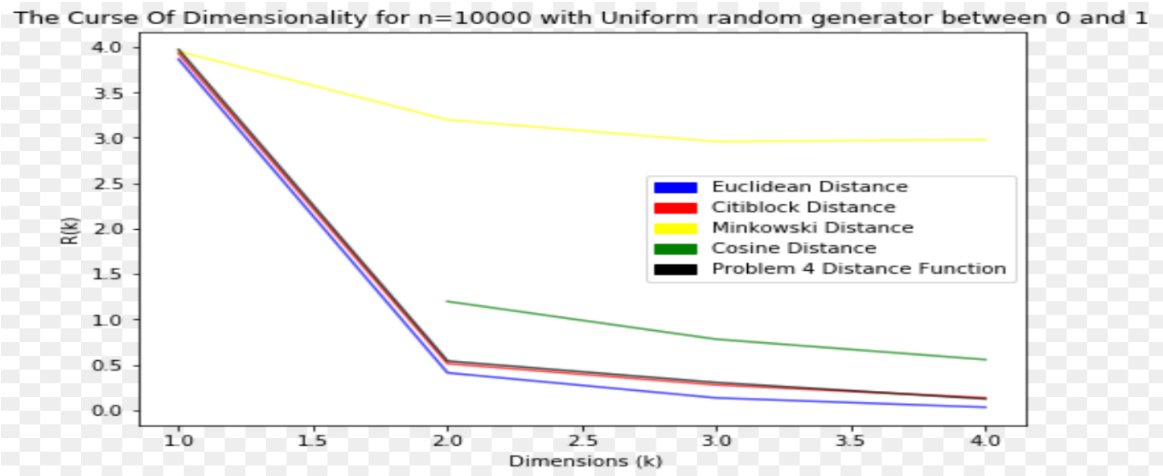
The Curse of Dimensionality

Part A

Analysing the Curse of Dimensionality problem when the data comes from a uniform random source

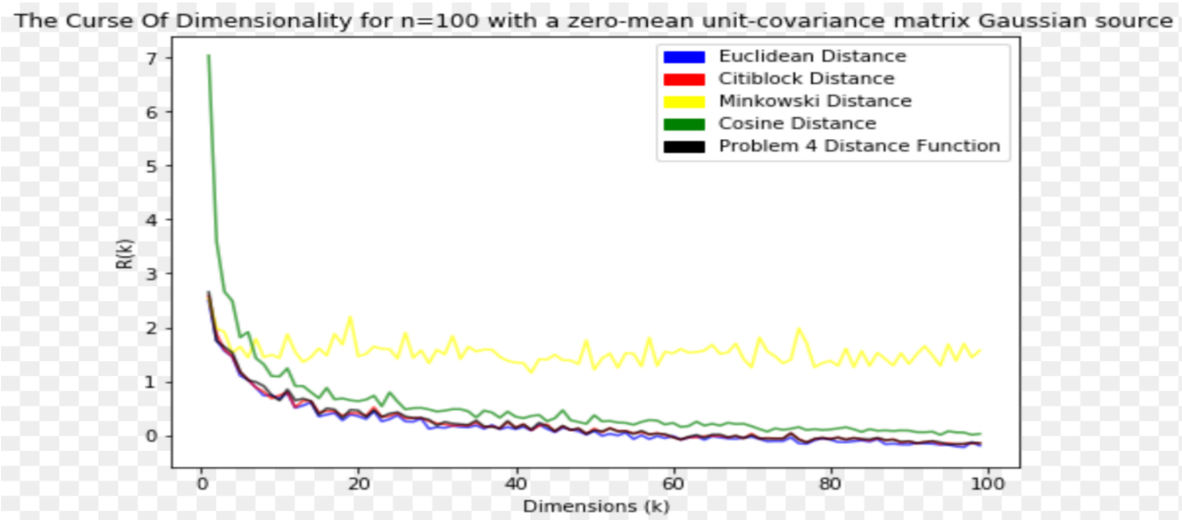


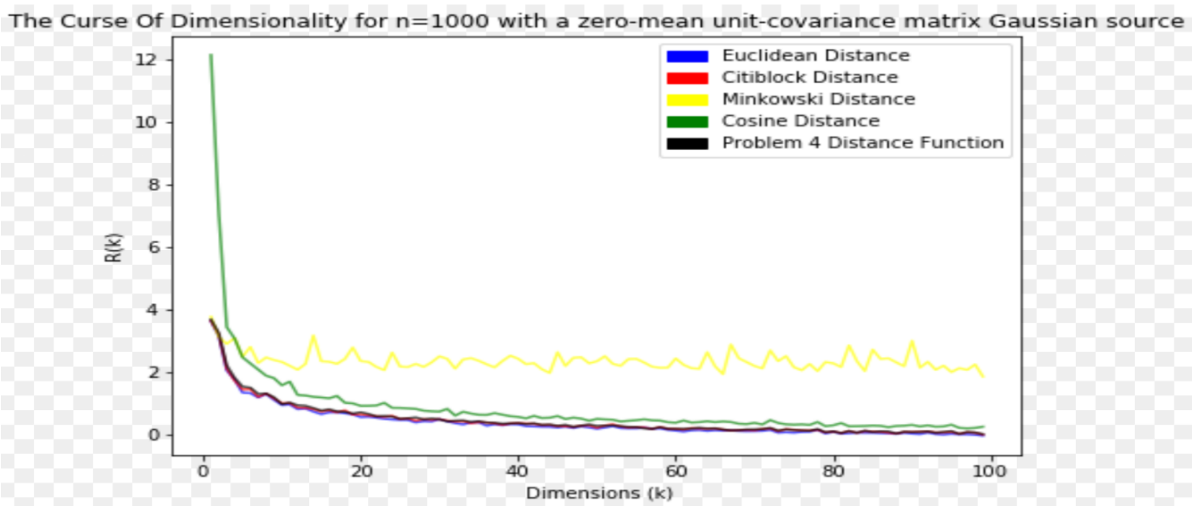


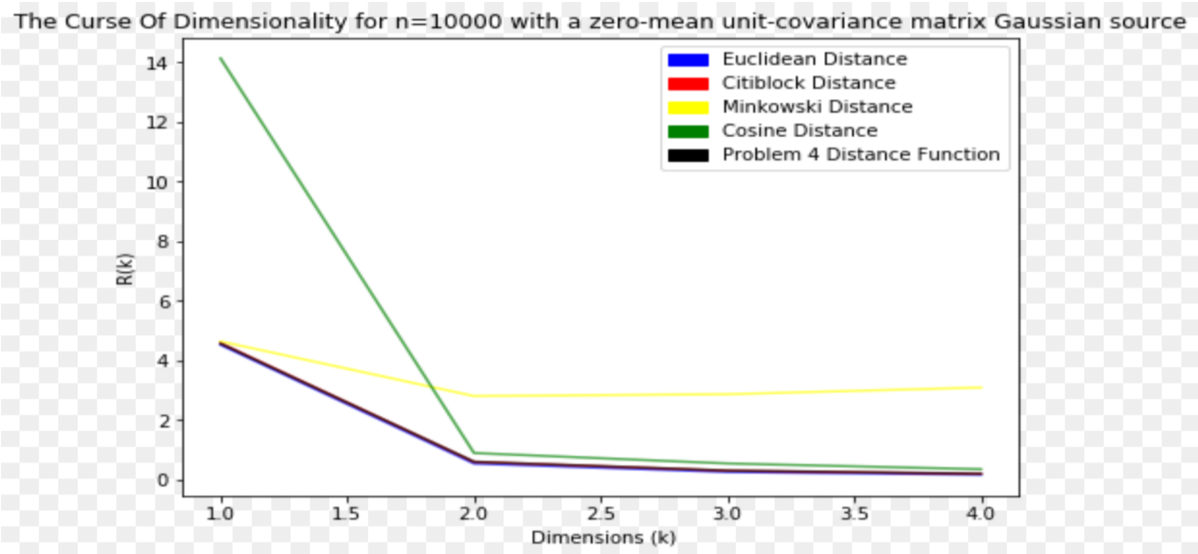


Part B

Repeated the experiment, with the data being generated from a gaussian source







Explanation

- It is observed that as the dimensionality increases all the distance metrics tend to drop in R(k) measure, thereby proving that in higher dimensions of space most of the distance metrics are performing slightly similar.
- It is inferred from the graphs that the Minkowski distance metric specified here is more applicable for high dimensional data compared to the other metrics. It is also notable that the P value (p=3) contributes to this effect.
- We can also notice that for increasing values of n the R(k) values are performing slightly better with less spikes or variations.