Kernel Density Estimation on spherical domain

Yash Shukla, Master of Science in Computer Science

University of Dublin, Trinity College, 2022

Supervisor: Dr. Athanasios Georgiadis

In this data-driven world, loads of data are generated every second. To get insights from this data, we need to process it through statistical algorithms or some pre-defined ML models. This is feasible only for data with any patterns or parameters that can be predicted. It is tough to get any insights from any dataset that has randomness. The solution for such data is a non-parametric estimation. This research attempts to get insights from such random data using kernel density estimation (a non-parametric estimation technique) in a spherical domain. The dataset used here consisted of earthquakes from 1950 to 2022 of magnitudes 7 and 6, which the U.S. Geological Survey collected under the Earthquake Hazards program. This dataset has random longitude and latitude coordinates of the places where earthquakes occurred. The data was first cleaned and analyzed using python scripts. Then, data was converted from two to three dimensions using trigonometric functions as the research is in a spherical domain. The kernel density estimation approach uses Legendre polynomials to estimate the randomness easily. The data was processed through a kernel for different bandwidths h = 0.01, 0.25, 0.5, and kernel levels k = 1, 10, plotted on a graph to visualize the estimated values. These graphs were compared to the 2D graphs, which were generated with just coordinates placed on the map to check if the estimated values were similar to those of the original values. The major challenge during this research was executing the code for huge data sets, but it was solved by making some changes in the sphere generation. This research also evaluates the estimator using the log-loss sample method and plots for multiple h-value and truncation point combination where $N = 50, s = 0.5$ gives the perfect estimation. The evaluation was done on the basis of both datasets.