Course name: RS: Spatiotemporal Data Analysis

Course code: 800880-M-3

Semester: Spring 2020

Assignment

The Individual take-home assignment is out of 100 points and is worth 40% of your grade. For any questions or clarifications regarding the assignment, please post your queries on Canvas.

Late submission policy: There is a 10% penalty per day; thus, zero credit is earned if submitted after 4 days past the due date. Start early, most assignments will take longer than you expect.

Your individual datasets for this assignment can be found in a zip file with your SSID on Canvas.

The assignment consist of two parts. You will submit code and a report in a zip file . For each part of the assignment, you should include the following:

1. A description of your submitted solution, including any data processing, algorithms used, etc.
2. References to any code, methods, or ideas that you used that are not your own. Remember, these must be publicly available and free to use.
3. Any special instructions that are required to run your code.

In your report, all figures must have captions and titles.

1 Periodicity Analysis of Trajectory Data

This section is worth a total of 35 points. You will be provided with a dataset ($XXXXXXX\_traj.csv$), which contains the time, longitude and latitude of a vehicle. This vehicle (equiped with a poor quality global positioning system (GPS)), travels from it's base (starting location) back again to two different destinations at two different frequencies in one day. The trajectories were modified from the GPS Trajectories Data Set from the UC Irvine Machine Learning Repository [1].

You will create code to explore and visualize the dataset and detect periodicity.

1. (2 points) Find the earliest date and the most recent date for the dataset.
2. (2 points) Find the time interval for the dataset. Is it constant?
3. (2 points) Plot the latitude vs longitude.
4. (4 points) Create a 3D plot of the latitude vs longitude vs time.
5. (5 points) Compute the temporal auto correlation. Display the autocorelation in a plot. (Hint: you can combine latitude and longitude into one value (e.g. $latitude^2 + longitude^2$)). Can you find the periodicity of the two trajectories.
6. (5 points) Evaluate the periodicity of the dataset with a periodogram.
7. (15 points) Denoise the trajectory data. Compare the performance of of autocorrelation plot and periodogram with the original data.

2 Spatial analysis of incidences of virus X in California

This section is worth a total of 35 points. You will be provided with a dataset (*XXXXXXX_spatial.csv*). This file contains the number of incidences of virus X over a number of years (depending on the disease). The results are by county in California. Also included are the names of the county seats in each county, their population and, their approximate latitude and longitude. This data was obtained from the California department of Public Health (https://www.cdph.ca.gov/) In addition, a shape file of the Californian counties was obtained from the California Open Data Portal (https://data.ca.gov/). It can be found in the folder *CA_Counties* in the Assignment folder on Canvas.

1. (5 points) Load the dataset. Display the latitude and longitude in a scatterplot with a colormap corresponding to incidences of virus X.
2. (10 points) Check if the data is normally distributed by a histogram and normal quantile-quantile plot (qqplot). If the data is NOT normally distributed, perform some transformation on this data so that the data is normally distributed. For example, you can apply a sine, cosine, exponential or logarithmic transformation.
3. (5 points) Compute pairwise distances between the two points. Visualize the spatial lag in a scatterplot. Try different lag values to find the point when the autocorrelation is negligible.
4. (10 points) Compute the variogram. Fits a variogram model over the data. Which variogram model (e.g. spherical, linear etc..) best fits the data?
5. (10 points) Compute the number of incidences of virus X at the following latitude and longtitudes:

-120,35 and

-121,37

   You can use interpolation techniques such as kriging to estimate the number of incidences.

6. (3 points) Load and display the shape file for the CA counties.
7. (18 points) Compute with different spatial weights. How does the global and local spatial autocorrelation differ when you use different weighting schemes?
8. (5 points) Compute the correlation with population density and the number of cases of Virus X. Is it positively correlated.?

References

[1] GPS Trajectories Data Set https://archive.ics.uci.edu/ml/datasets/GPS+Trajectories