Hannah Brine

UID: 805107983

CEE/MAE M20

December 12, 2020

Final Report: PCA Study of COVID-19 and Spatial SIR Modeling

This report will explain the PCA method of analysis and Spatial SIR simulations. PCA, or Principal Component Analysis, is a popular method of data reduction and finding which variables are most important for selecting data. Given a file of COVID-19 data from different countries, we will use this method to see which of the categories of data is the most important. To help simulate the spread of disease, we will then create an ordinary differential equation solver and show the spatial SIR model through plots and animations.

Principal Component Analysis

The given set of data consisted of eight categories from 27 different countries: Date, Countries, Infections, Deaths, Cures, Mortality rate, Cure rate, Infection rate. The first two columns of data, Date and Countries, were eliminated as they have no bearing on the analysis.

The main script, project\_805108983\_p1.m is simplistic in its execution. The data is read in using readtable and table2array, and the heading for each column are gathered using the Properties.VariableNames feature.

Our PCA script is then called and plotted using the biplot function.

The myPCA.m script is more complex. There are two variables we want at the end of the program: coeffOrth and pcaData. coeffOrth are the two eigenvectors with the largest eigenvalues that we will to plot the rest of the data against. pcaData is our normalized data multiplied by said eigenvectors

Data is normalized by finding the average of every column and subtracting it from each subsequent element of that column, then subtracting the standard deviation of that column.

for i = 1:columns

avg = mean(data(:,i));

stddev = std(data(:, i));

data\_normal(:,i) = (data(:,i) - avg) / stddev;

end

Once the data is normalized, the convariance is taken using MATLAB's cov function. The aforementioned eigenvectors and eigenvalues are found using the eig function.

Since we want to find the eigenvectors with the greatest eigenvalues, the sort functon is used with the following line, allowing us to sort the absolute value of the diagonal matrix easily.

[eigenvalues, index] = sort(diag(eigenvalues), 'descend', 'ComparisonMethod', 'abs');

Chart

Description automatically generatedOur coeffOrth variable is selected using the eigenvectors matrix, sorted according to the index created by the above line of code. The normalized data is then multiplied by coeffOrth, and the data is ready to plot.

Figure : PCA analysis of COVID data

Figure 1 is the result of the PCA method of dimension reduction. Instead of six different dimensions to plot, the PCA method chooses the most important two and plots the normalized data accordingly. The smaller the angle between the categories, the closer they are relate. Thus, Infections and Deaths have the smallest angle between them, and Infection Rate and Cure Rate are almost at a 180 angle.

Spatial S.I.R.

The second goal of this report is to explain and demonstrate the S.I.R. model of disease spreading among a fixed population. *S* is the number of people susceptible, *I* are those infected,