**Covid-19 Lockdown Management using K-means Clustering Method**

Supraja Rapuru

Masters in Data Science, Bellevue University

[DSC500-T301 Introduction to Data Science (2215-1)](https://cyberactive.bellevue.edu/webapps/blackboard/execute/courseMain?course_id=_500210_1" \o "DSC500-T301 Introduction to Data Science (2215-1))

Catherine Williams

April 23rd, 2021

**Covid-19 Lockdown Management using K-means Clustering Method**

The emergence of novel COVID-19 causing an overload in health system and high mortality rate. The key priority is to contain the epidemic and prevent the infection rate. In this context, many countries are now in some degree of lockdown to ensure extreme social distancing of entire population and hence slowing down the epidemic spread.

The best thing to do seems to be let people go out for their business, but any body tests positive of COVID-19, through contract tracing which is already employed by the government, we would be able, to trace everybody in contact with the confirmed case and managing the lockdown and mass quarantine. This will confirm preventing the spread of the virus to the rest of the people. Further, based on these predictions, specifics actions can be rolled out for positive impact on economy and behavioral aspects of people in the community- for example, relaxing the social distance rules to some extent, opening up the businesses etc.

We do not have a true target to predict instead we look at the data and then try to club similar observations and form different groups. Hence it is an unsupervised learning problem and K-means clustering would be able to identify clusters based on collected positions and distance measurements of the covid confirmed cases in the vicinity. Lets deep dive into application of k-means clustering algorithm for this problem (Sharma,Pulkit (2019,August)).

**Step1: Preparing the data set**

Identify the features from the sourced datasets(s), how the number of infects people may be depend on several factors (14. 10.22207/JPAM.14.SPL1.40.) so considering all those as features in the datasets would help to identify the right pattern. Those factors could be number of vulnerable people in the communities, The time takes to recover a person without symptoms, the social contacts and possibility of infecting them with coronavirus. Further, another factor will affect fast spreading of coronavirus is the frequency of visiting crowded places such as malls and mini markets.

Variables like number of confirmed covid cases,location coordinates of infects person, fatality rate, population density, area name, pre-existing health conditions, age, hospital occupancy, recreation accessibility, type of businesses in the neighbourhood etc

**Step2: Deciding the number of clusters**

Plot and map the data and decide the number of clusters using visual inspection, in this case it would be challenging as there are many factors but however to simplify for this assignment, lets use visual inspection such as grouping areas with similar background

**Step3: Select k random points from the data as centroids**

Next, we randomly select the centroid for each cluster.

**Step 4: Assign all the points to the closest cluster centroid**

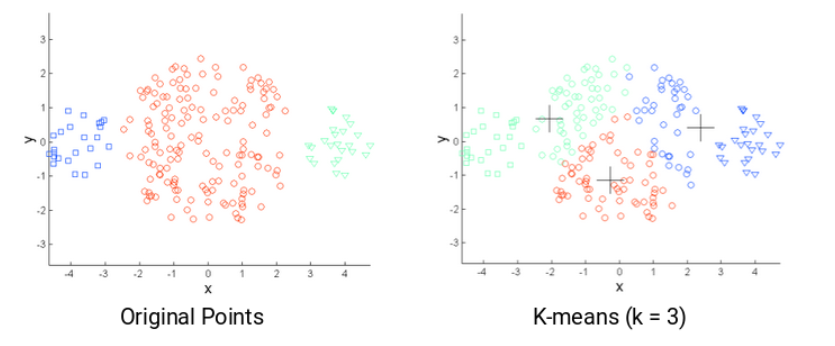
Once we have initialized the centroids, we assign each point to the closest cluster centroid:

**Step 5: Recompute the centroids of newly formed clusters**

Now, once we have assigned all of the points to either cluster, the next step is to compute the centroids of newly formed clusters:

**Step 6: Repeat steps 4 and 5**

We then repeat steps 4, 5 and can stop the algorithm if the centroids of newly formed clusters are not changing. Even after multiple iterations, if we are getting the same centroids for all the clusters, we can say that the algorithm is not learning any new pattern and it is a sign to stop the training.



**References**

Siddiqui, Mohammad Khubeb & Morales-Menendez, Ruben & Gupta, Pradeep & Iqbal, Hafiz & Hussain, Fida & Khatoon, Khudeja & Ahmad, Sultan. (2020). Correlation Between Temperature and COVID-19 (Suspected, Confirmed and Death) Cases based on Machine Learning Analysis. Journal of Pure and Applied Microbiology. 14. 10.22207/JPAM.14.SPL1.40.

Sharma,Pulkit (2019,August). The Most Comprehensive Guide to K-Means Clustering You’ll Ever Need. Website:

https://www.analyticsvidhya.com/