Delaney Curran

Machine learning

9 December 2022

Final Exam

**Executive summary**

For this project I use data from our world in data to look at civil liberties and human rights across the world and how they compared to the United States. The method I used were k-means clustering, hierarchical clustering and agnes-ward. I also used a dendrogram to show how my clusters were dispersed. The variables that I looked at were civil liberties on a variety of democracy, civil liberties with high democracy and civil liberties with low democracy.

**Problem**

The problem I was looking at was civil liberties across the world and how the United States compared to the rest of the world in civil liberties. The data was human rights distribution from 1789 to 2021. However the data from the United States start in 1790.The range is from 0 to 1 with 1 being the most democratic. From basic summary statistics, we can see that the United States outperforms the rest of the world in most summary statistic methods. Until the maximum**,** the United States falls just behind the rest of the world in 2021 when it comes to human rights and civil liberties. This is because countries like Sweden have very high human rights and civil liberties.

**Technique**

After looking through my data I decided the best method to use was k-means clustering. I decided this was the best method because I wanted to group years together based on their attributes. First, I clustered every year from all of the countries together but it was entirely too big and you could hardly make out the clusters. That is when I decided to look specifically at the United States. This way it would dramatically reduce the number of observations I would be clustering.

Next, I ran the “wws” and “silhouette” in order to find the optimal number of clusters. These tests both said the optimal number was 2. I then created the clusters with k = 2. That gave me the two clusters of the United States data. Cluster one has 31 observations while cluster to had 30 observations. The clusters follow a very similar pattern with the older years being clustered together and the more recent years being clustered together. Finally, I ran a dendrogram cluster that outlines the two major clusters.

**Conclusions**

After running all of my tests I am able to conclude that the United States has overall better civil liberties compared to the rest of the world. A major reason for this is that countries like North Korea, which have very poor civil liberties, are skewing the data. I can also conclude that civil liberties dramatically increase with time. As time goes on civil liberties and human rights get better in every country around the world. This is very promising to see and gives me great hope for the future.

**Appendix**

I have linked my GitHub which has my code and other information about this project.