Erdi Olmezogullari

K-Means Assignment

In this assignment, I implemented K-means algorithm in Java from scratch. You can take a look at it to see implementation details.

Before data was normalized to between [0-1], the results were obtained with respect to the initial centroids which you gave at left side as shown in table, below. At right side of table below, after being applied normalization data, K-means algorithm was performed on normalized data.

Before Normalization		After Normalization	
Country	**Group**	**Country**	**Group**
Brazil	0	Brazil	2
Argentina	0	Germany	2
South Africa	0	Mozambique	2
Turkey	0	Australia	2
Lithuania	0	China	2
Germany	1	Argentina	2
Australia	1	United Kingdom	2
United Kingdom	1	South Africa	2
Sweden	1	Zambia	2
Greece	1	Namibia	2
Italy	1	Georgia	2
Japan	1	Pakistan	2
Mozambique	2	India	2
China	2	Turkey	2
Zambia	2	Sweden	2
Namibia	2	Lithuania	2
Georgia	2	Greece	2
Pakistan	2	Italy	2
India	2	Japan	2

In our normalization phase, we found min and max values of the specific attribute (column). And then, to normalize our data to between [0-1], we applied a simple formula below on each attribute on our original data.

$$Y = X - \min(X') / (\max(X') - \min(X')$$