

Intro to Data Analytics and Visualizations

Lecture 32 – k-means Clustering: Assign New Points and Shiny
Fall 2014, November 7th

Outline

1. Assign new points to clusters
2. Protein data clusters visualization with Shiny
3. Inclass 10_3

Practice Problem

Are different countries separated into groups based on their consumption of protein?

Can we identify these groups?

Data: We have 25 countries and variables and their consumption in 9 food groups (different food groups contain protein).

Assignment of New Points

- Suppose that we have determined 5 clusters of countries using the kmeans procedure and the 9 food groups as features
- Sometimes we use the resulting clusters as categories/classes to classify future observations (countries in this case) in a supervised learning machine learning algorithm implementation

Note: Full code and steps explanation for protein example in `Cluster_kmeans_assignpoints.R`

Shiny App for kmeans Clusters

- Vary the number of clusters
- Use only 2 features at a time to determine and visualize the clusters
- Use scaled data, as usual
- Look at protein example
- Label the points with known countries

Note: See Shiny_clusters_protein.R, and App_3 folder with server.R and ui.R for implementation

Inclass Assignment 10_3

- Modify the protein data clustering Shiny App to work with the iris data and visualize clusters.
- Submit your ui.R, server.R, Shiny_cluster_iris.R to Dropbox by Monday, November 10th, at 1pm.
- Submit your Inclass10_12 to Dropbox by the same deadline.