

CSUMB 550 Homework 6b
Cluster Analysis

Using the data from Homework 6a, perform a cluster analysis on the mean values of the variables by entity to determine how entities (individual animals) group together. Use `clValid` and pick different clustering methods (`clMethods = "hierarchical", "kmeans", "diana", "fanny", "som", "model", "sota", "pam", "clara", and "agnes"`). Why did you choose those methods? Choose a distance metric (`metric = "euclidean", "correlation", and "manhattan"`) and explain why you chose that one. For hierarchical clustering, which method (`method = "ward", "single", "complete", and "average"`) is appropriate?

The assignment will require that you search for the meanings of the metrics and methods. A description of some of the clustering methods is provided in the "cluster_details" document on the website.

Based on the validation results, how many clusters would you keep? Why? What are the defining characteristics of those clusters?

Extra credit: Based on the results of the PCA you performed in the previous homework, pick the number of components you would keep, and run the cluster analysis using PC scores of the retained components. Is the answer different? Is the interpretation different?