Assignment 3: Clustering

New Attempt

Due Mar 5 by 11:59pm **Points** 5 **Submitting** a file upload **Attempts** 1
Allowed Attempts 2

Amyotrophic Lateral Sclerosis (ALS) Case-Study Cluster analysis:

Use the attached datasets, **15_ALS_CaseStudy_Docs.docx (https://unt.instructure.com/courses/63454/files/15507484/download?download_frd=1)** , **ALS_PROACT_DataDictionary.pdf (https://unt.instructure.com/courses/63454/files/15507485/download?download_frd=1)** , **ALS_TestingData_78.csv (https://unt.instructure.com/courses/63454/files/15507486/download?download_frd=1)** , **ALS_TrainingData_2223.csv (https://unt.instructure.com/courses/63454/files/15507491/download?download_frd=1)**

This case-study examines the patterns, symmetries, associations and causality in a rare but devastating disease, amyotrophic lateral sclerosis (ALS). A major clinically relevant question in this biomedical study is: *What patient phenotypes can be automatically and reliably identified and used to predict the change of the ALSFRS slope over time?*. This problem aims to explore the data set by unsupervised learning (you only need to work on K mean in this assignment).

- Load and prepare the data.
- Perform summary and preliminary visualization (i.e. show the clustering with selected features).
- Train a **k-Means** model on the data with selected features (3 or more), experiment at least two different k values, and explain which k value is a better choice.
- Evaluating the model performance by report the center of clusters.
- Visualize the final clustering result.

Submit Python code, report that explains the k experiment, performance evaluation, and visualizations.

Ref:
http://www.socr.umich.edu/people/dinov/2017/Spring/DSPA_HS650/notes/12_kMeans_Clustering_

(http://www.socr.umich.edu/people/dinov/2017/Spring/DSPA_HS650/notes/12_kMeans_Clustering_Assignme