

CSCI446/946 Big Data Analytics

Lab Week 4 Clustering

School of Computing and Information Technology
University of Wollongong Australia

Task One

1. Follow the example in the lecture notes to perform K-means cluster analysis using R on the provided dataset “grades_km_input.csv”.
2. Plot WSS of this data to choose the value of k
3. Follow the example as described in <https://dean.mba/portfolio-item/analyzing-student-performance-k-means-clustering/> to plot the identified student clusters in the 2D space.

Task Two

- Using the “ruspini” dataset provided with the `cluster` package in R, perform a k-means analysis. Document the findings and justify the choice of `k`. Hint: use `data(ruspini)` to load the dataset into the R workspace.

