COMP20008 2020 SM1 Workshop Week 8: Clustering and Linear Regression

Clustering

1. Consider the 1-dimensional data set with 10 data points {1,2,3,...10}. Show the iterations of the k-means algorithm using Euclidean distance when $k = 2$, and the random seeds are initialized to {1, 2}.

2. Repeat Exercise 1 using agglomerative hierarchical clustering and Euclidean distance, with single linkage (min) criterion.

Regression and Clustering visualisation

Questions 3 – 4, use the Python notebook *workshop-week8-2020S1-regression.ipynb*.