

Data Mining Fall 2021
Assignment 3

Question 1 (3 marks)

Discuss the advantages and disadvantages of having to specify the number of clusters that an algorithm needs to find. Describe different methods for finding the number of clusters and discuss the implications of the shape of the clusters.

Question 2 (4 marks)

Research semi-supervised methods and describe in your own words how supervised and unsupervised methods can be used together. Describe and compare at least two techniques (400 to 500 words). Include the references.

Question 3 (3 marks)

Describe (in your own words, roughly 300 words) one clustering technique that is not covered in the textbook or the videos that does not use a Euclidean distance measure. Include a reference to a publication that describes the method.

.