

# PATTERN RECOGNITION AND CLASSIFICATION

## WEEK 4 ASSIGNMENT ANSWERS

ARYA RAJIV CHALOLI

PES1201700253

1. Why do we use unsupervised learning and what are its benefits?

Answer:

Unsupervised learning is the procedure of designing a classifier by using unlabelled samples as the training data.

The benefits of using unsupervised procedures include:

- Collecting and labelling a large set of sample patterns can be costly. For instance, recorded speech is virtually free, but accurately labelling the speech — marking what word is being uttered at each instant — can be very expensive and time consuming. If a classifier can be crudely designed on a small set of labelled samples, and then “tuned up” by allowing it to run without supervision on a large, unlabelled set, much time and trouble can be saved.
- It could be useful to process data in the reverse direction: train with large amounts of (less expensive) unlabelled data, and only then use supervision to label the groupings found. This may be appropriate for large “data mining” applications where the contents of a large database are not known beforehand.
- In many applications the characteristics of the patterns can change slowly with time, for example in automated food classification as the

seasons change. If these changes can be tracked by a classifier running in an unsupervised mode, improved performance can be achieved.

- We can use unsupervised methods to find features, that will then be useful for categorization. There are unsupervised methods that represent a form of data-dependent "smart pre-processing" or "smart feature extraction."
- In the early stages of an investigation it may be valuable to gain some insight into the nature or structure of the data. The discovery of distinct subclasses or similarities among patterns or of major departures from expected characteristics may suggest we significantly alter our approach to designing the classifier.

2. Sort the following points using K-means clustering. Do two iterations. (You will get means three times, the initial assumption and once after each iteration)

A(1,1), B(1.5,2), C(3,4), D(5,7), E(3.5,5), F(4.5,5), G(3.5,4.5)?

Answer:

3. Sort the following using Fuzzy K-means clustering. Take A and H as initial means.

A(2,4), B(1,1), C(3,4), D(5,7), E(3,5), F(4,5), G(4,6), H(7,9) One iteration only.

Answer:

4. Sort the following using online clustering. A(2,4), B(1,1), C(3,4), D(5,7), E(3,5),

F(4,5), G(4,6), H(7,9)

Answer:

