Open your mind. LUT.
Lappeenranta University of Technology

BM40A0700 Pattern Recognition

Lasse Lensu

Exercise 11: Unsupervised pattern recognition

1. Proximity measures (1 point): Equidistance curve for a particular distance metric is the set of points having equal distance to a reference point.

   (a) Draw the 2-D equidistance curves for L_1, L_2, and L_∞ distances. Which are their geometric shapes? Is any pair of these identical to each other?

   (b) What would the geometric shapes of equidistance curves be in 3-D?

   Hints: You can choose to fix the reference point to the origin and find the set of points having unit distance to the origin using the metric.

2. Data normalization in clustering (1 point): Should the features used for clustering be normalized? Does this depend on the features? Give examples where normalization might be useful or not useful?

   Hints: There might not be unique "correct" answers. Remember that clustering tries to group the data into clusters where the distance of samples within a group is small and the distance between the groups is large.

3. K-means clustering (1 point): Use Matlab function `kmeans` to cluster the provided data. Use different number of clusters for the experimentation. Visualize the results.

   Repeat each clustering several times. Do you always get the same result?

   Additional files: `irisdata.mat`.

4. Fuzzy c-means clustering (FCM) (1 point): Experiment with FCM. The Fuzzy Logic Toolbox of Matlab includes the function `fcm` which implements this approach.

   Use different number of clusters for the experimentation. Compare your results to the ones produced by another clustering method.

   Additional files: `irisdata.mat`.

PO BOX 20, FI-53851 LAPPEENRANTA, FINLAND
E-mail: lasse.lensu@lut.fi