



Open your mind. LUT.

Lappeenranta University of Technology

LUT Machine Vision and Pattern Recognition

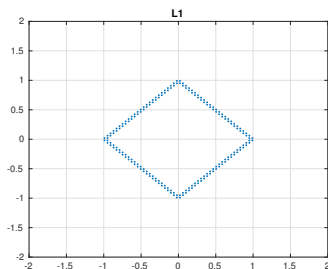
2015-11-30

BM40A0700 Pattern Recognition

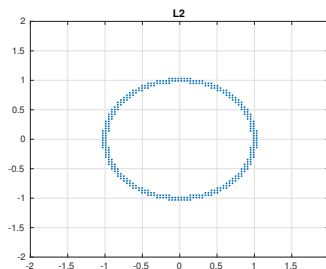
Lasse Lensu

Exercise 11 solutions: Unsupervised pattern recognition

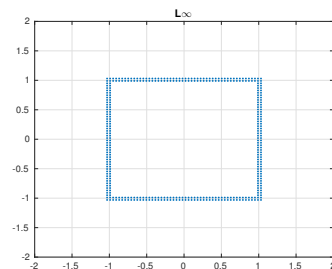
1. Proximity measures (1 point): Equidistance curve for a particular distance metric is the set of points having equal distance to a reference point.
  - (a) The equidistance curves for  $L_1$ ,  $L_2$  and  $L_\infty$  distances are a diamond, circle and square, see Fig.1. None of these is identical to another.
  - (b) The geometric shapes of equidistance curves in 3-D would be a diamond (two pyramids attached to each other), sphere and cube, see Fig.1.



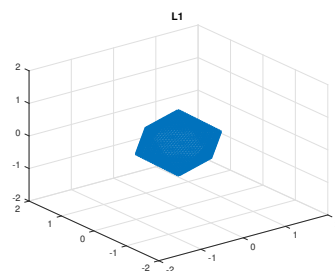
(a)



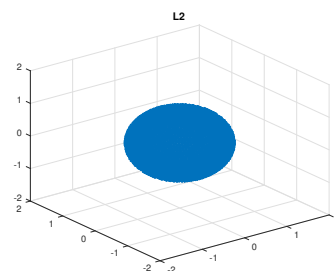
(b)



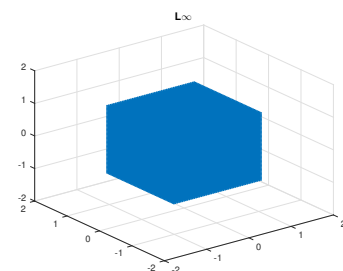
(c)



(d)



(e)



(f)

Figure 1: Approximate equidistance curves and shapes for the distances  $L_1$ ,  $L_2$  and  $L_\infty$ .

2. Data normalization in clustering (1 point): If the feature space is isotropic and the data are distributed approximately evenly in all dimensions, using Euclidean distance for defining the clusters causes the clusters to be invariant to rigid transformations such as translations and rotations. However, the clusters are not invariant to general linear transformations or other transformations affecting the distance relationships. If the assumptions above are not valid, data normalization such as standardization can be applied to achieve invariance to, for example, scaling. This prevents one or a subset of features dominating the distance calculations and this is appropriate for certain type of distributions. However, normalizing unevenly distributed data can significantly affect the result of clustering because of reducing the separation between the clusters.

3. K-means clustering (1 point): not published.
4. Fuzzy c-means clustering (FCM) (1 point): not published.